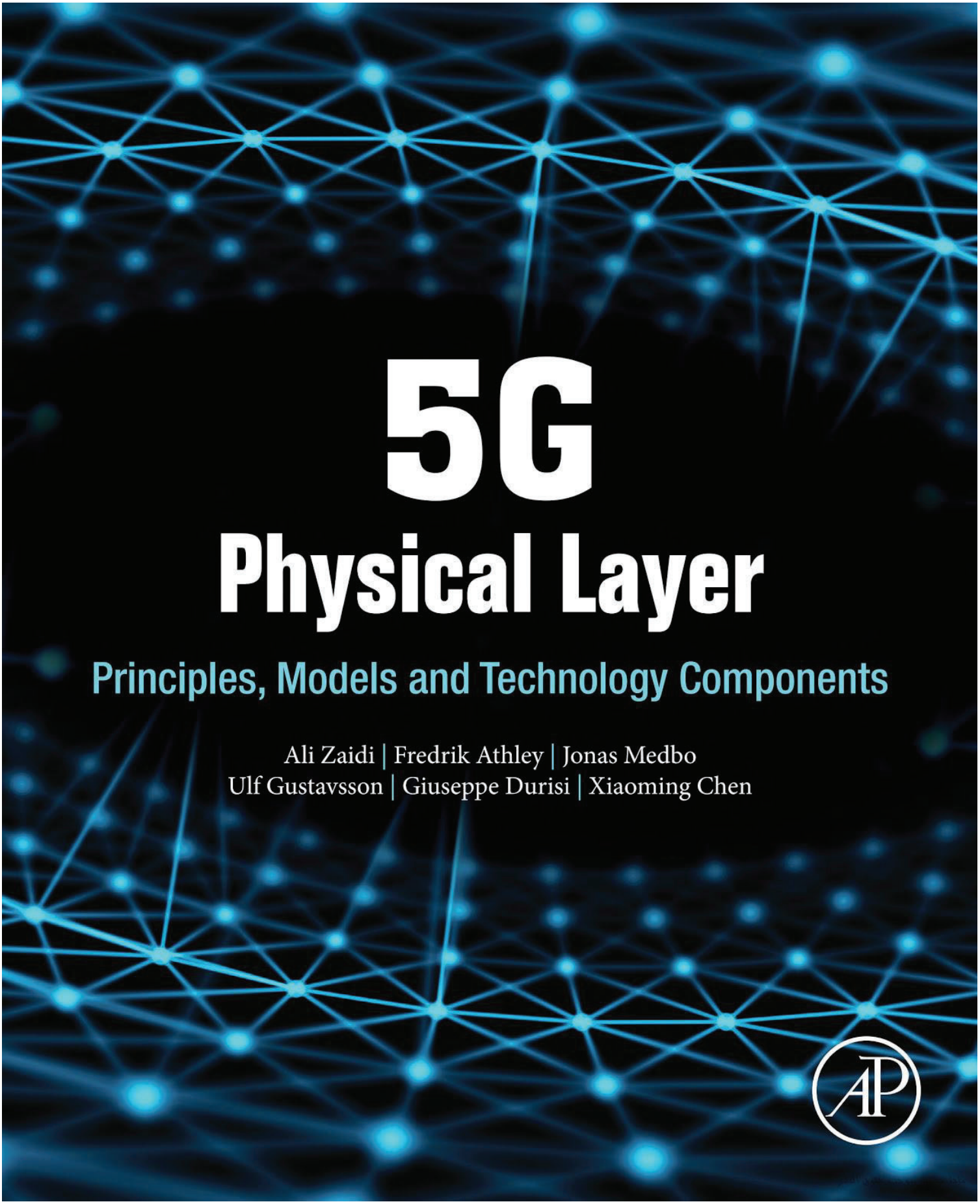

5G
Physical Layer
Principles, Models and Technology Components
Ali Zaidi | Fredrik Athley | Jonas Medbo
Ulf Gustavsson | Giuseppe Durisi | Xiaoming Chen
AP

5G Physical Layer

Principles, Models and Technology Components

Ali Zaidi
Fredrik Athley
Jonas Medbo
Ulf Gustavsson
Giuseppe Durisi
Xiaoming Chen





ACADEMIC PRESS
An imprint of Elsevier

To my loving wife, Marzieh, for bringing new inspirations.
(Ali Zaidi)

To Linnea, Simone, and my parents for constant support and patience.
(Ulf Gustavsson)

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1650, San Diego, CA 92101, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

Copyright © 2018 Elsevier Ltd. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-814578-4

For information on all Academic Press publications
visit our website at https://www.elsevier.com/books-and-journals



Publisher: Mara Conner
Acquisition Editor: Tim Pitts
Editorial Project Manager: Leticia Lima
Production Project Manager: R. Vijay Bharath
Designer: Christian J. Bilbow

Typeset by VTeX

Acknowledgments

We thank the following colleagues for their continuous support, collaboration, and inspiration: Robert Baldemair (Ericsson), Vicent Molés-Cases (former Ericsson, now UPV), Markus Ringström (Ericsson), Joakim Sorelius (Ericsson), Marie Hogan (Ericsson), Gianluigi Liva (German Aerospace Center), Hua Wang (Keysight), Kittipong Kittichokechai (Ericsson), Mattias Andersson (Ericsson), Erik Dahlman (Ericsson), Stefan Parkvall (Ericsson), Kristoffer Andersson (Ericsson), Sven Mattisson (Ericsson), Lars Sundström (Ericsson), Per Landin (Ericsson), Sven Jacobsson (Ericsson), Thomas Eriksson (Chalmers University), Christian Fager (Chalmers University), Erik G. Larsson (Linköping University), Christopher Mollén (former Linköping University), Katharina Hausmair (former Chalmers University, now Qamcom), Christoph Studer (Cornell University), Jian Luo (Huawei), Jaakko Vihriälla (Nokia), Andreas Wolfgang (Qamcom), Robin Gerzaguet (ENSSAT), Yinan Qi (Samsung), Ning He (Ericsson), Karl Werner (Ericsson), Sebastian Faxér (Ericsson), Shehzad Ali Ashraf (Ericsson), Eleftherios Karipidis (Ericsson), Peter von Wrycza (Ericsson), Miurel Tercero (Ericsson), Håkan Björkegren (Ericsson), Mikael Wahlén (Ericsson), Joakim Hallin (Ericsson), Arne Simonsson (Ericsson), Kjell Larsson (Ericsson), Göran Klang (Ericsson), Dennis Sundman (Ericsson), Henrik Asplund (Ericsson), and Satyam Dwivedi (Ericsson).

We acknowledge mmMAGIC (a European collaborative research project), 3GPP (a collaboration project between standard development organizations), and Ericsson Research for their tremendous contributions to the development of 5G New Radio. We are also grateful to Qamcom Research & Technology AB for publishing the source code that is part of this book.

Samsung Exhibit 1020, Page 5 of 314

List of Acronyms

3GPP third generation partnership project
2G second generation
3G third generation
4G fourth generation
5G fifth generation
ACEPR adjacent channel error power ratio
ACLR adjacent channel leakage ratio
ACE active constellation extension
ACLR adjacent channel leakage ratio
ADC analog-to-digital converter
ADSL asymmetric digital subscriber line
AMPS advanced mobile phone system
AMAM amplitude-to-amplitude modulation
AWGN additive white Gaussian noise
BICM bit-interleaved coded modulation
BP belief propagation
BEC binary erasure channel
BER bit error rate
BLER block error rate
BS base station
CCDF complementary CDF
CDD cyclic delay diversity
CDF cumulative distribution function
CDMA code division multiple access
CFO carrier frequency offset
CN check node
CPE common phase error
CTF channel transfer function
CIR channel impulse response
CoMP coordinated multipoint
CP cyclic prefix
CP-OFDM cyclic prefix OFDM
CPRI common public radio interface
CQI channel quality indicator
CRI CSI-RS resource indicator
CRC cyclic-redundancy check
CRS cell-specific reference signal
CSI channel state information
CSI-IM CSI interference measurement
CSIR channel state information at the receiver
CSI-RS channel state information reference signal
CSIT channel state information at the transmitter
CW continuous wave
D2D device-to-device
DAB digital audio broadcast
DAC digital-to-analog converter
DFT discrete Fourier transform

Samsung Exhibit 1020, Page 6 of 314

DFTS-OFDM DFT-spread OFDM
DLL delay-locked loop
DPD digital pre-distortion
DM-RS demodulation reference signal
DPC dirty paper coding
DVB-T digital video broadcast-terrestrial
EDGE enhanced data rates for GSM evolution
EGT equal gain transmission
EMF electromagnetic field
EPDCCH enhanced physical downlink control channel
eNB evolved Node B
eMBB enhanced MBB
EVM error vector magnitude
FBMC filter bank multicarrier
FCC federal communications commission
FDD frequency division duplex
FDM frequency division multiplexing
FD-MIMO full-dimension MIMO
FEC forward error correction
FFT fast Fourier transform
FIR finite impulse response
F-OFDM filtered OFDM
FS frequency spreading
FSTD frequency-switched transmit diversity
GaN Gallium-Nitride
GI guard interval
GMP generalized memory polynomial
GMSK Gaussian minimum shift keying
gNB the next generation Node B
GoB grid-of-beams
GPRS general packet radio service
GSM global system for mobile communications
HARQ hybrid ARQ
HetNet heterogeneous network
HPBW half-power beam width
HSPA high speed packet access
ICI intercarrier interference
ICT information and communication technologies
ID identity
IFDMA interleaved frequency division multiple access
IFFT inverse FFT
IID independent identically distributed
IMT international mobile telecommunications
IoT internet of things
IP internet protocol
ISI intersymbol interference
ITU international telecommunications union
ITU-R ITU-radiocommunications sector
KPIs key performance indicators
L1-RSRP layer-1 RSRP
LBT listen-before-talk
LDPC low density parity check
LFDMA localized frequency division multiple access

LLR log-likelihood ratio
LNA low-noise amplifier
LoS line-of-sight
LS least-square
LTE long term evolution
LUT look-up table
MBB mobile broadband
MBSFN multicast-broadcast single-frequency network
MAC medium access control
MCS modulation and coding scheme
MIMO multiple-input multiple-output
ML maximum likelihood
MP memory polynomial
MMS multimedia message service
MMSE minimum mean square error
mmMAGIC millimeter magic project
mMTC massive machine type communications
MRC maximum ratio combining
MRT maximum ratio transmission
MSE mean squared error
MSK minimum shift keying
MU multi user
MU-MIMO multiuser MIMO
NB Node B
NB-IoT narrow band-IoT
NMSE normalized mean squared error
NMT nordic mobile telephone
NLoS non line-of-sight
NR new radio
NSA non-stand-alone
NZP CSI-RS non-zero-power CSI-RS
OFDM orthogonal frequency division multiplexing
OSTBC orthogonal space–time block code
OOB out-of-band
OQAM offset QAM
OSD ordered-statistic decoding
PA power amplifier
PAPR peak-to-average power ratio
PBCH physical broadcast channel
PCM parity check matrix
PDCCH physical downlink control channel
PDCP packet data convergence protocol
PDSCH physical downlink shared channel
PHD polyharmonic distortion
PHY physical
PLL phase-locked loop
PN phase-noise
PMI precoding matrix indicator
PPN polyphase network
PQI PDSCH rate matching and quasicolocation indicator
PRACH physical random access channel
PSD power spectral density
PSK phase shift keying

PSS primary synchronization signal
PTS partial transmit sequence
PT-RS phase tracking reference signal
PUCCH physical uplink control channel
PUSCH physical uplink shared channel
QAM quadrature amplitude modulation
QoS Quality of Service
QPSK quadrature phase shift keying
RAN radio access network
RB resource block
RCFO residual CFO
RF radio frequency
RFICs radio frequency integrated circuits
RI rank indicator
RLC radio link control
RLF radio link failure
RMa rural macro
RMS root mean square
RRC radio resource control
RSRP reference signal received power
Rx receive
SA stand-alone
SAR specific absorption rate
SC-FDMA single-carrier frequency division multiple access
SDAP service data adaptation protocol
SDMA space division multiple access
SER symbol error rate
SFO sampling frequency offset
SFBC space frequency block coding
SFTD space frequency transmit diversity
SIC successive interference cancellation
SINDR signal-to-interference, distortion and noise ratio
SINR signal-to-interference-and-noise ratio
SIR signal-to-interference-ratio
SIMO single-input multiple-output
SISO single-input single-output
SLI strongest layer indicator
SLM selective mapping
SMS short message service
SNR signal-to-noise ratio
SORTD spatial orthogonal-resource transmit diversity
SQNR signal to quantization noise ratio
SRI SRS resource indicator
SRS sounding reference signal
SSB synchronization signal block
SSS secondary synchronization signal
SSBRI SSB resource indicator
STTD space–time transmit diversity
SU-MIMO single-user MIMO
SVD singular value decomposition
TB tail-biting
TCI transmission configuration indicator
TDD time division duplex

TDM time division multiplexing
TI tone injection
TM transmission mode
TPMI transmit precoder matrix indicator
TR tone reservation
TRI transmit rank indicator
TRS tracking reference signal
TS technical specification
TSG technical specification group
TTI transmit time interval
Tx transmit
QCL quasicolocation
UE user equipment
UF-OFDM universally filtered OFDM
ULA uniform linear array
UMa urban macro
UMTS universal mobile telecommunications system
UPA uniform planar array
URLLC ultra-reliable low-latency communications
V2I vehicle-to-infrastructure
V2N vehicle-to-network
V2P vehicle-to-pedestrian
V2X vehicle-to-anything
V2V vehicle-to-vehicle
VCO voltage controlled oscillator
VN variable node
WAP wireless application protocol
WAVA wraparound Viterbi algorithm
WiMAX worldwide inter-operability for microwave access
W-OFDM windowed OFDM
WRC world radiocommunication conference
WSS wide-sense stationary
VNA vector network analyzer
VS-GMP vector switched GMP
WCDMA wideband CDMA
WG working group
WLAN wireless local area network
ZF zero-forcing
ZP CSI-RS zero-power CSI-RS

Contents

Acknowledgments xiii
List of Acronyms xv

CHAPTER 1 Introduction: 5G Radio Access 1
1.1 Evolution of Mobile Communication 2
1.2 5G New Radio Access Technology 3
1.3 5G NR Global View 4
1.3.1 5G Standardization 4
1.3.2 Spectrum for 5G 6
1.3.3 Use Cases for 5G 8
1.3.4 5G Field Trials 10
1.3.5 5G Commercial Deployments 14
1.4 Preview of the Book 16
References 18

CHAPTER 2 NR Physical Layer: Overview 21
2.1 Radio Protocol Architecture 21
2.2 NR PHY: Key Technology Components 23
2.2.1 Modulation 23
2.2.2 Waveform 24
2.2.3 Multiple Antennas 24
2.2.4 Channel Coding 25
2.3 Physical Time-Frequency Resources 26
2.4 Physical Channels 26
2.5 Physical Signals 28
2.6 Duplexing Scheme 30
2.7 Frame Structure 30
2.8 PHY Procedures and Measurements 32
2.9 Physical Layer Challenges 32
2.9.1 Propagation Related Challenges 32
2.9.2 Hardware Related Challenges 33
References 34

CHAPTER 3 Propagation & Channel Modeling 35
3.1 Propagation Fundamentals 35
3.1.1 Electromagnetic Waves 36
3.1.2 Free-Space Propagation 37
3.1.3 Scattering and Absorption 39
3.2 Propagation Channel Characterization 40
3.2.1 Frequency-Delay Domain 42
3.2.2 Doppler-Time Domain 44
3.2.3 Directional Domain 46

مادة محمية بموجب حقوق النشر

Samsung Exhibit 1020, Page 11 of 314

3.3 Experimental Channel Characteristics 48
3.3.1 Measurement Techniques 48
3.3.2 Analysis Methods 50
3.3.3 Transmission Loss Measurements 55
3.3.4 Delay Domain Measurements 61
3.3.5 Directional Domain Measurements 66
3.4 Channel Modeling 74
3.4.1 5G Stochastic Channel Models 75
3.4.2 Geometry-Based Modeling 82
3.5 Summary and Future Work 82
References 84

CHAPTER 4 Mathematical Modeling of Hardware Impairments 87
4.1 RF Power Amplifiers 88
4.1.1 The Volterra Series 89
4.1.2 Common Subsets of the Volterra Series 90
4.1.3 Global vs. Local Basis Functions 92
4.1.4 Experimental Model Validation 93
4.1.5 Mutually Orthogonal Basis Functions 97
4.1.6 Multi-Antenna Environments and Mutual Coupling 100
4.2 Oscillator Phase Noise 104
4.2.1 Phase-Noise Power Spectrum and Leeson's Equation 104
4.2.2 Phase-Noise Modeling: Free-Running Oscillator 104
4.2.3 Phase-Noise Modeling: Phase-Locked Loop 106
4.3 Data Converters 107
4.3.1 Modeling of Quantization Noise 107
4.4 Statistical Modeling 109
4.4.1 The Bussgang Theorem and the System Model 109
4.5 Stochastic Modeling of Power Amplifiers 110
4.6 Oscillator Phase Noise 110
4.7 Stochastic Modeling of Data Converters 111
4.8 Model Concatenation and Simulations 112
4.8.1 Signal-to-Interference and Noise Ratio 113
4.8.2 Simulations 113
4.8.3 Simulation Results 114
References 118

CHAPTER 5 Multicarrier Waveforms 119
5.1 Multicarrier Waveforms 119
5.1.1 The Principle of Orthogonality 120
5.1.2 OFDM-Based Waveforms 123
5.1.3 Filter Bank-Based Waveforms 130
5.2 Single Carrier DFTS-OFDM 140
5.3 Waveform Design Requirements for 5G NR 142

مادة محمية بموجب حقوق النشر

5.4 Key Performance Indicator for NR Waveform Design 143
5.5 Waveform Comparison for NR 146
5.5.1 Frequency Localization 146
5.5.2 Power Efficiency 148
5.5.3 Time-Varying Fading Channel 149
5.5.4 Baseband Complexity 149
5.5.5 Phase-Noise Robustness Comparison 152
References 158

CHAPTER 6 NR Waveform 159
6.1 Suitability of OFDM for NR 159
6.2 Scalable OFDM for NR 162
6.2.1 Why 15 kHz as Baseline Numerology? 165
6.2.2 Why 15×2^n kHz Scaling? 166
6.3 OFDM Numerology Implementation 167
6.3.1 Phase Noise 167
6.3.2 Cell Size, Service Latency, and Mobility 170
6.3.3 Multiplexing Services 172
6.3.4 Spectral Confinement 173
6.3.5 Guard Band Considerations 176
6.3.6 Implementation Aspects 179
6.4 Improving Power Efficiency of NR Waveform 179
6.4.1 Techniques With Distortion 180
6.4.2 Distortion-less Techniques 182
6.5 Effects of Synchronization Errors 184
6.5.1 Effect of Timing Offset 184
6.5.2 Effect of Carrier Frequency Offset 186
6.5.3 Sampling Frequency Offset 187
6.6 Impairment Mitigation 188
6.6.1 A Phase-Noise Mitigation Scheme 188
6.6.2 CFO and SFO Mitigation 191
References 197

CHAPTER 7 Multiantenna Techniques 199
7.1 The Role of Multiantenna Techniques in NR 200
7.1.1 Low Frequencies 200
7.1.2 High Frequencies 200
7.2 Multiantenna Fundamentals 202
7.2.1 Beam-Forming, Precoding, and Diversity 202
7.2.2 Spatial Multiplexing 208
7.2.3 Antenna Array Architectures 215
7.2.4 UE Antennas 220
7.2.5 Antenna Ports and QCL 223
7.2.6 CSI Acquisition 224
7.2.7 Massive MIMO 229

مادة محمية بموجب حقوق النشر

7.3 Multiantenna Techniques in NR 230
7.3.1 CSI Acquisition 231
7.3.2 Downlink MIMO Transmission 235
7.3.3 Uplink MIMO Transmission 235
7.3.4 Beam Management 238
7.4 Experimental Results 245
7.4.1 Beam-Forming Gain 245
7.4.2 Beam Tracking 247
7.4.3 System Simulations 248
References 251

CHAPTER 8 Channel Coding 253
8.1 Fundamental Limits of Forward Error Correction 254
8.1.1 The Binary AWGN Channel 254
8.1.2 Coding Schemes for the Binary-AWGN Channels 254
8.1.3 Performance Metrics 254
8.2 FEC Schemes for the Bi-AWGN Channel 258
8.2.1 Introduction 258
8.2.2 Some Definitions 259
8.2.3 LDPC Codes 260
8.2.4 Polar Codes 264
8.2.5 Other Coding Schemes for the Short-Blocklength Regime 269
8.3 Coding Schemes for Fading Channels 271
8.3.1 The SISO Case 272
8.3.2 The MIMO Case 276
References 277

CHAPTER 9 Simulator 279
9.1 Simulator Overview 280
9.2 Functional Modules 280
9.2.1 Channel Model 281
9.2.2 Power Amplifier Model 281
9.2.3 Phase-Noise Model 282
9.2.4 Synchronization 282
9.2.5 Channel Estimation and Equalization 283
9.3 Waveforms 283
9.3.1 CP-OFDM 284
9.3.2 W-OFDM 284
9.3.3 UF-OFDM 284
9.3.4 FBMC-OQAM 285
9.3.5 FBMC-QAM 286
9.4 Simulation Exercises 286
9.4.1 Spectral Regrowth 286
9.4.2 Impairment of CFO 287
9.4.3 Impairment of PN 291

مادة محمية بموجب حقوق النشر

9.4.4 Impairment of Fading Channel 293
References 294
Index 297

مادة محمية بموجب حقوق النشر

CHAPTER

1 INTRODUCTION: 5G RADIO ACCESS

Information and communication technologies (ICT) have sparked innovations in almost every society. The ever-growing capabilities for instantly transferring and processing information are transforming our societies in many ways—online shopping, social interactions, professional networking, media distribution, e-learning, instant information access, remotely watching live events, audio and video communication, virtual offices and workforce, and so on. Various industries and corporations have also been evolving their processes and businesses based on technological advancements in information and communication technologies.

Fifth generation (5G) mobile communication is expected to enormously expand the capabilities of mobile networks. New technologies and functionalities are being introduced for 5G systems in various domains—wireless access, transport, cloud, application, and management systems [6]. These advancements are targeting traditional mobile broadband users as well as emerging machine-type users, so that new and superior services can be enabled for both consumers and industries at large, unleashing the potential of the internet of things (IoT), and virtual and augmented reality. According to a recent survey performed across 10 different industries [4], the global revenues driven by 5G technologies will be as high as 1.3 trillion USD by 2026 (see Fig. 1.1 for revenues per industrial segment). It is estimated that by 2023, there will be around 3.5 billion cellular IoT connections [24].

The backbone of any mobile communication system is its wireless access technology, which connects devices with radio base stations. As almost every society and industry is looking forward to the 5G revolution with its specific set of requirements, the design of the 5G wireless access is challenging. A 5G wireless access technology is expected to provide extreme data rates, ubiquitous coverage, ultra-reliability, very low latency, high energy efficiency, and a massive number of heterogeneous connections. The human-centric emerging applications are augmented reality, virtual reality, and online gaming—these demand extreme throughput and low latency. For machine-type communication there are two main segments: massive IoT and critical IoT. Massive IoT is characterized by a high number of low cost device connections, supporting small volumes of data per device with long battery life and deep coverage (for example, for underground and remote areas). The applications are in smart buildings, utilities, transport logistics, agriculture, and fleet management. The critical IoT is characterized by ultra-reliability and very low-latency connectivity, for example, to support autonomous vehicles, smart power grids, robotic surgery, traffic safety, and industrial control.

This book is about the upcoming 5G wireless access technology, and it focuses on its physical layer. A preview of the book is provided in Section 1.4. This chapter gives a holistic view of the 5G wireless access technology and its global development. We start with a brief history of mobile access technologies in Section 1.1 and introduce the 5G mobile access technology in Section 1.2. In Section 1.3, we provide a global picture of 5G wireless access—spectrum allocations, standardization, use cases and their requirements, field trials, and future commercial deployments.

5G Physical Layer. https://doi.org/10.1016/B978-0-12-814578-4.00006-0
Copyright © 2018 Elsevier Ltd. All rights reserved.

Samsung Exhibit 1020, Page 16 of 314

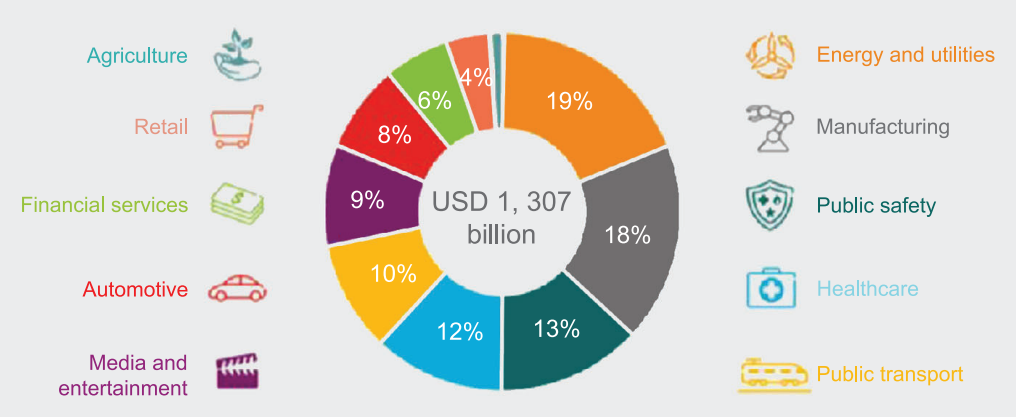


FIGURE 1.1

5G enabled industry digitalization revenues for ICT players, 2026 (Source: Ericsson [4]).

1.1 EVOLUTION OF MOBILE COMMUNICATION

In 1946, the US federal communications commission (FCC) approved a first mobile telephony service to be operated by AT&T in 1947. At this time, the equipment were bulky and had to be installed in a vehicle due to the weight and its excessive power consumption. From this point on, more than three decades of cellular communication technology evolution has led to a shift from analog to digital formats of communication, going from what was mainly voice to high-speed data communication.

Leading up from the mid-1980s, the first generation (1G) of cellular communication, which mainly carried voice, grew up using formats such as advanced mobile phone system (AMPS) in the USA and nordic mobile telephone (NMT) in Scandinavia. These analog formats were later replaced moving towards 2G with the first digital communication schemes around the mid to late-1990s—global system for mobile communications (GSM) in Europe and digital-AMPS for the USA. At this point, the short message service (SMS) was introduced, being one of the first widely used non-voice applications for cellular communication. Enhancement for 2.5G using enhanced data rates for GSM evolution (EDGE), general packet radio service (GPRS) and code division multiple access (CDMA) sparked the use of mobile data communication and early cellular internet connectivity in the early-2000s. This was an early enabler, which did, however, require a specific protocol, known as wireless application protocol (WAP).

Moving forward from 2G into 3G, in order to meet the increasing demand for cellular access data rates, universal mobile telecommunications system (UMTS) based on wideband CDMA (WCDMA) technology was introduced by third generation partnership project (3GPP) just around 2000. With advances in mobile user equipment technology, this enabled the user to not only communicate via multimedia message service (MMS), but also stream video content. Transitioning to 4G, long term evolution (LTE) was introduced, which does not only imply major changes on the air interface, but was moving from code division multiplexing to orthogonal frequency division multiplexing (OFDM) and time division duplex (TDD) or frequency division duplex (FDD).

Entering the era of 4G, there were mainly two competing technologies at an early stage. These were worldwide inter-operability for microwave access (WiMAX), based on IEEE 802.16m, and LTE Advanced, which is an extension of LTE. LTE-A introduced technology components such as carrier

aggregation and improved support for coordinated multipoint (CoMP) transmission and heterogeneous network (HetNet) deployments for improving Quality of Service (QoS) in hot-spots and coverage for cell-edge users. LTE-A prevailed as the dominant cellular access technology today and has served as the basis of the transition to 5G mobile communications. The transition from 4G to 5G is inspired by new human-centric and machine-centric services across multiple industries.

1.2 5G NEW RADIO ACCESS TECHNOLOGY

5G wireless access is envisioned to enable a networked society, where information can be accessed and shared anywhere and anytime, by anyone and anything [2]. 5G shall provide wireless connectivity for anything that can benefit from being connected. To enable a truly networked society, there are three major challenges:

- A massive growth in the number of connected devices.
- A massive growth in traffic volume.
- A wide range of applications with diverse requirements and characteristics.

To address these challenges, 5G wireless access not only requires new functionalities but also substantially more spectrum and wider frequency bands.

Fig. 1.2 illustrates the operational frequency ranges of existing (2G, 3G, 4G) and future (5G) mobile communication systems. The current cellular systems operate below 6 GHz. A large amount of spectrum is available in the millimeter-wave frequency band (30–300 GHz); however, there is no commercial mobile communication system operating in the millimeter-wave frequencies today. 4G LTE is designed only for frequencies below 6 GHz. There are some local area networks and (mostly) indoor communication systems based on the IEEE 802.11ad and 802.15.3c standards that operate in the unlicensed 60 GHz band. IEEE 802.11ay, a follow-up of 802.11ad, is under development. 3GPP is currently developing a global standard for new radio access technology, 5G new radio (NR), which will operate in frequencies from below 1 GHz up to 100 GHz. 5G NR shall unleash new frequencies and new functionalities to support ever-growing human-centric and machine-centric applications.

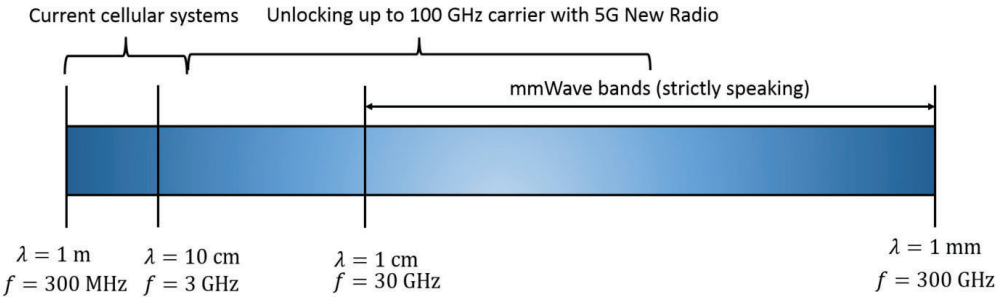


FIGURE 1.2

Frequency ranges of current and future mobile communication systems.

The vision of 5G wireless access is shown in Fig. 1.3. 5G wireless access comprises both 5G NR and LTE evolution. LTE is continuously evolving to meet a growing part of the 5G requirements. The evolution of LTE towards 5G is referred to as the LTE Evolution [13]. LTE will operate below 6 GHz and NR will operate from sub-1 GHz up to 100 GHz. 5G NR is optimized for superior performance; it is not backwards compatible to LTE, meaning that the legacy LTE devices do not need to be able to access the 5G NR carrier. However, a tight integration of NR and LTE evolution will be required to efficiently aggregate NR and LTE traffic.

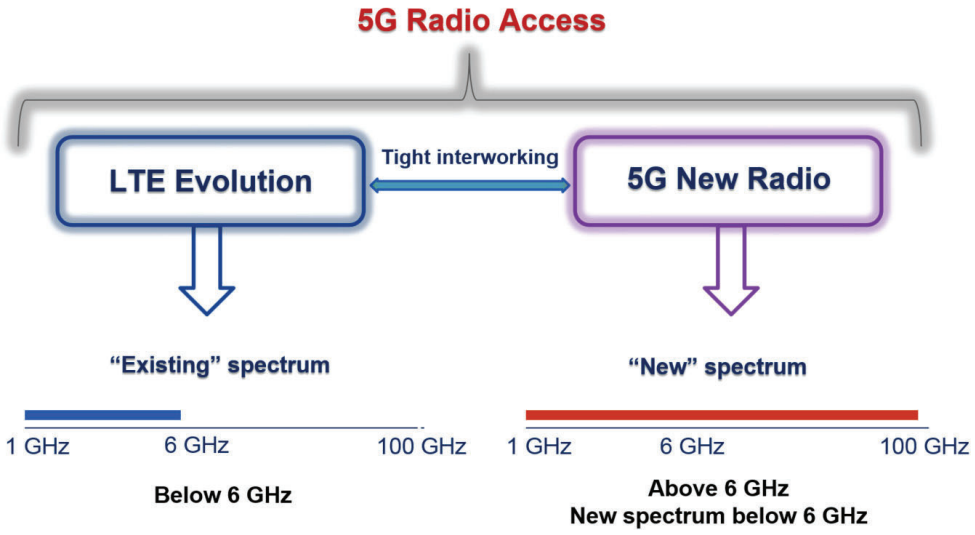


FIGURE 1.3

5G radio access vision.

1.3 5G NR GLOBAL VIEW

The research and concept development of the 5G wireless access technology (named 5G NR since 2016) started almost a decade ago with new inspiring applications and business cases in mind. The research efforts led to development of 5G test-beds in universities as well as industry. Like previous generations of cellular systems, the development of 5G NR is a well-coordinated global effort—addressing the new spectrum allocations on global and regional levels for 5G NR and the global 5G NR standardization in 3GPP based on the 5G requirements defined by international telecommunications union (ITU). To give a global picture of the 5G NR developments, we will in the following discuss spectrum allocations and regulatory aspects, the standardization process, major use cases and their requirements, some precommercial trials, and expected commercial deployments.

1.3.1 5G STANDARDIZATION

The specifications for NR and LTE are developed by the 3GPP which is a collaboration between seven regional and national standard development organizations from Asia, Europe and North America:

ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, and TTC. 3GPP produces technical specifications which are transposed into standards by the standardization bodies. 3GPP was started in 1998 with the initial goal of developing globally applicable specifications for the third generation (3G) of mobile communications. The scope has since then been widened and now includes the development and maintenance of specifications for the second generation (2G) GSM, 3G WCDMA/HSPA, fourth generation (4G) LTE, and 5G NR/LTE evolution.

The international radio frequency (RF) spectrum is managed by the ITU-radiocommunications sector (ITU-R). ITU-R has also a responsibility to turn technical specifications from, e.g., 3GPP into global standards, including also countries not covered by the standardization bodies in 3GPP. ITU-R also defines the spectrum for the so-called international mobile telecommunications (IMT) systems. The IMT systems correspond in practice to the different generations of mobile communications, from 3G onwards. The 3G and 4G technologies are included in the IMT-2000 and IMT-Advanced recommendations, respectively. A new ITU-R recommendation for 5G, called IMT-2020, is planned to be developed in 2019–20.

ITU-R does not produce the detailed technical specifications but defines IMT in cooperation with regional standardization bodies by specifying requirements that an IMT technology should fulfill. The actual technology is developed by others, e.g. 3GPP, and submitted to ITU-R as a candidate IMT technology. The technology is evaluated against the specified requirements and may then be approved as an IMT technology. ITU-R gives recommendations of radio interface technologies in radio interface specifications for a particular IMT system and provides references to the corresponding detailed specifications which are maintained by the corresponding standardization bodies. The IMT-2000 radio interface specifications include six different radio interface technologies while IMT-Advanced includes two. Unlike previous generations, for 5G it is not expected that competing technologies will be submitted as candidates for IMT-2020; only 3GPP-based technologies are anticipated. 3GPP will submit LTE evolution and NR together as their candidate system for IMT-2020.

Different generations of mobile communications have appeared around every ten years. However, the individual systems are continuously evolving with new features. The 3GPP specifications are divided into releases, where each release consists of a complete and self-contained set of specifications. This means that a particular release contains all components needed to build a complete cellular network, not just the newly added features. When a release is completed, the features are frozen and ready for implementation. When a release has been frozen, only essential corrections are permitted. Further functionalities will have to go into the next release. The work on different releases has some overlap, so that the work on a new release starts before the completion of the current release. The releases should be backwards compatible so that a user equipment (UE) developed for one release can also work in a cell that has implemented a previous release. The first version of LTE was part of release 8 of the 3GPP specifications. LTE release 10 was named *LTE-Advanced* since it was approved as an IMT-Advanced technology by ITU-R. In release 13, the marketing name for LTE changed to *LTE-Advanced Pro*.

The everyday 3GPP work is divided into study items and work items. Study items are feasibility studies of concepts, where the results are documented in a technical report (TR). The details of agreed concepts are worked out in the work items, where the features are defined and end up in a technical specification (TS). The 3GPP technical specifications (TSs) are organized in series and are numbered TS XX.YYY, where XX defines the series. The radio aspects of NR are defined in the 38-series. All 3GPP specifications are publicly available at www.3gpp.org.

Organizationally, 3GPP consists of three technical specification groups (TSGs) where TSG radio access network (RAN) is responsible for the radio access specifications. TSG RAN in turn consists of six working groups (WGs), where RAN WG1 deals with the physical layer specifications. The WGs meet regularly and come together four times a year in TSG plenary meetings. The decisions in 3GPP are based on consensus among the members.

Fig. 1.4 illustrates the ITU and 3GPP time lines as well as the expected development of commercial equipment for 5G. The standardization work of 5G NR started in 3GPP in April 2016, with the aim of making it commercially available before 2020. 3GPP is taking a phased approach in defining the 5G specifications. A first standardization phase, with limited NR functionality, NR release 15, was completed for non-stand-alone (NSA) operation at the end of 2017 and for stand-alone (SA) operation in mid-2018. The second standardization phase, NR release 16, is expected to meet all the requirements of IMT-2020 and be completed by 2019.

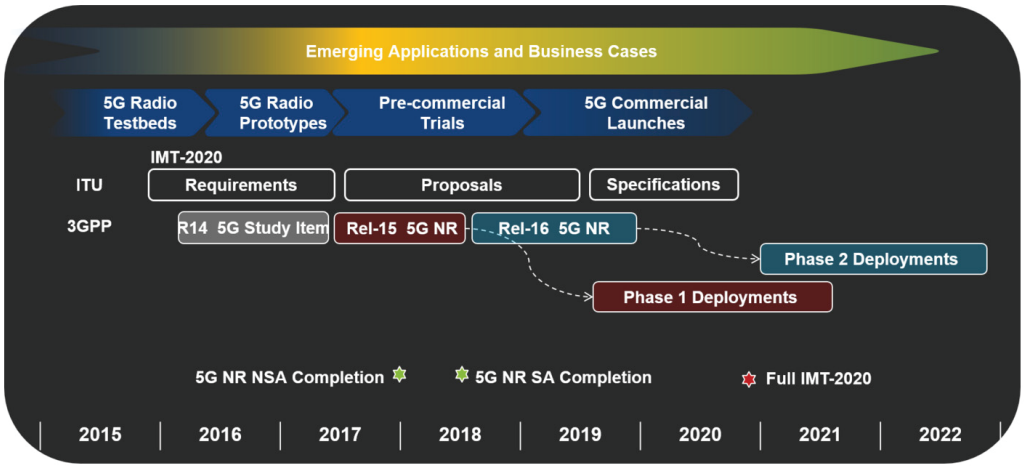


FIGURE 1.4

5G global view.

With 5G standardization in acceleration, 5G precommercial trials are also taking place worldwide. Based on the release 15 specifications, 3GPP-compliant base stations and devices are already under development. The commercial deployments are then expected in two phases. The first phase of NR commercial deployments is expected in 2019; it is based on the release 15 specifications. The second phase of NR commercial deployments is expected to start in the 2021 time frame, based on the release 16 specifications. It is likely that the NR specifications will continue to evolve in 3GPP beyond 2020, with a sequence of releases including additional features and functionalities.

1.3.2 SPECTRUM FOR 5G

One of the main changes when going from previous generations of mobile communications to the fifth generation is the spectrum use at radically higher frequencies in the millimeter-wave range. 3GPP has decided to support the range from below 1 GHz up to 52.6 GHz already from the first releases of NR [12]. A main reason for this change is the availability of large amounts of spectrum with very large

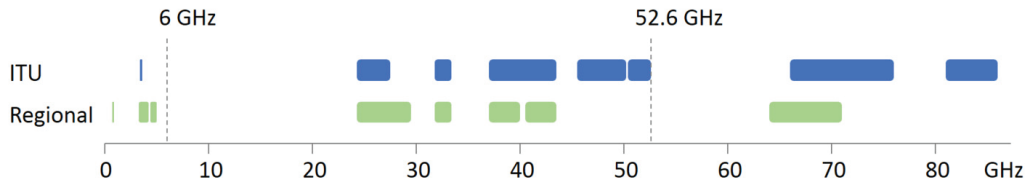


FIGURE 1.5

Global (ITU) and regional spectrum identifications for 5G.

bandwidths in the range of several GHz. Though the millimeter-wave spectrum seems very attractive, there are also many challenges:

- The transmission loss goes up substantially if multiantenna and beam-forming techniques are not used.
- RF hardware performance regarding e.g. phase noise and output power is degraded.
- There is no non-utilized spectrum, meaning that coexistence, with e.g. satellite systems, will be required where acceptable interference levels have to be guaranteed.

As a consequence, 5G is designed by 3GPP (NR) for flexible frequency use over the full frequency range. Joint operation at lower frequencies and higher frequencies is supported in order to provide both reliable coverage (utilizing e.g. frequencies below 6 GHz) and very high capacity and bitrates when possible (where millimeter-wave coverage is available).

An overview of the spectrum identified for 5G is given in Fig. 1.5. The 3300–3600 MHz band was identified as a global IMT band by ITU-R at world radiocommunication conference (WRC)-15.[1] The corresponding global allocation above 6 GHz is subject for WRC-19. There is a substantial additional regional spectrum identified below 6 GHz. Above 6 GHz the regional spectrum is largely overlapping with the global bands except for around 28 GHz and 65 GHz. It is evident that the amount of spectrum below 6 GHz is scarce and that spectrum above 6 GHz will be needed for fulfilling the requirements of 5G. For early deployments of 5G, bands below 6 GHz at 600–700 MHz, 3300–4200 MHz, and 4400–5000 MHz have drawn attention [11]. Except the 3300–3600 MHz band which was identified as a global IMT band at WRC15 these bands are subject to regional regulations.

In Europe the focus for early deployments is put on the 3600–3800 MHz band [3]. Also the United States, Japan, South Korea and China use different allocations in the 3300–4200 MHz band in their planning for early 5G systems, and 3GPP is developing a specification for this band.

The 4400–5000 MHz band is mainly promoted by China and Japan but could potentially also be adopted by other countries in the Asia–Pacific region.

Though the bandwidth is limited in the 600–700 MHz band, it is useful for providing coverage as the transmission losses are low without requiring any advanced multiantenna transmission techniques. There are already allocations for LTE at 700 MHz which could be migrated to 5G in 2020 or later. Moreover, the United States has identified 614–698 MHz for mobile use.

In the process of allocation of IMT frequencies above 6 GHz it was decided at WRC-15 to study 11 bands in the range 24.5 GHz to 86 GHz for decisions at WRC-19. ITU-R has estimated the maximum

[1] A WRC is held every 3–4 years where regulations are revised for the RF spectrum use across the world.

need of spectrum for 5G to be up to 20 GHz [8]. As the current situation is that most spectrum is already allocated, this amount of spectrum can hardly be allocated for IMT primary use, and coexistence will be required. For this reason, a crucial part of the spectrum studies prior to WRC-19 is to assess whether the different bands are suitable with respect to coexistence with incumbent systems, such as satellites.

The interference between IMT and other systems on earth, in air or in space, is mainly of long range. For this reason, mainly the aggregation of multiple IMT transmitters is of interest. In this case the specific pointing directions and antenna gain is averaged out, providing an effective omni-directional radiation. For base stations, however, the radiation in elevation angle is focused in down-tilted directions.

In order to accurately determine interference signals between IMT and other systems, proper propagation models are required. Such models are provided by the ITU-R study group 3 recommendations. There are earlier recommendations (P.619, P.2041, P.1409, P.452, P.2001, etc.) which are valid for atmospheric propagation paths and paths over the earth profile and for the full range of frequencies under consideration by WRC-19. However, two important cases have been missing for modeling of additional losses due to clutter (vegetation and man made structures like buildings) and penetration into buildings. In March 2017 ITU-R study group 3 succeeded in providing proper models for these cases [10], [9]. The corresponding urban clutter loss is up to 50 dB (median) for low elevation angles. For higher elevation angles the clutter loss may be very low, down to 0 dB, as no buildings obstruct the sky. However, for these cases the building penetration loss for indoor transmitters is in the range 40 dB to 60 dB (median) at 30 GHz. As 5G deployments are expected to be highly concentrated in urban environments below roof-top and indoors, the urban clutter and building penetrations losses greatly improve the prospects for IMT coexistence with incumbent systems.

In addition to the global ITU process there have been a lot of regional efforts for finding a suitable spectrum for 5G. The United States, Korea, Canada, Japan, Singapore, and Sweden are focusing on bands in the 26.5–29.5 GHz range. The United States is also targeting bands in the 37–40 GHz range and Canada is considering the very high end of frequencies in the 64–71 GHz band. Both Europe and China have identified bands in the range 24.25–27.5 GHz. In addition China is looking at 37–42.5 GHz, whereas Europe considers the ranges 31.8–33.4 GHz and 40.5–43.5 GHz. The 5G spectrum availability around the world is summarized in Fig. 1.6.

Except the obvious need for contiguous bands of large bandwidths, the allocation of spectrum for 5G will crucially depend on the frequency dependent performance of RF hardware, propagation channel and multiantenna techniques. All these aspects are addressed in depth in Chapter 3, Chapter 4, and Chapter 7.

1.3.3 USE CASES FOR 5G

Differently from its predecessors, 5G has the goal of providing optimized support for a number of drastically different types of services and user requirements. Conforming to the ITU nomenclature for international mobile communications for 2020 and beyond (IMT-2020) [7], 5G will target three use case families with very distinct features: enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable low-latency communications (URLLC). We will next describe the features of these three use case categories, which are depicted in Fig. 1.7

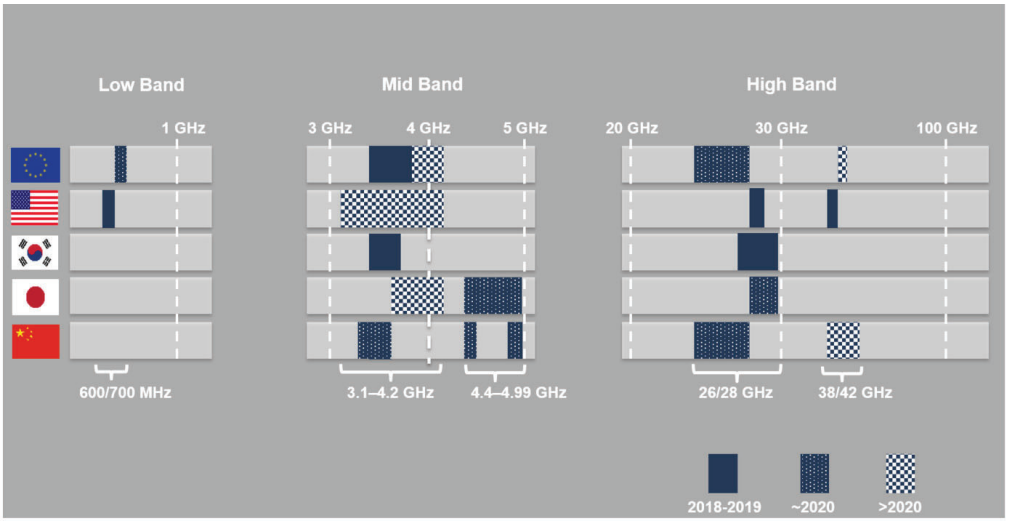


FIGURE 1.6

5G spectrum availability around the world.

eMBB:

This use case category is the natural extension of the classic mobile broadband connectivity scenario under current mobile telecommunication standards. It addresses human-centric connectivity, including access to multimedia content, services, and data. This is done by providing the high data rates that are required to support future multimedia services and the increasing traffic volume generated by these services. The eMBB use case covers a range of scenarios including:

- *Hotspot connectivity*, which is characterized by a high user density and extremely high data rates, and low mobility.
- *Wide-area coverage*, where the user density and data rates are lower, but the mobility is higher.

URLLC:

Stringent requirements on both latency and reliability are the distinctive features of this use case category, which targets mainly machine-type communications (MTC). Envisaged applications include wireless control of industrial manufacturing and production processes, remote medical surgery, driverless and/or remotely driven vehicles, and distribution automation in smart grids.

mMTC:

The growth of the Internet of Things (IoT) is causing a proliferation of wirelessly connected devices carrying MTC traffic. Indeed, the number of such devices is expected to exceed soon the number of devices carrying human-generated traffic. The focus of the mMTC is on providing connectivity to a massive number of devices, which are assumed to transmit sporadically a low amount of traffic, which is not delay critical. The mMTC devices are expected to have a very long battery lifetime to allow for remote deployments. A unique feature of this use case is that the MTC devices will be extremely heterogeneous in terms of capabilities, cost, energy consumption, and transmission power.

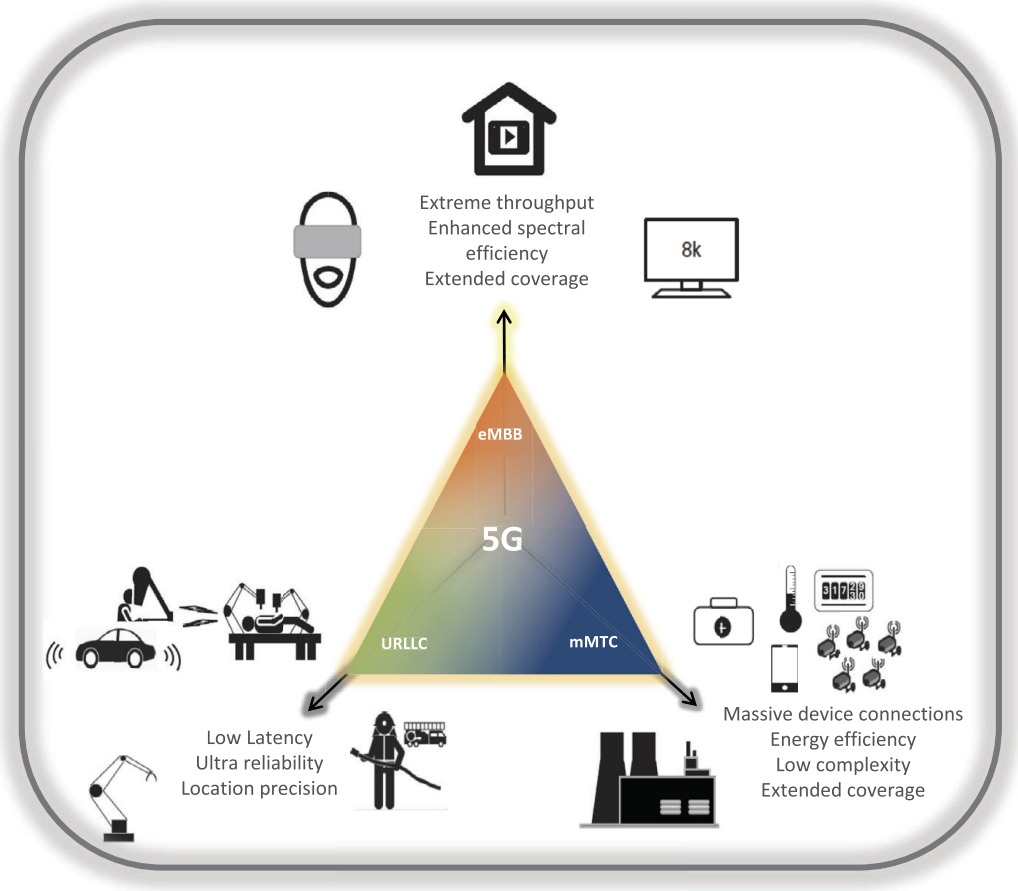


FIGURE 1.7

The three 5G use cases, according to the ITU nomenclature, and their main features.

Fig. 1.8 provides a graphical summary of some of the applications targeted by 5G, and their relation to the IMT-2020 use cases just reviewed.

The ITU has set key performance requirements for IMT-2020 to be able to address satisfactorily the specific needs of eMBB, URLLC, and mMTC traffic. Such requirements are summarized in Fig. 1.9, where a comparison with the key capabilities of the previous mobile generation—IMT-advanced—is provided. The requirements include a peak data rate of 20 Gbit/s, a latency below 1 ms, and the capability to support a connection density of 10^6 devices per square kilometer. As exemplified in Tables 1.1–1.2, NR is being designed to satisfy (in some cases, with large margins) all these requirements [1].

1.3.4 5G FIELD TRIALS

In the following, we will briefly present some of the world's most exciting 5G trials, which show the potential of 5G to unlock new and exciting opportunities. Some pictures from these trials are shown in Fig. 1.10.

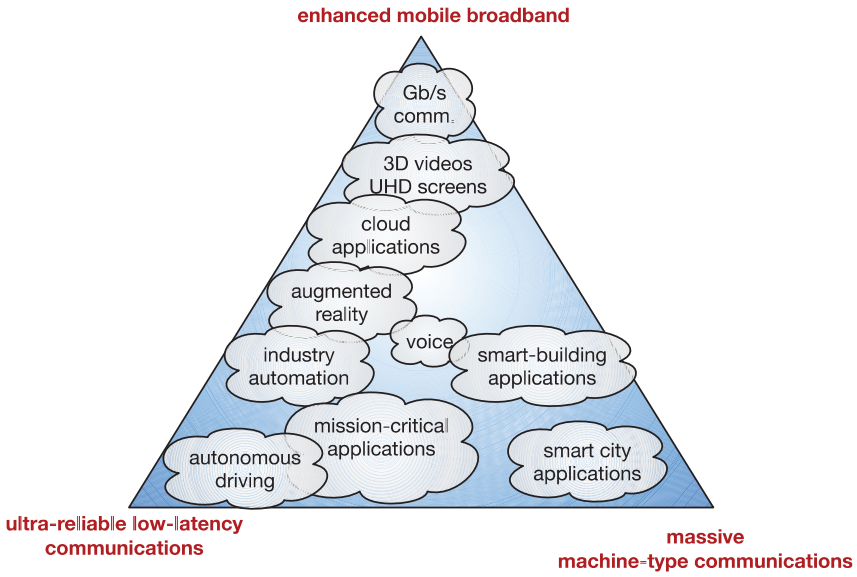


FIGURE 1.8

The three IMT-2020 use cases and some targeted applications.

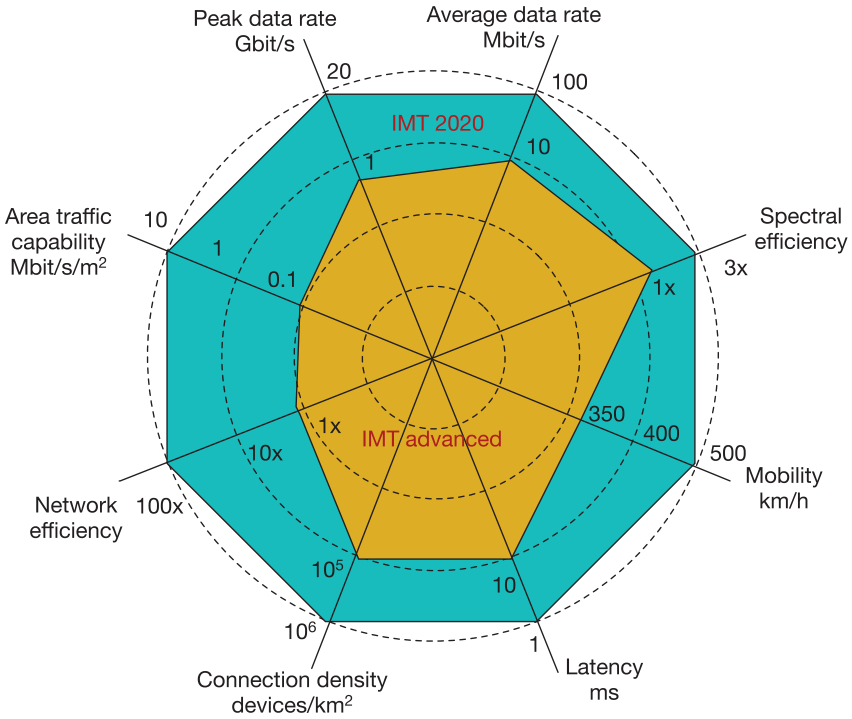


FIGURE 1.9

The key capabilities of IMT-2020 in relation to the ones of its predecessor, IMT-advanced.

Table 1.1 A subset of the performance requirements for high-data-rate and high-traffic-density scenarios. The complete list can be found in [1, Table 7.1-1]

Scenario	Data-rate DL	Data-rate UL	User density	Mobility
Urban macro	50 Mbit/s	25 Mbit/s	10^5 UEs per km^2	Up to 120 km/h
Indoor hotspot	1 Gbit/s	500 Mbit/s	25×10^4 per km^2	Pedestrian
High-speed train	50 Mbit/s	25 Mbit/s	10^3 per train	Up to 400 km/h

Table 1.2 A subset of the performance requirements for low-latency and high-reliability scenarios. The complete list can be found in [1, Table 7.2.2-1]

Scenario	Latency	Error probability	Data rate	Connection density
Motion control	1 ms	10^{-6}	Up to 10 Mbit/s	10^5 UEs per km^2
Transport system infrastructure backhaul	10 ms	10^{-6}	10 Mbit/s	10^3 UEs per km^2
Electricity distribution	25 ms	10^{-3}	10 Mbit/s	10^3 UEs per km^2

5G for flying drones: Ericsson and China Mobile successfully performed the world's first 5G trial for flying drones on a commercial network in 2016, with 5G enabled technologies on a cellular network [14]. A drone (an unmanned aerial vehicle) was flown with handovers across multiple sites. In order to demonstrate the concept's validity in a real-world setting, the handovers were performed between sites that were simultaneously in use by commercial mobile phone users. The applications of drones are emerging, for example, in agriculture, public safety, search and rescue, inventory management, and goods delivery.

5G connectivity for high-speed cars: Ericsson and Verizon tested the limits of 5G by demonstrating 6+ Gbps throughput and super-low latency inside a moving car in 2017 [20]. The car was completely backed out and the driver operated the car while wearing a set of virtual reality glasses, relying solely on video captured from a camera on the hood of the car. The trial showcased the massive potential of multiantenna technologies with beams tracking a high-speed car. This shows that 5G is capable of supporting 360-degree 4K video streaming with very low latencies, even in high mobility scenarios. This would, for example, give us the possibility to watch live sporting events with virtual reality.

A similar trial was also conducted by Ericsson, SK Telecom and BMW Group on a racetrack in Yeonjong-do, South Korea [22]. The test network consisted of four radio transmission points from Ericsson operating in the 28 GHz band at BMW Group Korea's driving center. Under this setup, test results achieved driving speeds of 170 km/h, while reaching downlink data speeds of 3.6 Gbps. This was enabled by advanced beam-forming and beam tracking, which allows the base station to transmit signals that follow the UEs.

5G connecting ships and ports: Tallink, Telia, Ericsson, and Intel have created a 5G test and exploration area at the Port of Tallinn in Estonia. Ericsson has set up a 5G base station in the port of Tallinn and Intel has placed equipment inside the ferries to receive signals. The 5G trial network is delivering high-speed internet connectivity to the commercial passenger cruise ships and their passengers while in the port area since 2017 [19].

5G delivering tens of Gbps worldwide: In 2016, Ericsson and Korea Telecom (KT) demonstrated the world's first 5G achieving 25.3 Gbps throughput [16]. The trial was performed in millimeter-wave frequency bands for 5G mobile communications. (KT was aiming to provide 5G trial services for the

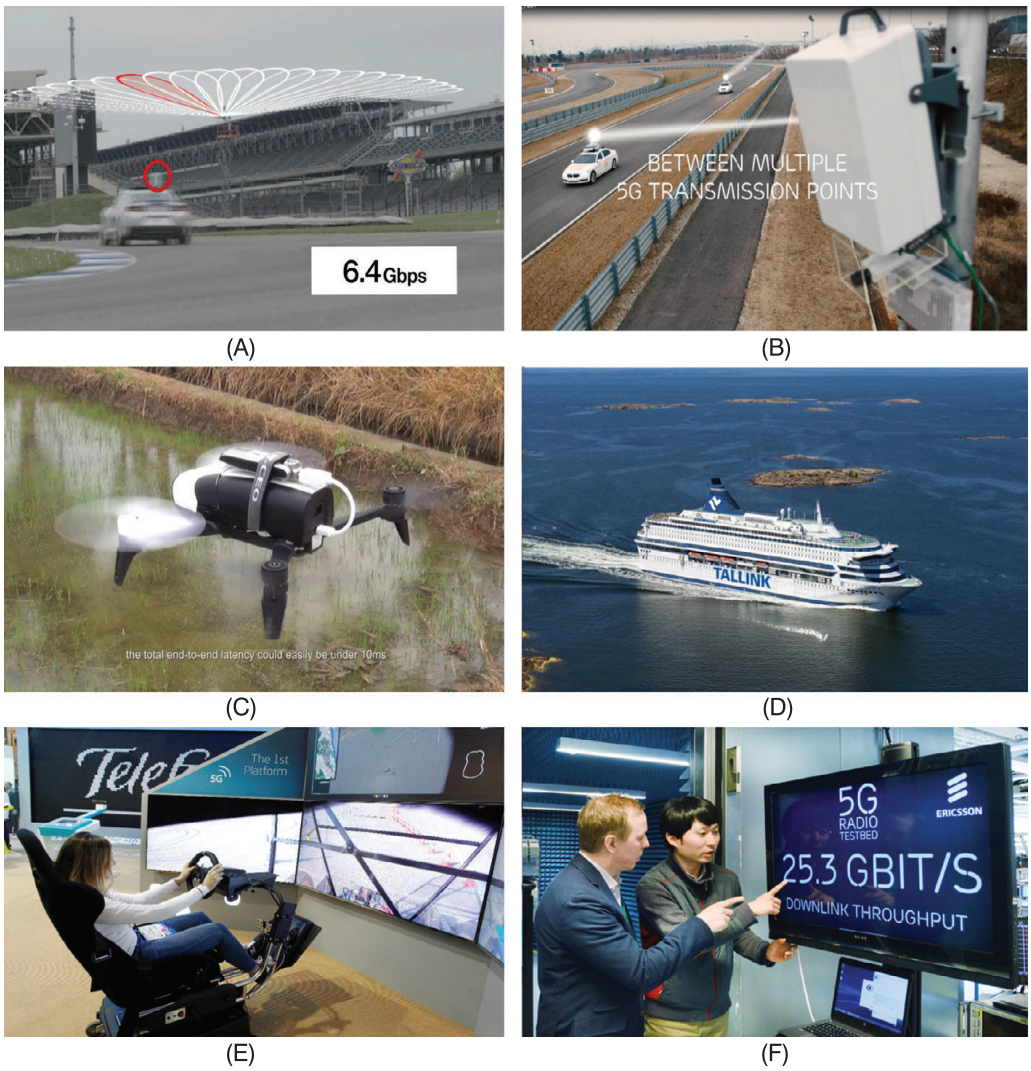


FIGURE 1.10

5G trials. (A) 5G connectivity for high-speed cars (Ericsson, Verizon); (B) 5G connectivity for high-speed cars (Ericsson, SK Telecom); (C) 5G for flying drones (Ericsson, China Mobile); (D) 5G connecting ships and ports (Ericsson, Telia, Tallink Grupp, Intel); (E) 5G for remotely driving cars (Ericsson, Telefonica); (F) 5G delivering tens of Gbps (Ericsson, Korea Telecom).

PyeongChang 2018 winter games.) In the same year (2016), Ericsson and Telstra conducted Australia's first live 5G trial, achieving 20+ Gbps and a latency of half of what is seen in Telstra's 4G networks [15]. The trial used 800 MHz of spectrum, which is 10 times more than what Telstra currently uses for its 4G networks.

Ericsson and Mobile Telesystems (MTS) have built a prototype 5G network in Moscow and successfully achieved 25 Gbps in 2017, with the use of advanced antenna systems on 5G base stations [21].

In January 2018, Mobile Telephony Network Group (MTN) and Ericsson performed the first 5G trial in Africa and achieved a throughput of more than 20 Gbps with less than 5-ms latency. This was the best performance ever achieved on a mobile network in Africa. The 5G trial was based on Ericsson's 5G prototype radios [23].

5G making remote driving a reality: Telefonica, Ericsson, the Royal Institute of Technology (KTH), and Applus Idiada developed a revolutionary demo to showcase the world's first 5G remote driving of a car on Applus Idiada race track during the Mobile World Congress 2017 [17,18]. A 5G base station provided connectivity to a car at 15 GHz carrier frequency. The driver in a remote location gets an in-car experience with 4K video streaming in the uplink. In the downlink, the 5G connectivity provides ultra-low latency and high reliability to communicate driving decisions to the car. One possible application of remote driving is remote parking, where a driver can leave a car in a drop-off zone and request assistance from a remote driver to ensure safe and efficient parking.

1.3.5 5G COMMERCIAL DEPLOYMENTS

The deployment of a 5G NR network should start from an existing 4G LTE network, which already has good coverage [5]. As shown in Fig. 1.3, NR can coexist and interwork with LTE, which reduces the time to market for NR. Commercial deployment of the 5G NR network is expected to occur in two phases:

1. The first phase is a non-standalone (NSA) mode, with tight interworking between 5G NR and LTE, as shown in Fig. 1.3. This option uses LTE as the control-plane[2] anchor for NR, and it uses either LTE or NR for user traffic (user-plane). The first 5G specification (3GPP Release 15, NSA) was finalized at the end of 2017. Fully standard-compliant radio systems are likely to be available at the end of 2018. We expect to see 5G NR to start being deployed during 2019.
2. The second phase is a standalone (SA) mode, with both control-plane and user-plane existing over 5G NR. The 5G specification for the SA mode was released in mid-2018. The commercial deployments for the SA mode are expected beyond 2020.

NR will need to coexist and interwork with LTE for many years to come, not only as a way of reducing the time to market but also ensuring good coverage and mobility. With 5G NR deployments, new use cases will pick up and 5G devices will come to market. We see the deployment of standalone NR beginning when devices are more widely available, new use cases (e.g., ultra-reliable and low-latency communications, industrial IoT) start to gain momentum, and NR has greater access to both new and existing spectrum. Ultimately, 5G NR is likely to become the mainstream cellular technology used to address multiple use cases across multiple industries. The evolution path of 5G NR deployments is sketched in Fig. 1.11.

There are three main 5G deployment scenarios considering that LTE operates below 6 GHz (i.e., low or mid frequency band) and NR can operate up to 100 GHz (i.e., in low, mid, and high bands) [5]:

1. **NR NSA in mid/low with LTE in low/mid band** In this deployment scenario, NR and LTE will have similar coverage per base station, since they are both deployed in similar frequencies. The

[2]The control-plane is mainly responsible for control signaling for connection setup, mobility (handovers), and security.

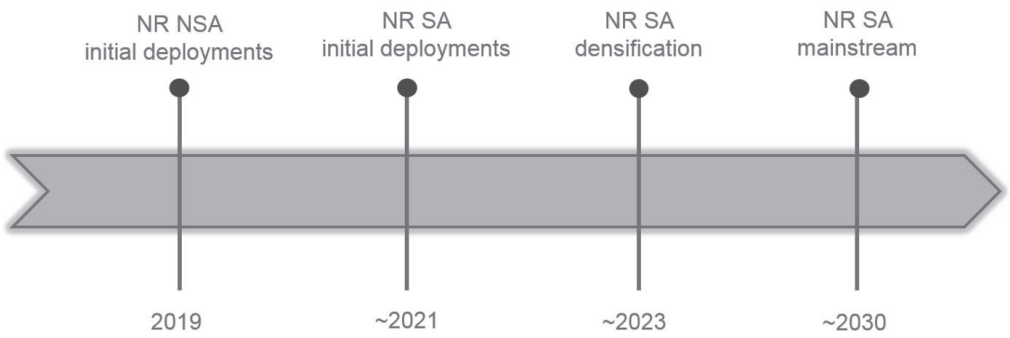


FIGURE 1.11

The evolution path of 5G NR deployments.

greater capacity and throughput enabled by the 5G NR spectrum implies that this is a more likely way to deployment for high user density areas, typically in city and urban areas.

2. **NR NSA in high band with LTE in low/mid band** In this deployment scenario, NR coverage area per base station will be smaller than LTE coverage area, due to higher transmission losses at higher frequencies. The main driver for this deployment is to enable high capacity and extreme throughput cells with wider channel bandwidths available in millimeter-wave frequencies. In addition to eMBB, fixed wireless access is an emerging use case for this deployment category.
3. **NR SA in mid/low/high band** NR standalone will be deployed in all frequency bands, depending on the spectrum availability. In addition to eMBB, standalone NR deployments are very likely to be used for private or enterprise networks to support industrial applications (for example, manufacturing). Although NR SA deployments do not rely on LTE for its operation, many NR SA deployments are likely to reuse the existing 4G physical infrastructure and transport network, thereby reducing the cost of deployment. NR NSA in low/mid bands will be suitable for massive IoT applications. Deployment of 5G NR in low bands is of interest to boost coverage. Deployments in high bands are more likely to be used for high-traffic areas and for private (industrial) IoT networks.

Although deployment for cost-efficient and well-performing 5G NR networks is crucial for 5G success, it is the availability of NR capable devices and their market uptake that will eventually determine growth driven by 5G NR connectivity. Fig. 1.12 shows approximate timing of NR capable devices. Early fixed wireless access devices are already available in some regions (e.g., in USA) for operation in high bands. The first 3GPP-compliant 5G smart-phones and tablets are likely to be launched in 2019. New NR capable devices for IoT are expected to be introduced in 2020 and beyond. For many IoT use cases, the cost of NR capable devices might have to significantly go down below eMBB devices in order to allow mass adoption. However, for critical IoT applications requiring high-reliability and very low-latency connections, device cost is not expected to be a barrier.

The United States will be among the first to experience 5G commercial services [24]. The major operators in the US have already announced that they will start providing 5G services between late 2018 and mid-2019. South Korea, Japan, and China are expected to be among early 5G service providers with significant volumes of 5G subscriptions. Ericsson forecasts over 1 billion 5G mobile broadband subscriptions by the end of 2023, accounting for 12 percent of all mobile subscriptions.

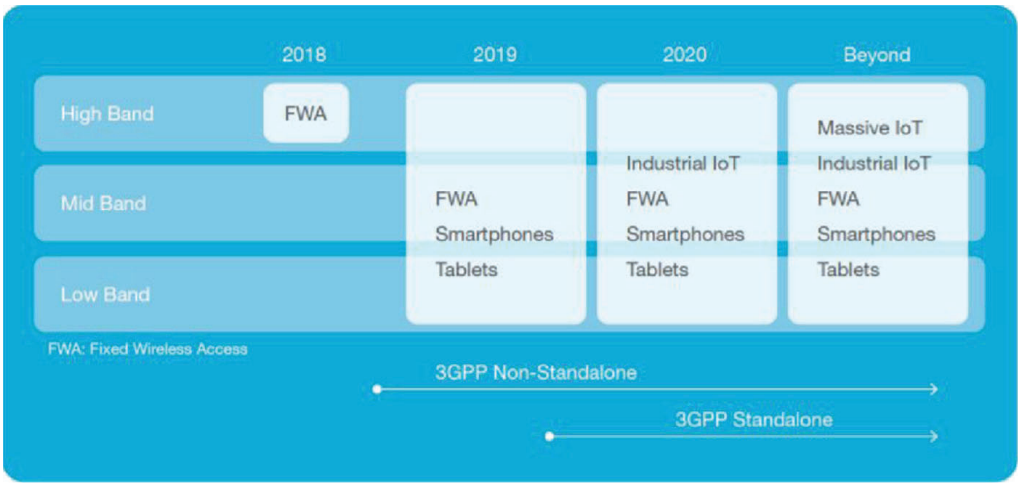


FIGURE 1.12

5G device availability (Source: Ericsson [5]).

1.4 PREVIEW OF THE BOOK

This book is written for researchers and system designers in the field of mobile radio communication, interested in understanding principles, models, and technology components for the 5G NR physical layer. Although the focus is on 5G NR, many concepts presented in the book are of a fundamental nature and are applicable beyond 5G. We assume that the reader has a basic understanding of digital wireless communications and signal processing; however, familiarity with cellular technologies (for example, 4G LTE standard) is not required. We will introduce relevant standard related concepts and terminologies in the book.

A preview of the book is sketched in Fig. 1.13, highlighting key aspects covered in various chapters. The book is composed of nine chapters covering various aspects of 5G NR—a global picture of its development, the physical layer overview based on the first 5G NR release in 3GPP, the physical limitations imposed by radio wave propagation and hardware impairments, the key physical layer technologies, and an open-source link-level simulator. In the following, we briefly outline the content of each chapter.

Chapter 1 introduces 5G NR and discusses global efforts in the development of 5G NR and its future impact on industry and society. We provide a holistic view on 5G use cases and their requirements, spectrum allocations, standardization, field trials, and future commercial deployments.

Chapter 2 provides an overview of the 5G NR physical layer based on the first 3GPP NR release. We will see that the physical layer components of NR are flexible, ultra-lean and forward-compatible. Moreover, we provide an overview of radio wave propagation and hardware impairment related challenges associated with enabling a high performing NR.

Chapter 3 presents state-of-the-art insights on radio wave propagation along with description of fundamental concepts and propagation characteristics. We focus on the frequency dependency of the channel properties for the full range of frequencies envisioned for 5G NR, with experimental examples.

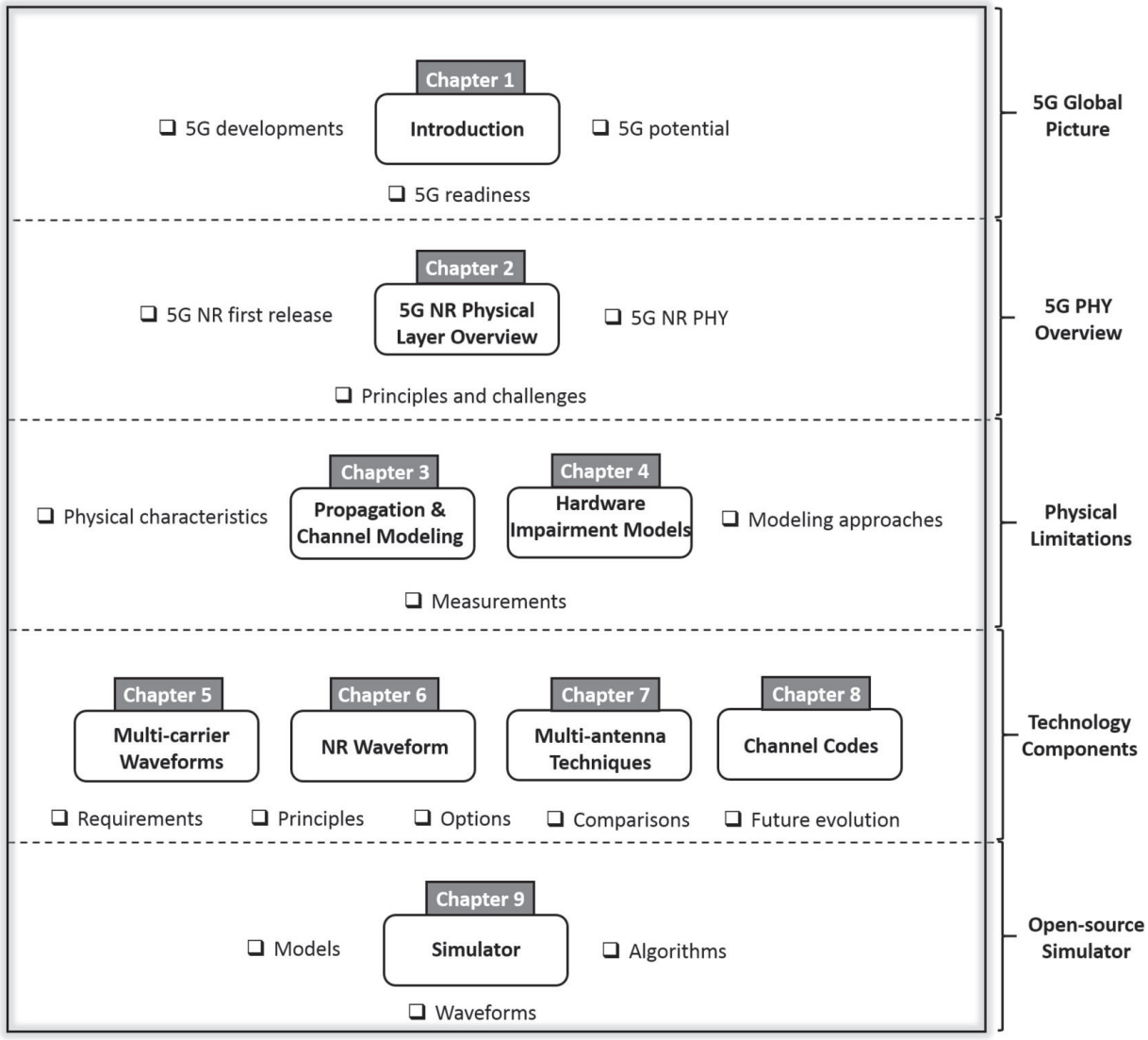


FIGURE 1.13

Preview of the book.

The channel modeling for 5G NR is discussed. Moreover, we point out both validated and non-validated (or deficient) aspects of the current 5G channel models defined by 3GPP and ITU-R.

Chapter 4 covers some of the traditional behavioral models for power amplifiers, local oscillators and data converters. These models may accurately predict the input–output relation of analog and mixed-signal components. Furthermore, a novel modeling approach is presented that provides the second order statistics of the errors caused by non-ideal components. This stochastic modeling framework provides a powerful tool for link-level evaluations and aids in making sound choices in terms of the radio performance versus energy efficiency trade-offs.

Chapter 5 presents state-of-the art multicarrier waveforms. Based on the design requirements for NR, the chapter provides an overall waveform comparison that has led to the down selection of CP-OFDM waveform for 5G NR. The multicarrier waveforms are compared as regards a number of key performance indicators: phase-noise robustness, baseband complexity, frequency localization, time localization, robustness to power amplifier nonlinearities, channel time selectivity and channel frequency selectivity.

Chapter 6 presents a flexible OFDM for 5G NR. Different factors involved in the implementation of OFDM-based NR modems are discussed, for example, quality of service requirements, type of deployment, carrier frequency, user mobility, hardware impairments, and implementation aspects. This chapter puts special focus on high carrier frequencies (e.g., millimeter-wave band), where robustness to hardware impairments (phase noise, synchronization errors) and power efficiency of the waveform is crucial.

Chapter 7 discusses the role of multiantenna techniques in 5G NR and the features included in the first release of the NR specifications. To provide an understanding and motivation of the features adopted by NR, the fundamental theory behind these features is provided. The viability of the multiantenna techniques presented is illustrated by several experimental examples.

Chapter 8 presents different channel coding schemes for 5G NR. The performance of the coding schemes is evaluated for different blocklength values. We review recently developed information-theoretic tools to benchmark the performance of these coding schemes. Looking beyond what is currently standardized in NR, we consider transmissions over multiantenna fading channels and highlight the importance of exploiting frequency and spatial diversity through the use of space-frequency codes, in applications that require high reliability.

Chapter 9 presents an open-source simulator that includes hardware impairment models (power amplifier, oscillator phase noise), a geometry-based stochastic channel model, and modulation/demodulation modules of state-of-the art waveforms. The chapter provides simulation exercises with various waveforms subject to different types of impairments.

REFERENCES

[1] 3GPP TS 22.261, 3rd generation partnership project: technical specification group services and system aspects; service requirements for the 5G systems; stage 1; (release 16), V16.2.0, 2017.
[2] E. Dahlman, G. Mildh, J. Peisa, J. Sachs, Y. Selén, S. Parkvall, 5G radio access, Ericsson Technology Review (2014, July), https://www.ericsson.com/assets/local/publications/ericsson-technology-review/docs/2014/er-5g-radio-access.pdf.
[3] ECC, Operational guidelines for spectrum sharing to support the implementation of the current ECC framework in the 3600–3800 MHz range, ECC Report 254, 2016.
[4] Ericsson, The 5G business potential, 2017, Oct.
[5] Ericsson, 5G deployment considerations, 2018.
[6] Ericsson White Paper, 5G systems, 2017, Jan.
[7] ITU-R, Recommendation ITU-R M.2083-0: IMT vision–framework and overall objectives of the future development of IMT for 2020 and beyond, 2015.
[8] ITU-R, Liaison statement to task group 5/1-spectrum needs and characteristics for the terrestrial component of IMT in the frequency range between 24.25 GHz and 86 GHz, ITU-R, WP 5D, Doc. TG5.1/36, 2017.
[9] ITU-R, Prediction of building entry loss, Recommendation ITU-R P.2109, 2017.
[10] ITU-R, Prediction of clutter loss, Recommendation ITU-R P.2108, 2018.
[11] J. Lee, E. Tejedor, K. Ranta-aho, H. Wang, K.T. Lee, E. Semaan, E. Mohyeldin, J. Song, C. Bergljung, S. Jung, Spectrum for 5G: global status, challenges, and enabling technologies, IEEE Communications Magazine 56 (3) (2018) 12–18.

[12] S. Parkvall, E. Dahlman, A. Furuskär, M. Frenne, NR: the new 5G radio access technology, IEEE Communications Standards Magazine (2017, Dec.) 24–30.
[13] O. Teyeb, G. Wikström, M. Stattin, T. Cheng, S. Fáxer, H. Do, Evolving LTE to fit the 5G future, Ericsson Technology Review (2017, Jan.), https://www.ericsson.com/assets/local/publications/ericsson-technology-review/docs/2017/etr_evolving_lte_to_fit_the_5g_future.pdf.
[14] Ericsson and China Mobile conduct world's first 5G drone prototype field trial, Ericsson, 2016, https://www.ericsson.com/en/news/2016/8/ericsson-and-china-mobile-conduct-worlds-first-5g-drone-prototype-field-trial-.
[15] Ericsson and Telstra conduct the first live 5G trial in Australia, Ericsson, 2016, https://www.ericsson.com/en/news/2016/9/ericsson-and-telstra-conduct-the-first-live-5g-trial-in-australia.
[16] KT demonstrates world's first 5G performing at 25 Gbps, Business Korea, 2016, http://www.businesskorea.co.kr/news/articleView.html?idxno=13896.
[17] 5G can make remote driving a reality, Telefónica and Ericsson demonstrate at MWC, Telefonica, 2017, https://www.telefonica.com/en/web/press-office/-/5g-can-make-remote-driving-a-reality-telefonica-and-ericsson-demostrate-at-mwc.
[18] 5G can make remote driving a reality, Telefónica and Ericsson demonstrate at MWC, Ericsson, 2017, https://www.ericsson.com/en/press-releases/latin-america/2017/2/5g-can-make-remote-driving-a-reality-telefonica-and-ericsson-demonstrate-at-mwc.
[19] 5G goes live in the Port of Tallinn, Ericsson, 2017, https://www.ericsson.com/en/news/2017/9/5g-goes-live-in-the-port-of-tallinn.
[20] The 5G race is on, Ericsson, 2017, https://www.ericsson.com/en/news/2017/5/ericsson-and-verizon-test-the-limits-of-5g.
[21] Ericsson and MTS test new 5G features, Ericsson, 2017, https://www.ericsson.com/en/press-releases/2017/4/ericsson-and-mts-test-new-5g-features.
[22] Ericsson, SK Telecom and BMW Group Korea reach new world record speed with 5G, Ericsson, 2017, https://www.ericsson.com/en/news/2017/2/ericsson-sk-telecom-and-bmw-group-korea-reach-new-world-record-speed-with-5g.
[23] MTN and Ericsson first in Africa to trial 5G technology, Ericsson, 2018, https://www.ericsson.com/en/news/2018/1/5g-in-south-africa.
[24] Ericsson Mobility Report, June 2018, https://www.ericsson.com/assets/local/mobility-report/documents/2018/ericsson-mobility-report-june-2018.pdf.

CHAPTER

NR PHYSICAL LAYER: OVERVIEW 2

Like for any wireless technology, the physical layer forms the backbone of 5G NR. The NR physical layer has to support a wide range of frequencies (from sub-1 GHz to 100 GHz) and various deployment options (pico cells, micro cells, macro cells). There are human-centric and machine-centric use cases with extreme and sometimes contradictory requirements. There may also be unforeseen applications with new requirements in the future. To successfully address these challenges, 3GPP is developing a flexible physical layer for NR. The flexible components can be properly optimized with an accurate understanding of radio wave propagation and hardware imperfections in networks and devices. This is a challenge, because these characteristics are less understood. NR is the first ever mobile radio access technology going into millimeter-wave frequency range (with frequencies as high as 100 GHz), targeting channel bandwidths in the GHz range, and enabling massive multiantenna systems.

The first release of NR (3GPP NR Release 15) was completed in June 2018. Any future releases of NR will be backwards compatible to its first release. This is usually referred to as forward compatibility of NR, that is, NR will be developed in such a manner that any future releases of NR will be backwards compatible to its initial release(s). For the reader interested in understanding what an NR physical layer is, in this chapter we provide an overview of the NR physical layer (based on the first NR release) and discuss the radio wave propagation and hardware impairment related challenges associated with enabling the NR physical layer. Our focus in the following chapters will be on fundamental principles, models, and technology components for the NR physical layer. If the reader is interested in the details of the NR specification, we strongly recommend [6].

The chapter is organized as follows. In Section 2.1, we briefly describe the protocol stack of the NR radio interface and the role of the physical layer therein. Due to brevity, this section may be hard to fully grasp if the reader is not familiar with 3GPP-based cellular technologies (e.g., 4G LTE [5]). The rest of the chapter focuses only on physical layer aspects. Section 2.2 gives an overview of the key physical layer technology components of NR—modulation, waveform, multiantenna, and channel coding schemes. These technology components are explained in detail in Chapters 5–8. Section 2.2 introduces the physical time-frequency resource structure of NR. Sections 2.4 and 2.5 describe how the time-frequency resources are allocated to different types of signals. Sections 2.6 and 2.7 explain flexible duplexing schemes and a flexible transmission structure (frame structure) available in NR. Finally, Section 2.9 briefly summarizes some challenges associated with radio wave propagation and hardware impairments. This section motivates Chapters 3 and 4, which, respectively, cover radio wave propagation and hardware impairments in depth.

2.1 RADIO PROTOCOL ARCHITECTURE

In 3GPP terminology, a base station is an implementation of a logical radio access network node. For example, in 3G UMTS and 4G LTE, the network node is termed Node B (NB) and evolved Node B

5G Physical Layer. https://doi.org/10.1016/B978-0-12-814578-4.00007-2
Copyright © 2018 Elsevier Ltd. All rights reserved.

Samsung Exhibit 1020, Page 35 of 314

(eNB), respectively. A 5G NR radio access network node has been named the next generation Node B (gNB) by 3GPP. It is important to stress that gNB is a logical entity and not a physical implementation of a base station. A base station can be realized in different ways based on a standardized gNB protocol. Similarly, any device is referred to as a UE in 3GPP specifications.

The radio protocol architecture for NR can be separated into control-plane architecture and user-plane architecture. The user-plane delivers user data, whereas the control-plane is mainly responsible for connection setup, mobility, and security. Fig. 2.1 illustrates the user-plane protocol stack of NR. The protocol is split into the following layers: physical (PHY) layer, medium access control (MAC) layer, radio link control (RLC) layer, packet data convergence protocol (PDCP) layer, and service data adaptation protocol (SDAP) layer. The main functionalities of these layers are briefly described now.

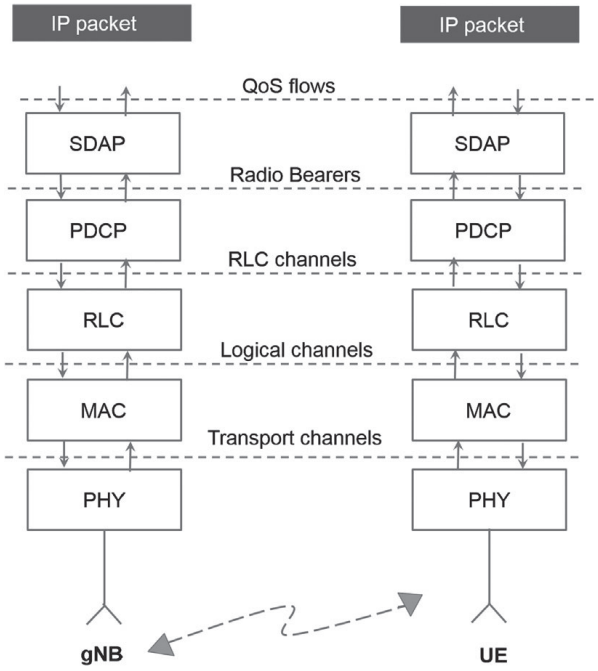


FIGURE 2.1

NR user-plane protocol stack.

- The SDAP layer handles the mapping between quality of service (QoS) flow and radio bearers.[1] IP packets are mapped to radio bearers according to their QoS requirements.
- The PDCP layer is primarily responsible for IP header compression/decompression, reordering and duplicate detection, ciphering/deciphering and integrity protection. The header compression mechanism reduces the number of bits to transmit over radio interface. The ciphering protects from

[1] A radio bearer can be seen as a pipe that carries internet protocol (IP) packets through a network and gets prioritized according to the specified QoS requirement.

eavesdropping and ensures message integrity. The reordering and duplication detection mechanisms allow in-sequence delivery of data units and removes duplicate data units.
- The RLC layer mainly performs error correction through an automatic repeat request[2] (ARQ) mechanism, segmentation/resegmentation of (header compressed) IP packets, and in-sequence delivery of data units to higher layers.
- The MAC layer is mainly responsible for error correction through hybrid ARQ[3] (HARQ) mechanism and uplink and downlink scheduling. The scheduler controls the assignment of uplink and downlink physical time-frequency resources for transmission. The MAC layer also takes care of multiplexing data across multiple component carriers when carrier aggregation is employed.
- The PHY layer handles coding/decoding, modulation/demodulation, multiantenna processing, and mapping of signals to physical time-frequency resources.

The control-plane is mainly responsible for control signaling for connection setup, mobility, and security. Control signaling originates either from a core network or from a radio resource control (RRC) layer in gNB. The main services from a RRC layer include broadcasting of system information, transmission of paging messages, security management including key management, handovers, cell-selection/reselection, QoS management, and detection of and recovery from radio link failures. The RRC messages are transmitted using the same PDCP, RLC, MAC, and PHY layers as for the user-plane. Therefore, from a physical layer perspective, there is not at all a fundamental technological difference in providing services to higher layers in control-plane and user-plane protocol stacks.

2.2 NR PHY: KEY TECHNOLOGY COMPONENTS

The key technology components of the NR physical layer are modulation, waveform, multiantenna transmission, and channel coding. In the following, we provide a brief overview of these physical layer components.

2.2.1 MODULATION

NR supports quadrature phase shift keying (QPSK), 16 quadrature amplitude modulation (QAM), 64 QAM and 256 QAM modulation formats for both uplink and downlink, as in LTE. Moreover, $\pi/2$-BPSK is supported in uplink to enable a further reduced peak-to-average power ratio and enhanced power amplifier efficiency at lower data rates, which is important for mMTC services. Since NR will cover a wide range of use cases, it is likely that the set of supported modulation schemes may be expanded. For example, 1024 QAM may become part of the NR specification, since fixed point-to-point backhaul already uses modulation orders higher than 256 QAM. Different modulation schemes for different UE categories may also be included in the NR specification.

[2] Automatic repeat request (ARQ) is an error-control method for data transmission that uses acknowledgments and timeouts to achieve reliable data transmission.
[3] Hybrid ARQ is a combination of forward error-correction coding and ARQ error-control.

Table 2.1 Scalable OFDM numerology for 5G NR (3GPP Release 15)

OFDM numerology	15 kHz	30 kHz	60 kHz	120 kHz
Frequency band	0.45–6 GHz	0.45–6 GHz	0.45–6 GHz 24–52.6 GHz	24–52.6 GHz
OFDM symbol duration	66.67 μs	33.33 μs	16.67 μs	8.33 μs
Cyclic prefix duration	4.69 μs	2.34 μs	1.17 μs	0.59 μs
OFDM symbol with CP	71.35 μs	35.68 μs	17.84 μs	8.91 μs
Maximum bandwidth	50 MHz	100 MHz	200 MHz	400 MHz

2.2.2 WAVEFORM

NR employs cyclic prefix OFDM (CP-OFDM) in both uplink and downlink up to at least 52.6 GHz. This is in contrast to LTE, where CP-OFDM is only used for downlink transmissions and DFT-Spread OFDM (DFTS-OFDM) is used for uplink transmissions. Having the same waveform in both directions simplifies the overall design, especially with respect to wireless backhauling and device-to-device (D2D) communications. Additionally, there is an option for DFT-spread OFDM (DFTS-OFDM) in uplink for coverage-limited scenarios, with single stream transmissions (that is, without spatial multiplexing). In practice, a gNB can select the uplink waveform (either CP-OFDM or DFTS-OFDM) and a UE should be able to support both OFDM and DFTS-OFDM. Any operation that is transparent to a receiver can be applied on top of the NR waveform, such as windowing/filtering to improve spectrum confinement.

NR has a scalable OFDM numerology to enable diverse services on a wide range of frequencies and deployments. The subcarrier spacing is scalable and specified as 15×2^n kHz, where n is an integer and 15 kHz is the subcarrier spacing used in LTE. In Chapter 6, we provide details on numerology design for NR. In 3GPP Release 15, four subcarrier spacings are specified: 15 kHz, 30 kHz, 60 kHz, and 120 kHz (i.e., $n = 1/2/3/4$), all with 7% CP overhead as in LTE (see Table 2.1). For the 60 kHz numerology, an extended CP is also defined. Different numerologies are specified for different frequency bands. Currently, there is no spectrum identified for NR between 6 and 24 GHz; therefore, the corresponding numerology has not been specified. When new frequencies will become available, the corresponding numerologies can be specified according to 15×2^n kHz. For all numerologies, the number of active subcarriers is 3300. Considering 3300 active subcarriers, the maximum bandwidths enabled by different numerologies are given in Table 2.1. To support even larger channel bandwidths, carrier aggregation can be employed. In Release 15, up to 16 component carriers are supported, where each component carrier can have up to 3300 active subcarriers.

The spectrum of an OFDM signal decays rather slowly outside the transmission bandwidth. In order to limit out-of-band emission, the spectrum utilization for LTE is 90 percent. For NR, it has been agreed that the spectrum utilization will be 94 to 99 percent. Windowing and filtering operations are viable ways to confine the OFDM signal in the frequency domain. In Chapters 5 and 6, we provide details on the spectrum confinement techniques for multicarrier waveforms.

2.2.3 MULTIPLE ANTENNAS

Multiantenna techniques were important already in LTE, but in NR they have a more fundamental role in the system design. The extension of the spectrum for mobile communication to include

also the millimeter-wave bands has led to a beam-centric design of NR in order to support analog beam-forming for achieving sufficient coverage. Furthermore, multiantenna techniques are crucial for fulfilling the performance requirements for 5G also for the traditional cellular frequency bands.

For low frequencies, multiantenna techniques are mainly enhancements of features developed in the later releases of LTE. The purpose with these enhancements is improved spectral efficiency driven by the ever-increasing quest for higher data rates and capacity in a congested spectrum. Advances in active array antenna technology have made it possible to have digital control over a large number of antenna elements, sometimes referred to as massive multiple-input multiple-output (MIMO). This enables higher spatial resolution in the multiantenna processing which can give higher spectral efficiencies. To this end, NR provides better support for multiuser MIMO (MU-MIMO) and reciprocity-based operation. A new framework for acquiring channel state information (CSI) has been developed to allow for more flexibility in the transmission of reference signals and to enable CSI with higher spatial resolution. This framework also provides a leaner system design and makes it easier to adapt to diverse use cases and to introduce new features in future releases of NR.

For high frequencies, obtaining coverage is the main challenge rather than obtaining high spectral efficiency. The reason for this is that transmission losses, when using legacy transmission techniques, are considerably higher, while there is a large amount of bandwidth available in the millimeter-wave spectrum. To overcome higher transmission losses and provide sufficient coverage, beam-forming is useful, particularly under LoS conditions and possibly both at the gNB and UE. With current hardware technology, analog beam-forming is expected to be prevalent at millimeter-wave frequencies. Therefore, procedures for supporting analog beam-forming in both the gNB and the UE have been developed in NR. Unlike previous generations of mobile communication systems, NR supports beam-forming not only for the data transmission but also for initial access and broadcast signals.

NR multiantenna techniques for both the gNB and the UE are discussed in detail in Chapter 7.

2.2.4 CHANNEL CODING

NR employs low density parity check (LDPC) codes for the data transmission for mobile broadband (MBB) services and polar codes for the control signaling. LDPC codes are attractive from an implementation perspective, especially at multigigabits-per-second data rates. Unlike the LDPC codes implemented in other wireless technologies, the LDPC codes considered for NR use a rate-compatible structure. This allows for transmission at different code rates and for HARQ operation using an incremental redundancy.

For the physical layer control signaling where the information blocks are small compared to data transmission and HARQ is not used, NR employs polar codes. By concatenating the polar code with an outer code and by performing successive cancellation list decoding, good performance is achieved at shorter block lengths. For the smallest control payloads, Reed–Muller codes are used.

NR LDPC and polar codes are discussed in detail in Chapter 8. For URLLC services, channel codes have not been agreed in 3GPP yet. In Chapter 8, a selection of coding schemes that exhibit a favorable performance/complexity tradeoff in the short block length regime is discussed. These schemes may be candidates for the future NR releases in addition to LDPC and polar codes.

2.3 PHYSICAL TIME-FREQUENCY RESOURCES

Physical time-frequency resources correspond to OFDM symbols and subcarriers within the OFDM symbols. The smallest physical time-frequency resource consists of one subcarrier in one OFDM symbol, known as a resource element. The transmissions are scheduled in group(s) of 12 subcarriers, known as physical resource blocks (PRBs). An example of NR physical time-resource structure is shown in Fig. 2.2.

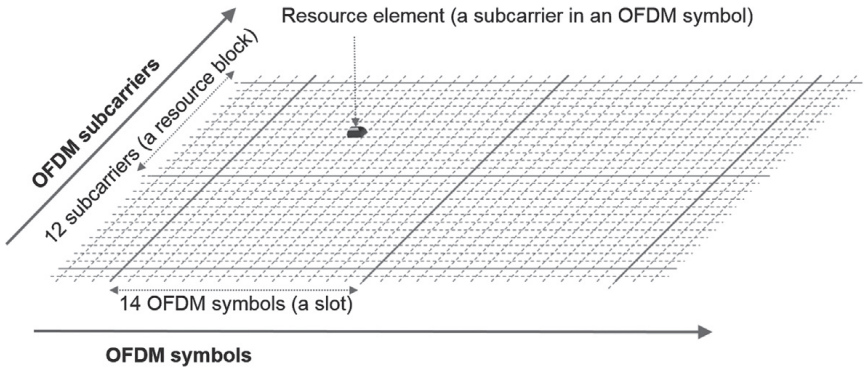


FIGURE 2.2

NR physical time-frequency structure.

In the time domain, the radio transmissions are organized into *radio frames*, *subframes*, *slots*, and *mini-slots*. As illustrated in Fig. 2.3, each radio frame has a duration of 10 ms and consists of 10 subframes with a subframe duration of 1 ms. A subframe is formed by one or multiple adjacent slots, each slot having 14 adjacent OFDM symbols. A mini-slot, in principle, can be as short as one OFDM symbol, but in Release 15 mini-slots are restricted to 2, 4, and 7 OFDM symbols. The time duration of a slot/mini-slot scales with the chosen numerology (subcarrier spacing) since the duration of an OFDM symbol is inversely proportional to its subcarrier spacing.

Physical layer uses time-frequency resources for transmission. In NR, the time-frequency resources (resource elements) represent either *physical channels* or *physical signals*, as in LTE. In the 3GPP terminology, a physical channel corresponds to a set of resource elements carrying information originating from higher layers, whereas a physical channel corresponds to a set of resource elements (used by physical layer) that do not contain information from higher layers. In the following sections we discuss both.

2.4 PHYSICAL CHANNELS

The time-frequency resources carrying information from higher layers (layers above PHY) are termed physical channels [1]. There are a number of physical channels specified for uplink and downlink:

- Physical downlink shared channel (PDSCH), used for downlink data transmission.

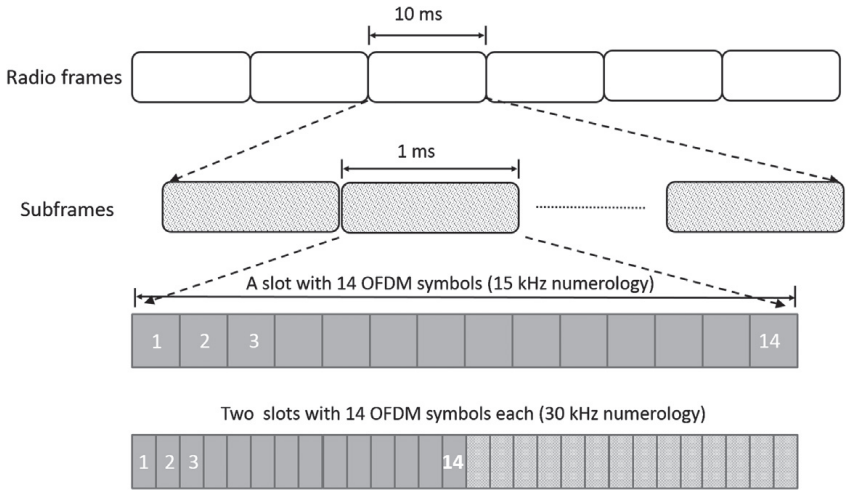


FIGURE 2.3

NR frame structure.

- Physical downlink control channel (PDCCH), used for downlink control information, which includes scheduling decisions required for downlink data (PDSCH) reception and for scheduling grants giving permission for uplink data (PUSCH) transmission by a UE.
- Physical broadcast channel (PBCH), used for broadcasting system information required by a UE to access the network.
- Physical uplink shared channel (PUSCH), used for uplink data transmission (by a UE).
- Physical uplink control channel (PUCCH), used for uplink control information, which includes: HARQ feedback acknowledgments (indicating whether a downlink transmission was successful or not), scheduling request (requesting time-frequency resources from network for uplink transmissions), and downlink channel-state information for link adaptation.
- Physical random access channel (PRACH), used by a UE to request connection setup referred to as random access.

On a high level, the downlink and uplink transmissions between gNB and UE work as follows. In downlink, a UE monitors the PDCCH, typically once per slot (PDCCH can also be configured more than once to enable ultralow-latency transmissions). Upon detection of a valid PDCCH, the UE receives one unit of data (named a transport block) on the PDSCH following the scheduling decision of gNB. Afterwards, the UE responds with a hybrid ARQ acknowledgment indicating whether the data was successfully decoded or not. In the case of unsuccessful decoding, retransmission(s) is scheduled. For uplink data transmission, a UE first requests gNB for physical time-frequency resources for the data awaiting transmission. This is termed a scheduling request and is sent over the PUCCH. In response, gNB sends a scheduling grant (over the PDCCH) which gives permission to a UE to use certain time-frequency resources for transmission. Following the scheduling grant, the UE schedules its data transmission over the PUSCH. gNB receives uplink data and sends hybrid ARQ acknowledgment indicating whether the uplink data transmission was successfully decoded or not. In the case of a decoding failure, retransmission(s) is scheduled. To enable ultralow-latency communication, the network

can also preconfigure data transmission resources for a UE to avoid the signaling involved in sending scheduling request and scheduling grants. This is known as grant-free transmission. A drawback of this scheme is that time-frequency resources can be unnecessarily reserved for a UE even if there is no data awaiting transmission at the UE.

2.5 PHYSICAL SIGNALS

The time-frequency resources that are used by the PHY layer but do contain information from higher layers (i.e., layers above the PHY layer) are termed physical signals [1]. The physical signals are reference signals used for different purposes, for example, demodulation, channel estimation, synchronization, and channel-state information. There are different physical signals in the uplink and the downlink. The downlink physical signals include:

- demodulation reference signal (DM-RS)
- phase tracking reference signal (PT-RS)
- channel state information reference signal (CSI-RS)
- primary synchronization signal (PSS)
- secondary synchronization signal (SSS)

Furthermore, the following uplink physical signals are defined:

- demodulation reference signal (DM-RS)
- phase tracking reference signal (PT-RS)
- sounding reference signal (SRS)

NR has an ultralean design, which minimizes always-on transmissions to enhance network energy efficiency, reduce interference, and ensure forward compatibility. In contrast to the setup in LTE, the reference signals in NR are transmitted only when necessary. Next, we briefly discuss four main reference signals: DM-RS, PT-RS, CSI-RS, and SRS.

DM-RS is used to estimate the radio channel for demodulation. DM-RS is UE-specific, can be beam-formed, confined in a scheduled resource, and transmitted only when necessary, both in the downlink and the uplink. The DM-RS design takes into account the early decoding requirement to support low-latency applications. For this reason, the DM-RS is placed in the beginning of a slot (known as a front-loaded DM-RS). For low-speed scenarios, DM-RS uses low density in the time domain (i.e., fewer OFDM symbols in a slot contain DM-RS). For high-speed scenarios, the time density of DM-RS is increased to track fast changes in the radio channel.

PT-RS is introduced in NR to enable compensation of the oscillator phase noise. Typically, phase noise increases as a function of the oscillator carrier frequency. PT-RS can therefore be utilized at high carrier frequencies (such as by millimeter wave) to mitigate phase noise. One of the main degradations caused by phase noise in an OFDM signal is an identical phase rotation of all the subcarriers, known as common phase error (CPE). (This is discussed in detail in Chapter 6 and Chapter 7.) PT-RS is designed so that it is sparse in the frequency domain and has a high density in the time domain and the reason is as follows. The phase rotation produced by CPE is identical for all subcarriers within an OFDM symbol, but there is low correlation of phase noise across OFDM symbols. The density of PT-RS in

frequency domain is one subcarrier in every PRB, every second PRB, or every fourth PRB. The density in time domain is every OFDM symbol, every second OFDM symbol, or every fourth OFDM symbol. Fig. 2.4 shows an example of DM-RS and PT-RS time-frequency structure. Like DM-RS, PT-RS is also UE-specific, it is confined in a scheduled resource, and it can be beam-formed. PT-RS is configurable depending on the quality of the oscillators, the carrier frequency, the OFDM subcarrier spacing, and on the modulation and coding schemes used for transmission.

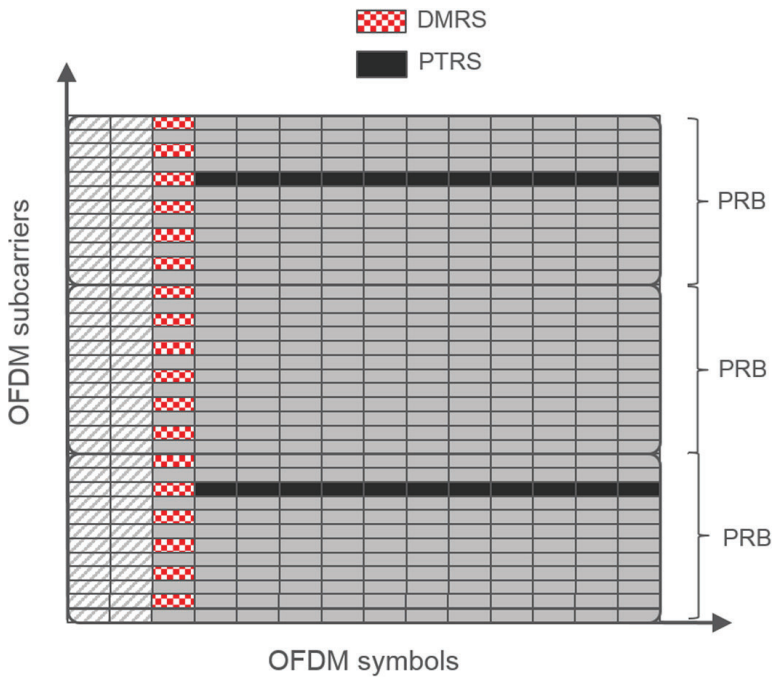


FIGURE 2.4

An example of time-frequency structure DM-RS and PT-RS.

CSI-RS is a downlink reference signal which is used mainly for CSI acquisition, beam management, time/frequency tracking and uplink power control. It has a very flexible design in order to support diverse use cases. CSI-RS for CSI acquisition is used for determining CSI parameters such as channel quality indicator (CQI), rank indicator (RI), and precoding matrix indicator (PMI), which are needed for link adaptation and for determining precoders. Furthermore, a so-called CSI interference measurement (CSI-IM) resource, which is a zero-power CSI-RS (ZP CSI-RS) resource, can be configured for interference measurements in the UE. For beam management, CSI-RS is used for evaluating candidate transmission beams by measuring reference signal received power (RSRP) for each beam. It can also be used for beam recovery purposes. A CSI-RS configured for time/frequency tracking is called a tracking reference signal (TRS). TRS can be used for fine time and frequency synchronization, and Doppler and delay spread estimation. This is needed for channel estimation and demodulation.

The SRS is transmitted in uplink to perform CSI measurements mainly for scheduling and link adaptation. For NR, it is expected that the SRS will also be utilized for reciprocity-based precoder design for massive MIMO and uplink beam management. SRS has a modular and flexible design to support different procedures and UE capabilities.

Table 2.2 Semi-static TDD configurations

OFDM numerology	Uplink/downlink switching periodicities [ms]
15 kHz	0.5, 1, 2, 5, 10
30 kHz	0.5, 1, 2, 2.5, 5, 10
60 kHz	0.5, 1, 1.25, 2, 2.5, 5, 10
120 kHz	0.5, 0.625, 1, 1.25, 2, 2.5, 5, 10

2.6 DUPLEXING SCHEME

NR supports TDD and FDD transmissions, as in LTE. The duplex scheme typically depends on spectrum allocation. At lower frequencies, the spectrum allocations are mostly paired, implying FDD transmission. At higher frequencies the spectrum allocations are often unpaired, implying TDD. In addition, NR supports dynamic TDD, where uplink and downlink allocations dynamically change over time. This is one of the key enhancements over LTE which is useful in scenarios with rapid traffic variations. The transmission scheduling decisions are made by the gNB scheduler and the UEs follow these scheduling decisions. The network can coordinate scheduling decisions between neighboring network sites to avoid interference, if necessary. TDD can also be semi-statically configured with certain uplink/downlink switching periodicities. There are uplink/downlink switching periodicities specified in Release 15 for different OFDM numerologies given in Table 2.2.

2.7 FRAME STRUCTURE

The NR frame structure follows three key design principles to enhance forward compatibility and reduce interactions between different functionalities. The first principle is that transmissions are self-contained. Data in a slot and in a beam is decodable on its own without dependency on other slots and beams. This implies that reference signals required for the demodulation of data are included in a given slot and a given beam. The second principle is that transmissions are well confined in time and frequency. Keeping transmissions confined makes it easier to introduce new types of transmissions in parallel with legacy transmissions in the future. The NR frame structure avoids the mapping of control channels across the full system bandwidth. The third principle is to avoid static and/or strict timing relations across slots and across different transmission directions. For example, asynchronous HARQ is used instead of a predefined retransmission time.

The NR frame structure supports TDD and FDD transmissions and operation in both the licensed and the unlicensed spectrum. It enables very low latency, fast HARQ acknowledgments, dynamic TDD, coexistence with LTE and transmissions of variable length (for example, short duration for URLLC and long duration for enhanced MBB (eMBB)). Considering the TDD operation, Fig. 2.5 provides examples of NR frame structure for different scenarios.

NR can also employ mini-slots to support transmissions with a flexible start position and a duration shorter than a regular slot duration. In principle, a mini-slot can be as short as one OFDM symbol and can start at any time. In Release 15, mini-slots are limited to 2, 4, and 7 OFDM symbols. Mini-slots can be useful in various scenarios, including low-latency transmissions, transmissions in unlicensed spectrum and transmissions in the millimeter-wave spectrum. In low-latency scenarios, transmission needs

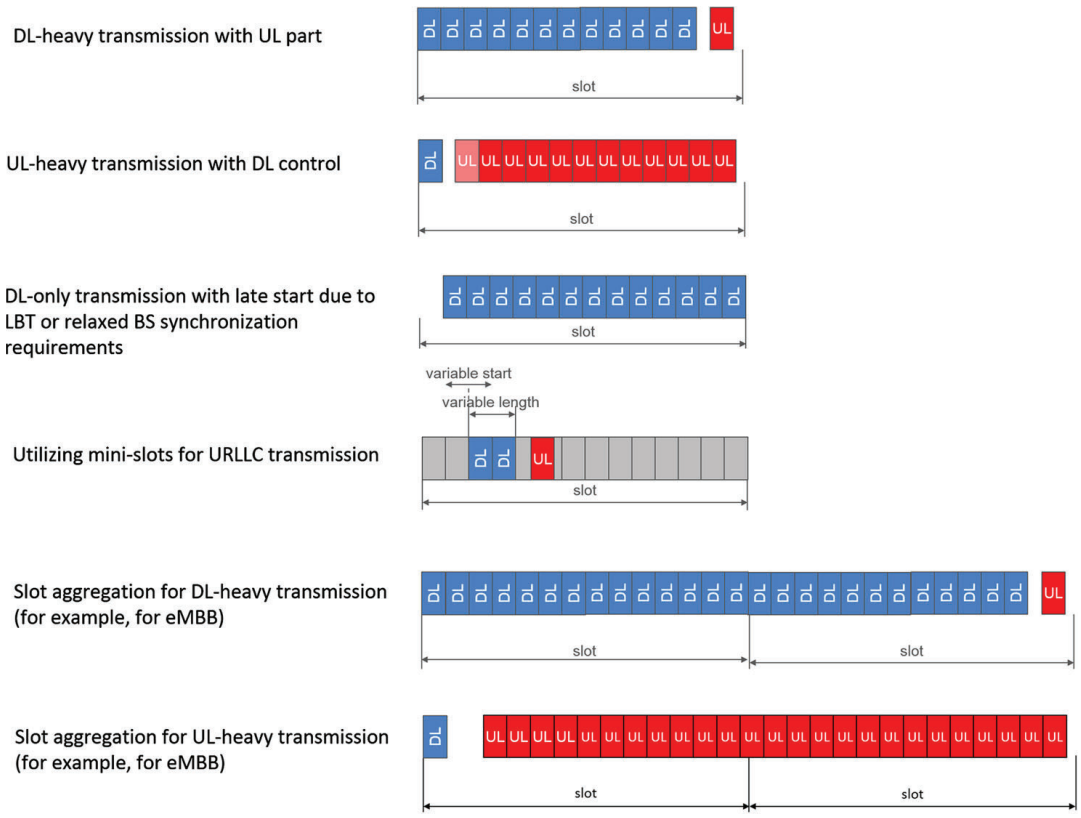


FIGURE 2.5

Examples of NR physical layer frame structure.

to begin immediately without waiting for the start of a slot boundary (ultra-reliable low-latency communications (URLLC), for example). When transmitting in the unlicensed spectrum, it is beneficial to start transmission immediately after the listen-before-talk (LBT) mechanism. When transmitting in the millimeter-wave band, the large amount of bandwidth available implies that the payload supported by a few OFDM symbols is large enough for many of the packets. Fig. 2.5 provides examples of URLLC- and LBT-based transmission in unlicensed spectrum via mini-slots and illustrates that multiple slots can be aggregated for services that do not require extremely low latency (eMBB, for example). Having a longer transmission duration helps to increase coverage or reduce the overhead due to switching (in TDD), transmission of reference signals, and control information.

The same frame structure can be used for FDD, by enabling simultaneous reception and transmission (that is, downlink and uplink can overlap in time). This frame structure is also applicable to device-to-device (D2D) communications. In that case, the downlink slot structure can be used by the device that is initiating (or scheduling) the transmission, and the uplink slot structure can be used by the device responding to the transmission.

To obtain low latency, a slot (or a set of slots in the case of slot aggregation) is front loaded with control signals and reference signals at the beginning of the slot (or set of slots), as illustrated in Fig. 2.6. An NR frame structure also allows for rapid HARQ acknowledgment, in which decoding

is performed during the reception of downlink data and the HARQ acknowledgment is prepared by the UE during the guard period, when switching from downlink reception to uplink transmission. NR supports very short duration uplink control signals (along with longer formats) to provide fast HARQ feedbacks from a device to the base station. Fig. 2.6 shows an example of a self-contained slot where the delay from the end of the data transmission to the reception of the acknowledgment from the device is on the order of just one OFDM symbol.

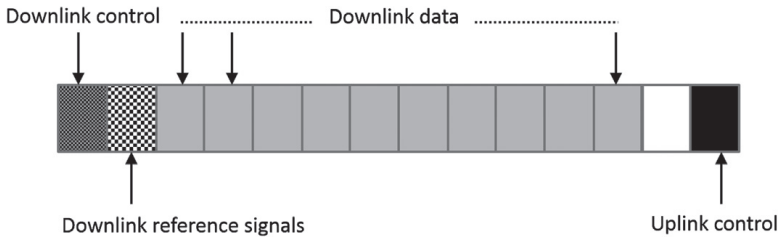


FIGURE 2.6

A self-contained slot.

2.8 PHY PROCEDURES AND MEASUREMENTS

There are a number physical layer procedures defined in the NR specifications, for example for cell search, power control, uplink synchronization and timing control, random access, and beam management, and there are channel-state information related procedures [2,3]. The NR specification also defines physical layer measurements for intra- and inter-frequency handover, inter RAT handover, timing measurements, and measurements for radio resource management [4]. These details (apart from the beam management procedure) are out of scope of this book. We strongly recommend [6] for understanding these details.

2.9 PHYSICAL LAYER CHALLENGES

5G NR is the first cellular technology to operate at millimeter-wave frequencies, support GHz of bandwidths, and utilize a massive number of antennas. These aspects impose a number of challenges for operation of NR physical layer mainly due to less understood radio wave propagation characteristics and hardware impairments (at the base stations and the devices). To enable a high performing NR, it is important to accurately understand the characteristics of radio propagation and hardware impairments. Chapters 3 and 4 address radio wave propagation, channel modeling, and hardware impairment modeling in great detail. In the following, we provide a brief discussion of some of these challenges.

2.9.1 PROPAGATION RELATED CHALLENGES

When going up in frequency the antenna related transmission loss increases as the square of the frequency, for any fixed receive antenna pattern, because the aperture is proportional to the square of the

wavelength. Some additional increase in loss due to propagation effects is likely, particularly under NLoS conditions. The challenge here is to utilize advanced multiantenna techniques to steer the signal into favorable directions and to increase the receive antenna aperture. At this point it is not fully clear to what extent the channel conditions may be improved at higher frequencies by means of these antenna techniques. In some scenarios, like LoS, it may even be favorable to go up in frequency. It is clear that the outdoor-to-indoor scenario will be challenging as the penetration loss goes up substantially with frequency. Another effect of using beam-forming and narrow beams is that the channel's dynamic changes will be larger and faster due to sudden blockage of the beam. Furthermore, the directional spread characteristics of the propagation channel is not well known. For channels which are highly dispersive in direction (rich scattering) high gain antennas are not very useful as they will only catch/direct a small fraction of the signal from the transmitter/to the receiver. In this case fully coherent antenna combining techniques are more suitable. However, such techniques are very complex and resource costly when the array antenna sizes are large in terms of the number of elements.

The main challenge for NR is to what extent novel multiantenna techniques may compensate for loss of, or even gain in, performance at higher frequencies comparing with legacy techniques used at lower frequencies. Under LoS conditions it is favorable to use higher frequencies and beam-forming as shown in Chapter 3. Moreover, the propagation losses in NLoS do not increase dramatically with frequency (in the range $0 \log f$ to $6 \log f$) except for outdoor-to-indoor transmissions. However, as early deployments of NR are expected to largely rely on analog beam-forming, the propagation channel is required to be highly directive. There are unfortunately only few highly directionally resolved channel measurements, implying that the possible beam-forming performance at millimeter-wave frequencies is largely unknown.

2.9.2 HARDWARE RELATED CHALLENGES

As regards the demands for increasingly high data rates moving into NR, the need for spectral efficiency and more bandwidth is rapidly increasing. The technology components used to address this increase involve advanced multiantenna techniques such as massive multi user (MU)-MIMO or, moving into the millimeter-wave regime, analog beam-forming. This raises new challenges in terms of efficient radio implementation as both the number of deployed transceivers and their operating frequencies and bandwidths increase.

As one of the important building-blocks in radio transceiver front-ends, the RF power amplifier (PA) continues to play a critical role as a major consumer of power. Operating in dense and highly integrated antenna arrays, the PA may also suffer the effects of mutual coupling unless sufficient isolation is implemented. This introduces another source of distortion caused by load-modulation as the active impedance presented to the amplifier changes over time. Another critical issue in antenna systems using directive transmission is assessing the distortion behavior in terms of how it is distributed over space. With conventional methods, the distortion follows the same pattern as the desired signal, but using MU-MIMO precoding or beam-forming techniques, this is generally not the case.

RF oscillators are another source of imperfections in the analog front-end, which, when moving toward the millimeter-wave bands, becomes a limiting factor. Maintaining a stable oscillation becomes more difficult at very high frequencies as the increasing losses decrease the quality factor of the resonator-tanks and there is a lack of power generation due to fundamental limitations in transistor technology. Thus, as the phase noise increases, coherent transmission may become increasingly difficult. Thus, phase noise needs to be accurately modeled in order to assess its impact and to provide a

basis for countermeasures such as tracking. Phase-locked loop (PLL) techniques may be used in order to decrease the narrow-band $1/f$ part of the phase noise, which may have more significant negative effects on OFDM-based transmission, but this comes at the cost of increased white noise levels.

As advanced signal processing hardware and algorithms make it possible to compensate for some of the more prominent impairments, data-converters such analog-to-digital converter (ADC) and digital-to-analog converter (DAC) are crucial components connecting the analog and digital domains. Unfortunately, data-converters do not enjoy the same beneficial scaling as described by Moore's law, as these are mixed signal components and generally require linear transistor technologies. Scaling down the geometries to increase the available processing speed is not necessarily a good option in this case, as it makes the transistors operate more as switches. As shown in the literature, massive MU-MIMO opens up possibilities to decrease the effective resolution. This will, however, inflict an increasing amount of quantization noise which, if left unchecked, may corrupt the signal to a large extent. In order to assess the impact of a coarser quantization, different modeling techniques based on either deterministic behavioral models or stochastic processes may aid us.

Overall, the need for improved mathematical tools capable of assessing the behavior of non-ideal radio components over time/frequency and space is larger than ever. Specifically this is so since these models may serve as a basis for advanced compensation techniques such as digital pre-distortion (DPD) or phase-noise tracking. Current developments in this field have further opened up possibilities of new analysis and understanding of how radio imperfections behave in the context of a large antenna array. Another relevant field of modeling which currently is in its cradle is that of stochastic modeling of radio impairments aimed to aid performance analysis on link- or system-level, as the kind of oversampled passband or baseband data most behavioral models are built upon may not always be available. This modeling framework will be further discussed in Chapter 4.

REFERENCES

[1] 3GPP, TS 38.211 NR Physical channels and modulation, June 2018.
[2] 3GPP, TS 38.213 NR Physical layer procedures for control, June 2018.
[3] 3GPP, TS 38.214 NR Physical layer procedures for data, June 2018.
[4] 3GPP, TS 38.215 NR Physical layer measurements, June 2018.
[5] E. Dahlman, S. Parkvall, J. Sköld, 4G LTE-Advanced Pro and The Road to 5G, Academic Press, 2016.
[6] E. Dahlman, S. Parkvall, J. Sköld, 5G NR: The Next Generation Wireless Access Technology, Academic Press, 2018.

CHAPTER

PROPAGATION & CHANNEL MODELING

3

Radiowave propagation knowledge is crucial for development of wireless communications. The reason is that propagation represents the two most important respects in which mobile and fixed telecommunications differ.

The first is the physical media in which the transmitted electromagnetic waves propagate. In the fixed case the medium is mainly transmission lines, for example optical fibers and coaxial copper cables. The media in the case of mobile communications is in contrast the complete environment in which the transmitter and receiver are embedded. As denser materials both absorb and reflect a substantial fraction of the energy of radio waves, the main medium for propagation is the free space between those materials. Though it is possible to steer the transmitted signal to the desired receiver by utilizing advanced antenna techniques, the propagation environment does not allow for full isolation between links, wherefore signals transmitted from different sources are normally mixed at the receivers, causing substantial interference.

The second respect in which mobile communications are different from the fixed case is that the channel may vary heavily in time and space. Good knowledge of the characteristics of these variations is essential when optimizing transmission techniques as well as when dimensioning and planning networks.

This chapter provides a comprehensive overview of the impact of the radio propagation channel on mobile communications, explaining some key aspects in detail. It is organized as follows. In Section 3.1 the fundamentals of radiowave propagation are explained based on electromagnetic theory. Section 3.2 presents commonly used quantities for the characterization of the propagation channel, whereas corresponding experimental results are exemplified in Section 3.3. Recent 5G channel modeling provided by mainly 3GPP and ITU-R is presented in Section 3.4. Focus is put on understanding the models and the extent to which the modeling is accurate and realistic. Moreover, model components that need further improvements are pointed out. Finally, the chapter is summarized and future work is proposed in Section 3.5.

3.1 PROPAGATION FUNDAMENTALS

This section deals with some fundamental aspects of propagation that are useful for understanding characterization and modeling of the radiowave channel in a mobile communications context. For a comprehensive description of electromagnetic theory the reader is referred to standard textbooks on the topic such as [3], [13]. Propagation concerns the radiowave pathway between the transmit antenna and the receive antenna. Sometimes the antenna is mixed up with propagation effects in propagation modeling, which is a problem which is addressed in Section 3.1.2. Antennas may, however, not

5G Physical Layer. https://doi.org/10.1016/B978-0-12-814578-4.00008-4
Copyright © 2018 Elsevier Ltd. All rights reserved.

Samsung Exhibit 1020, Page 49 of 314

be avoided in propagation research where they primarily are used for measuring the electromagnetic fields at specific space points or in specific directions of propagation. This section sheds light on some of the fundamental aspects of radiowave propagation. The first subsection explains what electromagnetic waves are. The second subsection deals with free-space propagation and the third with the basic propagation mechanisms.

3.1.1 ELECTROMAGNETIC WAVES

In the most fundamental respect electromagnetic waves are defined as the set of solutions of Maxwell's equations in space regions free from charges. This set of solutions may be represented by different series expansions. We will here focus on the plane wave expansion, which is very useful for describing propagation in a comprehensible way. Moreover, plane waves are the most common representation used in standard channel models. For plane waves, the electric $\mathbf{E}$ and magnetic $\mathbf{H}$ fields are equal and orthogonal to each other and to the Poynting vector,

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}^*, \tag{3.1}$$

which points in the direction of propagation. Defining an orthonormal reference system ($\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$) the electric field of the plane wave is given by

$$\mathbf{E}(\mathbf{x}, t) = \mathrm{Re}\left[(\mathbf{e}_1 E_1 + \mathbf{e}_2 E_2) e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t}\right] \tag{3.2}$$

where $\mathbf{x}$ is the space coordinate, t is the time, $\mathbf{k} = \mathbf{e}_3 2\pi/\lambda$ where λ is the wavelength, and ω is the angle frequency of the wave. The corresponding magnetic field is given by

$$\mathbf{H} = \mathbf{e}_3 \times \mathbf{E}\sqrt{\frac{\epsilon}{\mu}} \tag{3.3}$$

where ϵ and μ are the permittivity and permeability, respectively. Fig. 3.1 shows three cases of plane waves defined by the following electric fields:

$$\left.\begin{array}{lll} 1) & E_1 = 0, & E_2 = 1 \\ 2) & E_1 = \frac{1}{\sqrt{2}}, & E_2 = \frac{i}{\sqrt{2}} \\ 3) & E_1 = \frac{1}{\sqrt{4}}\left(e^{-i\pi/6} - 1\right), & E_2 = \frac{1}{\sqrt{4}}\left(e^{-i\pi/6} + 1\right) \end{array}\right\} \tag{3.4}$$

where 1) corresponds to linear polarization, 2) to circular polarization and 3) to elliptic polarization. It is clearly seen that the $\mathbf{E}$ and the $\mathbf{H}$ fields are orthogonal and equal in strength at each time instance. A plane wave is defined by its six corresponding degrees of freedom: 1) the polarization elliptic axial ratio, 2) the polarization ellipse main axis rotation angle, 3) the field amplitude, 4) the wave phase, 5) the Poynting vector polar angle, and 6) the Poynting vector azimuth angle. These degrees of freedom may also be represented by the electric and the magnetic vector fields at a space point. In this case the electromagnetic wave field is viewed as a sum of plane waves which results in that the magnetic and the electric fields are decorrelated. For any wave vector $\mathbf{k}$, however, a plane wave supports only two orthogonal polarization states.

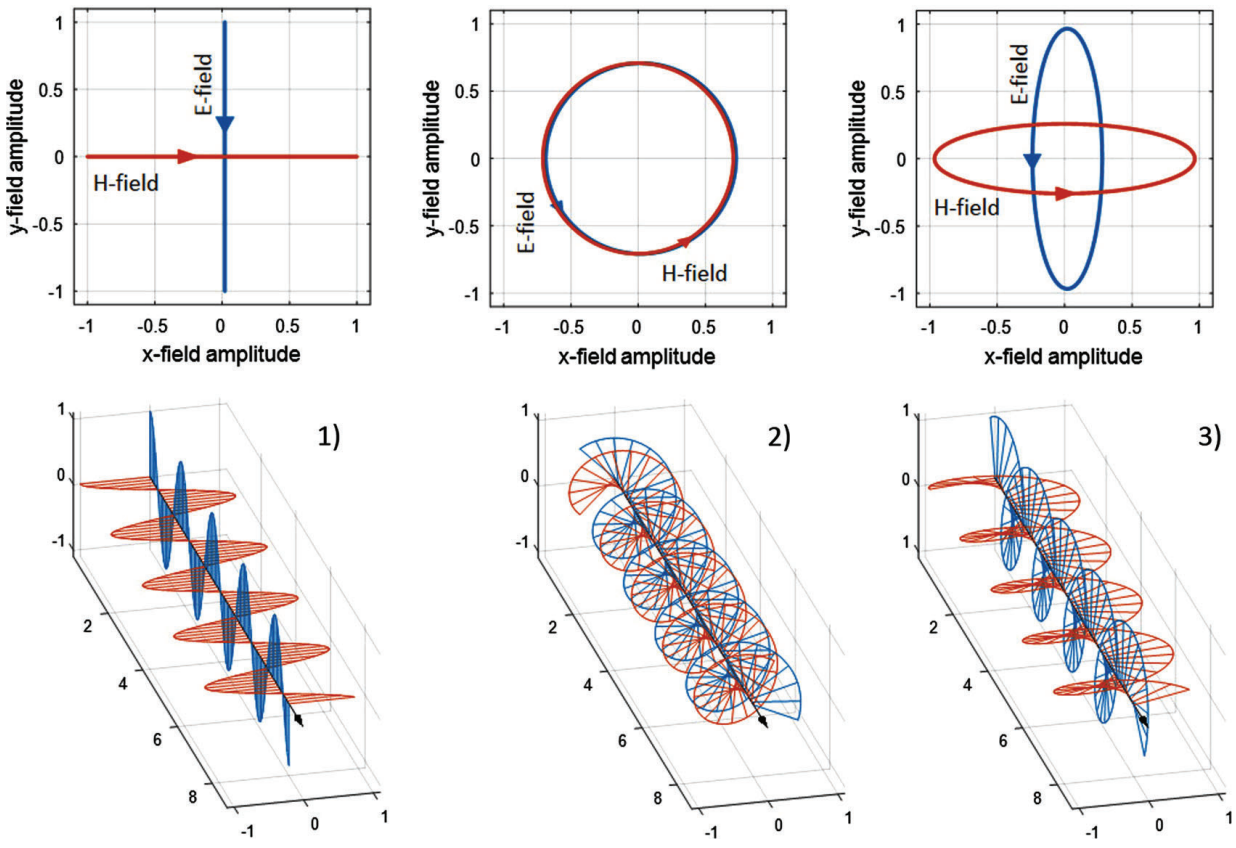


FIGURE 3.1

Plane waves with 1) linear, 2) circular, and 3) elliptic polarizations.

3.1.2 FREE-SPACE PROPAGATION

In order to understand propagation in general it is essential to have full insight into free-space propagation. In principle, propagation should exclude antenna effects. In practice, however, it is not possible to study propagation without the use of antennas. Here we will focus on the case of using some basic antennas being very clear on which effects are due to the antennas and which effects are due to propagation. The isotropically radiating antenna is commonly used as a reference. In free space the transmission loss between two such antennas (shown in Fig. 3.2) is referred to as *free-space basic transmission loss* (by ITU-R [8]) and is in dB units given by

$$L_{\text{bf}} = 20 \log \left(\frac{4\pi d}{\lambda} \right) \tag{3.5}$$

where d is the distance between the antennas and λ is the wavelength. Though the isotropic antenna is commonly used as a reference it has been shown that it is not very practical to realize [16], [33]. There is a corresponding expression for arbitrary antenna patterns,

$$L_{\text{f}} = L_{\text{bf}} - G_t - G_r \tag{3.6}$$

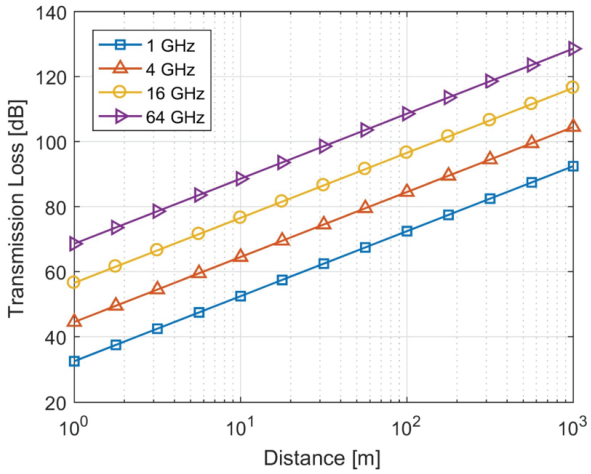


FIGURE 3.2

Free-space basic transmission loss versus distance for different carrier frequencies.

where G_t and G_r are the antenna gains at the transmitter and the receiver, respectively. It should be noted that (3.6) is based on the assumption that the transmit and receive antennas are fully matched in both direction and polarization.

It is clear from Fig. 3.2 that the free-space basic transmission loss is frequency dependent. This is the case, by (3.6), for any type of antenna (dipole, horn, patch, etc.) with a specified antenna pattern. As the shape of a specific type of antenna does not change with frequency the far field radiation does not change either. The size of a specific antenna is, however, proportional to the wavelength λ and thus the aperture (effective antenna area) is proportional to λ^2. As a consequence the received power for free-space transmission using any type of antennas (fixed gain and antenna pattern) depends on the frequency, f, as $-20\log f$ in dB units. This is indeed the result of fixing the antenna gains of (3.6). It should be noted that this frequency dependence is a pure antenna effect. From a propagation perspective there is no frequency dependence when transmitting with a fixed gain antenna as the power flux density at any far field distance is frequency independent in this case.

Instead of fixing the gain of the receive antenna the aperture may be kept constant over frequency. In this way the received signal becomes proportional to the power flux density at the receiver, which is frequency independent when a fixed gain transmit antenna is used. When a fixed antenna aperture is used also in the transmit end of a free-space link, there is actually a reduction of transmission loss with increasing frequency proportional to $-20\log(f)$. This is explained by that the antenna gain of a fixed aperture antenna increases with frequency providing corresponding increased signal strength at the receiver. This gain is, however, valid only in the far field of the antennas. For high frequencies in the millimeter-wave range the far field may be at fairly large distances depending on the sizes of the antenna apertures. In the near field region the transmitted lobe width may be substantially smaller than the receive antenna aperture. In this case the transmission loss may actually be negligible (0 dB). It should, however, be noted that, for physical reasons, there is no possibility of obtaining a transmission gain (i.e., loss $<$ 0 dB) [4]. Moreover, large aperture antennas have corresponding high gains and narrow lobes, which have to be directed towards the antenna in the other end of the link.

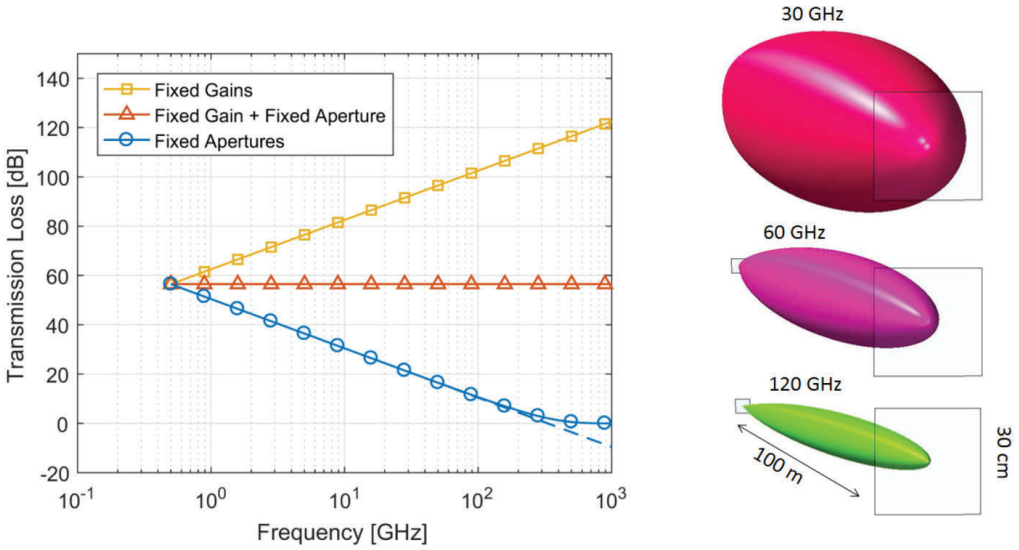


FIGURE 3.3

Free-space transmission loss versus frequency for 1) fixed gains, 2) fixed gain and fixed aperture, and 3) fixed apertures (left hand side), and corresponding lobes for the fixed aperture transmit antenna case (right hand side). The dashed line in the left hand graph shows the loss assuming far field when in the near field region.

All cases described above are illustrated in Fig. 3.3 where a 100-m free-space transmission link is simulated for fixed gain antennas ($G = 5$ dBi), and fixed square antenna aperture (30x30 cm). The fixed aperture gives 5 dBi gain at 500 MHz and 51 dBi gain at 100 GHz. The transition from the far field region to the near field region between 200 GHz and 500 GHz when using fixed aperture antennas in both ends of the link is clearly shown in the figure.

The *free-space basic transmission loss* is commonly referred to as *free-space path loss* in the literature. There are probably historical reasons for this. This use is, however, somewhat problematic as it gives the impression that free-space propagation might be frequency dependent, which is not the case as explained above. For this reason the notion of *path loss* is avoided throughout this book. Instead the notion of *transmission loss* is used when appropriate.

3.1.3 SCATTERING AND ABSORPTION

In a real environment the conditions are rarely similar to free space. Typically there is an abundance of structures and objects in the environment surrounding a transmission link which scatter the propagating electromagnetic waves. In an urban environment these objects are typically man made, like buildings and vehicles. In rural environments vegetation and geographical formations dominate the environment. Scattering can be categorized into a few basic types:

- ***Specular reflection and refraction*** At sufficiently flat and large surfaces electromagnetic waves are scattered fulfilling Fresnel reflection and refraction formulas. In optics this type of scattering results

in specular and refracted images. The size of a surface has to be considerably larger than one Fresnel zone[1] to scatter according to Fresnel's formulas.

- ***Diffraction*** When the size of a flat surface is at the order of one Fresnel zone or smaller, the scattering is most accurately described by diffraction theory. This is also the case in the shadow zone behind an object which is blocking a propagating wave.
- ***Diffuse scattering*** Any object or structure which is not smooth and homogeneous over several Fresnel zones will scatter a wave in a diffuse manner. This is typically the effect of rough surfaces such as stone or brick walls, vegetation and ground.
- ***Absorption*** Absorption is substantial in some scenarios. For example when transmitting through vegetation the absorbed energy may be substantial. This is also the case when transmitting from outdoor to locations inside a building. Moreover, at large distances the absorption in the atmosphere may be substantial.

3.2 PROPAGATION CHANNEL CHARACTERIZATION

When developing radio transmission techniques it is crucial to know well the characteristics of the radio propagation channel. Under conditions close to free space, like for microwave links, the main challenge is to determine the time-varying transmission loss of the direct path. In this case the loss is mainly varying due to rain and consequently the corresponding probability of rain intensity is of interest. For cellular communications, however, the channel is characterized by the scattering environment and the mobility of terminals and some of the scattering objects which are not stationary. As a consequence, the received signal may vary considerably over time and space. Moreover, in cellular scenarios the propagation is affected by multipath propagation, i.e., the received signal is a mix of a multitude of different waves scattered around in the environment arriving with different directions and delays. To account for the multipath, the channel is commonly modeled as a set of discrete (plane) waves which combine at the receive antenna. Each discrete wave is determined by its pathway from the transmit antenna, via scattering (except for the direct path), to the receive antenna. In typical mobile communications scenarios the number of paths needed to characterize the channel may be very large up to tens of thousands. For the plane wave expansion the channel H_{mn} between transmit antenna element n and receive antenna element m is mathematically characterized by

$$H_{mn} = \sum_{l=1}^{N} \mathbf{g}_m^{\mathrm{rx}}\left(-\mathbf{k}_l^{\mathrm{rx}}\right)^{\mathrm{T}} \cdot \mathbf{A}_l \cdot \mathbf{g}_n^{\mathrm{tx}}\left(\mathbf{k}_l^{\mathrm{tx}}\right) \cdot \exp\left[i\left(\mathbf{k}_l^{\mathrm{tx}} \cdot \mathbf{r}_n^{\mathrm{tx}} - \mathbf{k}_l^{\mathrm{rx}} \cdot \mathbf{r}_m^{\mathrm{rx}} + \omega\tau_l + \omega_{\mathrm{D}_l} t\right)\right] \tag{3.7}$$

where $\mathbf{A}_l$ is the complex polarimetric amplitude matrix of the lth of totally N plane waves, $\mathbf{g}_m^{\mathrm{rx}}\left(-\mathbf{k}_l^{\mathrm{rx}}\right)$ and $\mathbf{g}_n^{\mathrm{tx}}\left(\mathbf{k}_l^{\mathrm{tx}}\right)$ are the complex polarimetric antenna pattern vectors for the corresponding wave vectors $\mathbf{k}_l^{\mathrm{rx}}$ and $\mathbf{k}_l^{\mathrm{tx}}$, $\mathbf{r}_n^{\mathrm{tx}}$ and $\mathbf{r}_m^{\mathrm{rx}}$ are the position vectors of the receive and transmit antenna elements relative to corresponding antenna reference points, ω is the angular frequency ($\omega = 4\pi f$), τ_l is the wave propagation delay between transmit and receive antenna reference points, ω_{D_l} is the Doppler angular

[1] If the pathway via any point at the border of a reflective surface is more than one wavelength longer than the specularly reflected path, the reflecting surface is larger than one Fresnel zone.

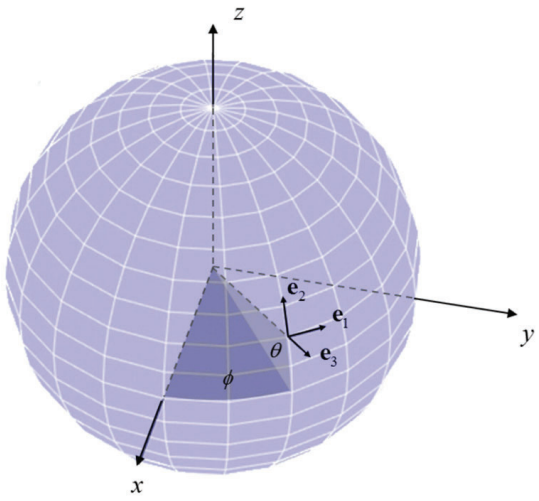


FIGURE 3.4

Local antenna reference system. The coordinate system for polarization is given by $\mathbf{e}_1 = \hat{\boldsymbol{\phi}}$, $\mathbf{e}_2 = \hat{\boldsymbol{\theta}}$ and $\mathbf{e}_3 = \hat{\mathbf{r}}$.

frequency and t is the time. The antenna patterns are defined based on local reference systems. For this purpose fixed local coordinate systems (x, y, z, or in polar coordinates, r, ϕ, θ) are used. Based on the fixed reference systems additional direction related reference systems are used for specifying the corresponding polarimetric pattern components in the different directions. The axes of these direction relative reference systems are defined by corresponding unit vectors pointing in the directions of increasing azimuth, $\mathbf{e}_1 = \hat{\boldsymbol{\phi}}$, and elevation, $\mathbf{e}_2 = \hat{\boldsymbol{\theta}}$, angles and increasing radius distance, $\mathbf{e}_3 = \hat{\mathbf{r}}$, as shown in Fig. 3.4. For free-space propagation and use of unit gain antennas the polarimetric amplitude is given by

$$\mathbf{A} = \frac{\lambda}{4\pi d}\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.8}$$

where the minus sign is due to that the receiver coordinate system is rotated 180 degrees relative to the transmitter coordinate system. With both transmit and receive antennas having any of the three polarization cases of Eq. (3.4) the channel gain is given by

$$\left|\left(\mathbf{g}^{\mathrm{rx}}\right)^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{g}^{\mathrm{tx}}\right| = \frac{\lambda}{4\pi d} \tag{3.9}$$

which, as expected, corresponds to the freespace-loss L_{bf} given by Eq. (3.5). When using receive antenna polarizations which are orthogonal to the transmit antenna polarizations, we have

$$\left|\left(\mathbf{g}^{\mathrm{rx}}\right)^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{g}^{\mathrm{tx}}\right| = 0. \tag{3.10}$$

With regard to (3.4) the corresponding orthogonal polarizations are given by

$$\left.\begin{array}{lll} 1) & E_1 = 1, & E_2 = 0 \\ 2) & E_1 = -\frac{1}{\sqrt{2}}, & E_2 = \frac{i}{\sqrt{2}} \\ 3) & E_1 = \frac{1}{\sqrt{4}}\left(e^{-i\pi/6}+1\right), & E_2 = \frac{1}{\sqrt{4}}\left(e^{-i\pi/6}-1\right) \end{array}\right\}. \tag{3.11}$$

3.2.1 FREQUENCY-DELAY DOMAIN

The channel response of Eq. (3.7) is specified in the frequency domain, i.e., it is a function of the radio frequency f. There is a complementary equivalent way to specify the channel in the delay domain providing the corresponding channel response of an impulse, $h(\tau)$, as a function of the delay τ. The relation between these two domains is given by the corresponding Fourier transforms,

$$\left.\begin{array}{rcl} h(\tau) & = & \int_{-\infty}^{\infty} H(f)\exp(i2\pi f\tau)\,df \\ H(f) & = & \int_{-\infty}^{\infty} h(\tau)\exp(-i2\pi f\tau)\,d\tau \end{array}\right\}. \tag{3.12}$$

For radio data transmissions, modulation of a continuous wave around a specific carrier frequency f_0 is utilized. The speed of the modulation and corresponding bitrate is proportional to the bandwith B used. For a channel, the band-limiting filter characteristics determine the shape of the corresponding impulse response. This may be exemplified by means of a channel having a single multipath component with a delay of τ_1 and a uniform bandpass filter

$$\left.\begin{array}{rcl} h(\tau) & = & \frac{B}{2}\mathrm{sinc}\left[\pi B(\tau-\tau_1)\right]\cdot\left[\exp(i2\pi f_0 t)+\exp(-i2\pi f_0 t)\right] \\ H(f) & = & \left\{\begin{array}{ll} 0 & \text{if } |f| > f_0+\frac{B}{2} \text{ and } |f| < f_0-\frac{B}{2} \\ \frac{1}{2} & \text{if } |f| < f_0+\frac{B}{2} \text{ and } |f| > f_0-\frac{B}{2} \end{array}\right. \end{array}\right\}. \tag{3.13}$$

The corresponding channel responses for $f_0 = 2$ GHz, $B = 200$ MHz, and $\tau_1 = 10$ ns are shown in Fig. 3.5. In the passband the fast oscillations of $h(\tau)$, due to the carrier frequency f_0, are evident. Moreover, both positive and negative frequencies of $H(f)$ are needed for providing the real channel response. For convenience the channel is commonly described in the baseband meaning that the frequency is translated to zero mean, i.e., $f' = f - f_0$

$$\left.\begin{array}{rcl} h(\tau) & = & B\,\mathrm{sinc}\left[\pi B(\tau-\tau_1)\right] \\ H(f') & = & \left\{\begin{array}{ll} 0 & \text{if } |f'| > \frac{B}{2} \\ 1 & \text{if } |f'| < \frac{B}{2} \end{array}\right. \end{array}\right\}. \tag{3.14}$$

In the following the baseband representation of the channel will be used, if the passband representation is not specifically indicated, with f representing the frequency in the baseband.

The uniform frequency filter in Eq. (3.14) results in substantial ringings, or side-lobes, around the main impulse in the delay domain. This can be mitigated by choosing a different frequency filter such

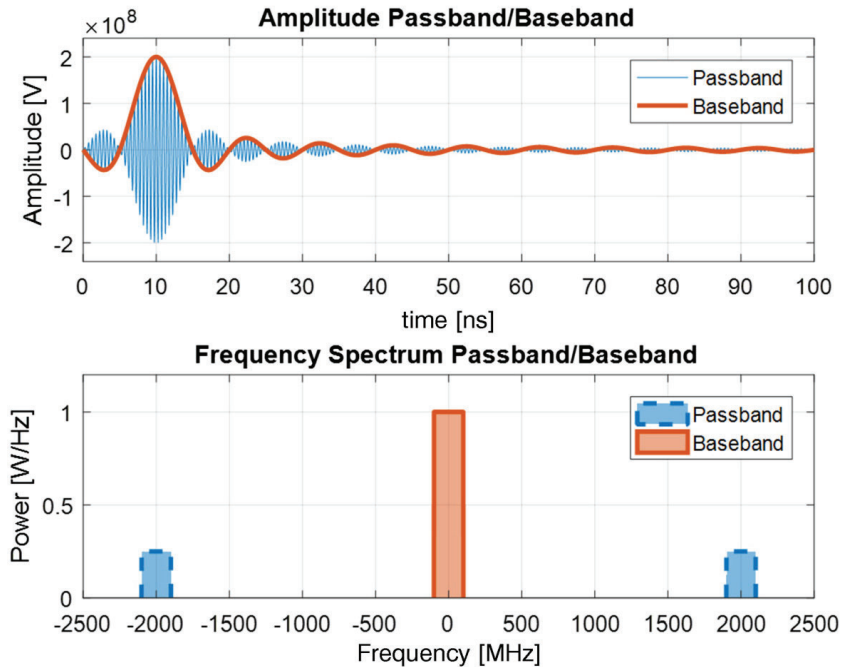


FIGURE 3.5

Example of a channel in passband and baseband with carrier frequency at 2000 MHz using 200 MHz bandwidth. The power of the baseband channel is twice the power of the passband channel to facilitate the comparison in the delay domain.

as the commonly used Hann filter

$$\left.\begin{array}{rcl} h(\tau) & = & B\,\mathrm{sinc}\left[\pi B\left(\tau-\tau_1\right)\right]+ \\ & & \frac{B}{2}\mathrm{sinc}\left[\pi B\left(\tau-\tau_1\right)+\pi\right]+\frac{B}{2}\mathrm{sinc}\left[\pi B\left(\tau-\tau_1\right)-\pi\right] \\ H(f) & = & \begin{cases} 0 & \text{if } |f|>\frac{B}{2} \\ \cos\left(\frac{\pi}{B}f\right) & \text{if } |f|<\frac{B}{2} \end{cases} \end{array}\right\}. \tag{3.15}$$

In Fig. 3.6, the channel responses of a uniform frequency filter and a Hann filter are shown. The side-lobes for the Hann filter are clearly smaller than for the uniform filter.

When modeling a channel in frequency domain a sum of different waves having different delays and amplitudes, as a result of scattering in the environment, is used. We will illustrate this by the typical exponentially decaying channel,

$$\left.\begin{array}{rcl} H(f) & = & \sum_{l=1}^{N} a_l \exp\left(i2\pi f\tau_l\right) \\ a_l & = & \exp\left(-\frac{\tau_l}{\sigma_\tau}+i\phi_l\right) \end{array}\right\} \tag{3.16}$$

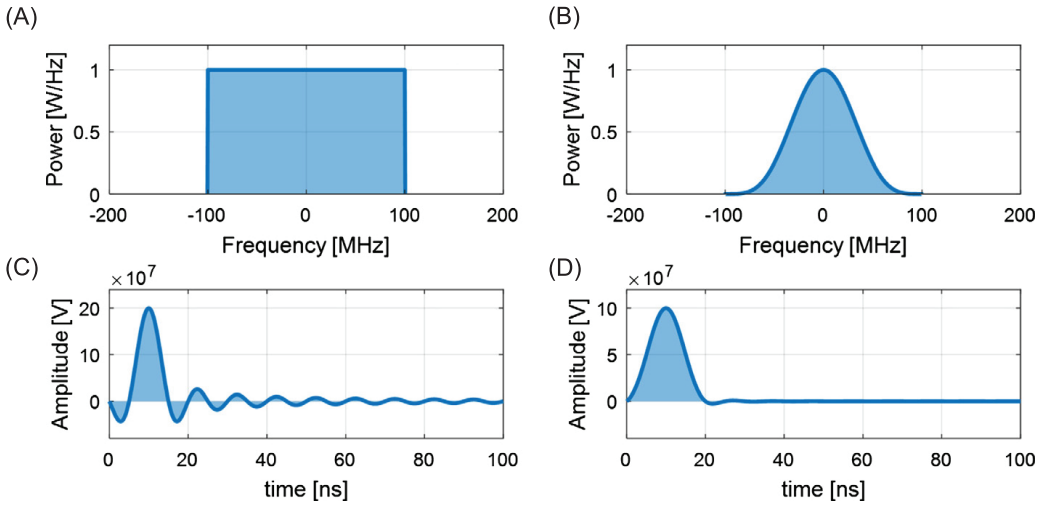


FIGURE 3.6

Channel responses for a uniform frequency filter (A) and (C), and a Hann filter (B) and (D).

where ϕ_l is a random phase and σ_τ is the root mean square (RMS) delay spread which is a useful measure defined by

$$\left.\begin{aligned} \mu_\tau &= \frac{\sum_{l=1}^{N} \tau_l |a_l|^2}{\sum_{l=1}^{N} |a_l|^2} \\ \sigma_\tau &= \sqrt{\frac{\sum_{l=1}^{N} (\mu_\tau - \tau_l)^2 |a_l|^2}{\sum_{l=1}^{N} |a_l|^2}} \end{aligned}\right\}. \tag{3.17}$$

For delay spreads substantially larger than the one over the bandwidth, i.e., $\sigma_\tau > \frac{1}{\mathrm{B}}$ the channel becomes frequency selective, meaning that the signal is fading substantially over this band. In Fig. 3.7 the corresponding channels are shown for $\sigma_\tau = 20$ ns and $\sigma_\tau = 100$ ns. Two cases of bandwidths are shown, $B = 10$ MHz and $B = 100$ MHz, applying Hann filtering in the delay domain to suppress the side-lobes. As the delay domain resolution is proportional to the used bandwidth, more multipath components are resolved at the higher bandwidth. An important channel property is that the channel is frequency flat over a bandwidth which is less than that needed to resolve multipath components as shown in Fig. 3.7D. A measure commonly used to characterize the frequency selectivity is the coherence bandwidth defined as the bandwidth over which the channel is correlated at some level, typically over 0.9. For the channels in Fig. 3.7 the corresponding coherence bandwidths are 1 MHz for $\sigma_\tau = 100$ ns, and 5 MHz for $\sigma_\tau = 20$ ns.

3.2.2 DOPPLER-TIME DOMAIN

The channel characteristics in Doppler-time domain are fully analogous with frequency-delay domain. In this case the channel variations in time t are specified by the Doppler frequencies f_D of the corre-

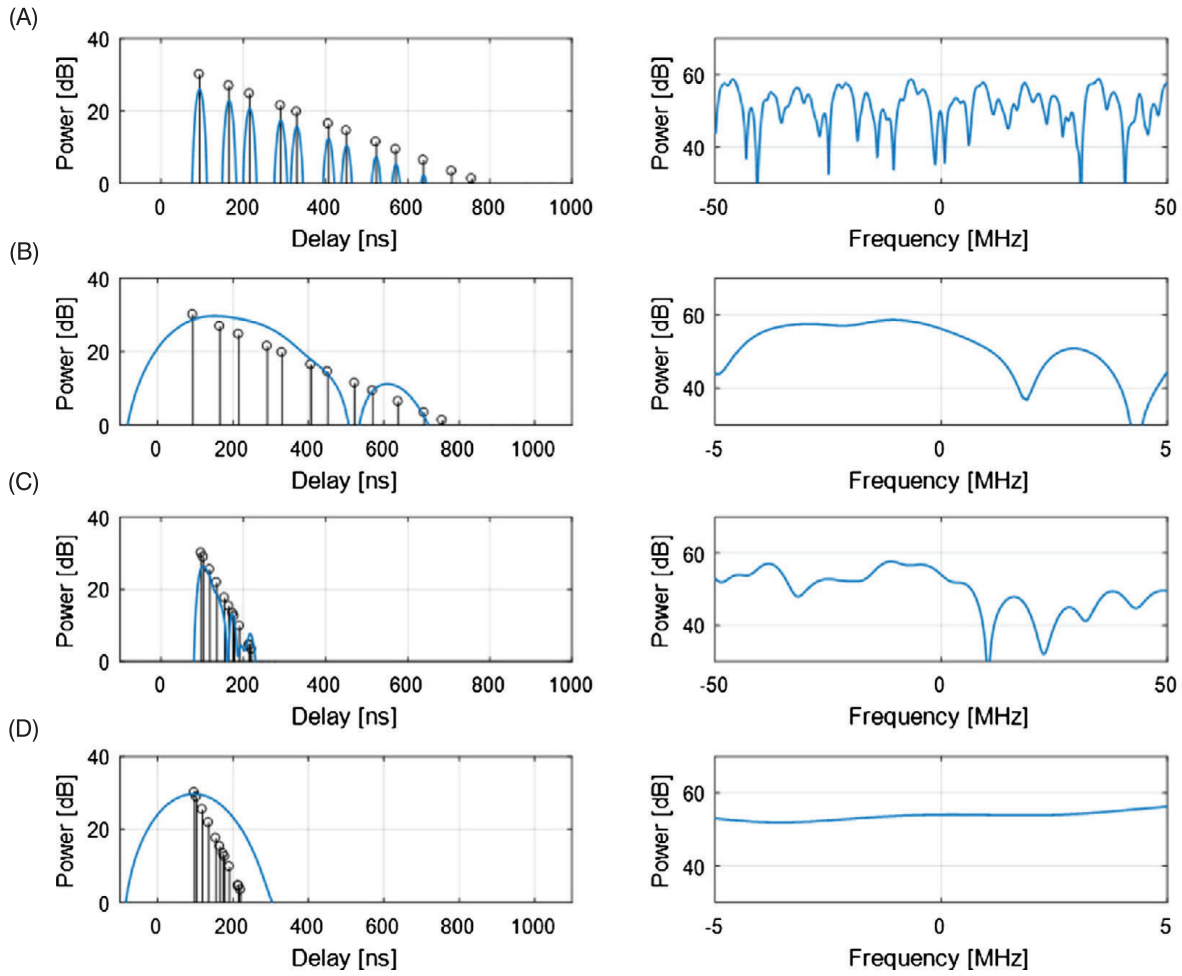


FIGURE 3.7

Channel responses in delay and frequency domains for $\sigma_\tau = 100$ ns, (A) and (B), and $\sigma_\tau = 20$ ns, (C) and (D), for $B = 100$ MHz, (A) and (C), and $B = 10$ MHz, (B) and (D).

sponding multipath components,

$$\left.\begin{array}{rcl} h(t) & = & \int_{-\infty}^{\infty} H(f_{\mathrm{D}}) \exp(i2\pi f_{\mathrm{D}} t)\, df_{\mathrm{D}}, \\ H(f_{\mathrm{D}}) & = & \int_{-\infty}^{\infty} h(t) \exp(-i2\pi f_{\mathrm{D}} t)\, dt \end{array}\right\}. \tag{3.18}$$

When modeling a channel in the time domain a sum of different waves having different Doppler frequencies is used. The main reason for having different Doppler frequencies is that either end, or both ends, of a radio link are moving causing a Doppler up-shift (down-shift) in frequency when the antenna is moving towards (away from) a wave. Another reason is that significant scatterers in the environment, like vehicles and trees, might be moving. We will illustrate this by a channel having a uniform directional distribution of waves in the horizontal plane around a moving terminal. The corresponding

Doppler distribution is referred to as the classical Doppler distribution; we have

$$\left.\begin{array}{lcl} h(t) & = & \sum_{l=1}^{N} a_l \exp\left(i2\pi t f_{\mathrm{D}_l}\right) \\ f_{\mathrm{D}_l} & = & \frac{v}{\lambda}\cos\left(\frac{2\pi l}{N}\right) \\ a_l & = & \exp\left(i\phi_l\right) \end{array}\right\} \tag{3.19}$$

where v is the terminal velocity. In analogy with the frequency-delay domain the coherence time is inversely proportional to the RMS Doppler spread $\sigma_{f_{\mathrm{D}}}$,

$$\left.\begin{array}{lcl} \mu_{f_{\mathrm{D}}} & = & \frac{\sum_{l=1}^{N} f_{\mathrm{D}_l}|a_l|^2}{\sum_{l=1}^{N}|a_l|^2} \\ \sigma_{f_{\mathrm{D}}} & = & \sqrt{\frac{\sum_{l=1}^{N}\left(\mu_{f_{\mathrm{D}}}-f_{\mathrm{D}_l}\right)^2|a_l|^2}{\sum_{l=1}^{N}|a_l|^2}} \end{array}\right\}. \tag{3.20}$$

In Fig. 3.8 a classical Doppler channel with maximum frequency $f_{\mathrm{D}_{\max}} = 100$ Hz is shown. The corresponding coherence time is 1 ms. In Fig. 3.8C and D the case with a stationary added path, having 10 dB higher power than the sum of the other paths, is also shown. This case corresponds to a scenario with stationary transmitter and receiver in an environment with substantial amount of moving scatterers, like a street with heavy traffic. For this case the coherence time is infinite, as the correlation in time never goes below 0.9.

3.2.3 DIRECTIONAL DOMAIN

The directional domain relates directly to the Doppler domain by

$$f_{\mathrm{D}_l} = \frac{\mathbf{v}\cdot\mathbf{k}_l}{2\pi} = \frac{v}{\lambda}\cos\left(\theta_l\right) \tag{3.21}$$

where θ_l is the angle between the velocity $\mathbf{v}$ and the wave vector $\mathbf{k}_l$. The direction information may be obtained from Doppler information in terms of the $\mathbf{k}$ vector,

$$\left.\begin{array}{lcl} k_{x,y,z} & = & 2\pi\frac{f_{\mathrm{D}_{x,y,z}}}{v_{x,y,z}} \\ u_{x,y,z} & = & \lambda\frac{f_{\mathrm{D}_{x,y,z}}}{v_{x,y,z}} \end{array}\right\} \tag{3.22}$$

where $\mathbf{u}$ is the direction unit vector. In practice, the corresponding directional spectrum is obtained by performing a Fourier transformation of three dimensional spatial channel samples as described in Section 3.3.2.1.

The directional spread is commonly characterized by means of the angular spread in azimuth and elevation (or polar) angles. Angle spreads are problematic, however, as they are both cyclic and non-Euclidian. To overcome the problem that angles are cyclic the cut in angle range may be placed at the point that minimizes the spread. The other problem—that angles are non-Euclidian variables—is more serious. When characterizing the channel by azimuth and elevation spreads the channel is not invariant under rotation of the coordinate system.

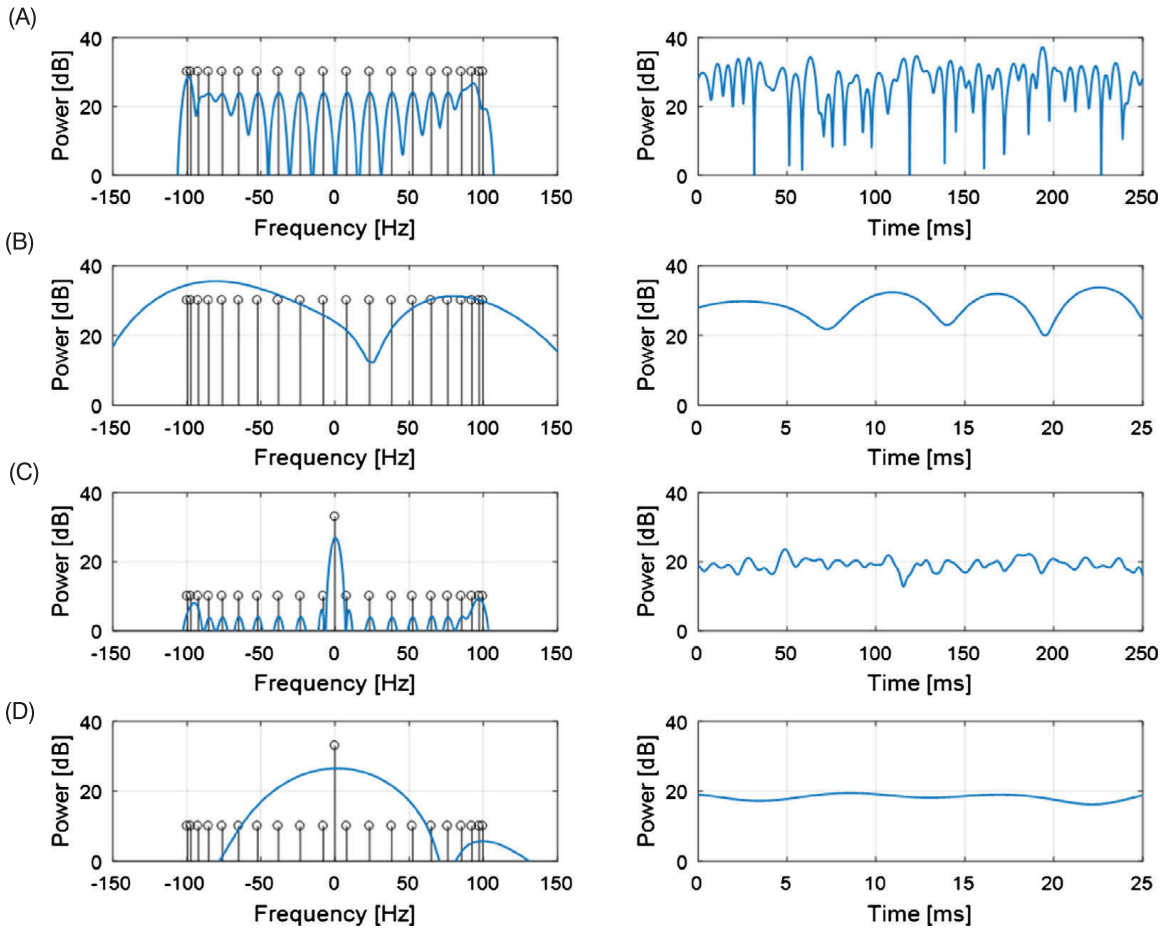


FIGURE 3.8

Channel responses in Doppler and time domains for classical Doppler with $\sigma_{f_D} = 70$ Hz, (A) and (B), and for classical Doppler with a stationary added path with $\sigma_{f_D} = 20$ Hz, (C) and (D), for $T = 250$ ms, (A) and (C), and $T = 25$ ms, (B) and (D).

An alternative directional spread definition, proposed in [19], does not suffer from these problems. It is based on a Doppler spread in three dimensions normalized so that it is equal to the angular spread when the spread is small. This directional spread σ_{dir} is defined by

$$\left.\begin{aligned} \mu_{u_n} &= \frac{\sum_{l=1}^{N} u_{n,l}|a_l|^2}{\sum_{l=1}^{N}|a_l|^2} \\ \sigma_{\text{dir}_n} &= \frac{180}{\pi}\sqrt{\frac{\sum_{l=1}^{N}\left(\mu_{u_n}-u_{n,l}\right)^2|a_l|^2}{\sum_{l=1}^{N}|a_l|^2}} \\ \sigma_{\text{dir}} &= \sqrt{\sigma^2_{\text{dir}_x}+\sigma^2_{\text{dir}_y}+\sigma^2_{\text{dir}_z}} \end{aligned}\right\} \tag{3.23}$$

where n corresponds to the three spatial components (x, y, z) representing the spreads along the corresponding directions. It should be noted that the total spread σ_{dir} is invariant with respect to coordinate system rotations.

3.3 EXPERIMENTAL CHANNEL CHARACTERISTICS

The previous section laid the theoretical basis of radiowave propagation. This section deals with the experimental results needed for understanding and characterizing real mobile communications propagation channels. For this purpose one needs to understand both measurement techniques and analysis methods. For a complete characterization of the channel both frequency-delay and Doppler-time domains are required. Moreover, to understand and utilize antenna characteristics like pattern, lobewidth and MIMO, the channel needs to be characterized in the spatial/directional domain. However, this domain is actually equivalent to the Doppler-time domain as shown in the previous section. All these aspects are addressed in some depth in this section.

3.3.1 MEASUREMENT TECHNIQUES

There are several different techniques for propagation measurements which are more or less advanced. The corresponding hardware designs may be highly complex. Here the most commonly used types of equipment and techniques used for measuring a radio channel response and transmission loss are briefly described.

3.3.1.1 Continuous Wave

The continuous wave (CW) method is commonly used for measuring transmission loss only. It is based on transmitting a sine wave signal at a fixed frequency using a narrow frequency filter at the receiver. Combined with a high transmit power and a low noise amplifier at the receiver a very high sensitivity may be obtained. The hardware solutions are typically relatively compact and simple, enabling extensive sampling of the signal strength over large areas in a fast and convenient way. One drawback is, however, that the multipath of the channel is not resolved resulting in substantial spatial fading.

Fig. 3.9 shows a measured signal CW signal at 5.1 GHz from a measurement route in a street microcell under non line-of-sight (NLoS) conditions. In order to reduce the fading due to the multipath effect a sliding average over 1.7 m has been applied.

3.3.1.2 Vector Network Analyzer

In contrast to CW measurements, a vector network analyzer (VNA) allows maximum possible measurement bandwidths. The basic principle is to perform a frequency swept sampling of the channel over a predefined bandwidth. For providing corresponding channel responses in the delay domain, it is convenient to use discrete Fourier transformation (DFT) methods. As the VNA measurement principle is based on measuring fully coherent ratios between the transmitted and received signal both the transmitter and the receiver antennas have to be connected to the VNA with RF cables. The advantage is that the received signal is fully synchronized with the transmitted signal, which enables absolute delay measurements and coherent averaging over long times for the suppression of noise. A substantial drawback is that the mobility is limited by the RF cabling. Moreover, a single frequency sweep may

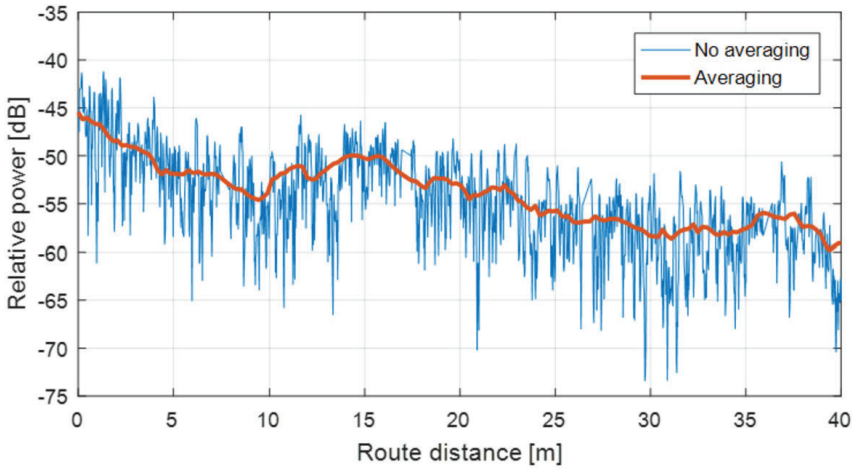


FIGURE 3.9

Continuous wave measurement at 5.1 GHz in a street microcell under NLoS condition.

take several seconds depending on the SNR at the receiver location which requires that both ends of the link are stationary. Another drawback is that the range is limited due to signal attenuation by the RF cables, which may be several dB per meter depending on the radio frequency. The range may be extended considerably to many kilometers, however, by replacing the RF cables by optical fibers using RF-to-opto and opto-to-RF converters. In Fig. 3.10 the measured channel response at 58.7 GHz using 2 GHz bandwidth is shown for a NLoS microcellular street scenario [22]. In this specific measurement an optical fiber was used to extend the range to more than 100 m. The large bandwidth results in a large fraction of the rich multipath of this channel being resolved.

3.3.1.3 Correlation-Based Channel Sounding

The main advantage with correlation-based channel sounders is that they are both mobile and wideband. To achieve this, a dedicated sounding signal is transmitted periodically and then correlated as a function of delay at the receiver. Typically OFDM or pseudo-random sequences are used together with a sliding correlator in delay. This technique allows both mobile and wideband measurements. The trade-off is between channel sampling frequency and noise suppression. Moreover, as the transmitted signal is modulated in both phase and amplitude, there are limitations due to amplifier non-linearities. More details of this type of equipment are presented in [32] and [15].

3.3.1.4 Directional Characteristics

The directional characteristics of the propagation channel are of particular interest when going up in carrier frequency into the millimeter-wave range. At these frequencies the use of omnidirectional antennas limits the possible range due to the substantial increase of transmission loss as a result of the decreasing antenna apertures. For this reason beam-forming techniques are required for focusing the transmission and reception in propagation directions which minimize the loss. There are basically two methods used for this purpose in propagation measurements.

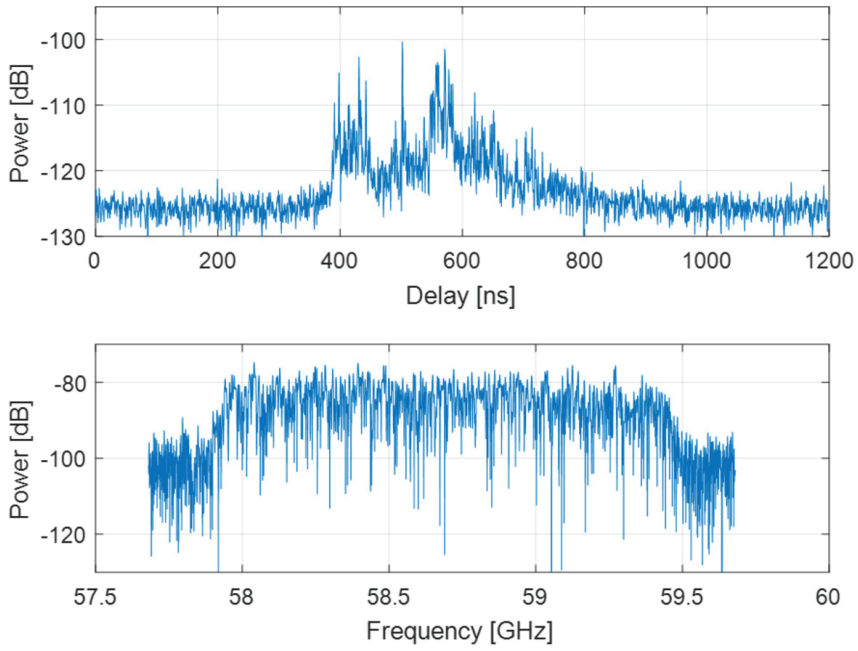


FIGURE 3.10

VNA channel measurement at 58.7 GHz and 2 GHz BW in a street microcell under NLoS condition. The baseband power delay profile is shown together with the channel response in the passband.

The first method is to use physical directive antennas, such as horn and parabolic reflector antennas, which are rotated in elevation and azimuth to scan the space angle. This method is not sensitive to time-varying channel conditions and is therefore suitable for CW and correlation-based channel sounding.

The second method is to use the so-called virtual antenna method. This method utilizes space sampling of the propagation channel where a single physical antenna is moved to different space positions by means of a robotic antenna positioning system. The directional characteristics of the measured channel are then determined off-line by means of array antenna techniques. The advantage of this method is that very high resolution and suppression of side-lobes may be achieved. The drawback is that the spatial sampling means long measurement times, up to many hours for large arrays, and therefore also it requires phase locked transmitter and receiver and stationary channel conditions. This method is suitable for VNA-based channel sounding.

3.3.2 ANALYSIS METHODS

Acquiring good quality measurement data requires considerable effort and skill. The raw measurement data is, however, of little use without thorough and accurate analysis. Providing reliable and accurate analyzed results takes even more effort than performing the actual measurements. In this section some commonly used analysis methods and their corresponding advantages and disadvantages are described. Moreover, requirements for providing comparability between different measurements and/or frequency ranges are provided.

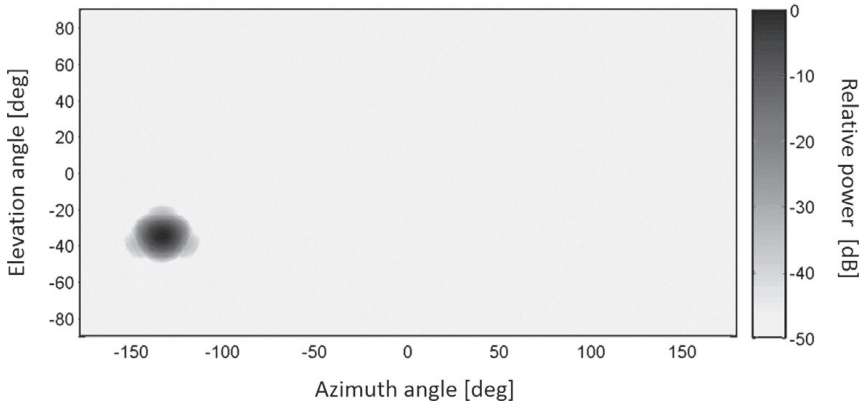


FIGURE 3.11

Simulated response of a plane wave impinging from $-135°$ in azimuth and $-35°$ in elevation for a cubic virtual array.

3.3.2.1 Spectral Analysis

The spectral methods are based on direct analysis of the measured signal and corresponding power distributions using Fourier techniques for transformations between frequency and delay, and between Doppler and time domains, as described in Sections 3.2.2 and 3.2.1. For directional analysis, direct angular channel sampling using directive physical antennas or virtual antennas may be used.

For a stationary channel, the spatial samples are analogous to time samples of a moving terminal. When the channel is sampled in three spatial dimensions it is possible to determine the corresponding power spectrum for all three components of the wave vector $\mathbf{k}$.

One example of this method using a cubic virtual antenna of $25^3 = 625$ samples is shown in Fig. 3.11. The space samples are transformed to the $\mathbf{k}$ domain by DFT using a Hann filter over all three space dimensions for reducing the side-lobes. The corresponding directional spectrum is obtained by filtering out values from the $\mathbf{k}$ domain cube with a fixed radius $\sqrt{k_x^2 + k_y^2 + k_z^2} = |\mathbf{k}|$. The performance of this method is impressive, providing side-lobe suppression of more than 50 dB. This is in contrast to the use of physical antennas for which the side-lobe level typically is suppressed by less than 30 dB. To bring the measured channel to the discrete format of (3.7) each peak of the multidimensional spectrally measured channel data is identified as a multipath component with corresponding phase and amplitude [21]. An example for an indoor NLoS scenario is shown in Fig. 3.12, where the directly measured channel is shown together with corresponding synthesized channel using 400 estimated multipath components. It is clear that the modeled channel agrees well with the directly measured channel.

3.3.2.2 Superresolution Methods

Assuming that the discrete plane wave model (3.7) is valid, the theoretically possible accuracy is limited only by the signal-to-noise ratio. There are many superresolution methods reported in the literature, of which the most popular ones are based on maximization of the likelihood P. The corresponding log-

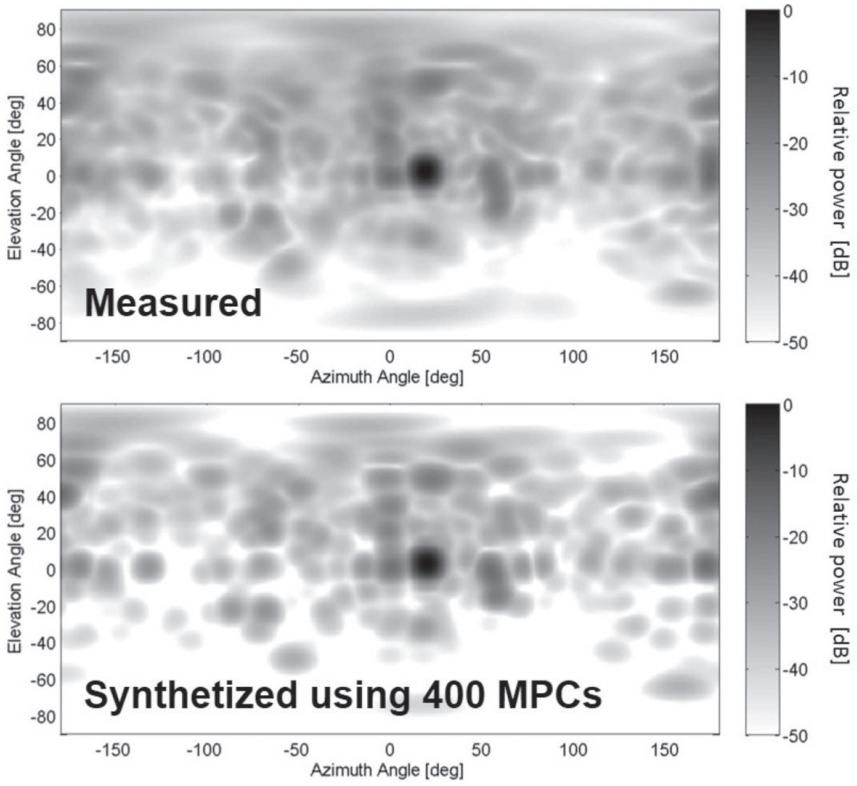


FIGURE 3.12

Measured indoor LoS channel at 60 GHz and corresponding synthesized channel using 400 MPCs for a cubic virtual array.

likelihood function to minimize is given by

$$-\log P = N_{\mathrm{rx}} N_{\mathrm{tx}} N_{\mathrm{f}} \log(\pi\sigma^2) + \frac{1}{\sigma^2} \sum_{m=1}^{N_{\mathrm{rx}}} \sum_{n=1}^{N_{\mathrm{tx}}} \sum_{k=1}^{N_{\mathrm{f}}} \left| \tilde{H}_{\mathrm{mnk}} - H_{\mathrm{mnk}} \right|^2 \tag{3.24}$$

where m and n are the space samples in the receive and transmit ends, respectively, k is the index over frequency, $\tilde{H}_{\mathrm{mnk}}$ and H_{mnk} are the modeled and measured channel responses, respectively, and σ^2 represents the power of the noise which is assumed to follow a zero mean complex Gaussian distribution over the channel samples. In order to find the most probable set of plane waves, which would mimic the measured channel best, the log-likelihood function is minimized with respect to the model parameters.

A free search over all model parameters and measurement samples is practically not possible due to the huge computational effort. One common method for reducing this effort is SAGE [5], which is based on maximizing the likelihood for one parameter at a time and iterate until the minimum is found. The problem with this method is that plane waves, which are closely spaced in angle or delay, are strongly correlated. As a consequence the increase in computational effort is very large and therefore the convergence is very slow. This problem has been addressed by means of gradient methods, like

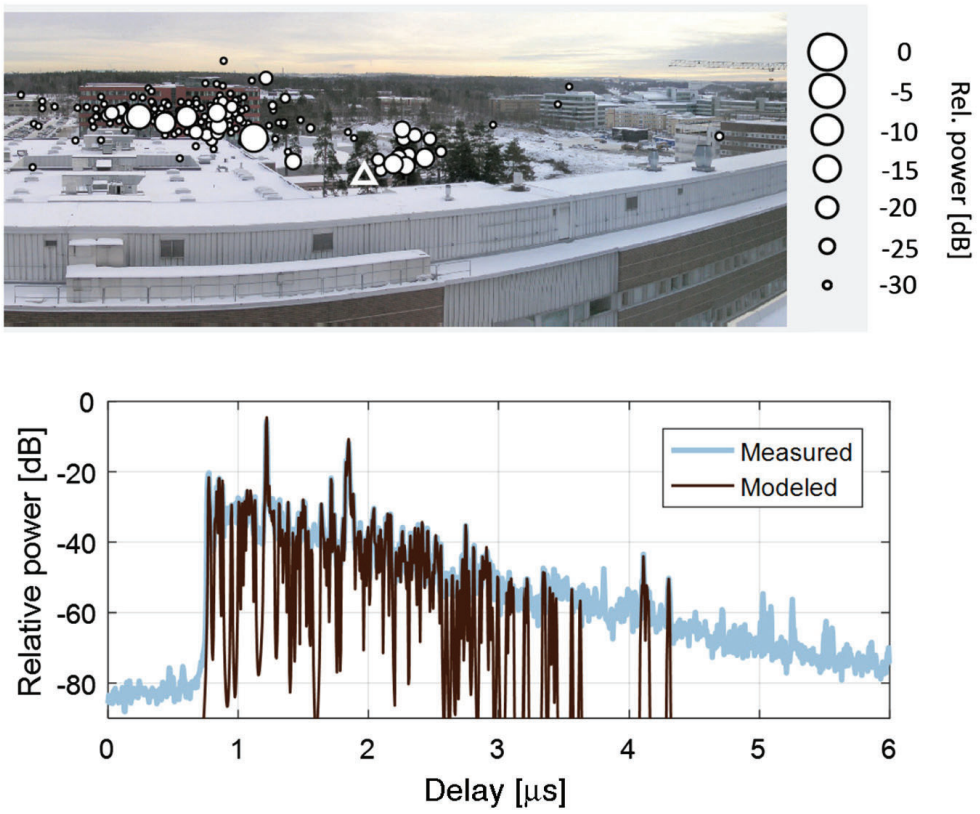


FIGURE 3.13

The lower graph shows power delay profiles from a NLoS urban macrocell scenario at 5 GHz. Both the directly measured profile (upper light) and the profile based on the super-resolved estimates (lower dark) are shown. The upper graph shows the directions of the estimated plane waves as viewed at the BS location where the triangle indicates the direction of the mobile station.

RIMAX [34], [30], which utilize differentiation around local likelihood maxima to achieve fast convergence of the correlated parameters. The problem with this method is, however, that some correlation or coupling between most of the multipath components remains, meaning that the likelihood needs to be maximized simultaneously for all waves. A method to decouple the multipath components is proposed in [17]. As the method also provides reduction of the data size and parameter space, many orders of magnitude better computational efficiency is obtained. An example of the application of this method is shown in Fig. 3.13 which is from an urban macrocell scenario at 5.1 GHz where a virtual planar array of 10x25 elements with 2-cm spacing was used at the BS. It is clear that the super-resolved estimates account for the main part of the received power. Moreover, the main directions of the waves, as viewed from the BS, show that the most significant scatterers are trees and exterior building walls.

Though superresolution methods may provide a high accuracy of significant discrete multipath components, the method suffers from incapability of estimating the diffuse or dense multipath component of the channel. As this component may be substantial, superresolution methods commonly

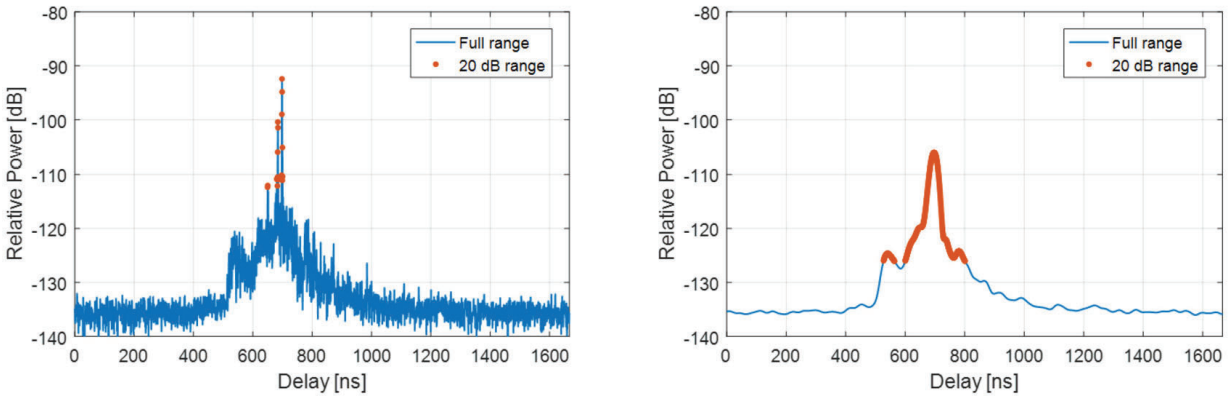


FIGURE 3.14

Power delay profiles using 2 GHz BW (left) and 80 MHz BW (right) in a NLoS microcellular scenario. Corresponding 20-dB dynamic ranges below the main peak are indicated. The resulting delay spreads are 7 ns for 2 GHz BW and 28 ns for 80 MHz BW.

give substantially inaccurate output. For this reason, the spectral methods are preferred for providing reliable output.

3.3.2.3 Measurement Comparability

The amount of propagation measurement campaigns performed worldwide is indeed extensive. Many of the corresponding results are compared and summarized in statistical analyses performed by the research community. In those analyses it is, however, important to be thorough and to fulfill the requirements for ensuring that the measurements are comparable. For this purpose a list of critical requirements which need to be fulfilled, for different campaigns and frequency bands, has been identified [28]:

- Equal measurement bandwidth (providing equal delay resolution)
- Comparable antenna pattern, either physical or synthesized
- Equal dynamic power range in the respective domain of analysis (e.g., delay, angle)
- Same environment and same antenna locations (for comparing different frequency bands)

It has e.g. been found that the requirement of equal bandwidths is critical for avoiding a fictitious decrease of delay spread with increasing frequency. The basic problem is that substantially larger measurement bandwidths are available at higher frequencies in the millimeter-wave range, wherefore those measurements typically are performed using substantially larger bandwidths. The need for equalizing the bandwidths used in the analysis is illustrated in Fig. 3.14 where a 20-dB dynamic range below the main peak has been used for determining the RMS delay spread. The impact of the different used bandwidths is considerable, giving delay spreads of 7 ns and 28 ns for 2 GHz and 80 MHz bandwidths, respectively. The reason why the equalization of bandwidth is required is that any strong discrete multipath component that is resolved has a peak power proportional to the used bandwidth, while the non-resolved components stay at a constant level irrespective of the bandwidth used. As a fixed dynamic power range below the main peak typically is used in the analysis, the difference of using different bandwidths is substantial.

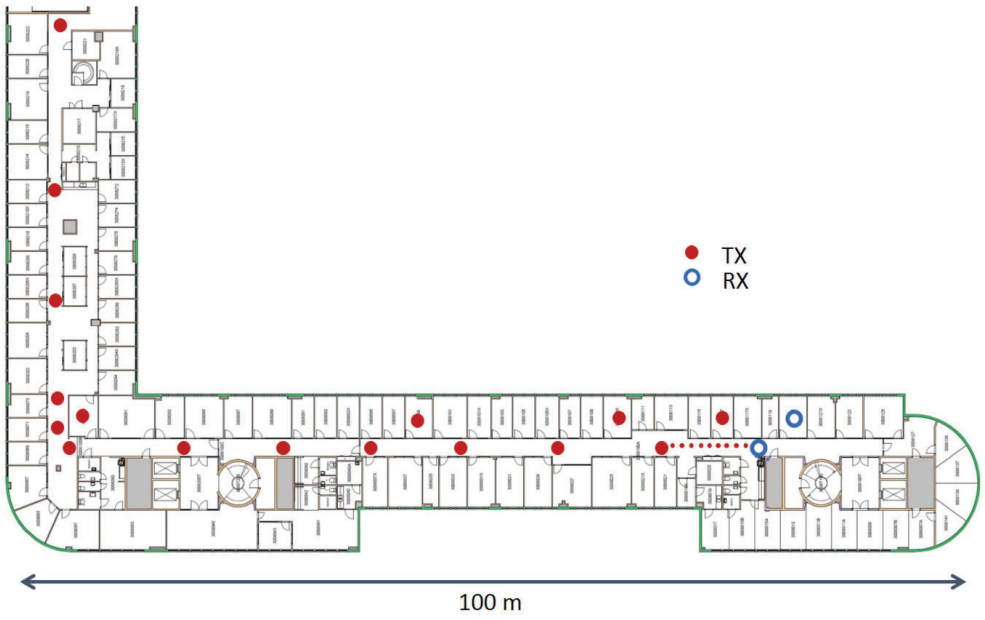


FIGURE 3.15

Floorplan of indoor office measurement scenario.

3.3.3 TRANSMISSION LOSS MEASUREMENTS

The loss in received signal strength due to propagation effects is one of the most basic and one of the most important characteristics of the radio propagation channel. In this section the current empirical understanding of transmission loss is presented and illustrated with a number of measurement examples from real cellular scenarios with focus on the frequency dependency in the range 1–100 GHz. Most of the measurements have been performed using vertical omni dipole antennas having very similar patterns for all frequencies. Additionally, vertical patch antennas or open waveguides have been used in outdoor-to-indoor measurements at the outdoor transmitter location. As measurements performed around 60 GHz are subject to oxygen absorption of about 1.5 dB/100 m, this loss has been compensated for in those measurements. The reason for this is the desire to facilitate the modeling by making possible smooth interpolation/extrapolation over the full frequency range and, depending on the need, adding the oxygen absorption to the baseline model. Moreover, all measurement data is provided in terms of the loss in excess of free-space loss in order to avoid any impact of antenna frequency dependence and to focus on pure propagation effects. For this purpose, all measurement data is carefully calibrated by line-of-sight (LoS) short range (0.1–1.0 m) measurements.

3.3.3.1 Indoor Office Scenario

This measurement example is from an indoor office environment. The basic layout shown in Fig. 3.15 is a corridor with office rooms along both sides. At the end of the corridor there is a 90 degrees turn. The receive (Rx) antenna is placed at two locations, one in the corridor and one inside an adjacent office room. The transmit antenna is placed at different locations both in the corridor and inside office

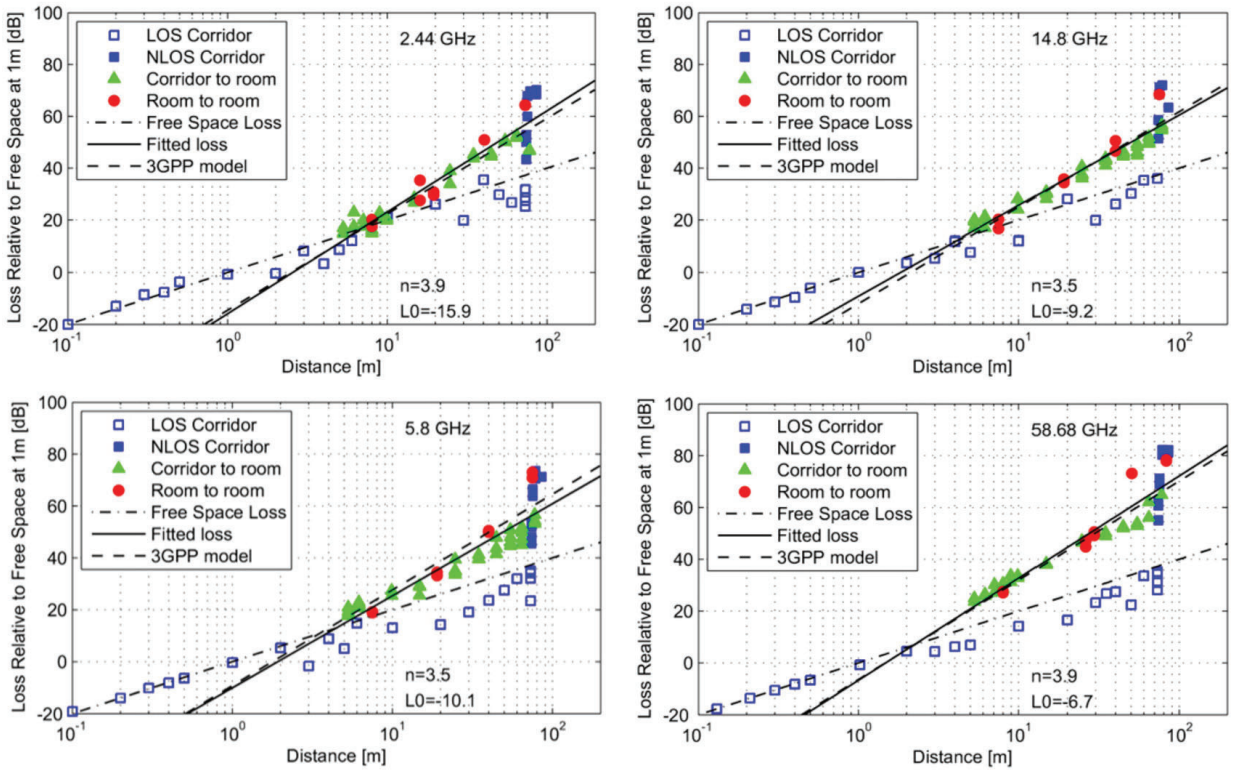


FIGURE 3.16

Transmission loss, in excess free-space loss, versus distance for 2.44, 5.8, 14.8, and 58.68 GHz measured in the indoor office environment.

rooms. The exterior walls of the building are made of brick and the interior walls of plasterboard and glass.

Fig. 3.16 summarizes the main results of the transmission loss analysis. The loss in dB units, L, relative to free-space power at 1-m distance is plotted for the frequencies, 2.44 GHz, 5.8 GHz. 14.8 GHz, and 58.68 GHz, for the different LoS and NLoS scenarios. A two parameter exponent model,

$$L = 10n\log(d) + L_0, \tag{3.25}$$

has been fitted to the measured loss, in dB units, where d is the distance between transmitter and receiver in meters. The corresponding model by 3GPP [2] accounts also for the frequency dependence,

$$L_{3GPP} = 38.3\log(d) - 15.1 + 4.9\log f \tag{3.26}$$

where f is the carrier frequency in GHz. The two last terms correspond to the frequency dependent term L_0 in (3.25). It is clear from the figure that the measurement data agrees very well with the 3GPP model. Moreover, the frequency trend is clear as the propagation loss increases about 5 dB per decade in addition to the free-space loss.

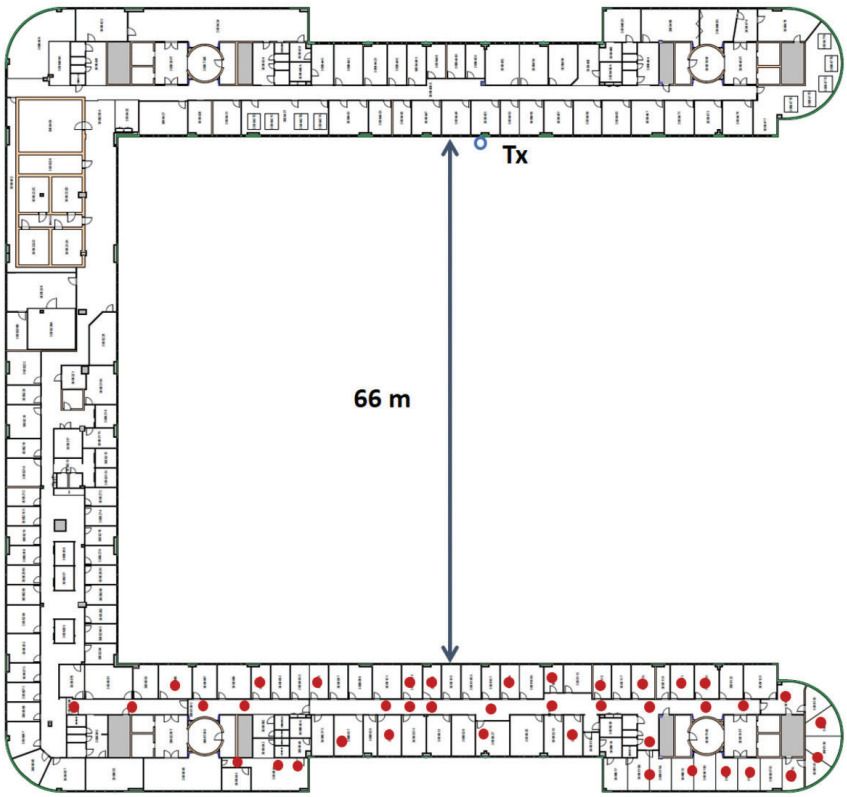


FIGURE 3.17

Outdoor-to-indoor measurement scenario. The indoor Rx locations are marked with filled circles and the Tx location is marked with an open circle.

3.3.3.2 Outdoor-to-Indoor Scenario

An outdoor-to-indoor multifrequency measurement campaign has been performed in an eight floors office building in an urban environment as depicted in Fig. 3.17. The transmitter is located in an open window at the top floor of the building and the received signal is measured at two slightly shifted (30 cm) positions at 40 indoor locations across the inner yard on the same floor. At the top floor of the building the exterior wall is covered with metal. The windows are, however, pure glass without metalization.

Between 2.44 GHz and 14.8 GHz the building penetration loss ranges from around 0 dB up to 30 dB (Fig. 3.18). The lower end of penetration loss around 0 dB is similar for all frequencies, while the highest losses around 45 dB occur only at 58.68 GHz. The minimum loss, due to penetration of the exterior wall/window only, is in the range 0–5 dB with the highest values for 5.8 GHz and 58.68 GHz. This non-monotonic dependence on frequency may be explained assuming that the three layers of glass, in the window frames, cause constructive or destructive interference, as an effect of multiple reflections, resulting in periodic varying attenuation as a function of frequency. Subsequent measurements show that the window loss is about 2, 10, 0, and 6 dB at 2.44, 5.8, 14.8, and 58.68 GHz, respectively, which confirms this effect and explains the measured minimum penetration loss. Moreover, it is clear

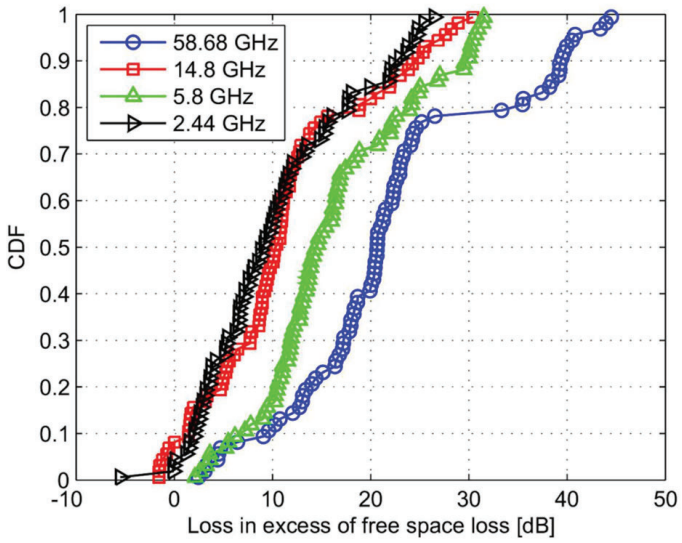


FIGURE 3.18

CDFs of measured outdoor-to-indoor loss, in excess of free-space loss, for the different frequencies.

that the spread of penetration loss is substantially larger for the higher frequencies. This may partly be the result of the venetian blinds, in some of the windows, which block the vertically polarized waves at the higher frequencies but are transparent at the lower frequencies. Regarding median loss, the measurement results agree very well with the corresponding 3GPP model [2]. It should be noted that the 3GPP model does not account for any non-monotonic frequency effects. Moreover, there is an increasing spread of the measured loss with increasing frequency, which also is not accounted for in the 3GPP model. Further details on outdoor to indoor propagation modeling are given in the channel modeling section, Section 3.4.1.1.

3.3.3.3 Outdoor Street Scenario

Outdoor street measurements have been made in an urban area consisting of mainly modern office building blocks of about 100 m length and 25 m height (Fig. 3.20). The measurements were performed in both LoS and NLoS in a street canyon, of about 20 m width, with both transmitter and receiver antennas located about 1.5 m above ground. In Fig. 3.19 the excess loss is shown for all frequencies. In LoS a multipath gain of up to 5 dB (relative to free space) is observed, which is similar at all frequencies. This gain is due to additional paths from reflections off the ground and exterior walls. In the NLoS region behind the corner of the building a substantial increase in the excess loss is observed. This loss is substantially lower than what is expected by knife edge diffraction at the corner, as indicated in Fig. 3.19. Further, the frequency dependence is much weaker than what is expected from diffraction. This result suggests that the dominating propagation mechanisms in NLoS must be different from diffraction, e.g. specular and/or diffuse scattering by objects or rough exterior walls. Moreover, it is clear that the oxygen compensation at 60 GHz is substantial, up to 4 dB, for the NLoS data, which is more than what is expected from the link distance only. This is, however, explained by the lengths of significant reflected propagation paths being substantially larger than the link distance. The measured

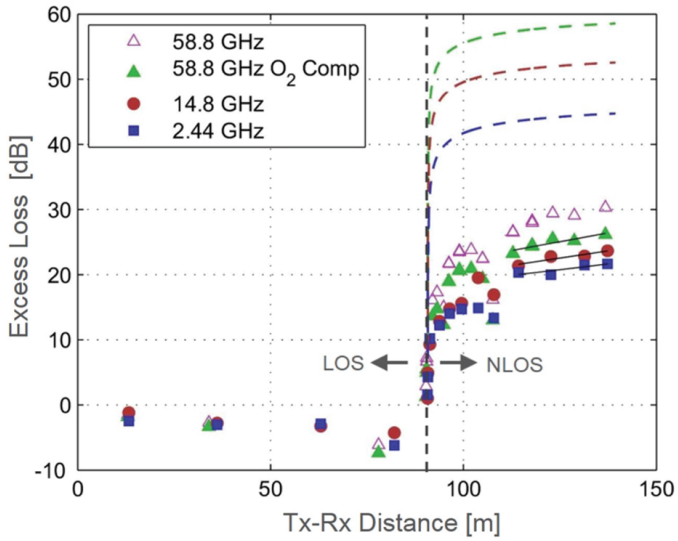


FIGURE 3.19

Measured loss versus distance for RX2 for the scenario shown in Fig. 3.20. The corresponding loss by knife edge diffraction is marked with dashed lines.

frequency dependence of the excess loss (about 3 log f [dB]) is clearly less than what is expected by knife edge diffraction (about 10 log f [dB]). It is, however, somewhat higher than that of the 3GPP channel model (1.3 log f [dB]) reported in [2].

In order to get some further insight into the propagation mechanisms, visual ray tracing has been performed for two measurement locations at 60 GHz as shown in Fig. 3.20. The first transmitter location (TX1) is in NLoS but very close to LoS. The first arriving path is attenuated by diffraction. It is possible to reconstruct the pathway of the strongest path assuming one specular reflection off an exterior building wall along the street. The second transmitter location (TX2) is substantially further down the street into the NLoS region. At the delay corresponding to the path diffracted around the corner, no signal above the noise floor is observed. The first cluster of weak paths is observed at propagation distances substantially longer than the diffraction path length. This cluster is likely to be caused by scatterers and/or rough surfaces in the area of the street corner. The strongest peak stands out having around 20 dB higher power level than the rest of the power delay profile. A plausible corresponding pathway (matching the propagation length of the measured peak) is possible to reconstruct assuming four specular reflections off exterior building walls. This shows that specular paths may be important even far into the NLoS region. However, for most of the NLoS locations such pronounced peaks were not observed.

3.3.3.4 Outdoor Urban Over Rooftop Scenario

To measure transmission loss for an urban outdoor over rooftop propagation scenario (macrocell scenario) is very challenging at higher frequencies, particularly in the millimeter-wave range as the transmission loss increases substantially when using practical omni antennas. This problem may be mitigated somewhat by using directive antennas at the BS location above the roof tops. However, at the UE locations on the ground, substantial angle spread is expected both in elevation and azimuth due to the UE being embedded in the urban clutter. To solve the problem by using high transmit power is

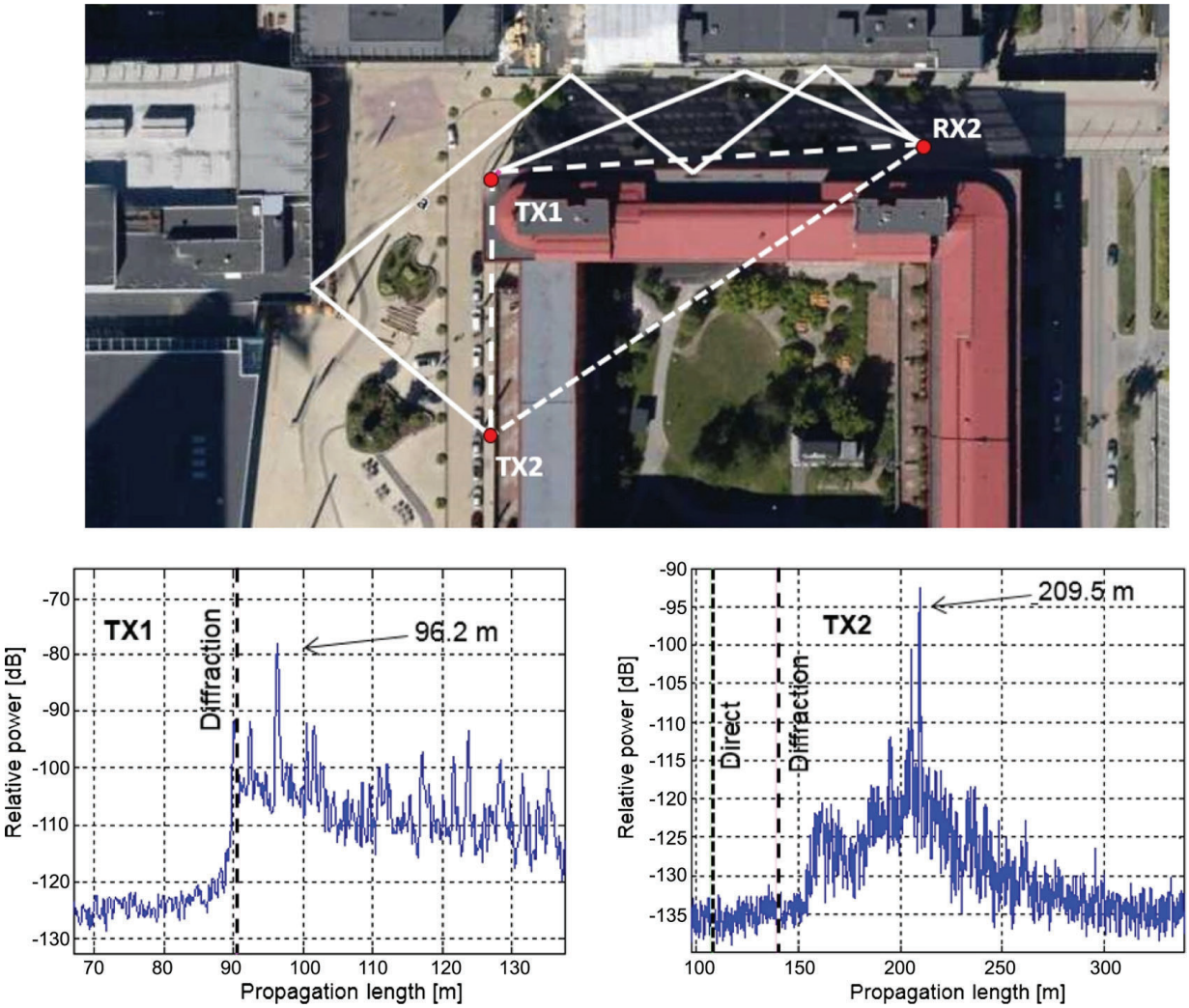


FIGURE 3.20

Power delay profiles in NLoS locations TX1 and TX2 at 60 GHz and corresponding reconstructed pathways.

practically not possible at millimeter-wave frequencies. A more practical solution is to use very high performance LNAs. Moreover, for CW measurements a very high suppression of noise is possible by means of Doppler filtering. This technique has been used in two independent multifrequency measurement campaigns in Aalborg [27] and Tokyo [23]. In the Aalborg campaign no frequency dependency was observed. This may, however, be the effect of a limited dynamic range at the highest frequency 28 GHz. In the Tokyo campaign the measurement sensitivity was better at the higher frequencies, allowing measurements up to 67 GHz.

The 3GPP modeling of urban macrocell transmission loss [2] is largely based on the Aalborg measurement campaign and is, relative to free-space, given by

$$L_{\mathrm{UMa}} = 19.08 \log(d) - 18.9 \ [\mathrm{dB}] \tag{3.27}$$

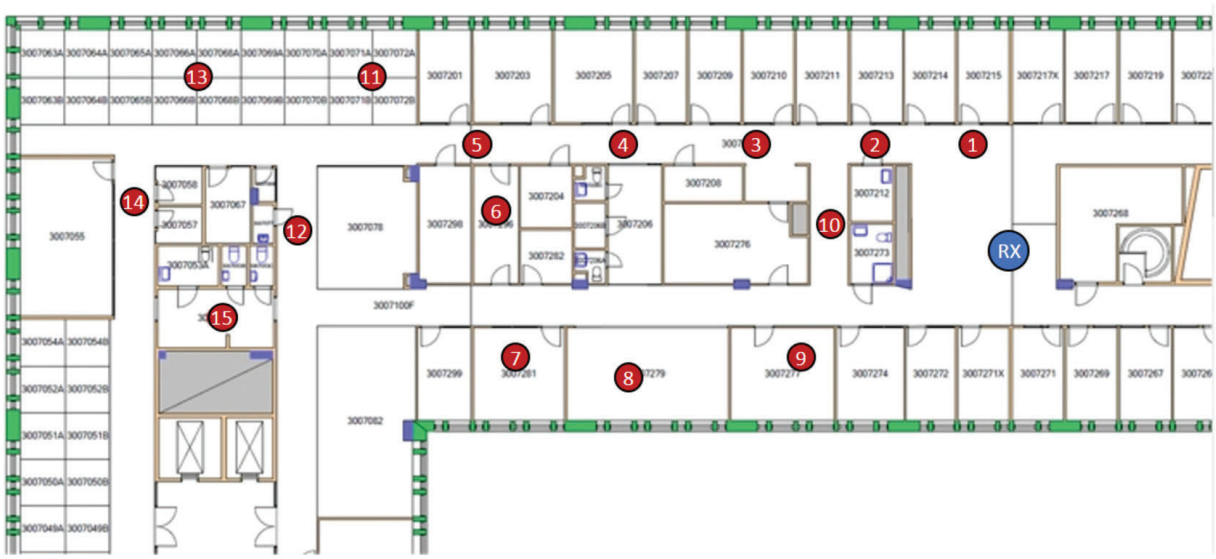


FIGURE 3.21

Indoor channel impulse response measurement scenario.

where d is the distance in meters and f the frequency in GHz. The corresponding model provided by ITU-R in [12] is based on the Tokyo measurements and is given by

$$L_{\text{UMa}} = 23.9\log(d) - 38.7 + 3.0\log f \text{ [dB]}. \tag{3.28}$$

Except that the ITU-R model is frequency dependent and the 3GPP model is not, they are similar. The empirical basis for whether the loss for urban macrocell scenarios in general is frequency dependent or not is a subject for further investigation.

3.3.4 DELAY DOMAIN MEASUREMENTS

The delay domain is important for providing the characterization of the frequency selectivity of the channel as described in Section 3.2.1. Moreover, it is critical for optimizing transmission waveforms with respect to delay spread (see Chapter 6). As 3GPP has chosen OFDM for NR, cyclic prefix length optimization is directly related to the delay spread of the channel. In this section measured delay domain properties for a wide frequency range and important propagation scenarios are presented. The general frequency trends and comparisons with the 3GPP channel model are saved for Section 3.3.4.4.

3.3.4.1 Indoor Office

A multifrequency measurement campaign has been performed in an indoor office scenario as shown in Fig. 3.21. The receiver was placed at a fixed location and the transmitter at 15 different locations mainly in NLoS. All requirements for measurement comparability over different frequencies (provided in Section 3.3.2.3) are fulfilled. The channel is measured at 2.4, 5.8, 14.8, and 58.7 GHz. In Fig. 3.22 corresponding power delay profiles of two exemplary Tx locations are shown together with the average

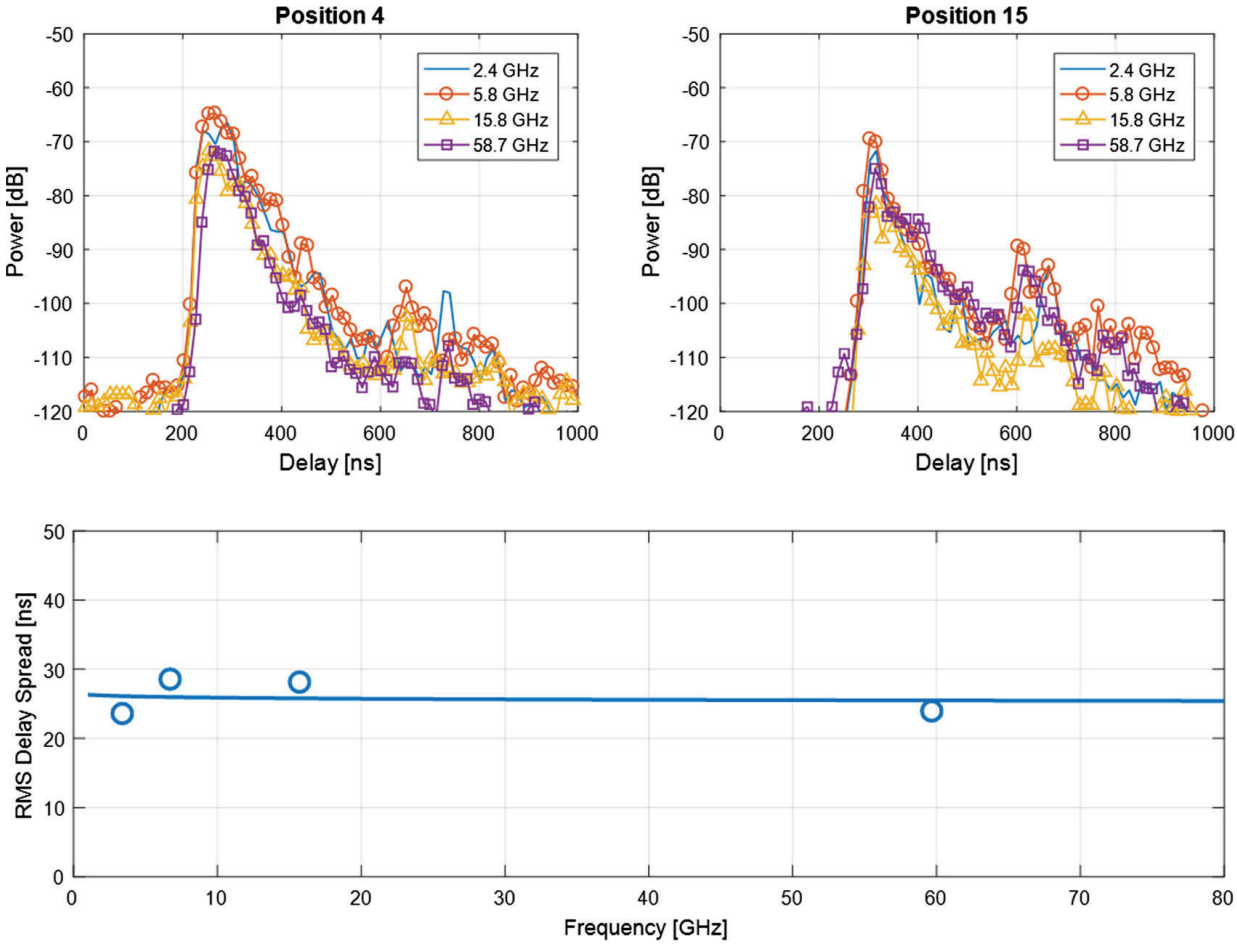


FIGURE 3.22

Power delay profiles for TX positions 4 and 14 (upper graphs) and average delay spread as a function of frequency together with fitted 3GPP type of model (lower graph).

RMS delay spread, σ_{DS}, as a function of frequency. No obvious frequency trend is observed, neither in power delay profiles nor in the average delay spread values. The independence is in fact well inside the 95% confidence range when fitting the 3GPP type of the model

$$\sigma_{\mathrm{DS}} = 10^{\beta}(1+f)^{\alpha} \tag{3.29}$$

where α and β are model parameters and f is the carrier frequency in GHz. The fitted values for the curve shown in Fig. 3.22 are $\alpha = -0.01 \pm 0.05$ and $\beta = -7.58$.

3.3.4.2 Outdoor-to-Indoor

The outdoor-to-indoor measurements described in Section 3.3.3.2 have been further analyzed to determine delay spread and the corresponding frequency dependency. It is here important to point out that the dynamic range of the impulse responses in many locations is limited, down to below 10 dB.

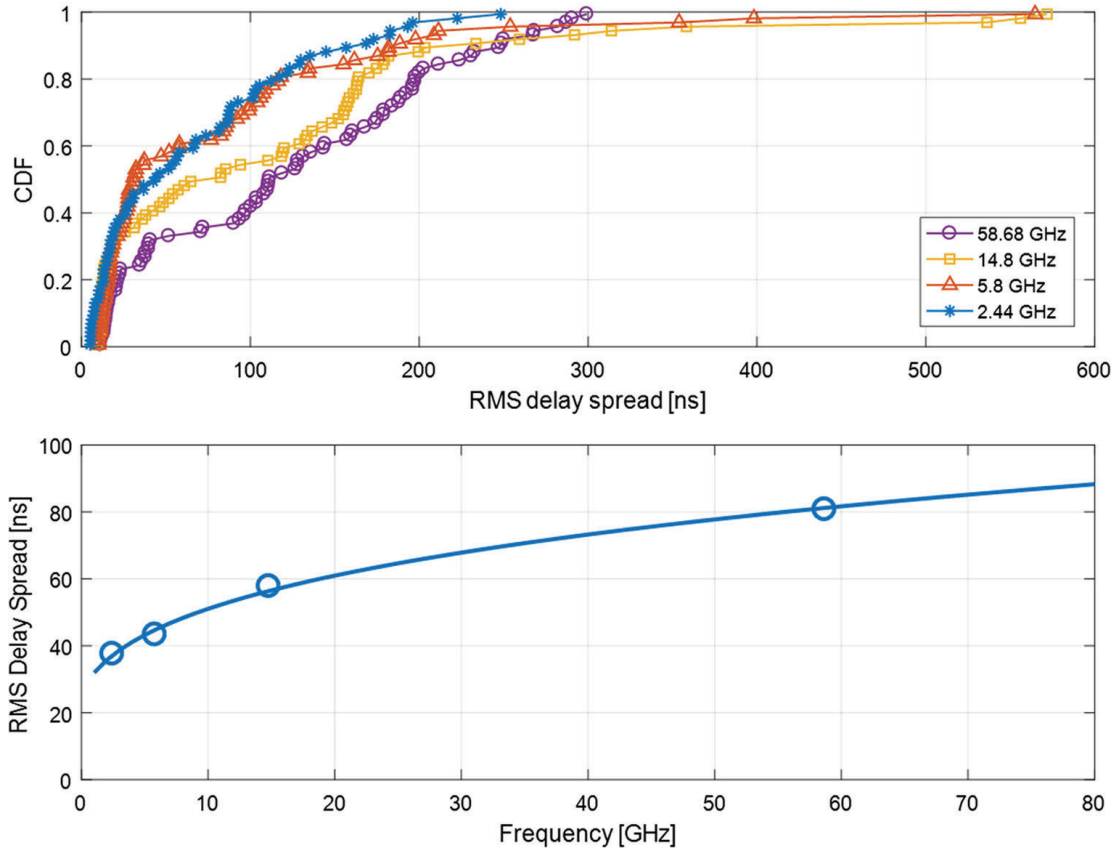


FIGURE 3.23

Outdoor-to-indoor CDFs of RMS delay spread for the different frequencies (upper graphs) and average delay spread as a function of frequency together with fitted 3GPP type of model (lower graph).

This means that the absolute delay spread sometimes may be underestimated. However, as the same dynamic range was used for all frequencies at each location no bias in the frequency dependency is introduced. The trend seems to be that the delay spread goes up with frequency as shown in Fig. 3.23. It is likely that the longer delays are caused by multiple reflections over the inner yard. Strong specular reflections are expected due to windows and/or metal tiles covering the exterior walls. However, these tiles and/or windows are smaller than one Fresnel zone at the lower frequencies, resulting in that the corresponding reflections are non-specular and therefore attenuated at the lower frequencies. It should be noted that this observed frequency trend is not general, but specific to the geometry of the scenario, as pointed out later in this section.

3.3.4.3 Outdoor Street Canyon Scenario

The outdoor street measurements described in Section 3.3.3.3 have also been analyzed to determine the corresponding delay spread characteristics with focus on the NLoS region. In this region the trend is that the delay spread seems to be independent of frequency, as shown in Fig. 3.24. The fitted model shows indeed that there is no significant frequency dependence.

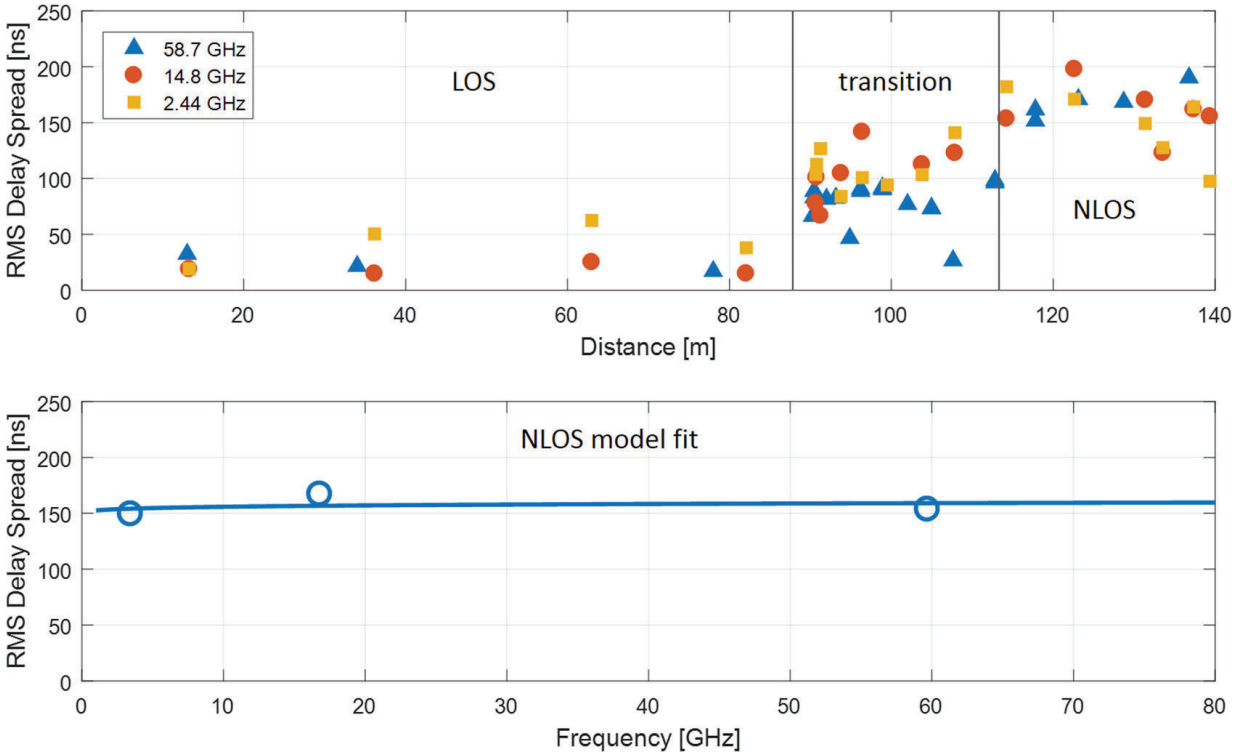


FIGURE 3.24

Outdoor street canyon RMS delay spread versus distance for the different frequencies (upper graph) and average delay spread as a function of frequency together with fitted 3GPP type of model (lower graph).

3.3.4.4 General Frequency Trend in Delay Domain

The experimental delay domain results presented so far indicate, except for the outdoor-to-indoor scenario, that there is no clear frequency trend. Previous results by e.g. 3GPP [2] indicate that the delay spread generally decreases when the frequency increases. However, when developing the 3GPP model the requirements for comparability between different frequency bands (provided in Section 3.3.2.3) were not thoroughly fulfilled, wherefore the corresponding results might be questioned. The EU funded project mmMAGIC [25], [26] has undertaken extensive channel measurements in which the requirements for comparability between different frequency bands were carefully seen to be fulfilled. Corresponding model parameters for five scenarios have been determined by statistically combining 15 independent measurement campaigns by six organizations. In Fig. 3.25 the 3GPP type of model (3.29) fitted to the mmMAGIC measurement data as well as the corresponding 3GPP model values are shown. There is a clear discrepancy between the two model fits, where the 3GPP model parameters show a strong general decrease of delay spread as a function of frequency, which is absent in the mmMAGIC data. Only street canyon LoS and indoor office in LoS show a slightly decreasing trend, within the 95% confidence range, for the mmMAGIC data.

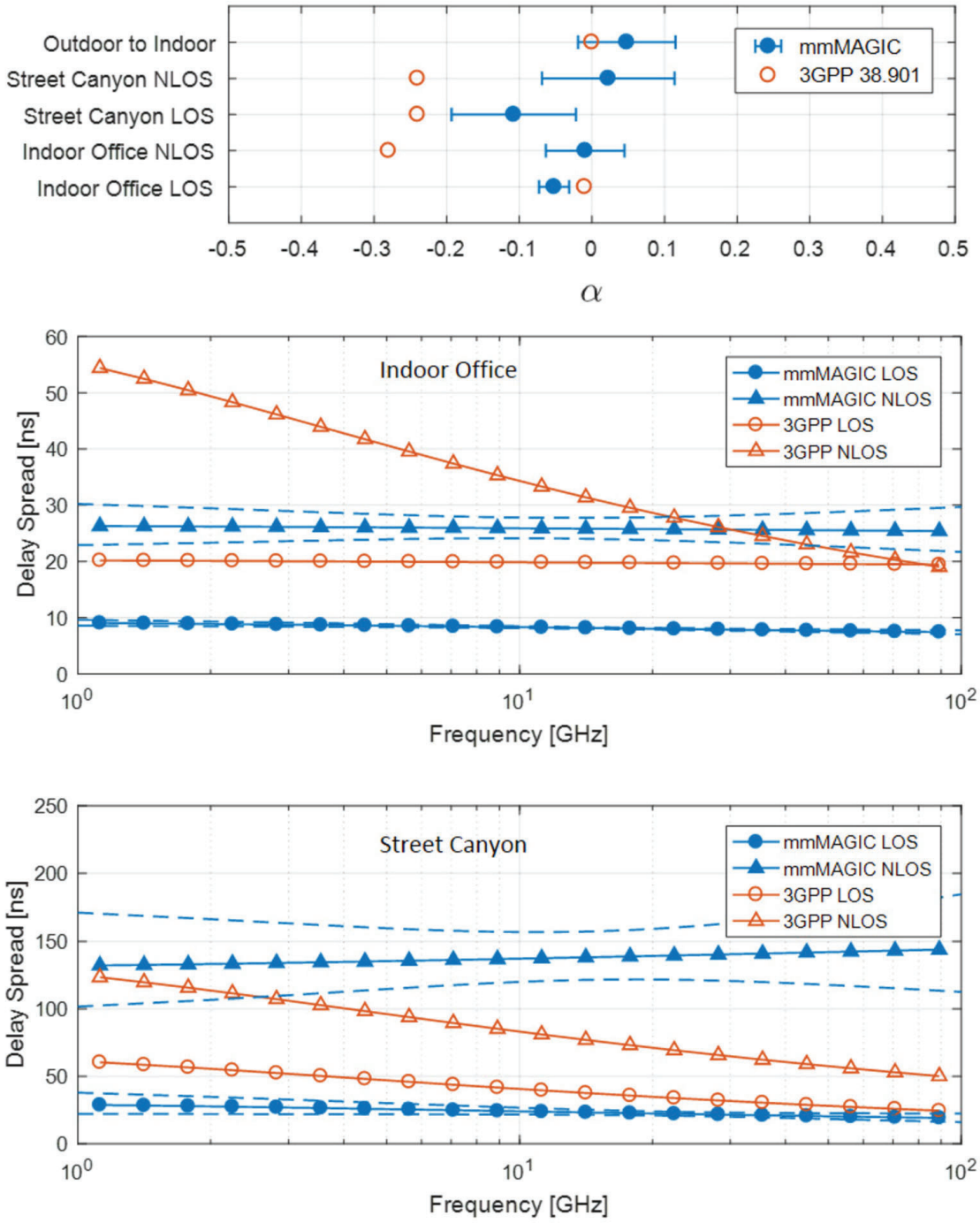


FIGURE 3.25

Delay spread model (3.29) fitted to mmMAGIC channel measurements and corresponding 3GPP model. The dashed lines indicate 95% the confidence limits for the mmMAGIC fits.

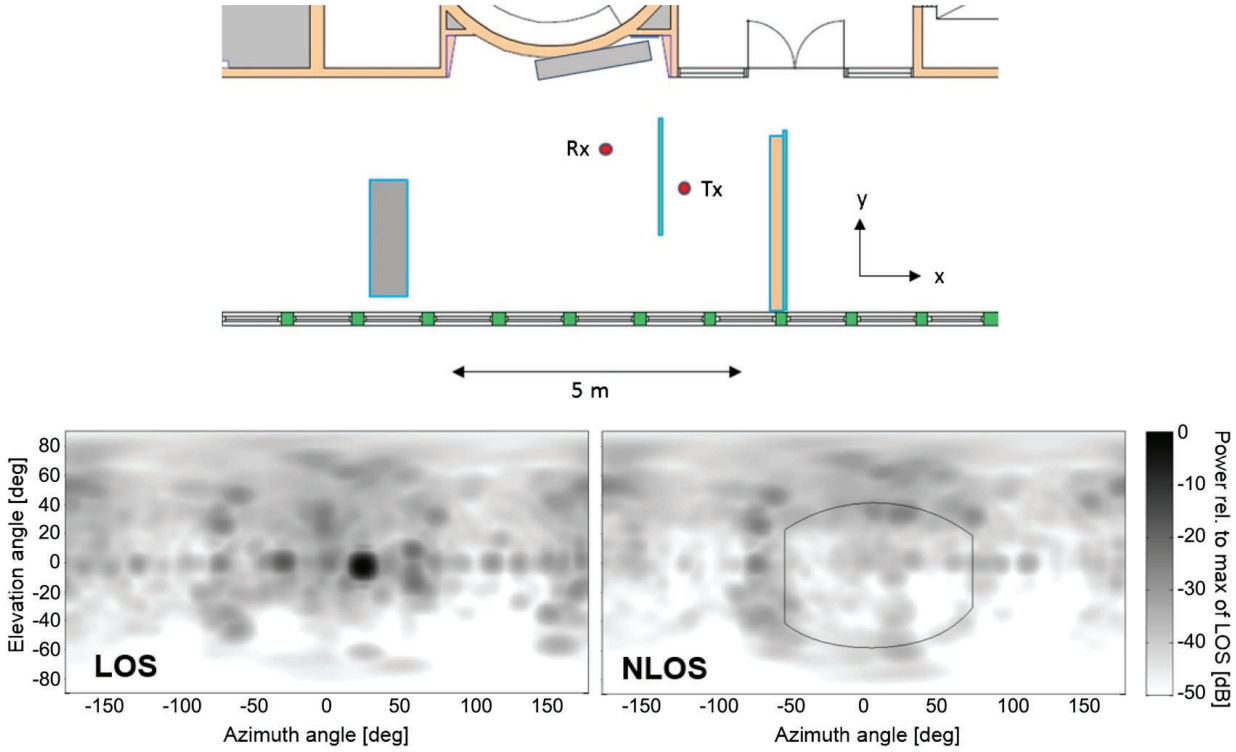


FIGURE 3.26

Measurement scenario (upper graph) where the whiteboard used for NLoS measurements is marked with a vertical line between the Rx and Tx antennas, and directional power spectra (lower graphs) for the LoS and NLoS measurements. The whiteboard contour is indicated with a black line in the NLoS graph.

3.3.5 DIRECTIONAL DOMAIN MEASUREMENTS

As pointed out in Sections 3.1.2 and 3.3.1.4 techniques to point the transmit and receive antennas in favorable directions will be required for mobile communications at high frequencies in the millimeter-wave range. This is due to the fundamental property that the aperture of a receiver using an omni-antenna is proportional to the square of the carrier wavelength and therefore corresponding transmission losses allow only very short link distances. For this reason it is essential to have in depth knowledge about the directional properties of radio propagation channels, particularly at the higher frequencies. This section presents highly resolved experimental characterization of the radio channel over a wide frequency range for some selected scenarios.

3.3.5.1 Indoor Office Wideband Results at 60 GHz

Channel measurement data at 58.7 GHz from an indoor office scenario, depicted in Fig. 3.26, have been analyzed using the spectral analysis method described in Section 3.3.2.1. For this purpose the channel has been sampled using a vertical dipole (2 dBi gain) antenna in both the transmit and the receive end of the link. The spatial samples were obtained by means of a 3D antenna positioning robot providing a

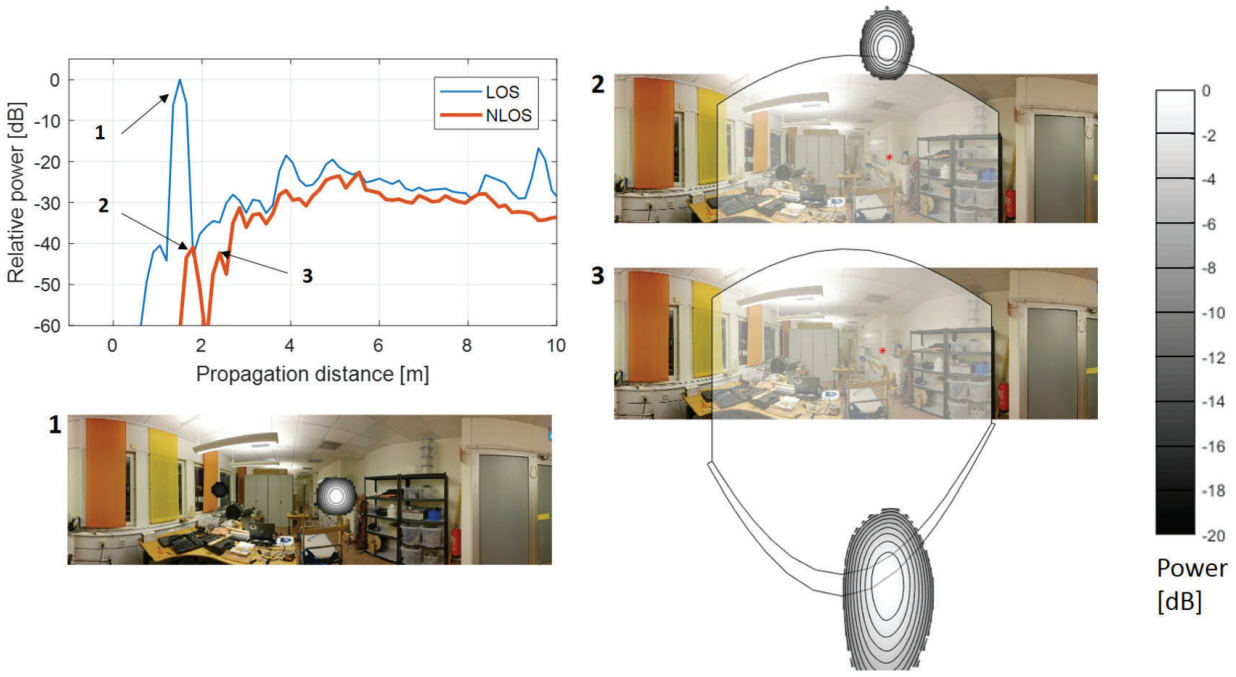


FIGURE 3.27

Power delay profiles for both LoS and NLoS (upper left graph) together with directional distributions of the first arriving paths on top of panoramic photographs.

virtual antenna array the size of 25x25x25 = 15625 elements using a spatial sampling distance of 0.4 wavelengths.

The measurements were performed in an indoor office environment at 1.5 m Tx–Rx distance under both LoS and NLoS conditions in the 57.68–59.68 GHz band. For the NLoS measurement a 2 m x 1.2 m large metal whiteboard was placed between the Tx and Rx antennas. The locations of Tx and Rx were the same in both the LoS and the NLoS measurements. In Fig. 3.26 the full directional spectra for both the LoS and the NLoS measurements are shown. Due to the short distance the LoS measurement is clearly dominated by the direct path. In contrast to LoS, the NLoS measurement is spatially much richer, having around ten strong paths in different directions. However, except for the direct path, and a few other strong paths blocked by the whiteboard, the two measurements show very similar directional characteristics. The channel seems to be composed of some distinct directions on top of a more smooth (diffuse) background. There seems to be a rich distribution of diffuse paths in all directions, except for those directions which correspond to the parts of the floor which are empty (no furniture).

In Fig. 3.27 corresponding power delay profiles for the LoS and the NLoS cases are shown for the first arriving paths. The first path is clearly dominant in the LoS case whereas the later reflected paths dominate in the NLoS case. There is, however, an early path also in the NLoS case which is suppressed by 40 dB relative to the LoS path due to diffraction over the upper edge of the whiteboard. There is also a path arriving somewhat later which is diffracted at the lower edge of the whiteboard. Both diffracted paths are shown on top of a panoramic photograph in Fig. 3.27.

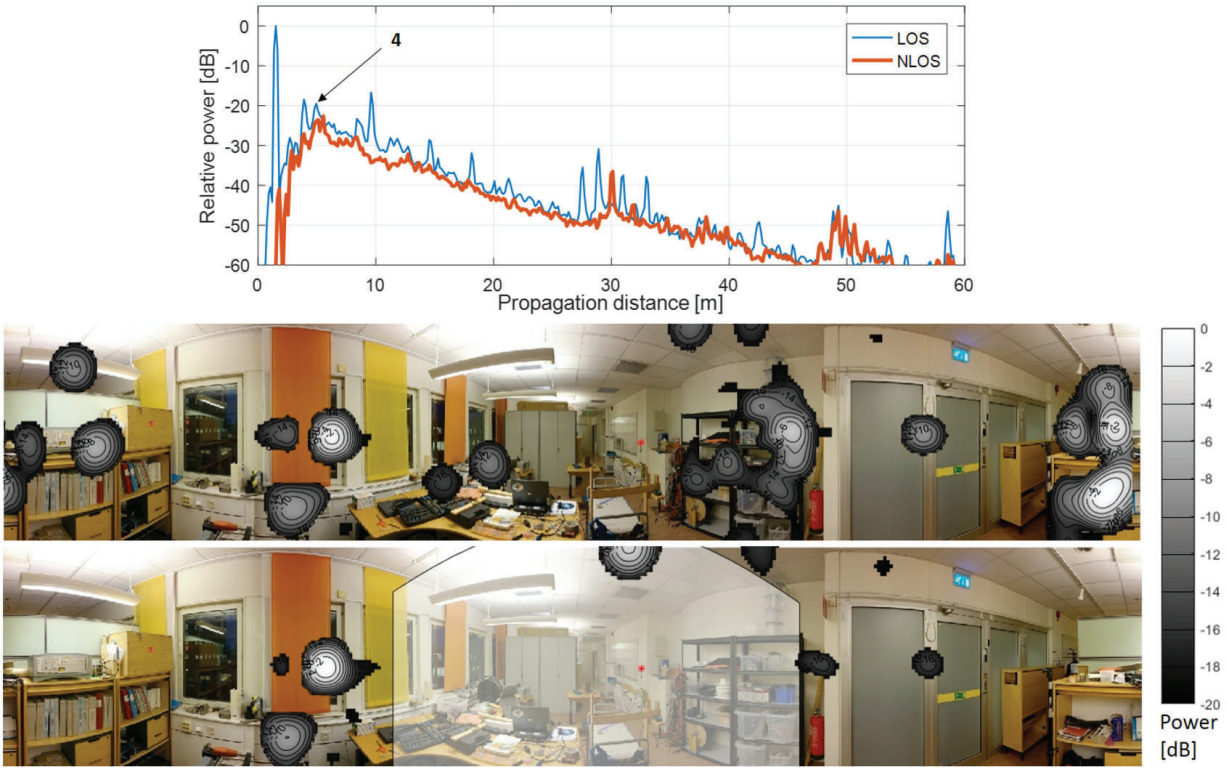


FIGURE 3.28

Power delay profiles for both LoS and NLoS (upper graph) together with directional distributions of the strongest path in NLoS on top of panoramic photographs.

When analyzing the strong peak number 4 of the power delay profile, shown in Fig. 3.28, it is clear that there are some significant scatterers such as objects on the table and the bookshelves. There are also a couple of strong reflections off a window and a wall segment. Comparing the LoS and the NLoS cases it is clear that some of the high power directions are absent in the NLoS graph due to the corresponding pathways being blocked by the whiteboard. An important conclusion that can be drawn is that simple ray tracing, based on the geometry of the room, would not succeed to model the rich scattering caused by furniture and other objects in the room.

3.3.5.2 Indoor Office Multifrequency Results

The measurements of previous section have been extended, adding two lower frequencies at 5.8 GHz and 14.8 GHz. In order to provide comparable characteristics over all measured frequencies the requirements of Section 3.3.2.3 are fulfilled in both measurements and analysis. For this purpose the measurement bandwidths have been equalized in the analysis, meaning that the 14.8 GHz and the 58.7 GHz measurement data are reduced to the bandwidth of the 5.8 GHz measurement data of 150 MHz. This equalization is important to avoid the effect that specular spikes, of the power delay profile, are amplified at higher frequencies due to the much larger bandwidth being available. Furthermore, the measurements at 58.7 GHz are affected by attenuation due to oxygen absorption. In order to provide re-

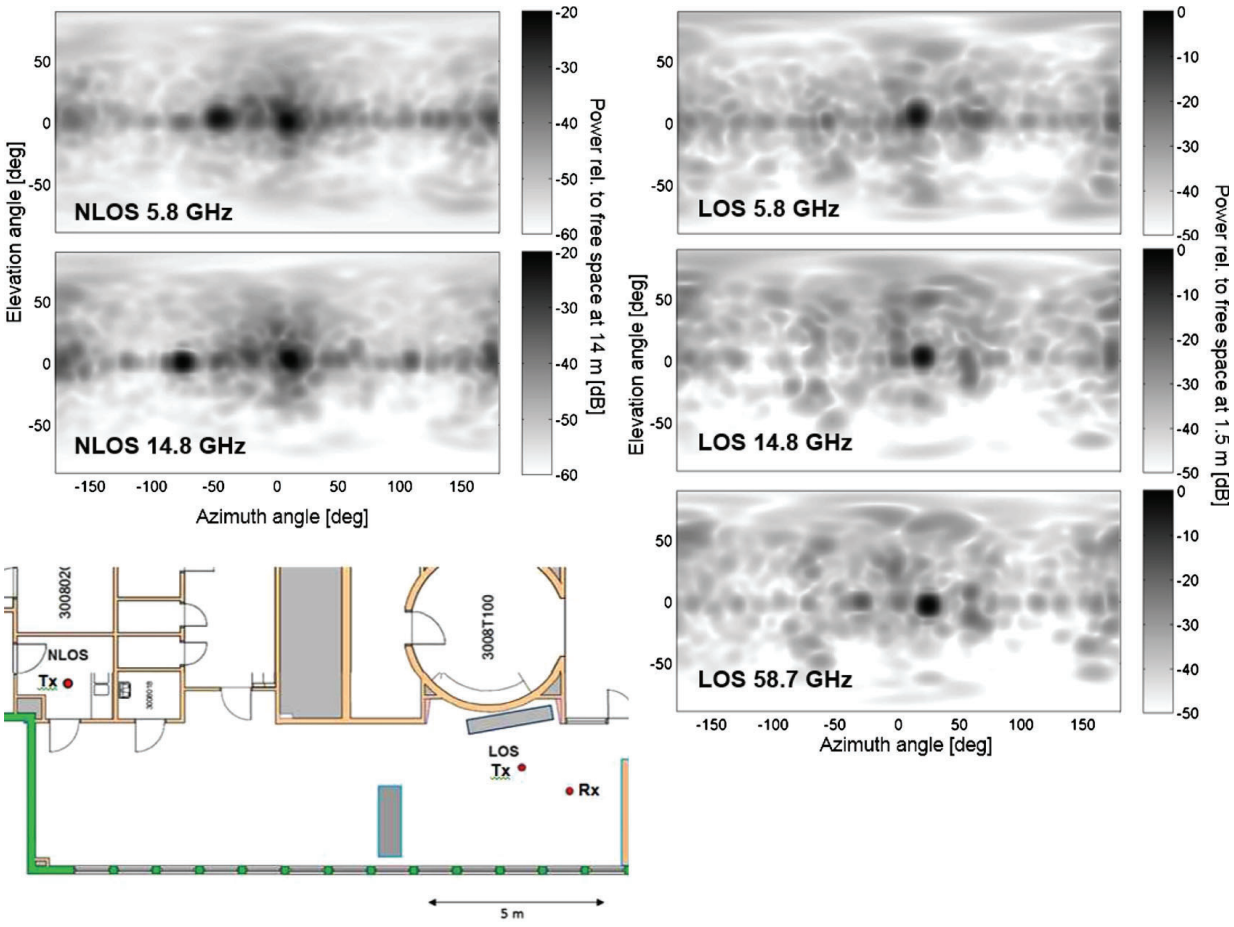


FIGURE 3.29

Measurement scenario (lower left graph), and directional power spectra for the NLoS (upper and middle left graphs) and LoS (right graphs) measurements for the different frequencies.

sults suitable for frequency consistent channel modeling and interpolation, this attenuation is removed in the analysis by compensating the power delay profiles with 1.5 dB per 100 m propagation distance at this frequency. It should be noted that though the link distance itself does not motivate this compensation, scattered paths may have substantially longer propagation distances, as seen in Fig. 3.31.

The LoS scenario is the same as in the previous section. In the NLoS measurements, the Rx antenna was placed in a small kitchen at the end of the office space (see Fig. 3.29). The distance between the Tx and the Rx antennas in the NLoS case was 14 m. It should be noted that the NLoS scenario was measured only at 5.8 GHz and 14.8 GHz due to the suspension of access to the indoor environment in the middle of the campaign. The directional power spectra are strikingly similar for all frequencies. One small frequency dependent difference is observed in the LoS graphs where a band of higher signal power around zero degree elevation angle is most pronounced at 5.8 GHz and least pronounced at 14.8 GHz. This difference is explained by the reflectivity of the windows which was found to be substantially higher at 5.8 GHz than at the other frequencies. The NLoS graphs differ from the LoS

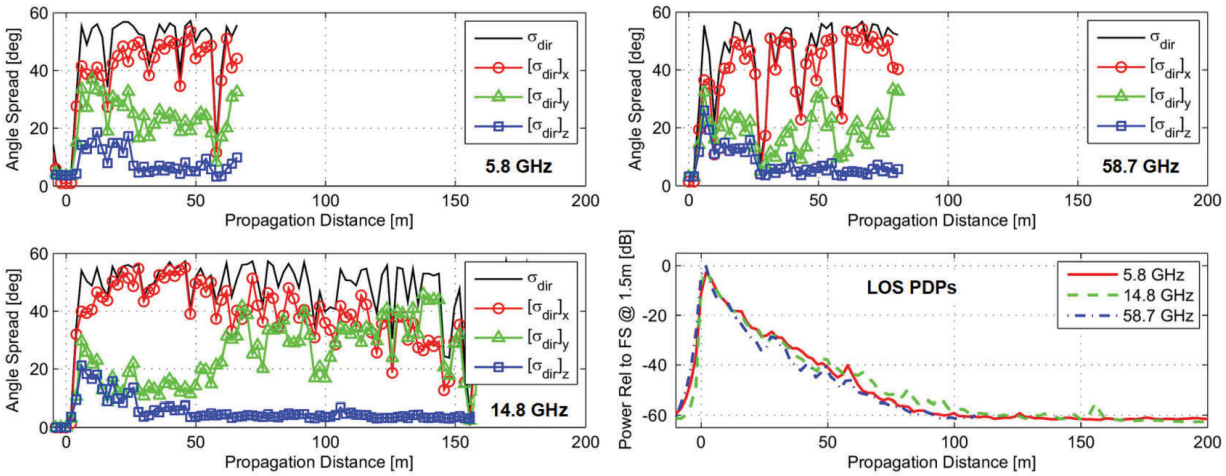


FIGURE 3.30

Measured directional spread versus propagation distance for the LoS scenario and the three different frequencies (left and upper right) and corresponding power propagation distance profiles (lower right). The power propagation distance profiles are normalized to the free-space power at 1.5 m.

graphs in that the diffuse cluster around the main peak is more focused and that there are not one but a couple of strong directions. Furthermore, the received power is attenuated about 20 dB relative to free-space propagation. The graphs of the two measured frequencies remain strikingly similar, as in the LoS case. One observed difference is the peak at −50 degrees in azimuth, which is strong at 5.8 GHz and weak at 14.8 GHz. The opposite effect is observed for the peak at −75 degrees in azimuth where the power is strong at 14.8 GHz and weak at 5.8 GHz. This is again the effect of frequency dependent window attenuation/reflection.

Directional spreads, σ_{dir}, according to the rotation invariant definition of Section 3.2.3 have been determined for the different scenarios and frequencies. In Fig. 3.30 the directional spreads for the LoS scenario versus propagation distance together with corresponding PDPs are shown. Again it is striking how similar the profiles for the different frequencies are. The directional spread is shown only for the part of the power propagation distance profiles for which there is sufficient signal above the noise floor. Basically the same characteristics are observed at all frequencies. For the LoS spike the spread is small, around 5 degrees. At other delays the spread is typically saturated at 57.3 degrees, which by definition is the maximum possible spread, as shown in Section 3.2.3. At a few delays where strong reflections occur the directional spread goes down. Another striking observation is that the elevation spread decays fast to very small values. For longer delays the directional spread in the x-dimension, which is the longest dimension of the room, dominates. Except for short delays the directional spreads in the different dimensions seem to be proportional to the corresponding lengths of the room. A likely explanation is that the power decays faster for smaller room dimensions, due to more frequent interactions with the corresponding walls, floor and ceiling, resulting in a smaller directional spread in those dimensions.

In Fig. 3.31 the corresponding graphs are shown for the NLoS scenario. The characteristics are very similar to those of the LoS scenario. One of the main observations is that there is a strong echo at about

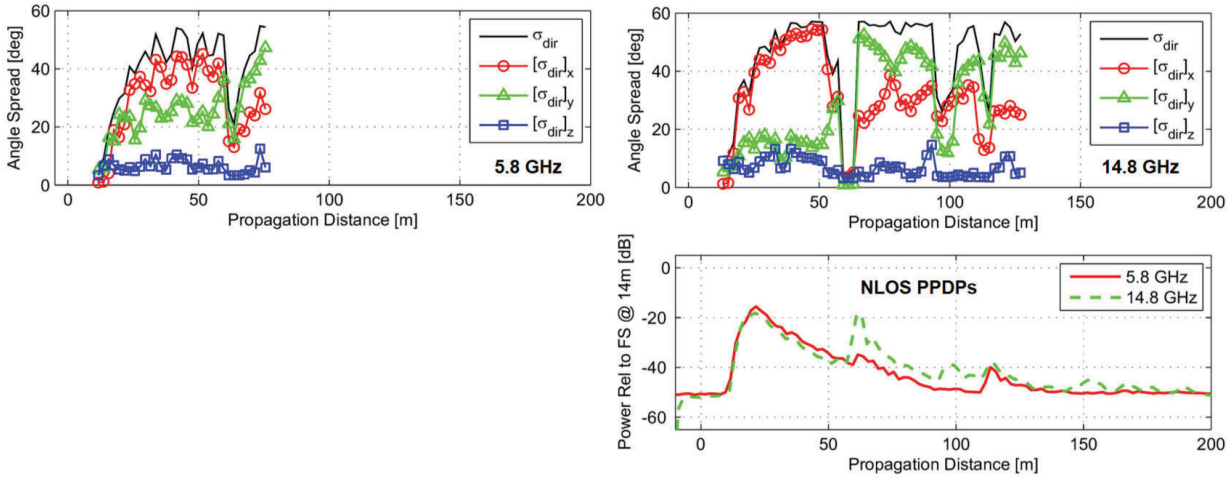


FIGURE 3.31

Measured directional spread versus propagation distance for the NLoS scenario and the three different frequencies (left and upper right) and corresponding power propagation distance profiles (lower right). The power propagation distance profiles are normalized to the free-space power at 14 m.

60 m propagation distance at 14.8 GHz, which is not observed at 5.8 GHz. The reason for the difference is that the windows reflect/attenuate differently at different frequencies. This is an effect of the three layers of glass (non-metalized) in the windows of the room. Due to multiple reflections between these layers different frequencies are attenuated differently when transmitted through the windows. Dedicated window attenuation measurements, performed in conjunction with the channel measurements, show that the window attenuation at 14.8 GHz is negligible, while it is around 10 dB at 5.8 GHz. The assumption that the strong echo at 14.8 GHz is due to a pathway going out through a window reflected off a neighboring building and in again through another window explains exactly the power difference of about 20 dB (relative to 5.8 GHz) which is twice the window attenuation at 5.8 GHz. It also explains why the directional spread in the y-direction increases at delays when strong radio waves enter the room from outside, by reflection off the adjacent building, as the power of waves propagating along the y-dimension of the room then is increased.

The total directional spreads, corresponding to summing the power of all delays (using angular distributions shown in Fig. 3.29), are shown in Fig. 3.32. Due to the shorter LoS link distance at 58.7 GHz (1.5 m) the relatively stronger LoS peak introduces a corresponding bias in directional spread (smaller spread) as compared with 5.8 and 14.8 GHz. In order to remove this bias the LoS peak has been decreased by 2.5 dB at 58.7 GHz in the analysis. No obvious frequency trend is observed. The characteristics are very similar for all frequencies where the elevation spread is small around 10 degrees and the directional spreads in x- and y-dimensions are substantially larger, between 20 and 40 degrees. Moreover, the directional spread in the y-dimension is substantially larger for the NLoS scenario due to the pathway which goes out of the building and in again, after reflecting off the adjacent building at 14.8 GHz and due to a strong window reflection at 5.8 GHz.

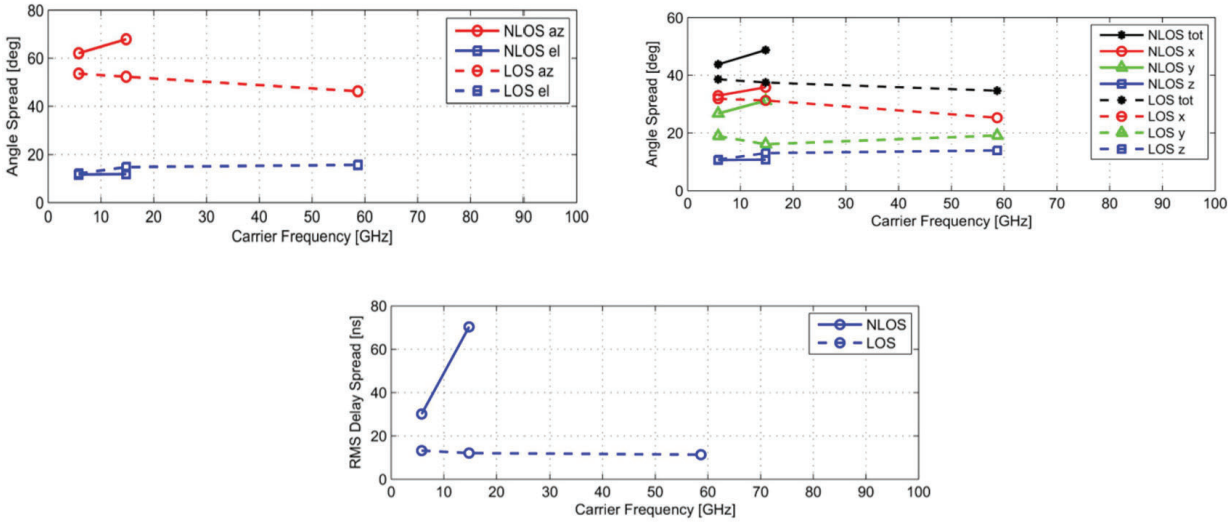


FIGURE 3.32

Total directional spreads (upper two) and delay spreads (lower) versus carrier frequency for the different measurement scenarios. The upper left plot show ordinary r.m.s. azimuth and elevation spreads, while the upper right show directional spreads according to the new improved definition according to the definition of Section 3.2.3.

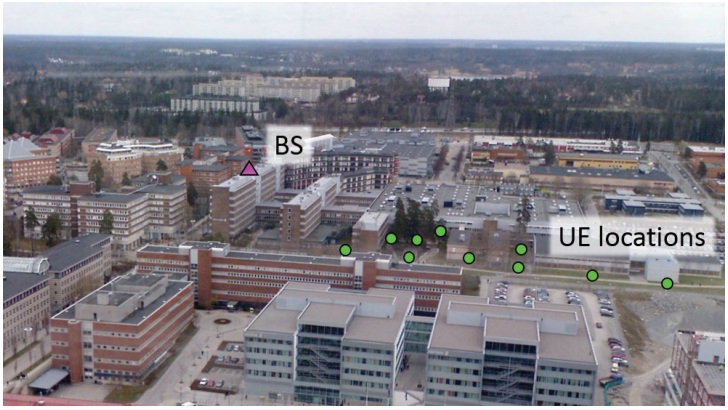


FIGURE 3.33

Photograph of measurement scenario.

3.3.5.3 Urban Macrocell Outdoor Results at 5 GHz

A measurement campaign in an urban outdoor macrocell scenario, depicted in Fig. 3.33, has been performed at 5.25 GHz using 200 MHz bandwidth [18]. At the BS a directional patch antenna with 7 dBi gain (90° beamwidth) and vertical polarization was used. A virtual planar array of 10x25 elements ($N_{\text{horizontal}} \times N_{\text{vertical}}$), with 2 cm (0.35λ) spacing was formed by means of a robotic antenna positioning system providing a spatial accuracy better than 0.1 mm. In the user equipment (UE) end an

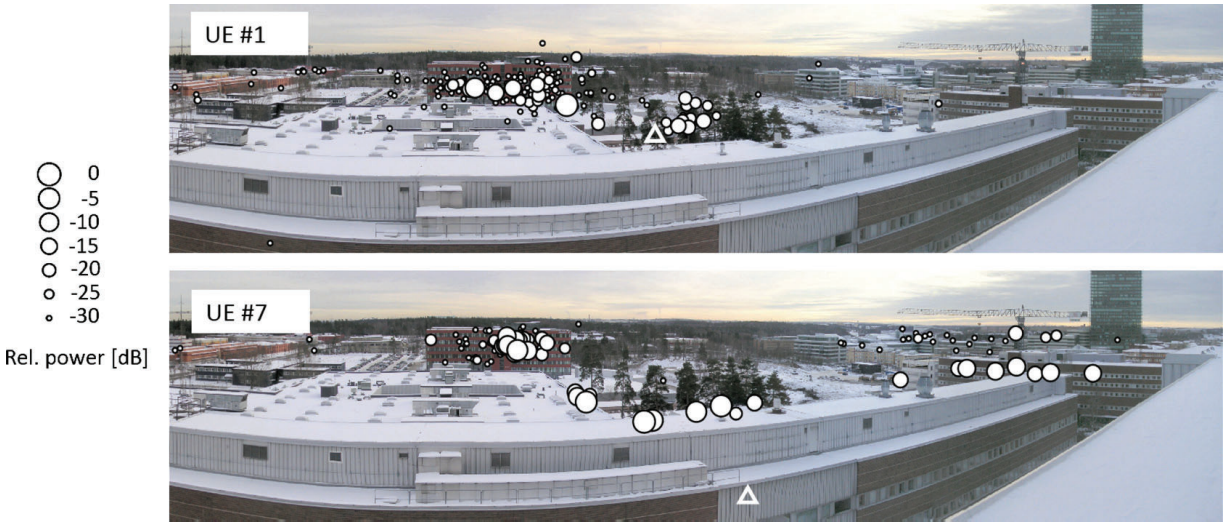


FIGURE 3.34

Estimated paths superimposed on a panoramic photograph taken from the BS location at the roof. The directions of the UE locations are indicated with triangles. It should be noted that the UEs are under NLoS conditions so that they are not visible from the BS.

ordinary vertical dipole antenna was used. The measurement data analysis is based on the superresolution method described in Section 3.3.2.2 and Ref. [17].

The total measurement time for a single UE location was around 7 min, since the antenna positioning system took about 1 s for each change of antenna position. Moreover, as it was a bit windy during the measurements, the movement of some significant trees affected the measurement data. At specific delays and directions of arrival and departure the corresponding radio waves are distorted due to time variations created by the moving trees. The result of these time variations is that the power of waves scattered in those trees will largely appear as randomly distributed in angle, i.e., as noise. The remaining coherent power scattered off the trees is reduced correspondingly, meaning that the significance of the trees is underestimated.

For each UE location $N = 500$ waves have been estimated. As described in [18] the estimates are initialized by finding peaks in the angle domain above the noise floor. Moreover, the requirement that the standard deviation of the errors of the estimates is less than 40 degrees in angle and 20 m in propagation distance is set. The corresponding result of a power delay profile is shown in Fig. 3.13 of Section 3.3.2.2 demonstrating that the estimates account for most of the power of the measured channel. To ensure a high reliability of the presented results the analysis is based on paths having estimation errors with standard deviations less than 2 degrees in elevation and 4 degrees in azimuth. The corresponding estimated plane waves superimposed on a panoramic photograph are shown in Fig. 3.34. It is clear that the dominant paths are diffracted over the roof tops and/or reflected off neighboring buildings. It seems that the main propagation mechanisms are reflections off neighboring buildings which are under LoS conditions from both the BS and the UE. At some UE locations (e.g. UE location 7) diffraction over the rooftops seems to be important.

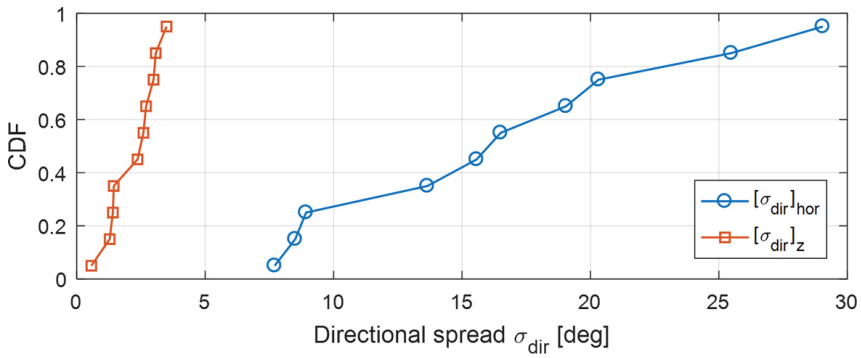


FIGURE 3.35

CDFs of directional spreads in z-direction and horizontal (x/y) plane.

The directional spreads according to the definition in Section 3.2.3 are shown in Fig. 3.35. In this case the spread in the z-direction is very similar to ordinary elevation spread and the spread in the horizontal plane is very similar to ordinary azimuth spread. The spread in z-direction (elevation) is very small, in the range 1–4 degrees, whereas the spread in the horizontal plane (azimuth) is between 7 and 30 degrees.

3.4 CHANNEL MODELING

This section deals with commonly used channel models specified in standardization bodies like 3GPP, ITU-R, and also international research projects and networks like METIS and mmMAGIC. The focus is put on the recent models with particular emphasis on 5G specific features. All these models are geometric stochastic channel models (except the METIS map based model) for which each channel realization is Monte Carlo generated according to each parameter's probability distribution. To achieve continuity in space and time the random distributions are auto-correlated in those dimensions. The provisioning of the full correlation matrix between all channel parameters and all autocorrelation functions is a huge challenge. Moreover, to provide values within proper confidence ranges requires a massive amount of measurement data. Even if all needed correlations are provided it is a considerable computational task to perform all autocorrelations needed for providing spatial consistency. The alternative approach employed in the METIS map based model is based on simplified geometry ray tracing. In this model, all correlations and spatial consistency are provided automatically as a result of the model being based fully on deterministic and physical principles. It accounts for all important scattering mechanisms, i.e., specular reflections, diffraction, scattering by rough surfaces and objects, and outdoor-to-indoor penetration. The drawback of this model is its corresponding implementation complexity.

The complexity of channel models has been increasing with each new generation of mobile communications. This is a result of the fact that each new generation has been providing better performance by utilizing more of the physical propagation channel's degrees of freedom. In the second generation (GSM) the complexity was relatively low as narrow bandwidths and low degree of spatial diversity

were utilized. Through the third and the fourth generations both bandwidths and spatial diversity techniques were extended by e.g. introducing soft handover and MIMO transmission.

The main focus of the section is on the recent model for 5G developed by 3GPP [2] and ITU-R [9]. This model provides a successful extension of transmission loss modeling into the mm-Wave range up to 100 GHz. Moreover, extended bandwidth (up to 2 GHz) and highly resolved directional characterization of paths is provided. Though some of the very highly resolved characteristics are not fully validated the model is likely to be accurate enough for the early 5G networks which do not utilize the largest bandwidths and narrowest antenna beams.

3.4.1 5G STOCHASTIC CHANNEL MODELS

In this section the 3GPP channel model for the development of 5G in the frequency range 0.5–100 GHz [2], which also has been adopted by ITU-R for IMT-2020 [9], is described. The model is largely based on previous generation modeling [1], [6] and is of the geometry-based stochastic type of channel model. The mmMAGIC channel model [26] is largely harmonized with the 3GPP/ITU-R model with some improvements and additions. The channel H_{mn} is defined by a set of plane waves according to (3.7) for which the polarimetric amplitudes $\mathbf{A}_l$, directions $\mathbf{k}_l^{\text{tx}}$ and $\mathbf{k}_l^{\text{rx}}$, delays τ_l, and Doppler frequencies ω_{D_l} are empirically characterized with the corresponding stochastic distributions. Only the line-of-sight component is deterministically defined by the geometry.

3.4.1.1 Transmission Loss Modeling

The transmission loss modeling is based on assuming isotropic antennas in both ends of the link using the basic transmission loss definition of ITU-R Recommendation P.341 [8]. Any dependence on specific antenna patterns is modeled separately by combining the antenna pattern with corresponding multipath distributions as described in the next section. The transmission loss (referred to as path-loss in the ITU-R and 3GPP specifications) properties of the models are summarized in Table 3.1. These models are largely validated by the measurements presented in Section 3.3.3 for the indoor office and street canyon scenarios. The losses for the line-of-sight cases are close to the free-space loss up to a break point, after which the loss is proportional to $40\log(d)$. Under non-line-of-sight conditions, there is no frequency dependence of the loss relative to free space for the urban macro (UMa) and the rural macro (RMa) scenarios. However, for the indoor and street canyon scenarios there is a small increase of the loss relative to free space with increasing frequency.

Further, it is assumed that there is a log-normal stochastic location variation of the loss around the distance dependent functions. The corresponding standard deviations range between 4 and 8 dB. These variations are correlated in space by the exponential autocorrelation function

$$R = \exp\left(-\frac{|\Delta d|}{d_{\text{cor}}}\right) \tag{3.30}$$

where Δd is the distance between two points in space and d_{cor} is a constant.

The building penetration loss, L_{O2I}, is modeled by

$$L_{\text{O2I}} = 5 - 10\log\left(\alpha 10^{\frac{-L_{\text{glass}}}{10}} + \beta 10^{\frac{-L_{\text{concrete}}}{10}}\right) + \gamma d_{\text{2D-in}} \tag{3.31}$$

Table 3.1 3GPP transmission loss models (with permission from 3GPP)

Scenario	LOS/NLOS	Pathloss [dB], f_c is in GHz and d is in meters, see note 6	Shadow fading std [dB]	Applicability range, antenna height default values
RMa	LOS	$PL_{RMa-LOS} = \begin{cases} PL_1 & 10\text{m} \le d_{2D} \le d_{BP} \\ PL_2 & d_{BP} \le d_{2D} \le 10\text{km} \end{cases}$, see note 5 $PL_1 = 20\log_{10}(40\pi d_{3D} f_c / 3) + \min(0.03h^{1.72}, 10)\log_{10}(d_{3D})$ $\quad - \min(0.044h^{1.72}, 14.77) + 0.002\log_{10}(h)d_{3D}$ $PL_2 = PL_1(d_{BP}) + 40\log_{10}(d_{3D} / d_{BP})$	$\sigma_{SF} = 4$ $\sigma_{SF} = 6$	$h_{BS} = 35\text{m}$ $h_{UT} = 1.5\text{m}$ $W = 20\text{m}$ $h = 5\text{m}$ h = avg. building height W = avg. street width The applicability ranges: $5\text{m} \le h \le 50\text{m}$ $5\text{m} \le W \le 50\text{m}$ $10\text{m} \le h_{BS} \le 150\text{m}$ $1\text{m} \le h_{UT} \le 10\text{m}$
	NLOS	$PL_{RMa-NLOS} = \max(PL_{RMa-LOS}, PL'_{RMa-NLOS})$ for $10\text{m} \le d_{2D} \le 5\text{km}$ $PL'_{RMa-NLOS} = 161.04 - 7.1\log_{10}(W) + 7.5\log_{10}(h)$ $\quad - (24.37 - 3.7(h/h_{BS})^2)\log_{10}(h_{BS})$ $\quad + (43.42 - 3.1\log_{10}(h_{BS}))(\log_{10}(d_{3D}) - 3)$ $\quad + 20\log_{10}(f_c) - (3.2(\log_{10}(11.75h_{UT}))^2 - 4.97)$	$\sigma_{SF} = 8$	
UMa	LOS	$PL_{UMa-LOS} = \begin{cases} PL_1 & 10\text{m} \le d_{2D} \le d'_{BP} \\ PL_2 & d'_{BP} \le d_{2D} \le 5\text{km} \end{cases}$, see note 1 $PL_1 = 28.0 + 22\log_{10}(d_{3D}) + 20\log_{10}(f_c)$ $PL_2 = 28.0 + 40\log_{10}(d_{3D}) + 20\log_{10}(f_c)$ $\quad - 9\log_{10}((d'_{BP})^2 + (h_{BS} - h_{UT})^2)$	$\sigma_{SF} = 4$	$1.5\text{m} \le h_{UT} \le 22.5\text{m}$ $h_{BS} = 25\text{m}$
	NLOS	$PL_{UMa-NLOS} = \max(PL_{UMa-LOS}, PL'_{UMa-NLOS})$ for $10\text{m} \le d_{2D} \le 5\text{km}$ $PL'_{UMa-NLOS} = 13.54 + 39.08\log_{10}(d_{3D}) +$ $\quad 20\log_{10}(f_c) - 0.6(h_{UT} - 1.5)$	$\sigma_{SF} = 6$	$1.5\text{m} \le h_{UT} \le 22.5\text{m}$ $h_{BS} = 25\text{m}$ Explanations: see note 3
		Optional $PL = 32.4 + 20\log_{10}(f_c) + 30\log_{10}(d_{3D})$	$\sigma_{SF} = 7.8$	
UMi - Street Canyon	LOS	$PL_{UMi-LOS} = \begin{cases} PL_1 & 10\text{m} \le d_{2D} \le d'_{BP} \\ PL_2 & d'_{BP} \le d_{2D} \le 5\text{km} \end{cases}$, see note 1 $PL_1 = 32.4 + 21\log_{10}(d_{3D}) + 20\log_{10}(f_c)$ $PL_2 = 32.4 + 40\log_{10}(d_{3D}) + 20\log_{10}(f_c)$ $\quad - 9.5\log_{10}((d'_{BP})^2 + (h_{BS} - h_{UT})^2)$	$\sigma_{SF} = 4$	$1.5\text{m} \le h_{UT} \le 22.5\text{m}$ $h_{BS} = 10\text{m}$
	NLOS	$PL_{UMi-NLOS} = \max(PL_{UMi-LOS}, PL'_{UMi-NLOS})$ for $10\text{m} \le d_{2D} \le 5\text{km}$ $PL'_{UMi-NLOS} = 35.3\log_{10}(d_{3D}) + 22.4$ $\quad + 21.3\log_{10}(f_c) - 0.3(h_{UT} - 1.5)$	$\sigma_{SF} = 7.82$	$1.5\text{m} \le h_{UT} \le 22.5\text{m}$ $h_{BS} = 10\text{m}$ Explanations: see note 4
		Optional $PL = 32.4 + 20\log_{10}(f_c) + 31.9\log_{10}(d_{3D})$	$\sigma_{SF} = 8.2$	

(continued on next page)

Table 3.1 (*continued*)

InH - Office	LOS	$PL_{\text{InH-LOS}} = 32.4 + 17.3\log_{10}(d_{3D}) + 20\log_{10}(f_c)$	$\sigma_{SF} = 3$	$1\text{m} \le d_{3D} \le 150\text{m}$
	NLOS	$PL_{\text{InH-NLOS}} = \max(PL_{\text{InH-LOS}}, PL'_{\text{InH-NLOS}})$ $PL'_{\text{InH-NLOS}} = 38.3\log_{10}(d_{3D}) + 17.30 + 24.9\log_{10}(f_c)$	$\sigma_{SF} = 8.03$	$1\text{m} \le d_{3D} \le 150\text{m}$
		Optional $PL'_{\text{InH-NLOS}} = 32.4 + 20\log_{10}(f_c) + 31.9\log_{10}(d_{3D})$	$\sigma_{SF} = 8.29$	$1\text{m} \le d_{3D} \le 150\text{m}$

Note 1: Breakpoint distance $d'_{BP} = 4\,h'_{BS}\,h'_{UT}\,f_c/c$, where f_c is the centre frequency in Hz, $c = 3.0\times10^8$ m/s is the propagation velocity in free space, and h'_{BS} and h'_{UT} are the effective antenna heights at the BS and the UT, respectively. The effective antenna heights h'_{BS} and h'_{UT} are computed as follows: $h'_{BS} = h_{BS} - h_E$, $h'_{UT} = h_{UT} - h_E$, where h_{BS} and h_{UT} are the actual antenna heights, and h_E is the effective environment height. For UMi $h_E = 1.0$m. For UMa h_E=1m with a probability equal to 1/(1+C(d_{2D}, h_{UT})) and chosen from a discrete uniform distribution uniform(12,15,…,(h_{UT}-1.5)) otherwise. With C(d_{2D}, h_{UT}) given by

$$C(d_{2D}, h_{UT}) = \begin{cases} 0 & , h_{UT} < 13\text{m} \\ \left(\frac{h_{UT}-13}{10}\right)^{1.5} g(d_{2D}) & , 13\text{m} \le h_{UT} \le 23\text{m} \end{cases},$$

where

$$g(d_{2D}) = \begin{cases} 0 & , d_{2D} \le 18\text{m} \\ \frac{5}{4}\left(\frac{d_{2D}}{100}\right)^3 \exp\left(\frac{-d_{2D}}{150}\right) & , 18\text{m} < d_{2D} \end{cases}.$$

Note that h_E depends on d_{2D} and h_{UT} and thus needs to be independently determined for every link between BS sites and UTs. A BS site may be a single BS or multiple co-located BSs.

Note 2: The applicable frequency range of the PL formula in this table is $0.5 < f_c < f_H$ GHz, where f_H = 30 GHz for RMa and f_H = 100 GHz for all the other scenarios. It is noted that RMa pathloss model for >7 GHz is validated based on a single measurement campaign conducted at 24 GHz.

Note 3: UMa NLOS pathloss is from TR36.873 with simplified format and $PL_{\text{UMa-LOS}}$ = Pathloss of UMa LOS outdoor scenario.

Note 4: $PL_{\text{UMi-LOS}}$ = Pathloss of UMi-Street Canyon LOS outdoor scenario.

Note 5: Break point distance $d_{BP} = 2\pi\, h_{BS}\, h_{UT}\, f_c/c$, where f_c is the centre frequency in Hz, $c = 3.0 \times 10^8$ m/s is the propagation velocity in free space, and h_{BS} and h_{UT} are the antenna heights at the BS and the UT, respectively.

Note 6: f_c denotes the center frequency normalized by 1GHz, all distance related values are normalized by 1m, unless it is stated otherwise.

where α and β are, respectively, the fractions of glass and concrete area of exterior building walls, L_{glass} and L_{concrete} are the corresponding losses of the multipane windows and concrete, γ is the loss per meter horizontal penetration depth $d_{\text{2D-in}}$ into the building. In the 3GPP and ITU models $\gamma = 0.5$ dB/m. A more thorough analysis by mmMAGIC shows that γ is uniformly distributed between 0.5 and 1.5 dB/m. It has been found that buildings may be grouped in a high-loss and a low-loss category. Buildings in the high-loss category are referred to as "thermal efficient," as the corresponding construction materials and methods result in considerable penetration losses. Particularly the thin metal coatings used for blocking heat radiation through windows also attenuates radio waves in the order of 20–30 dB. The low loss category corresponds to traditional buildings for which non-coated windows are used. The loss caused by the exterior wall construction material, L_{m}, is given by

$$L_{\text{m}} = \text{a} + \text{b}\,f. \tag{3.32}$$

The corresponding material parameters are given in Table 3.2.

It is assumed that there is a log-normal location variability of the penetration loss due to internal irregularities of the buildings such as furniture, interior walls elevator shafts, etc. The corresponding standard deviations are $\sigma_{\text{low}} = 4.4$ dB and $\sigma_{\text{high}} = 6.5$ dB for the low-loss and the high-loss categories, respectively. The mmMAGIC model has refined the variability by introducing a frequency dependence

Table 3.2 3GPP penetration loss material parameters

Material	a [dB]	b [dB/GHz]
Standard multipane glass	2.0	0.2
IRR glass	23.0	0.3
Concrete	5.0	4.0

of the corresponding standard deviation,

$$\sigma = 4 + \mathrm{k}_\sigma f \tag{3.33}$$

where k_σ has been estimated to be 0.08 dB/GHz for the low loss category.

In Fig. 3.36 the building penetration loss of the 3GPP model as a function of frequency is shown for the two categories of buildings (traditional and thermal efficient) for 5%, 50%, and 95% levels of the location variability distributions. For comparison the corresponding losses of the ITU-R Recommendation P.2109 for building entry loss are shown. The ITU-R model is empirical and is based on extensive measurement data whereas the 3GPP model is based on simplified physical principles. At the 50% level the two models agree well for frequencies below 50 GHz. However, the frequency trend is substantially stronger for the 3GPP model. This may be explained by the fact that the values of Table 3.2 correspond to a summed glass layer thickness of 24 mm for traditional and 36 mm for thermal efficient windows [7]. This is about three times the thicknesses found in real buildings. It is also clear that the ITU-R model [10] reflects that the spread increases with frequency, which the 3GPP mode does not.

In the 3GPP and ITU-R IMT2020 models the dependence of the propagation angle of incidence relative to the exterior wall is accounted for by the addition of a constant of 5 dB in (3.31). As a consequence the spread of the loss distributions is reduced. This is clearly observed in Fig. 3.36 where the 5% and 95% levels are shifted substantially more for the model of the ITU-R Recommendation P.2109 [10]. As this Recommendation was developed to support spectrum sharing studies between e.g. IMT and satellites the dependence of the elevation angle is accounted for by

$$L(\theta) = L(\theta = 0) + \alpha_\theta \theta \tag{3.34}$$

where θ is the elevation angle of the path relative to the exterior wall, and α_θ is a constant which is estimated to be about 20 dB/90 degrees.

3.4.1.2 Multipath Directional and Delay Modeling

The distributions of amplitudes, delays and directions of multipath components are generated based on first and second order moments of closed form stochastic distributions. Furthermore, the distributions are specified at two levels: the intercluster level and the intra-cluster level. This means that at the higher level the stochastic distributions of clusters of multipath components are generated. At the lower level the corresponding distribution within clusters are generated. The motivation for this two level clustering is historical and based on observations of power delay profiles [31]. It may, however, be questioned if this topology really can be justified when comparing with highly resolved empirical data.

In the delay domain, both the probability and the power of clusters follow distributions which are exponentially decaying with a log-normal shadowing added. In angle domains, i.e., elevation and

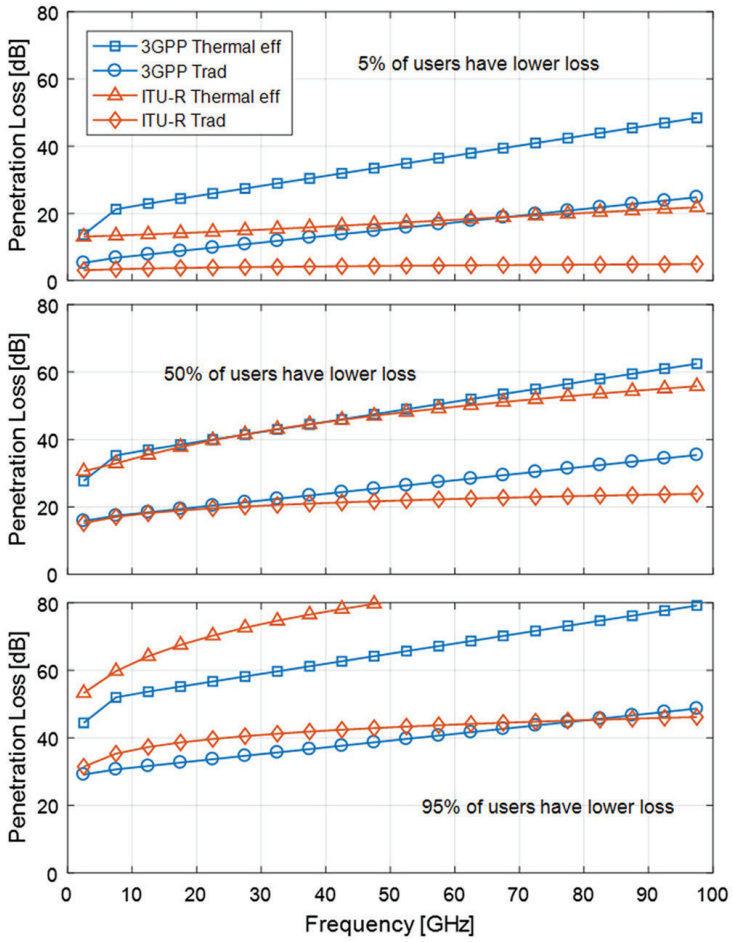


FIGURE 3.36

Building penetration loss versus frequency for the microcell scenario and the 3GPP and ITU-R P.2109 models at 5%, 50%, and 95% probability levels.

azimuth, the distributions are wrapped Gaussian functions of the cluster power. This is somewhat problematic, as the probability of having many clusters in the same direction becomes low which is in contrast with empirical data. In the baseline model each cluster consists of 20 multipath components. These are at fixed delays cluster-wise except for the two strongest clusters which are subdivided into three sublevels of fixed delays. All 20 subpaths have fixed power within each cluster and have a tabulated distribution to provide a Laplacian power distribution in angle. In Fig. 3.37 multipath component distributions in azimuth and propagation distance are shown for the indoor office scenario (same environment as in Section 3.3.4.1) in NLoS at 60 GHz. Both very high resolution measurement data [14], [26] and a corresponding realization of the 3GPP model are shown. The measurements were performed in an office environment using a 50 cm wide and 12.5 cm high planar array and 2 GHz bandwidth providing extreme resolution in direction. It is clear that the measured distributions do not show the effect

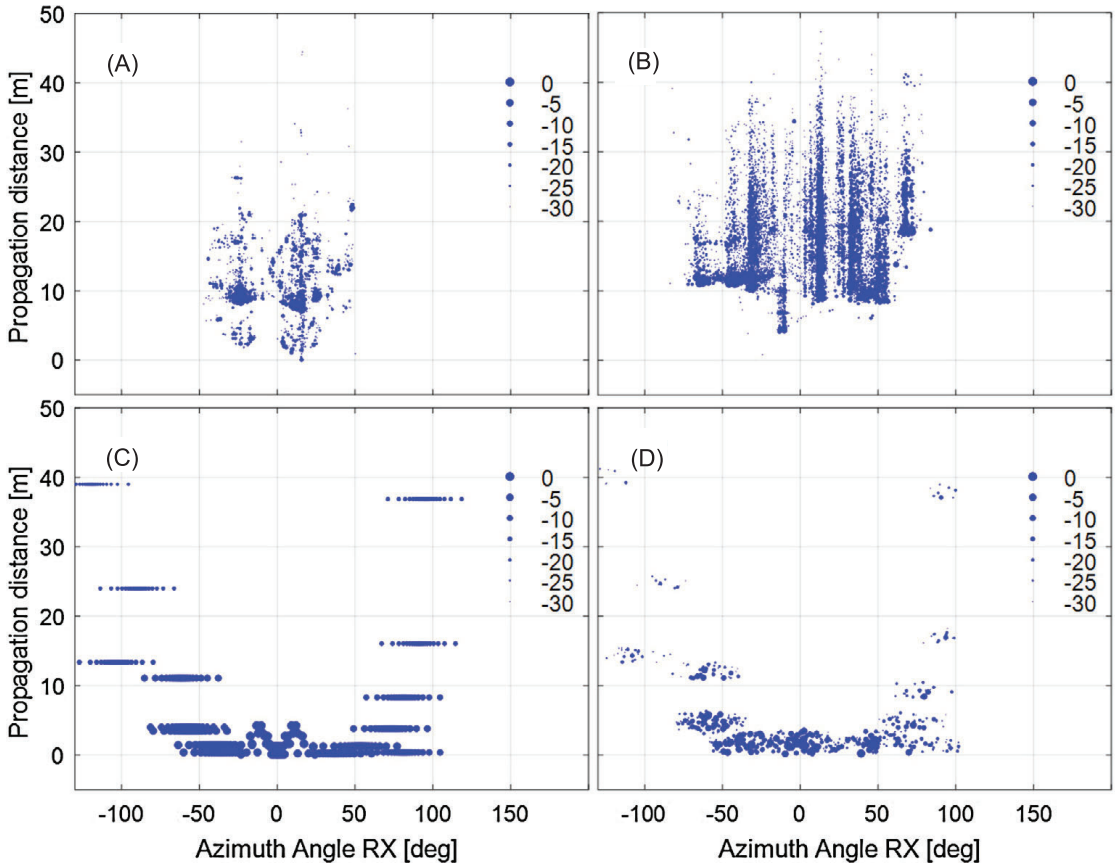


FIGURE 3.37

Multipath component distributions in azimuth and propagation distance for the office indoor scenario at 60 GHz under NLoS conditions. Measurement data are shown in (A) for a medium obstructed NLoS scenario, in (B) for a heavily obstructed NLoS scenario, and a random realization of the 3GPP model for the base line case in (C); and for the high resolution option in (D).

that clusters at longer delays are shifted from the main direction. This model property might lead to a non-realistic decrease of delay spread when narrow-beam antennas are used.

It is clear that the baseline model which uses 20 subpaths per cluster at fixed delays provides a synthetic output which does not agree well with the highly resolved measurements. This effect is even more pronounced when looking at the power ordered distributions of the multipath components, as shown in Fig. 3.38. In the measurements the power of the MPCs decay substantially with the power ordered number. At MPC number 20 the measured power is between 7 dB and 15 dB below the maximum power whereas there is no corresponding decay for the baseline 3GPP model. This is problematic to the extent that large arrays or very narrow beams are used as each MPC then may be resolved. The corresponding spatial multiplexing performance, of e.g. interference suppression or MIMO capacity, would then be unrealistically good as shown in the channel modeling chapter of [29] and [20]. For this reason, 3GPP provides an optional way to model the multipath components to be used in the case

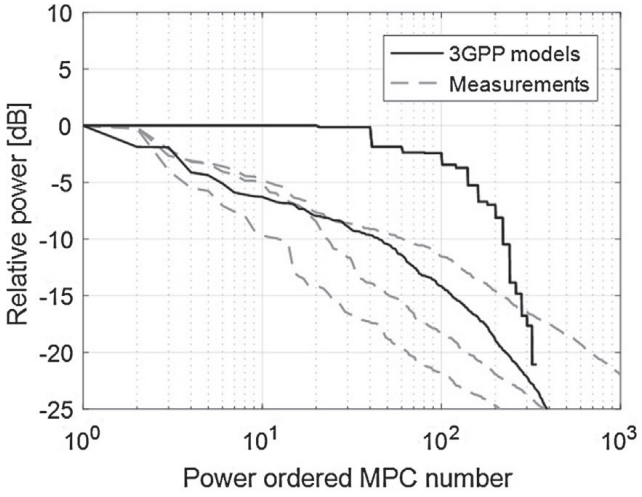


FIGURE 3.38

Power ordered multipath component distributions for the office indoor scenario at 60 GHz under NLoS conditions for the 3GPP baseline model and the high resolution option and corresponding distributions for three measurement locations.

of very large antenna arrays and/or large bandwidths. The distributions of the clusters are kept the same as for the baseline model. Within each cluster, however, a considerably larger number of MPCs is allowed. Corresponding distributions are uniform in angle and delay. The components are, however, power weighted using a Laplacian function in angle and an exponential function in delay. The result of this optional method is shown in Figs. 3.37 and 3.38. It is clear that the distribution of MPC powers is much more realistic with the high resolution option. Also the angle and delay distributions within clusters appear to be much more realistic.

The 3GPP model shows a strong frequency dependent decrease of both directional and delay spreads for most of the scenarios. This frequency trend is not confirmed by measurements presented in Sections 3.3.4 and 3.3.5. The explanation for the observed trend might be that all of the requirements for ensuring comparability between different frequency bands provided in Section 3.3.2.3 were not fulfilled in the measurements which the 3GPP model was based on. As pointed out in Section 3.3.4.4 a very thorough analysis performed by mmMAGIC showed no or very weak frequency trends of the delay spread.

3.4.1.3 Spatial Consistency

In order to provide realistic model output in the case of mobility, i.e., when the UE is moving or in the case of MU-MIMO, a spatial distribution of channel realizations is correlated using function (3.30). In the case of the baseline model only the intercluster parameters are correlated. However, for the high resolution option also the intra-cluster parameters are correlated. The corresponding correlation distances range from 10 to 50 m depending on the parameter and scenario. This method will indeed make the channel variations continuous as the UE moves. There is, however, no support for ensuring that changes are realistic in e.g. Doppler and death and birth processes. For e.g. an outdoor user, the channel conditions may be quite stationary until the user moves around a corner in a street crossing.

The same effect is likely for an indoor user when moving from one room to another. This may have some critical impact on the optimization of beam-tracking techniques as such techniques will be based on the dynamics of the radio channel. For this reason a more realistic geometry based option is provided by the blockage model described in next section.

3.4.2 GEOMETRY-BASED MODELING

In order to provide realistic dynamic channel output the METIS project developed an alternative channel model which is based on the 3D geometry of the environment combined with electromagnetic material properties and simple ray tracing. One component of this model—the blockage model—is particularly useful for describing the dynamics of death and birth process of paths in a realistic way.

3.4.2.1 Blockage

As has been pointed out previously, high antenna gain and corresponding narrow beams are needed for compensating for decreasing antenna apertures when going up in frequency. As a consequence, any moving object that at some time instant blocks the main beam will result in a dramatic reduction of the received signal. For this purpose METIS developed a model for blockage [24], which later was adopted as an additional feature of the 3GPP model. This model is based on 3D diffraction by a rectangular screen where the signal of each MPC is attenuated based on the geometry of the corresponding path. The model is based on standard closed form mathematical expressions, making it simple and computationally efficient.

The METIS blockage model was later considerably improved by mmMAGIC by accounting for the phase differences of the pathways over the four edges providing accurate output for nearly all geometries. This is in contrast with the standard Fresnel approximation which provides an accurate output only for the case when Tx and Rx are at large distances and perpendicularly oriented relative to the screen. Due to its general validity and good accuracy, the mmMAGIC model has been adopted by ITU-R in Recommendation P.526 [11].

In Fig. 3.39 the outputs of both models are shown for a 4x4 m screen at 4 GHz. It is clear that the mmMAGIC model provides a highly accurate output, as it follows the exact Kirchoff integral solution almost perfectly. The output of the METIS blockage model follows closely the peaks of the mmMAGIC model, meaning that it underestimates the loss somewhat. It is, however, likely that the average signal strength of the mmMAGIC model is within 3 dB of the METIS model. Given that the METIS model is substantially less complex, this model might be preferred in many cases.

3.5 SUMMARY AND FUTURE WORK

In the past decade there have been extensive measurement and modeling efforts for understanding and characterization of propagation for development and optimization of 5G mobile communications. One of the major challenges has been to provide understanding of how the propagation characteristics might change when going from legacy carrier frequencies around 2 GHz up to frequencies higher than 80 GHz. Even if many of the propagation characteristics are fairly similar at those higher frequencies, the antenna sizes scale with wavelength. Consequently, the aperture of any type of antenna is proportional to the square of the wavelength, meaning that the received power scales with $-20\log f$ [dB]

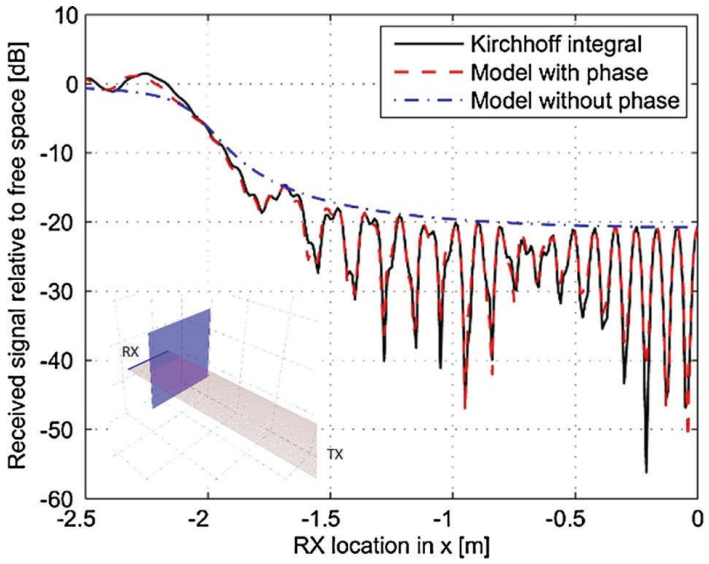


FIGURE 3.39

METIS (without phase) and mmMAGIC (with phase) blockage models at 4 GHz for a 4x4 m screen with Rx 1 m behind the screen and Tx 100 m away in front of the screen. The signal for the RX moving from the center behind the screen to LoS is shown.

relative to the power of the impinging waves. The corresponding increase in loss is more than 30 dB when going from 2 GHz to 80 GHz. However, by utilizing the available area using array antennas or other beam-forming techniques this increase in transmission loss may be compensated for and even turned to a gain when the frequency goes up as shown in Section 3.1.2. In order to optimize advanced antenna transmission techniques utilizing the directional domain, it is crucial that corresponding channel modeling is realistic. As shown in this chapter the knowledge in this area has indeed made progress in the past years. This knowledge is largely implemented in 5G propagation modeling, provided 3GPP and ITU-R, ensuring successful development and optimization of initial 5G cellular communications systems. As higher frequencies and more narrow beams are utilized in later releases of 5G systems higher requirements are put on channel modeling accuracy. For this reason a number of areas in which propagation model improvements would be valuable in order to ensure successful development and optimization of mobile communications systems in the longer term have been identified and summarized as follows:

- Current modeling of highly resolved channel characteristics, in both the delay and the direction domains, is largely arbitrary. Comparing with corresponding measurement data it is clear that there are substantial differences in the structure of multipath distributions and their clustering. Particularly, joint angular and polarimetric distributions for both ends of the link are poorly characterized for outdoor and outdoor-to-indoor scenarios.
- The importance to have accurate knowledge of dynamic variations of the channel including transmission loss is increasingly important with increasing carrier frequency, to make possible optimization of methods to quickly adapt beam-forming to multipath component death and birth processes.

Experimental data supporting such modeling is largely lacking for important scenarios and frequency ranges.

- The understanding of the dependence of transmission loss of different propagation mechanisms, particularly non-specular scattering, is largely lacking. It is clear that diffraction is not very significant under most non-line-of-sight conditions. However, what the main scattering is caused by is not well known. Moreover, it is not well understood to what extent there are frequency dependent trends of loss in excess of free-space loss, particularly for macrocell scenarios. Here, the height dependence is also poorly known.
- The outdoor-to-indoor scenario is very important as most cellular networks are deployed outdoors, while about 80% of the users are located indoors. Quite extensive measurement campaigns have been undertaken for characterizing the additional loss due to building penetration. As there is a vast distribution of different building types worldwide there is still a substantial need for more measurement data. Moreover, angle of incidence and polarization effects are poorly characterized, motivating further experimental investigations.
- New propagation scenarios are becoming increasingly important for 5G. For example machine type communications in e.g. factories or between drones or between drones and ground have recently drawn attention. These scenarios are not well characterized and need more empirical data.

These are only examples of identified areas for future work. Many additional areas, which are not easily foreseen today, are likely to be important in the future. It is, however, a fact that profound knowledge of propagation in general is increasingly important, as more of the degrees of freedom of electromagnetic propagation are utilized in mobile communications, and that all empirical data provided as a basis for modeling will be highly valuable also in the future.

REFERENCES

[1] 3GPP, Study on 3D channel model for LTE, Technical Report 36.873, 2017.
[2] 3GPP, Study on channel model for frequencies from 0.5 to 100 GHz, Technical Report 38.901, 2018.
[3] L. Barclay, Propagation of Radiowaves, 2, Iet, 2003.
[4] G.v. Borgiotti, Maximum power transfer between two planar apertures in the Fresnel zone, IEEE Transactions on Antennas and Propagation 14 (2) (1966) 158–163.
[5] B.H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, K.I. Pedersen, Channel parameter estimation in mobile radio environments using the sage algorithm, IEEE Journal on Selected Areas in Communications 17 (3) (1999) 434–450.
[6] ITU-R, Guidelines for evaluation of radio interface technologies for IMT-Advanced, Report M.2135, 2009.
[7] ITU-R, Effects of building materials and structures on radiowave propagation above about 100 MHz, Recommendation ITU-R P.2040, 2015.
[8] ITU-R, The concept of transmission loss for radio links, Recommendation ITU-R P.341-6, 2016.
[9] ITU-R, Guidelines for evaluation of radio interface technologies for IMT-2020, Report M.2412, 2017.
[10] ITU-R, Prediction of building entry loss, Recommendation ITU-R P.2109, 2017.
[11] ITU-R, Propagation by diffraction, Recommendation ITU-R P.526, 2018.
[12] ITU-R, Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz, Recommendation ITU-R P.1411, 2018.
[13] J.D. Jackson, Electrodynamics, Wiley Online Library, 1975.
[14] N. Jaldén, J. Medbo, H. Asplund, N. Tompson, D. Sundman, Indoor high-resolution channel characterization, in: Antennas and Propagation (ISAP), 2016 International Symposium on, 2016, pp. 618–619.
[15] V.M. Kolmonen, P. Almers, J. Salmi, J. Koivunen, K. Haneda, A. Richter, F. Tufvesson, A.F. Molisch, P. Vainikainen, A dynamic dual-link wideband mimo channel sounder for 5.3 GHz, IEEE Transactions on Instrumentation and Measurement 59 (4) (2010) 873–883.

[16] H. Matzner, M. Milgrom, S. Shtrikman, A study of finite size power isotropic radiators, in: Electrical and Electronics Engineers in Israel, 1995, Eighteenth Convention of, 1995, pp. 1–4.
[17] J. Medbo, F. Harrysson, Efficiency and accuracy enhanced super resolved channel estimation, in: Antennas and Propagation (EUCAP), 2012 6th European Conference on, 2012, pp. 16–20.
[18] J. Medbo, H. Asplund, J.E. Berg, N. Jalden, Directional channel characteristics in elevation and azimuth at an urban macrocell base station, in: Antennas and Propagation (EUCAP), 2012 6th European Conference on, 2012, pp. 428–432.
[19] J. Medbo, H. Asplund, J.E. Berg, 60 GHz channel directional characterization using extreme size virtual antenna array, in: Personal, Indoor, and Mobile Radio Communications (PIMRC), 2015 IEEE 26th Annual International Symposium on, 2015, pp. 176–180.
[20] J. Medbo, P. Kyosti, K. Kusume, L. Raschkowski, K. Haneda, T. Jamsa, V. Nurmela, A. Roivainen, J. Meinila, Radio propagation modeling for 5G mobile and wireless communications, IEEE Communications Magazine 54 (6) (2016) 144–151.
[21] J. Medbo, N. Seifi, H. Asplund, Frequency dependency of measured highly resolved directional propagation channel characteristics, in: European Wireless 2016; 22nd European Wireless Conference; Proceedings of, 2016, pp. 1–6.
[22] J. Medbo, D. Sundman, H. Asplund, N. Jaldén, S. Dwivedi, Wireless urban propagation measurements at 2.44, 5.8, 14.8 & 58.68 GHz, in: General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS), 2017 XXXIInd, 2017, pp. 1–4.
[23] J. Medbo, C. Larsson, B.E. Olsson, F.S. Chaves, H.C. Nguyen, I. Rodriguez, T.B. Sorensen, I.Z. Kovács, P. Mogensen, K.T. Lee, J. Woo, M. Sasaki, W. Yamada, The development of the ITU-R terrestrial clutter loss model, in: Antennas and Propagation (EUCAP), 2018 12th European Conference on, 2018.
[24] METIS, METIS channel models, Deliverable D1.4, 2015.
[25] mmMagic, Measurement campaigns and initial channel model for the frequency range 6–100 GHz, Deliverable D2.1, 2016.
[26] mmMagic, Measurement results and final mmMAGIC channel models, Deliverable D2.2, 2017.
[27] H.C. Nguyen, I. Rodriguez, T.B. Sorensen, L.L. Sanchez, I. Kovacs, P. Mogensen, An empirical study of urban macro propagation at 10, 18 and 28 GHz, in: Vehicular Technology Conference (VTC Spring), 2016 IEEE 83rd, 2016, pp. 1–5.
[28] S.L.H. Nguyen, J. Medbo, M. Peter, A. Karttunen, K. Haneda, A. Bamba, R. D'Errico, N. Iqbal, C. Diakhate, J.M. Conrat, On the frequency dependency of radio channel's delay spread: analyses and findings from mmMAGIC multi-frequency channel sounding, in: Antennas and Propagation (EUCAP), 2018 12th European Conference on, 2018.
[29] A. Osseiran, J.F. Monserrat, P. Marsch, 5G Mobile and Wireless Communications Technology, Cambridge University Press, 2016.
[30] A. Richter, Estimation of radio channel parameters: Models and algorithms, Ph.D. thesis, Technischen Universität Ilmenau, 2005.
[31] A.A. Saleh, R. Valenzuela, A statistical model for indoor multipath propagation, IEEE Journal on Selected Areas in Communications 5 (2) (1987) 128–137.
[32] S. Salous, Multi-band multi-antenna chirp channel sounder for frequencies above 6 GHz, in: Antennas and Propagation (EuCAP), 2016 10th European Conference on, 2016, pp. 1–4.
[33] W. Scott, K.S. Hoo, A theorem on the polarization of null-free antennas, IEEE Transactions on Antennas and Propagation 14 (5) (1966) 587–590.
[34] R. Thomä, M. Landmann, A. Richter, Rimax-a maximum likelihood framework for parameter estimation in multidimensional channel sounding, in: Proceedings of the International Symposium on Antennas and Propagation (ISAP'04), 2004, pp. 53–56.

CHAPTER

4 MATHEMATICAL MODELING OF HARDWARE IMPAIRMENTS

As Gordon Moore accurately predicted more than 50 years ago [19], the number of transistors in a dense integrated circuit doubles approximately every two years. In retrospect we have seen the industry follow Moore's observation and it is clear that this development has paved the way for high speed digital communication systems with highly integrated radio transceivers. This development has led to a movement of the border between analog and digital domains in which we today see radio transceivers which are becoming predominantly digital in design.

Through the development of high speed, real-time processing, techniques for compensating imperfections in analog hardware are today commonly used extensively due to stringent requirements imposed by the physical layer design. At the basis for any effective compensation algorithm lies a behavioral model mimicking the behavior of an imperfect analog component. Algorithms such as DPD for linearization of power amplifiers and phase-noise tracking all rely on accurate behavioral modeling in order to perform accurate mitigation.

One aspect that underlines the importance of radio hardware impairment modeling is the constant increase of throughput which is enabled by either using large channel bandwidths at millimeter-wave frequencies or by increasing the spectral efficiency by a denser modulation format and spatial multiplexing. In both of these directions, new challenges face the radio hardware design and the importance of accurate impairment models grows. One particular technology component important for 5G is that of multiantenna techniques such as massive MU-MIMO for spatially multiplex users at sub-6 GHz, along with different methods for analog or hybrid beam-forming techniques aimed toward higher frequencies in the millimeter-wave frequency bands. In both of these areas, impairment models may not only serve as a basis for advanced compensation mechanisms, but also provide valuable insights in how the impairments behave in the spatial domain.

Impairment modeling has also in recent years gained interest from researchers working in the field of communication systems design. Until now, however, may of the models used have been oversimplified for analytical convenience, [3]. However, under some circumstances these simplified models, which are quite often just additive and independent noise, may provide misguiding insights as they do not accurately represent the statistical properties of the impairments, [17].

Conventional behavioral models are however lacking in this aspect as they require time-domain signals, often oversampled. Since knowing the statistical properties of the distortion suffices for conventional link-level analysis which often relies on asymptotics, alternative approaches may be considered in which the impairments are modeled as a correlated noise process. As will be shown in this chapter, the properties of this noise process may be determined using parameterizations of the corresponding behavioral model and knowledge of the transmit covariance.

On a final note—the underlying mathematics behind most common behavioral models used to capture effects of hardware impairments may be more than a century old but, as we will illustrate in

5G Physical Layer. https://doi.org/10.1016/B978-0-12-814578-4.00009-6
Copyright © 2018 Elsevier Ltd. All rights reserved.

Samsung Exhibit 1020, Page 100 of 314

this chapter, the modeling work is still evolving in order to support the design of future radio access technologies and advanced antenna systems.

In this chapter, we will overview and discuss the current state-of-the-art in behavioral modeling of some important radio hardware subsystems. We will start with the power amplifier and the Volterra series commonly used to model their behavior. We will also discuss a novel extension used to model power amplifiers in antenna systems which may be under the influence of mutual coupling, something which is becoming increasingly important with the current drive toward compact antenna arrays operating at millimeter-wave frequencies. This is followed by a discussion on oscillator phase noise and the underlying Wiener process commonly used for modeling. We then move on to discuss quantization noise in data converters and how to model the granular noise from the quantization process and the distortion arising from clipping.

In the second part of this chapter, we will introduce and discuss a stochastic framework for modeling of radio hardware impairments in terms of second-order statistics. Analytical expressions for the second order statistics is derived using the Bussgang theorem and the moment theorem for Gaussians. This approach is aimed toward applications in which oversampled time-domain signals are either not available or simply not useful, such as link- or system-level simulations. An example of numerical simulations for an OFDM-based massive MU-MIMO case is presented.

4.1 RF POWER AMPLIFIERS

Power amplifiers are known as the most nonlinear subsystem component of any radio transmitter. As such, modeling and understanding their behavior in order to compensate for the nonlinear distortion it produces in order not to violate the requirements set by the agreed upon standards is crucial.

3GPP defines two important metrics commonly used to quantify the linearity performance, which are error vector magnitude (EVM) and adjacent channel leakage ratio (ACLR). EVM specifically accounts for the performance between a base station (BS) and its scheduled users as it measured the in-band quality of the links. ACLR accounts for the interference leaked into the adjacent frequency channels which is crucial for the co-existence between networks to function. ACLR is a crucial metric for which the regulatory requirements are quite stringent in order to reduce the interference between networks and other applications in the same frequency band.

As power efficiency is highly important for reducing the environmental impact and the total cost of operation and the power amplifier being one of the major consumers of power, different efficiency enhancement technologies have been developed over the years. Amplifier topologies such as the Chireix [6] or the Doherty [8] amplifier, are commonly used, although the latter is more favored in recent scientific literature and in most industrial applications. This is in great measure due to its rather low implementation complexity compared to the Chireix amplifier and the performance in terms of power efficiency.

In more recent years, generalizations of the Doherty concept has been developed, supporting high efficiency over larger dynamic range. This further strengthens the Doherty amplifier in its position as the most popular amplifier topology. However, the improved efficiency performance comes at the cost of nonlinear behavior, [15], which may impact the overall system design depending on the transmit power.

The modeling needs vary in different applications and amplifier topologies, but in general the effects commonly covered in the literature are:

- Nonlinear transfer function and compression.
- Dispersive effects from impedance matching and bias networks.
- Electron traps from material artifacts.
- Long-term thermal memory effects due to heating and cooling.
- Load-modulation via mutual coupling in between antennas.

We will now briefly introduce the Volterra series and discuss some of the more common subsets of the Volterra series used in the scientific literature. Strengths and weaknesses of each model are also discussed and illustrated using measurements. Further on, we will discuss some of the more non-conventional models using orthogonal basis functions. We will end the section on power amplifier modeling with a discussion of global and local basis functions. It should be noted that the models discussed here are all acting in complex baseband and, thus, this carries no information regarding RF harmonics and other passband specific aspects.

4.1.1 THE VOLTERRA SERIES

The Volterra series dates back to 1887 [24] and may be considered as one of the most powerful tools for modeling not only radio power amplifiers, but nonlinear time invariant systems in general. The full Volterra series may be viewed as a superset of polynomial-based nonlinear models containing permutations of the input signal. As many of these are not always relevant in all modeling scenarios, many efforts have been made on pruning the full Volterra series in order to tailor it for specific amplifier topologies.

Much of the very early work on the Volterra series was performed over real-valued variables for passband systems. However, due to the quadrature nature of digital communication systems and the analytical manners by which these are modeled, the natural step of extending the Volterra series to complex variables was taken, [2].

The discrete-time, complex-valued Volterra series for causal systems with symmetric kernels ($\theta_{i,j} = \theta_{j,i}$) of nonlinear order P and memory-depth M can be written as

$$\begin{aligned} y[n] \;=\; & \sum_{p=1,3,5,\ldots}^{P} \sum_{m_1=0}^{M} \sum_{m_2=m_1}^{M} \cdots \sum_{m_{(p+1)/2}=m_{(p-1)/2}}^{M} \\ & \times \sum_{m_{(p+3)/3}=0}^{M} \cdots \sum_{m_p=m_{p-1}}^{M} \theta_{p,m_1,\ldots,m_p} \times \prod_{l=1}^{(p+1)/2} x[n-m_l] \prod_{j=(p+3)/2}^{p} \bar{x}[n-m_j]. \end{aligned} \tag{4.1}$$

We may note here that in the Volterra series, as stated in Eq. (4.1), there will always be an even number of conjugate terms accompanied by an odd amount of non-conjugate counterparts. This causes the frequent occurrence of terms in the form of $x[n-m_i]|x[n-m_j]|^p$, which plays an important role in power amplifier modeling, as the behavior is for a large part dependent on the signal envelope.

One beneficial aspect of any model based on the Volterra series is that they are linear in the parameters. This means that the model may be written as a simple matrix-vector multiplication,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta}, \tag{4.2}$$

in which

$$\mathbf{X}(n,l) = \prod_{l=1}^{(p+1)/2} x[n-m_l] \prod_{j=(p+3)/2}^{p} \bar{x}[n-m_j] \tag{4.3}$$

is the nth row and lth column of the regression matrix and $\boldsymbol{\theta}$ is a vector containing the coefficients $\theta_{p,m_1,\ldots,m_p}$. This linear system of equations yields $\mathbf{y} = \left[y[0],\ldots,y[N]\right]^T$, which is the vector of output samples. Due to these models being parameter linear, the identification process is generally quite simple, provided that the regression matrix is numerically sound and not ill-conditioned.

The mean squared error (MSE)-optimal parameter estimate can be found as

$$\hat{\boldsymbol{\theta}} = \mathbf{X}^{+}\mathbf{y} \tag{4.4}$$

where we are making use of the Moore–Penrose pseudo-inverse, denoted $\mathbf{X}^{+}$.

4.1.2 COMMON SUBSETS OF THE VOLTERRA SERIES

The Volterra series as presented in Eq. (4.1) contains many terms which may not be providing useful degrees of freedom for accurate modeling. In order to avoid over- or noise-fitting by introducing unnecessary basis functions, some reductions of the Volterra series have been proposed over the years. We will now cover some of the more important ones.

4.1.2.1 Static Polynomial

The smallest and simplest subset of the Volterra series is that of the static polynomial. This model contains only terms at time n and can therefore not accurately model systems which are dispersive in nature. This excludes numerically accurate modeling of power amplifiers operating at a large fractional bandwidth, as these often exhibit dynamic memory effects, making the nonlinear behavior of the amplifier strongly frequency dependent. For narrow-band applications, however, the static polynomial model may very well be sufficient.

The static polynomial as it occurs in most literature is simply defined as

$$y[n] = \sum_{p=1}^{P} \theta_p x[n]|x[n]|^{2p-1} \tag{4.5}$$

where θ_p are the model parameters, sometime referred to as *kernels*. One very important observation regarding the static model, is that its asymmetries in the distortion spectrum may not be captured with this model as there is no memory involved. The output distortion spectrum will therefore always be symmetric around the center frequency.

Third-Order Static Polynomial

Most models covered in this chapter are aimed at accurate numerical predictive simulations or at providing a basis for DPD-algorithms. However, at certain points, simplified models may provide valuable insight in terms of analytical results. One special case of the static polynomial most suitable for this is that of the third-order static polynomial model. This is simply written as

$$y[n] = \theta_1 x[n] + \theta_2 x[n]|x[n]|^2. \tag{4.6}$$

Examples on how to use this model in conjunction with the moment theorem for Gaussians in order to analyze the impact of power amplifier distortion in massive MIMO arrays are shown in [18]. It may seem simplistic, but for amplifier operating in a reasonably nonlinear region, the third-order term is most commonly the more dominant one, even if higher-order terms are present. Thus, this may serve as a decent first-order approximation of a PA for analytical exercises, leading to valuable insight without having to struggle with tedious analytical manipulation.

4.1.2.2 A Note on Odd–Even and Odd Orders

Before moving on to more advanced behavioral models, we need to examine an issue which often occurs throughout the scientific literature. In regards to power amplifier behavioral modeling, there are mainly two different conventions used when defining the nonlinear orders of the model. These two conventions are commonly referred to as *odd–even-order* or just simply *odd-order* models. We will discuss this here using the static polynomial as an example, but the same reasoning applies for most Volterra-based models. The model which is written as

$$y[n] = \sum_{p=1}^{P} \theta_p x[n]|x[n]|^{2(p-1)} \tag{4.7}$$

is most commonly referred to as odd order, both since $2(p-1)$ is an odd number and since $x[n]|x[n]|^{2(p-1)}$ is a permutation of $x[n]$ an odd number of times. Another, slightly different convention commonly which is slightly less likely to occur in the scientific literature is written as

$$y[n] = \sum_{p=1}^{P} \theta_p x[n]|x[n]|^{p-1}, \tag{4.8}$$

in which the basis functions are referred to as odd–even, since the term $x[n]|x[n]|^{p-1}$ always is an odd permutation over $x[n]$, even if the power of $|x[n]|$ is allowed to be odd instead of solely even.

It has been claimed in the past that an odd–even model provides more degrees of freedom, [7], and should therefore provide a richer description of the amplifier behavior. However, this is a claim that has been rebuked by the authors of [13], who show that the odd–even- and odd-order models have the same modeling properties and the odd–even orders do not provide any richer description of the amplifier. This is a result of expanding the term $|x|$ in terms of $|x|^{2p}$, and showing that the series is uniformly convergent. The difference between orders in terms of modeling accuracy is in fact proportionality to the model truncation error.

Moving on, we will in this chapter stick to the more common odd-order version in the following sections as this is the conventional ordering in most publications.

4.1.2.3 Memory Polynomial

The first natural extension of the static polynomial is to add memory terms in order to model dispersive effects introducing a frequency selective behavior. The memory polynomial (MP) model as described in [12], can be written in its complex baseband form as

$$y[n] = \sum_{p=1}^{P} \sum_{m=1}^{M} \theta_{p,m} x[n-m]|x[n-m]|^{2(p-1)}, \tag{4.9}$$

where P is the nonlinear order and M is the memory depth. From Eq. (4.9), we see that the MP is in principle a nonlinear finite impulse response (FIR) filter. As described in [20], the MP is an extension to the common Hammerstein model in which a nonlinear function is followed by a linear filter.

In the MP model only diagonal memory terms from the Volterra series are included, which in some cases may limit the modeling capabilities, particularly in the case that the amplifier in question displays a strong, nonlinear memory for which cross-terms will be necessary. For simpler applications and for analytical manipulation, this model is often good enough and easy to use.

4.1.2.4 Generalized Memory Polynomial

The generalized memory polynomial (GMP) was first presented in [20]. GMP extends the aforementioned MP by introducing cross-terms in the memory-domain [12], which is written as

$$\begin{aligned} y[n] &= \sum_{p=1}^{P} \sum_{m=1}^{M} \alpha_{p,m} x[n-m]|x[n-m]|^{2(p-1)} \\ &+ \sum_{p=1}^{P} \sum_{m=1}^{M} \sum_{l=1}^{L} \beta_{p,m,l} x[n-m]|x[n-m-l]|^{2(p-1)} \\ &+ \sum_{p=1}^{P} \sum_{m=1}^{M} \sum_{l=1}^{L} \gamma_{p,m,l} x[n-m]|x[n-m+l]|^{2(p-1)} \end{aligned} \tag{4.10}$$

where P is the nonlinear order, M is the memory depth and L is the number of leading and lagging cross-terms. In this model, the memory terms span an L-dimensional strip in the memory-space of the Volterra series. Reference [20] also introduces a more general way of creating the model using index arrays in order to facilitate pruning of the parameter space in order to achieve the desired complexity/performance trade-off.

4.1.3 GLOBAL VS. LOCAL BASIS FUNCTIONS

All of the models examined so far are using so-called global basis functions. This is in the sense of basis functions acting upon the entire range of the envelope. This may in some cases cause numerical issues, in particular when using high-order polynomial basis functions in conjunction with signals, which has a large dynamic range combined with low probabilities of occurring peaks. This will end up causing a poor numerical fit as the number of samples collected around the peak power will be small, which tends to have a negative impact particularly on the high-order terms as their estimator variance rapidly increases with order.

One straightforward solution to this numerical issue is to use so-called local basis functions. By this, we mean dividing the envelope-space into segments which then get associated with their own specific set of basis functions. Let $\mathcal{P}$ be the envelope-range which we divide into Ω-ordered sections as

$$\mathcal{P} = \bigcup_{\omega=1}^{\Omega} \mathcal{P}_u \tag{4.11}$$

where $\mathcal{P}_u \cap \mathcal{P}_{u'} = \varnothing$ for $u \neq u'$. We can now use this to form modified versions of the models previously discussed. For example, examining the static polynomial, we can write

$$y[n] = \sum_{p=1}^{P} \sum_{\omega=1}^{\Omega} \mathbf{I}_{|x[n]| \in \mathcal{P}_\omega} \theta_p x[n] |x[n]|^{2(p-1)}, \tag{4.12}$$

in which $\mathbf{I}$ is the indicator function. It may be difficult to see from the equation above, but this yields a regression matrix which is largely sparse, e.g. the majority of elements will be zero. This may increase the overall numerical complexity of the model, but as the regression matrix will have a sparse structure, this complexity may be reduced using sparse matrix methods for computation.

In Fig. 4.1, we see one example of a static polynomial fitted with local basis functions on data generated via a memoryless polynomial. As the number of sections increases, we increase the modeling accuracy at the cost of added complexity.

Quite simple examples of local basis functions may be in the form of a look-up table (LUT) in which elements of the table only act upon its own specific region of the envelope. Another example which allows for switching between local cross-terms in the memory-space, is the so-called vector-switching model [1].

4.1.4 EXPERIMENTAL MODEL VALIDATION

Concluding the discussion on power amplifier modeling, the numerical accuracy will be examined using a set of measurements performed on a Gallium-Nitride (GaN)-based class-AB amplifier using the web-lab system have been used [14]. The models considered in this chapter will be the static polynomial, the GMP, and a vector-switching extension of the GMP. For easy quantification of the model performance, we need a set of metrics.

4.1.4.1 Quantifying Modeling Performance

In order to make a sound selection of the model to suit the application in mind, we need to examine how these behave using measurement data. For this purpose, we need to consider a couple of important figures of merit. One of the most commonly used figures of merit is the normalized mean squared error (NMSE). Here, we compare the measured and modeled output over the entire measurement bandwidth. The modeling error is then normalized with the measured output variance in order to make the figure

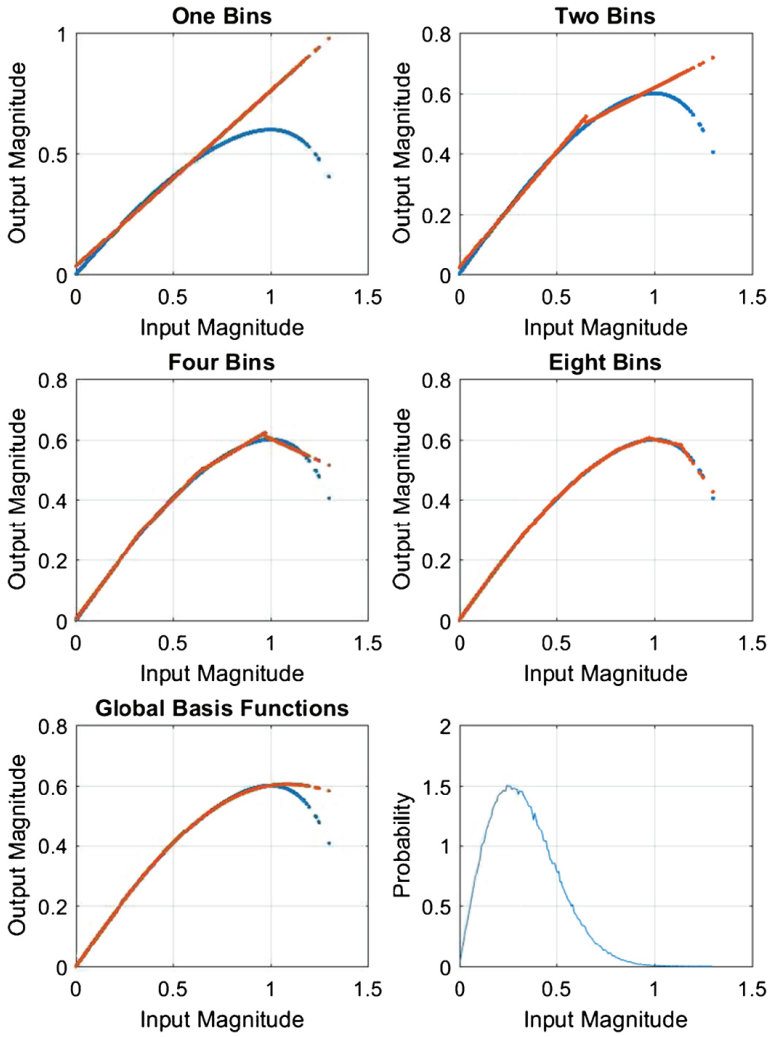


FIGURE 4.1

Illustration of local vs. global basis functions. An illustration of how local basis functions fit to a static model reference. Here, piecewise linear functions are used over different numbers of envelope bins.

power independent. NMSE is commonly defined as

$$\text{NMSE} = \frac{\sum_{n=0}^{N} |y_{\text{meas}}[n] - y_{\text{mod}}[n]|^2}{\sum_{n=0}^{N} |y_{\text{meas}}[n]|^2} \tag{4.13}$$

Table 4.1 Model validation results

PA model	NMSE (dB)	ACEPR (dB)
Static polynomial	−28.5	−29.0
GMP	−32.5	−33.5
VS-GMP	−37.0	−38.0

where y_{meas} is the measured output and y_{mod} is the modeled output. Compared to conventional MSE, the NMSE removes any power dependencies, enabling comparison over different power ranges if necessary.

The second figure of merit useful to assess the modeling performance is the adjacent channel error power ratio (ACEPR), which provides a measure of how accurate the model predicts the generated adjacent channel distortion. The ACEPR is defined as

$$\text{ACEPR} = \frac{\int_{\text{Adj. band}} |y_{\text{meas}}(f) - y_{\text{mod}}(f)|^2 \, df}{\int_{\text{Adj. band}} |y_{\text{meas}}(f)|^2 \, df}. \tag{4.14}$$

We may in ACEPR observe similarities to NMSE, but with the emphasis put on modeling of the distortion outside of the assigned channel.

In order to evaluate the modeling performance, we need two sets of measurement data—one for identification and one for evaluation. This is required to provide a fair comparison such that the model is not fitted to one specific signal realization. In this experiment, we use filtered complex Gaussian noise which commonly has statistical properties very similar to any wide-band OFDM-based system due to the central limit theorem. Since we are not evaluating a postdecoding metric such as bit error rate (BER), Gaussian noise is sufficient.

Fig. 4.2 depicts the measured power spectral density (PSD) and amplitude-to-amplitude modulation (AMAM) of the identification and evaluation data in terms of instantaneous normalized gain.

The model parameters are then found using the Moore–Penrose pseudo-inverse as described in Eqs. (4.3) and (4.4). After performing the identification, each model runs with the evaluation data which is compared to the measurements. The results in terms of NMSE and ACEPR are presented in Table 4.1.

In order to further illustrate the modeling capacities for each case, we may examine the AMAM characteristics and PSD from measurement data. Fig. 4.3 illustrates the measured and simulated AMAM. Fig. 4.4 shows the corresponding simulated PSD for the models.

As we may observe, the static polynomial shows no dispersion effects due to its inherent lack of memory. The GMP however, is capable of modeling these effects as shown. As this particular amplifier is quite nonlinear, Fig. 4.3 illustrates the manner in which with both the static polynomial and the GMP with global basis functions are having certain difficulties fitting to the data near peak power. This is however not the case for the VS-GMP, which further improves the modeling accuracy by applying local basis functions. In each region of the envelope, a low-order GMP ($P = 3$) is applied. Further

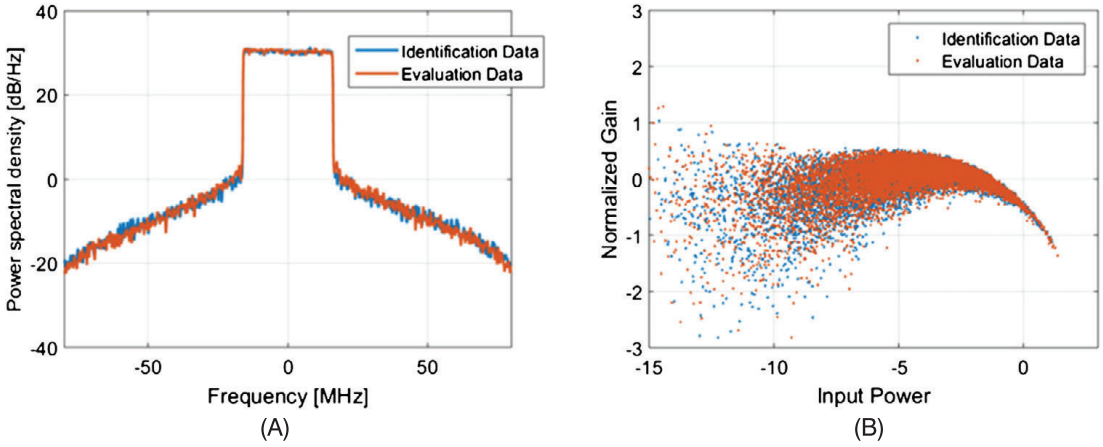


FIGURE 4.2

Measured PSD (A) and AMAM (B) characteristics of the identification and evaluation data used in the comparison between models.

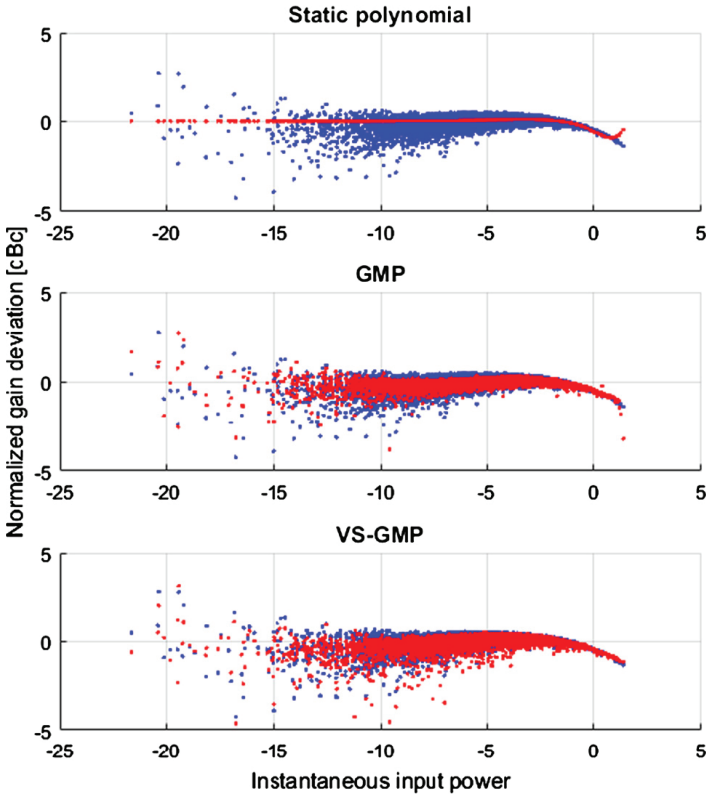


FIGURE 4.3

Comparison between measurement and models. Comparative simulations between the static polynomial model and the GMP, both with global and local basis functions. Blue (dark gray in print version) curves denote measurement data, red (light gray in print version) curves denote predicted output from the models.

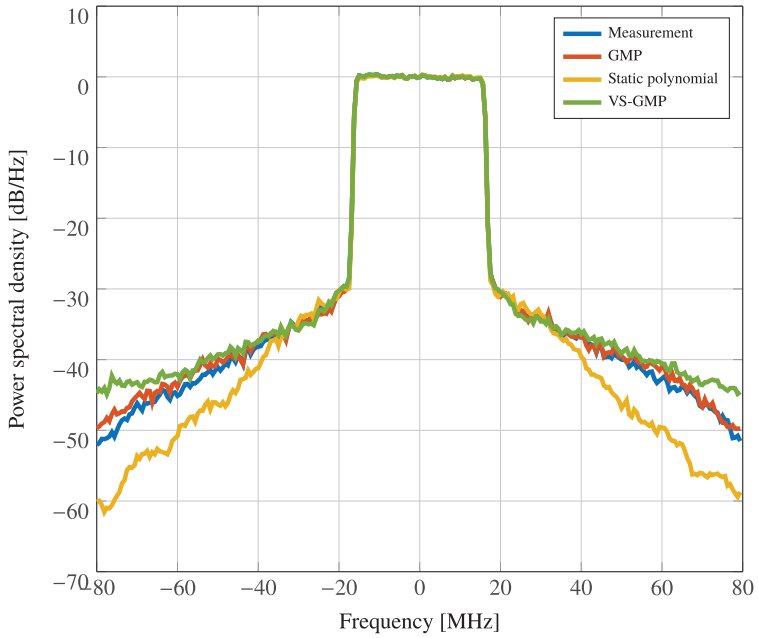


FIGURE 4.4

Comparative simulations between the static polynomial model and the GMP using global and local basis functions.

practical comparisons between models using global and local basis functions may be found in [1]. The numerical results are summarized in Table 4.1.

4.1.5 MUTUALLY ORTHOGONAL BASIS FUNCTIONS

A less common, yet useful model for power amplifier modeling is constructed by using mutually orthogonal basis functions. Consider the expansion of a nonlinear function using the basis functions H_p,

$$y[n] = \sum_{p=1}^{P} \theta_p H_p(x[n]). \tag{4.15}$$

These basis functions can be designed to be mutually orthogonal with respect to the statistics of our input signal. That implies, for all p, k, that the property

$$\mathbb{E}\left[H_p(x[n])H_k^*(x[n])\right] = 0 \tag{4.16}$$

needs to hold. For the case of a complex Gaussian distributed input signal, we can achieve this mutual orthogonality by using the complex Itô-generalization of the Hermitian polynomials, at least for a given input power σ_x^2.

The mutual orthogonality between each basis function may also be beneficial as it provides a description in which the linear term, e.g. the useful signal part, is orthogonal to the distortion as this is a

linear combination of the remaining, mutually orthogonal components. We write this as

$$y[n] = s[n] + d[n] \tag{4.17}$$

where

$$s[n] = \theta_1 H_1(x[n]) \tag{4.18}$$

is the linearly scaled input signal and

$$d[n] = \sum_{p=2}^{P} \theta_p H_p(x[n]) \tag{4.19}$$

is the distortion which is expressed using the complex Itô polynomials; now these are mutually orthogonal to the input signal,

$$\mathbb{E}[s[n]d^*[n]] = 0. \tag{4.20}$$

The property of mutual orthogonality provides a clearer path to determining the statistical properties of the distortion in terms of its covariance matrix, in a similar manner as the linear stochastic model discussed in Chapter 4.4.

This approach provides a simple way of analyzing of the spatial properties of the distortion, in the context of MU-MIMO [17]. Here, it is shown that

$$\frac{K^3 + K^2}{2} L^2 v(f) \tag{4.21}$$

where K is the number of users or transmit-directions, L is the number of significant taps of the channel and $v(f)$ is a constant depending on the excess bandwidth. A necessary condition for the distortion to behave omni-directionally is

$$\frac{K^3 + K^2}{2} L^2 v(f) \geq M. \tag{4.22}$$

As Fig. 4.5 illustrates, we see that, as the number of users increases from $K = 1$ to $K = 3$ and 10, the number of dominant directions increases with $\frac{K^3+K^2}{2}$ (as this is modeled using a frequency-flat, narrow-band channel with $L^2 v(f) = 1$). A few further observations may be made from Fig. 4.5 is that the radiated adjacent power is at worst in parity with the single-antenna case. This occurs only in the case of $K = 1$ and only towards the user served. Everywhere else, any potential victim user will receive less distortion compared to the corresponding single-antenna case.

As K then increases, the adjacent channel power is divided over more directions than that of the in-band signal. Thus, it never reaches the same radiated power as in the single-antenna case. This indicates that, in most practical cases, the impact of the radiated distortion will be less in a large array than in the case of a single-antenna system. More discussion and rigorous derivation of the radiation patterns can be found in [17].

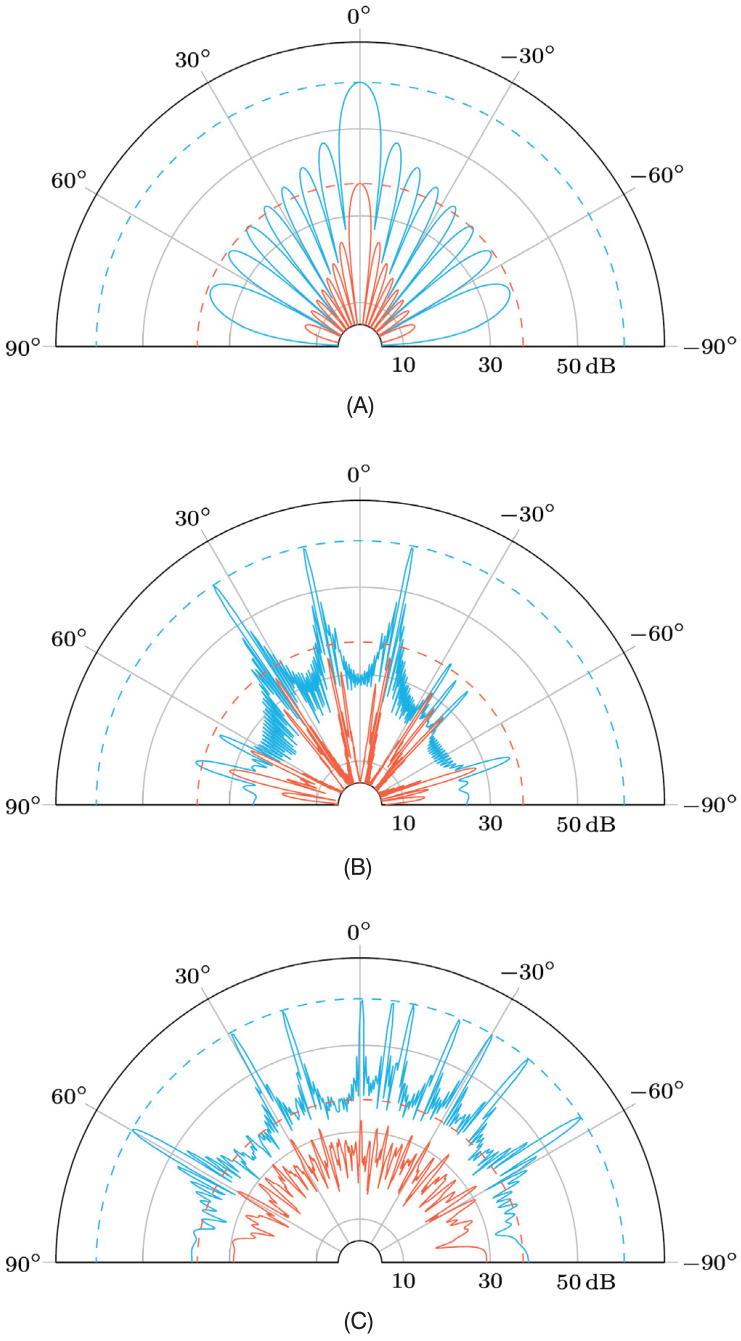


FIGURE 4.5

Radiation patterns from uniform linear arrays. Blue (dark gray in print version) curves indicate the in-band radiated power. Red (light gray in print version) curves indicate the power radiated in the adjacent channel. The dashed lines correspond to the single-antenna case with the same ACLR. (A) Radiated pattern for in- and adjacent-band in the case of $K = 1$ and $M = 16$, (B) radiated pattern for in- and adjacent-band in the case of $K = 3$ and $M = 128$, (C) radiated pattern for in- and adjacent-band in the case of $K = 10$ and $M = 128$.

4.1.6 MULTI-ANTENNA ENVIRONMENTS AND MUTUAL COUPLING

Another aspect of power amplifier modeling which has become increasingly important in multiantenna applications is modeling of the mutual coupling between the antennas and its impact on the power amplifier behavior. This coupling can create a signal-dependent and time-varying load-impedance seen by the amplifier, which in turn changes the behavior in terms of the instantaneous transducer gain. Such effects may be seen by plotting the normalized instantaneous gain, as shown in Fig. 4.7. The large variation of instantaneous gain, which may be seen in Fig. 4.7, is caused by load-modulation of the power amplifier via mutual coupling between branches.

In order to model the effects of mutual coupling, we need to view the power amplifier as a dual-input system in which the second added input is the reflected wave stemming from the transmitted signals from neighboring transmitters, transfered via mutual coupling between antennas as illustrated in Fig. 4.6. Here, $x_1, \ldots, x_M$ are the M transmit signals and $y_1, \ldots, y_M$ the M output signals affected by the nonlinear distortion of the power amplifiers. We also need the variables $y_{r,1}, \ldots, y_{r,M}$, which are mixtures of transmit signals being reflected back into each amplifier via the coupling.

At every time instant n, we may compute the reflected wave to the antenna port m as the scalar product

$$y_{r,m} = \mathbf{y}^T \boldsymbol{\lambda}_m, \tag{4.23}$$

in which $\boldsymbol{\lambda}_m = [\lambda_m 1m, \ldots, \lambda_{Mm}]$ is the parameter vector describing the mutual coupling and $\mathbf{y} = \left[y_1, \ldots, y_M\right]$ is the vector of the transmitted signals.

The first approach to modeling such effects was proposed by Root et al. [22], through a model referred to as the polyharmonic distortion (PHD) model. As a large-signal extension to the S-parameters, and thus nonlinear, this model is, however, quasistatic in nature. This has led to the development of an extension based on the memory polynomial, as derived and validated in [9]. We have

$$\begin{aligned} y_m[n] &= \sum_{p=1}^{P} \sum_{m=1}^{M_1} \alpha_{p,m} x_m[n-m_1] |x_m[n-m_1]|^{2p-1} \\ &+ \sum_{p=1}^{P} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \beta_{p,m,l} y_{r,m}[n-m_2] |x_m[n-m_1]|^{2p-1} \\ &+ \sum_{p=1}^{P} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \gamma_{p,m,l} x_m^2[n-m_1] y_{r,m}^*[n-m_2] |x_m[n-m_2]|. \end{aligned} \tag{4.24}$$

It shares some features with the model in [22], namely the occurrence of the conjugate terms in the reflected wave ($y_{r,m}^*$). The conjugate term occurs due to non-analyticity of the describing function. This goes against conventional Volterra theory in which the number of conjugates in each basis function is always even and accompanied by an uneven and larger number of non-conjugated terms, forming basis functions in the form $x[n]|x[n]|^p$.

The model stated in Eq. (4.24) is parameter linear in the time domain and is suitable for wide-band applications as it supports memory in both incident and reflected waves. As memory is included also in the reflected term, the description of the crosstalk network stated in Eq. (4.23) is no longer sufficient. The modification is, however, rather small as the coupling coefficients $\boldsymbol{\lambda}$ are linear and may be modeled by conventional FIR-filters.