5G Vision, Requirements, and Enabling Technologies

2016

5G White Paper

5G Forum
5G Vision, Requirements, and Enabling Technologies

March 2016
Copyright Notice

Copyright © 2016 by 5G Forum, Republic of Korea.
All rights reserved

Parts of this document may represent opinions of some individuals or organizations.
4.1.4. Operational Model ................................................................. 20

4.2. 5G Core Network Enabling Technologies ................................. 21
4.2.1. Highly Flexible 5G Core Infra ............................................... 22
4.2.2. Flat & Distributed Network .................................................. 31
4.2.3. Converged Access Control & Transport ............................... 39

4.3 Access Network Enabling Technology ........................................ 46
4.3.1. 5G Multi-connectivity .......................................................... 46
4.3.2. Radio Access Network Virtualization .................................... 53
4.3.3. Self-Organizing Network ...................................................... 59
4.3.4. Xhaul technologies ............................................................. 65
4.3.5. Vehicular Communication Network (V2X) ............................ 71
4.3.6. Network for IoT ................................................................. 78
4.3.7. Network Technologies for Massive connectivity .................... 84

Chapter 5. 5G Wireless Technologies ............................................. 89

5.1. Wide and Flexible Bandwidth Technology ................................ 91
5.1.1. Millimeter-wave Band ........................................................ 91
5.1.2. Heterogeneous Multi-RAT Integration ................................ 104
5.1.3. Cognitive Radio and Spectrum Sharing ................................. 114

5.2. Advanced Modulation and Coding .......................................... 122
5.2.1. Advanced Modulation (FQAM) ........................................... 122
5.2.2. Advanced high-speed Channel Coding ................................. 127

5.3. Duplexing ................................................................................. 136
5.3.1 Full Duplex Radio ............................................................... 136
5.3.2. Dynamic TDD ................................................................. 145

5.4. Multiple Access and Waveform ............................................... 154
5.4.1. Multiple Access ............................................................... 154
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4.2. Waveform design</td>
<td>177</td>
</tr>
<tr>
<td>5.5 Large Scale Antenna</td>
<td>206</td>
</tr>
<tr>
<td>5.5.1. Large Scale Antenna Below 6GHz</td>
<td>206</td>
</tr>
<tr>
<td>5.5.2. Large Scale Antenna System Above 6 GHz</td>
<td>219</td>
</tr>
<tr>
<td>5.6. Advanced Interference Management</td>
<td>230</td>
</tr>
<tr>
<td>5.6.1. Advanced Receiver for Simultaneous Non-Unique Decoding</td>
<td>230</td>
</tr>
<tr>
<td>5.6.2. Advanced Interference Management via Sliding-Window Superposition Coding (SWSC)</td>
<td>233</td>
</tr>
<tr>
<td>5.7. Access Architecture-related Radio Technologies</td>
<td>238</td>
</tr>
<tr>
<td>5.7.1. Advanced Small Cell</td>
<td>238</td>
</tr>
<tr>
<td>5.7.2. Enhanced Fixed Wireless Backhaul</td>
<td>244</td>
</tr>
<tr>
<td>5.7.3. Moving Network</td>
<td>252</td>
</tr>
<tr>
<td>5.7.4. Device-to-Device (D2D) Communication</td>
<td>263</td>
</tr>
<tr>
<td>5.7.5. Edgeless Cellular Network</td>
<td>273</td>
</tr>
<tr>
<td>5.8. Application Specific Radio Technologies</td>
<td>277</td>
</tr>
<tr>
<td>5.8.1. Massive Connectivity</td>
<td>277</td>
</tr>
<tr>
<td>5.8.2. V2X Communications</td>
<td>289</td>
</tr>
<tr>
<td>Chapter 6. Concluding Remarks</td>
<td>297</td>
</tr>
<tr>
<td>Appendix 1 Acronyms</td>
<td>298</td>
</tr>
<tr>
<td>Appendix 2 Contributors</td>
<td>310</td>
</tr>
<tr>
<td>Appendix 3 Revision History</td>
<td>312</td>
</tr>
</tbody>
</table>
Table 2-1 5G Forum KPIs .................................................................................................................................................. 4
Table 3.1-1 Core Network Requirements .................................................................................................................. 6
Table 3.1-2 HO Delay Latency ..................................................................................................................................... 7
Table 3.2-1 Summary of Wireless Network Requirements .......................................................................................... 10
Table 3.2-2 Mobility classes ......................................................................................................................................... 13
Table 3.2-3 The Handover Interruption Times ............................................................................................................. 14
Table 4.2-1 The 5G Core Network Enabling Technologies (F1~F4, A1~A4, O1~O2 in Table 3.1-1) ................. 21
Table 4.2-2 Comparison of implementation between traditional EPC and SDN based EPC ......................................................... 24
Table 4.2-3 Comparison of vendor’s approach toward SDN based EPC ................................................................. 26
Table 4.2-4 Comparison of the three approaches ....................................................................................................... 28
Table 4.2-5 Summary of Centralized vs. Distributed Architecture ........................................................................... 35
Table 4.2-6 Step-by-step trends of transmission speed and system capacity ............................................................ 44
Table 4.3-1 Benefits of virtualized RAN ..................................................................................................................... 57
Table 4.3-2 Main candidate wireless technologies for on-the-road communications .......................................... 74
Table 4.3-3 An Example of the main research topics and the considerations to support V2X [4] .......................................................... 75
Table 5-1 5G Wireless Network Enabling Technologies (R1~R10 in Table 3.2-1) ......................................................... 90
Table 5.1-1 LTE Aggregation Bands [2] ..................................................................................................................... 105
Table 5.3-1 Target Requirement for IBFD System (BW=20MHz) [2] ................................................................. 139
Table 5.3-2 Performance comparison of Self-IC schemes for IBFD systems. ............................................................ 141
Table 5.4-1 Summary of Wireless Network Requirements .............................................................................................. 202
Table 5.5-1 Parameters for link budget calculation ...................................................................................................... 219
Table 5.5-2 Comparison of singular and modular typed antennas ......................................................................................................................... 222
Table 5.7-1 Classification of various small cells (A typical deployment scenario is assumed) ................................. 239
Table 5.7-2 Examples of non-ideal backhaul ................................................................. 245
Table 5.7-3 Link budget calculation in an MHN system .................................................. 259
Table 5.7-4 Comparison of CoMP and pCell technology ................................................ 275
Figure 2-1 The 5G Technology Vision ........................................................................................................... 2
Figure 4.1-1 Logical model of 5G network ..................................................................................................... 18
Figure 4.1-2 Reference model of 5G network ................................................................................................ 19
Figure 4.1-3 Operational model of 5G network (an example) ................................................................. 21
Figure 4.2-1 The SDN based EPC Virtualization ......................................................................................... 23
Figure 4.2-2 Three approaches toward SDN based 5G core network ......................................................... 28
Figure 4.2-3 Problem and Solution in the centralized network architecture ........................................... 31
Figure 4.2-4 Overview of the Fully Distributed architecture .................................................................... 33
Figure 4.2-5 The Overview of distributed mobility control ....................................................................... 35
Figure 4.2-6 The transport network technology roadmap .......................................................................... 44
Figure 4.2-7 POTN network based on T-SDN architecture ....................................................................... 45
Figure 4.3-1 Multi-RAT connectivity ........................................................................................................... 46
Figure 4.3-2 Radio Interface C-plane architecture alternatives for dual connectivity ....................... 47
Figure 4.3-3 Bearer split options ................................................................................................................. 48
Figure 4.3-4 Alternative 1A ......................................................................................................................... 49
Figure 4.3-5 Alternative 3C ......................................................................................................................... 49
Figure 4.3-6 5G Multi-connectivity under ultra dense deployment environment ........................................... 51
Figure 4.3-7 Open RAN Lx Function Splits [1] ............................................................................................ 54
Figure 4.3-8 RAN functional splitting alternatives ...................................................................................... 56
Figure 4.3-9 Radio Resource pooling with the virtualized baseband .......................................................... 57
Figure 4.3-10 Plug & Play Configuration .................................................................................................... 60
Figure 4.3-11 ANR process ......................................................................................................................... 60
Figure 4.3-12 Example of MLB .................................................................................................................. 61
Figure 4.3-13 Example of MRO operation .................................................................................................. 61
Figure 4.3-14 Example of CCO operation ................................................................................................... 62
Figure 4.3-15 Example of Self-healing of Board Fault Function operation ............................................. 62
Figure 4.3-16 Self-healing of Cell Outage Process ..................................................................................... 63
Figure 4.3-17 Evolution of SON .......................................................... 64
Figure 4.3-18 Effect of SON .............................................................. 64
Figure 4.3-19 RoF concept ................................................................. 66
Figure 4.3-20 RoE concept ................................................................. 67
Figure 4.3-21 LTE backhaul network structure using the TDM-PON OLT and ONU .... 68
Figure 4.3-22 Mobile fronthaul structure using the P-t-P WDM-PON ...................... 69
Figure 4.3-23 Mobile backhaul structure using the P-t-P WDM PON OLT and ONU .... 70
Figure 4.3-24 Type of V2X communication [1] ........................................ 71
Figure 4.3-25 MEC-based V2X operation ............................................... 77
Figure 4.3-26 IoT deployment scenarios with Cellular network .......................... 78
Figure 4.3-27 Deployment model with MTC gateway device
    communicating with MTC server ............................................. 82
Figure 4.3-28 Requirements for massive IoT ........................................... 85
Figure 4.3-29 Network architecture for massive IoT .................................... 86
Figure 5.1-1 Results of verification measurements of propagation
    loss predicted by Friis equation ............................................. 92
Figure 5.1-2 Daejeon street canyon environments conducted
    measurement and the ray-tracing ........................................... 94
Figure 5.1-3. Block diagram of a hybrid beamforming architecture ....................... 97
Figure 5.1-4 Configuration of mmWave beamforming prototype .......................... 98
Figure 5.1-5 Outdoor coverage test results of mmWave beamforming prototype ........ 100
Figure 5.1-6 Outdoor to Indoor penetration test results of
    mmWave beamforming prototype .................................... 101
Figure 5.1-7 Cognitive radio network based on TVDB .................................. 115
Figure 5.1-8 Google’s spectrum database ........................................... 116
Figure 5.1-9 The architecture of LSA systems [8] .................................... 117
Figure 5.1-10 SAS Architecture [11] .................................................. 117
Figure 5.1-10 SAS Architecture [11] .......................................................... 118
Figure 5.1-11 Qualcomm’s carrier aggregation using unlicensed spectrum [12] .......... 119
Figure 5.2-1 16-ary FQAM signal constellation ........................................... 124
Figure 5.2-2 Minimum required SNR versus the BICM and the
CM channel capacity (achievable code rate) for FQAM, FSK
and QAM-based systems under AWGN channels .............................. 124
Figure 5.2-3 Binary and M-ary Turbo coded FER versus Eb/No for FQAM,
FSK and QAM-based systems under AWGN channels .................... 125
Figure 5.2-4 Sample histogram of ICI for downlink LTE systems with
32-ary FQAM and 4-ary QAM, ITU Pedestrian B, 3km/h ...................... 125
Figure 5.2-5 FER performance versus information bit rate of downlink LTE systems
with FQAM and QAM where NF =240 bits,
ITU Pedestrian B (3 km/h)I lI .......................................................... 126
Figure 5.2-6 Comparison of power consumption between Turbo code and LDPC code .... 128
Figure 5.2-7 Comparison of chip area between Turbo code and LDPC code ........... 129
Figure 5.2-8 Polar code has a significant improvement against turbo code .......... 131
Figure 5.2-9 Frame error rate test on FPGA of (1024, 512) polar code ............ 131
Figure 5.2-10 The code rate and code length of polar codes
can be flexibly adjusted. ................................................................. 132
Figure 5.2-11 FER performance of (2304, 1224) polar code under
16QAM with four transmissions ....................................................... 132
Figure 5.2-12 FER performance of (2304, 1224) polar code
under 16QAM with four transmissions ........................................... 133
Figure 5.3-1 Performance Comparison of IBFD system with HD system .......... 136
Figure 5.3-2 Example of radio sub-frame for 5G radio: METIS proposal [4] ........ 147
Figure 5.3-3 Implicit signaling of DL/UL switching .................................... 148
Figure 5.3-4 Indoor deployment with high and low isolation
(AP’s isolated by walls)[8]. .......................................................... 149
Figure 5.3-5 Exemplary DL and UL packet delay cumulative distribution for packet size 128 kB and high (dashed) and low (solid) cell isolation scenarios and four scheduling options [8].

Figure 5.3-6 Experienced user data rate vs. total load in the network for fixed and flexible DL/UL switching / local and centralized scheduling for 6 AP in the area = density 0.5 AP/room [8].

Figure 5.4-1 Resource allocation using OFDMA vs. NOMA

Figure 5.4-2 Sum rate of two users(20dB and 0dB) with OFDMA vs. NOMA

Figure 5.4-3 Received structure in NOMA

Figure 5.4-4 BLER and throughput curve depending on the different code rates of 64QAM, HP QPSK and LP 16QAM in AWGN channel.

Figure 5.4-5 An example of cooperation in NOMA

Figure 5.4-6 Performance variation by power rate and SNR of UEs with two beams.

Figure 5.4-7 Channel capacity comparison of OMA and NOMA in an uplink AWGN channel.

Figure 5.4-8 Illustration of a multi-user SCMA system

Figure 5.4-9 SCMA codebook illustration: bit-to-codeword mapping.

Figure 5.4-10 SCMA codebook illustration: bit-to-codeword mapping

Figure 5.4-11 SCMA Multi-dimensional codebook gain over LDS

Figure 5.4-12 An example of SCMA codebook with low projection

Figure 5.4-13 Examples of application scenarios for SCMA.

Figure 5.4-14 Examples of application scenarios for SCMA

Figure 5.4-15 Examples of application scenarios for SCMA

Figure 5.4-16 Examples scenarios for SCMA based open-loop CoMP

Figure 5.4-17 UFMC transceiver block diagram

Figure 5.4-18 Side-lobe behavior of OFDM and UFMC

Figure 5.4-19 Block diagram of Ff-OFDM

Figure 5.4-20 F-OFDM symbol block
Figure 5.4-21 (a) CP-based vs. (b) zero-tailed DFT-s-OFDM signals
(with different tail lengths)[6] ................................................................. 183
Figure 5.4-22 Zero-tail DFT-s-OFDM signal generation [6] ........................................ 185
Figure 5.4-23 A zero-tail DFT-s-OFDM signal [6] .................................................. 185
Figure 5.4-24 PAPR performance of zero-tail DFT-s-OFDM,
Figure 5.4-25 OOB emissions of zero-tail DFT-s-OFDM [6] .................................. 187
Figure 5.4-26 Transmit data structure using CPs and UWs ....................................... 188
Figure 5.4-27 Block diagrams of OFDM and FBMC .............................................. 191
Figure 5.4-28 An Example of prototype filter in time and frequency domain:

PHYDYAS filter with K=4 ........................................................................ 192
Figure 5.4-29 Overlap and sum structure of FBMC ............................................ 193
Figure 5.4-30 An example of burst signal of FBMC ............................................. 194
Figure 5.4-31 WCP-COQAM transceiver block diagram ..................................... 195
Figure 5.4-32 Overlap & sum structures of (a) FBMC-QAM,
(b) FBMC-COQAM, and (c) WCP-COQAM .............................................. 195
Figure 5.4-33 GFDM transceiver block diagram ................................................ 196
Figure 5.4-34 View of GFDM frame structure: (a) impulse responses for transmit filter,
(b) transmit matrix where it is possible to observe
the time slot blocks with K subcarriers. ................................................. 197
Figure 5.4-35 FBMC-QAM transmitter with two prototype filters ....................... 198
Figure 5.4-36 FBMC-QAM transmitter with two prototype filters ....................... 199
Figure 5.4-37 FBMC-QAM transmitter with multiple prototype filter .................. 199
Figure 5.4-38 OOB performance comparison ..................................................... 200
Figure 5.5-1 Concept of large-scale antenna system below 6GHz ....................... 208
Figure 5.5-2 Possible deployment scenarios for large-scale antenna system .......... 211
Figure 5.5-3 Active array and feed network example .......................................... 212
Figure 5.5-4 Antenna calibration in an anechoic chamber ................................. 214
Figure 5.5-5 Magnitude of reflection coefficients of 8 sub-arrays ........................................ 214
Figure 5.5-6 Magnitude of mutual coupling coefficients between adjacent sub-arrays ... 215
Figure 5.5-7 Co-pol and Cross-pol radiation patterns of one sub-array ............................... 215
Figure 5.5-8 Received power level with the distance ............................................................... 220
Figure 5.5-9 Modular typed antenna structure ................................................................. 221
Figure 5.5-10 Singular typed antenna structure ................................................................. 221
Figure 5.5-11 Analog/Hybrid/Digital Beamformers ............................................................... 224
Figure 5.5-12 Performance comparison of Digital, Analog, and Hybrid Beamformers ... 224
Figure 5.5-13 Prospect on 5G network configuration ......................................................... 226
Figure 5.5-14 Multiple 3D Beam Spots ............................................................................. 226
Figure 5.5-15 Antenna configuration for fixed 3D beamforming ........................................ 227
Figure 5.5-16 Single Beam Transmission ............................................................................ 227
Figure 5.5-17 Coordinated Multi-Beam Transmission-SM/-CoMP ........................................ 228
Figure 5.5-18 Beam coverage extension .............................................................................. 228
Figure 5.6-1 Achievable rate regions: a) treating interference as noise;  
   b) simultaneous decoding; c) simultaneous non-unique decoding  
   (advanced receivers); d) jointly achievable rate region of  
   two advanced receivers. ................................................................................................. 231
Figure 5.6-2 The encoding and decoding operation of the SWSC scheme for b blocks. ... 235
Figure 5.6-3 Sum rates of SWSC, IASD, IAD, and IAN for the  
   Gaussian interference channel with SIR = -0.2/0.2dB ............................................. 236
Figure 5.6-4 BLER for the 2/2-layer SWSC scheme with bit mapping rule  
   optimization and iterative soft decoding, the 2/1-layer  
   SWSC scheme with successive hard decoding (basic SWSC),  
   IASD, IAD, and IAN in the SISO symmetric Ped-B interference  
   channel with average received INR = 10 dB.  
   Sum-rate = 1.2 bits/D. Modulation: 4-PAM. ................................................................. 236
Figure 5.7-1 Ultra-dense heterogeneous network deployment with various small cells in 5G ............................................ 239
Figure 5.7-2 Architecture of mobile hotspot network for the trains ............................................ 257
Figure 5.7-3 Beam radiation pattern (inset: 3-dimensional image) ............................................ 259
Figure 5.7-4 (a) Test cars vehicles for a TX and an RX, 
(b) Test environment, (c) Test results ............................................ 260
Figure 5.7-5 Synchronization propagation using two resources. ............................................ 265
Figure 5.7-6 Synchronization signal and channel. ............................................ 266
Figure 5.7-7 Device-to-Network relay. ............................................ 267
Figure 5.7-8 Discovery transmission and resource allocation. ............................................ 267
Figure 5.7-9 D2D communication period, control and data pool. ............................................ 269
Figure 5.7-10 HARQ transmission of D2D communication. ............................................ 269
Figure 5.8-1 Three different modes of operation for NB-IoT ............................................ 270
Figure 5.8-2 Collision Avoidance with legacy LTE signals for in-band operation ........ 278
Figure 5.8-3 Straightforward application of compressed sensing for multiple access .... 282
Figure 5.8-4 Variable frame structure for CS-based random traffic transmission [3] ... 283
Figure 5.8-5 Single-sequence Spreading (a) vs. Multi-sequence spreading (b) ............. 286
Figure 5.8-6 Pedestrian warning [1] ............................................................................. 291
Figure 5.8-7 Platooning ......................................................................................... 292
Figure 5.8-8 Hidden vehicle problem .......................................................................... 292
Figure 5.8-9 High rate in-vehicle infotainment ......................................................... 293
Figure 5.8-10 Multi-cell coordinated broadcast ....................................................... 295
Figure 5.8-11 Various locations and types of vehicle antennas .................................. 296
Chapter 1 Introduction

Ever since the formation of 5G Forum in May 2013, the 5G Forum Technology Sub-committee has been playing an essential role in identifying the key capabilities of 5G systems and furthermore, promoting the enabling technologies. After the first version of white paper was published in February 2015, the Sub-committee has been restructured into two separate groups, Wireless Technology Sub-committee and Network Technology Sub-committee. Such effort was made to focus on more essential issues in core network, and in access-related architecture on top of radio access technologies.

In 2015, we reached a global agreement on the vision for 5G and the overall objective of the development of 5G network. Furthermore, at the 3GPP RAN Workshop on 5G in September 2015, we came to a consensus that there will be a new, non-backward compatible, radio access technology as part of 5G, supported by the need for LTE-Advanced evolution in parallel. Towards this end, two phases for the specification work had been identified—Phase 1 to be completed by H2 2018 (End of 3GPP Release 15) and Phase 2 to be completed by December 2019 for the IMT-2020 submission (End of 3GPP Release 16). Motivated by further establishments in the vision and requirements of 5G, Wireless Technology and Network Technology Sub-committees of 5G Forum had looked into the details of key technologies and network architecture centered on the global consensus. In particular, our earlier outlook on the 5G key capabilities has been revisited to be aligned with the views in Recommendation ITU-R M.2083, e.g., in the peak data rate and user-experienced data rate, as an effort to advocate global collaboration and 5G standardization.

This white paper intends to serve as a guideline for understanding 5G vision and establishing performance requirements, while identifying the key enabling technologies in 5G wireless and core networks. All these efforts will encourage universal standardization, and accelerate research toward the technology innovations by cooperating with all participants in the global 5G ecosystem.
Since the advent of the first generation analog mobile communications in 1980s, a new generation of mobile communication system has appeared roughly every 10 years. Development of each generation has been driven either by service or technology requirement. For example, when the 1G analog system went on to 2G digital system, the digital technology which can support more voice users was the driving force with its main target being still the same voice service as in 1G. From 2G to 3G, data service played a major role of the change, yet with the limited data rate. In fact, this generation-to-generation transition was motivated by a new paradigm centered on video telephony and mobile contents. From 3G to 4G, however, increasing a data rate was an essential design objective, especially driven by the emergence of smart phones, which requires a data rate of several Mbps for mobile broadband (MBB) internet, supporting uninterrupted web browsing and seamless video streaming. As an enabling technology for 4G, orthogonal frequency division multiple access (OFDMA) was successfully standardized and currently employed widespread as an essential mobile broadband infrastructure. We expect the further enhancement to MBB in the next generation system, possibly enabling the new user experience with the high-quality (e.g., ultra high definition) video and/or virtual reality (VR)-based immersive services.
In 5G Forum, the initial 5G vision has been depicted by ‘1Gbps/user anytime anywhere with hyper-connectivity in 2020s, towards multi-gigabit services for hologram, pervasive wireless connecting both human and things, and realistic experience for immersive service (see Figure 2-1) [1]. In recent recommendation on framework and overall objectives of the future development of IMT for 2020 and beyond by ITU-R Working Party 5D (WP5D) [2], three different usage scenarios have been identified for 5G as enhanced mobile broadband (eMBB), ultra-reliable & low latency communications (uMTC), and massive machine type communication (mMTC). The eMBB scenario will realize new user experience with UHD, hologram, and virtual reality contents, especially in a hotspot area with high user density yet low mobility. For the wide area coverage, meanwhile, much higher data rate will be provided compared to existing data rates, even if its data rate may be still lower than in hotspot. The uMTC scenario corresponds to the usage cases such as industrial manufacturing, remote medical surgery, and safety in autonomous vehicle, which require more stringent capabilities in throughput, latency, and availability. For the mMTC scenario, low-cost IoT devices with long battery lifetime will be abundant to support the machine-type communication (MTC) traffic with a short burst, which may be relatively less sensitive to delay. In other words, massive and efficient connectivity to support an enormous number of IoT devices at low cost will be one of the essential usage cases in 5G.

The specific technical requirements have been identified from the high-level service perspectives derived by 5G Service Sub-committee [1]. Through these requirements, we have extracted 11 key performance indicators (KPIs) for 5G, most of them defined for wireless network and one of them, i.e., reliability, defined for core network. First of all, we expect much more capable wireless access, e.g., a peak data rate beyond 20 Gbps. Furthermore, a minimum data rate of 100 Mbps to 1 Gbps must be provided for more enhanced user experience. It implies that a limited user experience, especially at the cell edge, must be overcome by guaranteeing at least 100 Mbps data rate. Meanwhile, more system capacity must be provided, e.g., areal capacity of 10 Mbps/m2 for eMBB service and massive connectivity of 106 connections/km2 for IoT services. Toward this end, spectrum efficiency must be still further improved to overcome the limitation in spectrum availability, especially in the frequency band below 6 GHz, It can be at least 10 bps/Hz, i.e., roughly 3 times as much as...
4G. In some uMTC scenarios, ultra-low latency of 1 ms must be guaranteed over the wireless access (front-haul) segment, e.g., from application/service layer of user device to baseband unit (BBU) in base station, possibly through remote radio head (RRH). Other than these KPIs, we also consider some other system-level performance such as intra-frequency handover latency and position accuracy. Table 2-1 summarizes all KPIs that we have identified in our early stage of establishing 5G vision. We note that all these are now pretty much aligned with key capabilities in [2], except that some KPIs, e.g. reliability and positioning, are still subject to further discussion.

**Table 2-1 5G Forum KPIs**

<table>
<thead>
<tr>
<th>Wireless Network</th>
<th>Key performance indicators</th>
<th>Values</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral efficiency</td>
<td>bps/Hz</td>
<td>3 times higher than 4G</td>
<td>Approximately 10bps/Hz</td>
</tr>
<tr>
<td>Peak data rate</td>
<td>bps</td>
<td>20Gbps</td>
<td></td>
</tr>
<tr>
<td>User experienced data rate</td>
<td>bps</td>
<td>100Mbps (outdoor) 1Gbps (indoor)</td>
<td></td>
</tr>
<tr>
<td>Latency</td>
<td>ms</td>
<td>1ms</td>
<td>Latency over radio interface</td>
</tr>
<tr>
<td>Mobility</td>
<td>km/h</td>
<td>500km/h</td>
<td></td>
</tr>
<tr>
<td>Handover interruption time</td>
<td>ms</td>
<td>10ms</td>
<td></td>
</tr>
<tr>
<td>Areal capacity</td>
<td>bps/m²</td>
<td>10Mbps/m²</td>
<td></td>
</tr>
<tr>
<td>Energy efficiency</td>
<td>Joules/bit</td>
<td>100 times higher than 4G</td>
<td></td>
</tr>
<tr>
<td>Connectivity</td>
<td>connections</td>
<td>10⁶connections/km²</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>cm</td>
<td>[TBD]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Core Network</th>
<th>Reliability</th>
<th>[TBD]</th>
<th>[TBD]</th>
</tr>
</thead>
</table>

- Radio link failure (RLF)
- Service availability
- Recovery time
References


Chapter 3 Requirements

3.1 Network Requirements

Core Network Requirements are investigated in three aspects to resolve 5G service requirements.

- Functional Requirements (F)
- Architectural Requirements (A)
- Operational Requirements (O)

<table>
<thead>
<tr>
<th>Brief Description</th>
<th>Technical Requirements for 5G Core Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 Seamless mobility</td>
<td>Shall support seamless mobility regardless of the cell types and RATs in the environment where the macro BS, small cell BS, personal cell, type 1/2 WLAN, and relay station are mixed and overlapped, Shall guarantee the service continuity when the change of anchoring GW occurs frequently in the distributed architecture</td>
</tr>
<tr>
<td>F2 Wired/wireless terminal switching</td>
<td>Shall support terminal and/or session mobility to provide fast handover between wireless and wired terminals</td>
</tr>
<tr>
<td>F3 Context aware best connection</td>
<td>Shall utilize the various context information (device context, user context, environment context, network context) to provide always best connection/service</td>
</tr>
<tr>
<td>F4 Single ID for multiple access</td>
<td>Shall recognize a mobile terminal as a single entity regardless of its access network</td>
</tr>
<tr>
<td>A1 Distributed architecture</td>
<td>Shall support the distributed network architecture to accommodate anticipated 1000 times of traffic</td>
</tr>
<tr>
<td>A2 Lightweight Signaling</td>
<td>Shall have lightweight signaling to support a variety of terminals such as massive MTC terminal</td>
</tr>
<tr>
<td>A3 Multiple RAT interworking</td>
<td>Shall have architecture to support ‘Flow over Multi-RAT’ to provide the high volume service with low cost and guarantee the service continuity in spite of the bandwidth deficiency in a wireless access</td>
</tr>
<tr>
<td>A4 Fine grained location tracking</td>
<td>Shall have function to trace the mobile terminal location in a fine granularity in order to provide advanced location based service</td>
</tr>
<tr>
<td>O1 Flexible reconfigure &amp; upgrade</td>
<td>Shall provide virtualization environment and support to reconfigure and upgrade the core network at low cost without changing the physical network infrastructure</td>
</tr>
<tr>
<td>O2 Network on-demand</td>
<td>Shall be able to build the network based on the QoS/QoE, charging, and service characteristics</td>
</tr>
</tbody>
</table>
3.1.1. Functional Requirements

3.1.1.1. (F1) Seamless Mobility

5G core network shall support seamless mobility regardless of the cell types and RATs in the environment where the macro BS, small cell BS, personal cell, Type 1 WLAN, Type 2 WLAN, and relay station are mixed and overlapped. Current mobile core network is mainly designed for macro cell. Thus, seamless HO is supported only between macro cells. The seamless HO between WLAN and macro cell and HO between relay stations are not supported. Since the future mobile core network will support various wireless access in the common platform, the seamless HO among different types should be supported.

5G core network shall guarantee the service continuity when the change of anchoring GW occurs frequently in the distributed architecture. In the distributed architecture, it is expected that inter GW mobility will occur more frequently, but it is impossible to change P-GW and IP address of the terminal in current network architecture and hard to provide inter GW mobility and active mode traffic steering. Following table shows the HO delay latency we expect in the 5G network.

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>5G Macro Cell BS</th>
<th>5G relay</th>
<th>Type 1 WLAN</th>
<th>RRH</th>
<th>5G Small Cell BS</th>
<th>Type 2 WLAN</th>
<th>Wire Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>5G macro Cell BS</td>
<td>F</td>
<td>F</td>
<td>M</td>
<td>M</td>
<td>F</td>
<td>S</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>5G relay</td>
<td>F</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>F</td>
<td>S</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Type 1 WLAN</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>RRH</td>
<td></td>
<td>M</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>5G small Cell BS</td>
<td></td>
<td>F</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Type 2 WLAN</td>
<td></td>
<td>M</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Wire Line</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td></td>
</tr>
</tbody>
</table>

- F (Fast HO): 20msec, M (Moderate HO): 100msec, S (Slow HO): 150msec
- HO performance is assumed symmetrical
- Explanation of the network elements is provided in the example architecture in section 3.1.4

1) We assume the WLAN will be two types. Type 1 WLAN connected to the macro BS provides more tight operation such as seamless mobility between macro BS and WLAN. The example architecture will be explained in section Chapter 4.
3.1.1.2. (F2) Wired/wireless Terminal Switching
5G core network shall support terminal and/or session mobility to provide fast handover between wireless and wired terminals. Current 3GPP focuses on wireless access including 3GPP and non-3GPP access. For wire-line access, only the QoS interworking with wire-line access is defined. Therefore, it is difficult to provide network level fast mobility between wireless terminal and wired terminal. For example, terminal switching in current multi-screen service is slow because the service is provided in server based.

3.1.1.3. (F3) Context Aware Best Connection
5G core network shall utilize the various context information (device context, user context, environment context, and network context) to provide always best connection/service. In 5G, various types of access points are available to the user. In current architecture, network makes the cell selection based on the radio signal quality in the same access technology. When multiple access technologies are available, user selects an access based on his/her personal decision. It is difficult to select the best quality access point for the user’s situation.

3.1.1.4. (F4) Single ID for Multiple Access
5G core network shall recognize a mobile terminal as a single entity regardless of its access type. Terminal with multiple wireless interfaces will be common in 5G. Currently, each wireless access uses its own ID to identify the terminal, therefore, it is difficult to support the service continuity and mobility between different access networks.

3.1.2. Architectural Requirements

3.1.2.1. (A1) Distributed Architecture
5G core network shall support the distributed network architecture to accommodate anticipated 1000 times of traffic. Current network has a weakness in terms of scalability on traffic explosion because mobile core network has hierarchical and centralized architecture with fixed location of P-GW in the middle.
3.1.2.2. (A2) Lightweight Signaling
5G core network shall provide lightweight signaling to support massive MTC service. Massive MTC service tends to be implemented in low cost, low power consuming devices. Current network is designed to support high speed communication and to employ power demanding expensive components. Current network signaling information may be bigger than massive MTC information, characterized short and burst, to be carried through the current core network. 5G core network is required to support lightweight signaling for varying characteristics of terminals.

3.1.2.3 (A3) Multiple RAT interworking
5G core network shall have architecture to support ‘Flow over Multi-RAT’ to provide the high volume service in low cost and guarantee the service continuity in spite of the bandwidth deficiency in a wireless access. Currently, the multi-RAT related technology is used for offloading purpose, and consideration on service continuity and QoS support is not being made.

3.1.2.4. (A4) Fine Grained location Tracking
5G core network shall have function to trace the mobile terminal location in a fine granularity in order to provide advanced location based service. Current network only has low granularity location tracking capability using IP address based and cell based location tracking.

3.1.3. Operational Requirements

3.1.3.1. (O1) Flexible Reconfigure & Upgrade
5G core network shall provide virtualization environment and support to reconfigure and upgrade the core network at low cost without changing the physical network infrastructure. Control plane and data plane functions are combined in the network equipment of current mobile core network, and therefore, it pays high cost in case of the network upgrade and capacity expansion. SDN based control and virtualization is initially applied from data center network, and it is expected to expand its area to the mobile core network (EPC: Evolved Packet Core). In order to achieve flexible reconfiguration and upgrade of the core network, the virtualization method such as SDN/NFV will be used to implement the 5G core network.
3.1.3.2. (O2) Network On-demand

5G core network shall be able to build the network based on user’s need in terms of the QoS/QoE level, charging scheme, and service characteristics. Currently, only network providers can control the QoS and charging policy of the network. The scope of control is limited to the services they provide.

3.2. Wireless Network Requirements

Table 3.2-1 summarizes the wireless network requirements and the detailed descriptions are as follows.

<table>
<thead>
<tr>
<th>Index</th>
<th>Requirement</th>
<th>Brief Definition</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>Spectral efficiency</td>
<td>The aggregate throughput of all users divided by the channel bandwidth divided by the number of cells</td>
<td>3 times higher than 4G (bps/Hz)</td>
</tr>
<tr>
<td>R2</td>
<td>Peak data rate</td>
<td>The maximum theoretically achievable data rate which can be assigned to a single mobile station assuming error-free conditions when all the available radio resources are utilized for the corresponding link</td>
<td>20 Gbps</td>
</tr>
<tr>
<td>R3</td>
<td>User experience data rate</td>
<td>Guaranteed cell edge user data rate that corresponds to 5% point of the cumulative distribution function (CDF) of the user data rate</td>
<td>100 Mbps / 1 Gbps (indoor)</td>
</tr>
</tbody>
</table>
| R4    | Latency                        | Control plane latency: typically measured as transition time from different connection modes, e.g., from idle to active state. User plane latency: the one-way transit time between an SDU packet being available at the IP layer in the user terminal/base station and the availability of this packet (PDU) at IP layer in the base station/user terminal. | Control plane: 50 ms  
User plane: 1 ms |
| R5    | Mobility                       | A mobility class is supported if the traffic channel link can be maintained when the user is moving at the maximum speed in that mobility class. | 500 km/h                   |
| R6    | Handover interruption time     | The time duration during which a user terminal cannot exchange user plane packets with any base station. | 10 ms                      |
| R7    | Areal capacity                 | In order to accommodate the explosive increase of future mobile data traffic, 5G RAN should be able to scale-up system capacity by adding more cells in a target area. | 10 Mbps/m²                 |
3.2.1. Spectral Efficiency

Spectral efficiency ($\eta$) is defined as the aggregate throughput of all users (the number of correctly received bits, i.e. the number of bits contained in the service data units (SDUs) delivered to Layer 3, over a certain period of time) divided by the channel bandwidth divided by the number of cells. The channel bandwidth for this purpose is defined as the effective bandwidth times the frequency reuse factor, where the effective bandwidth is the operating bandwidth normalized appropriately considering the uplink/downlink ratio.

The spectral efficiency is measured in bit/s/Hz.

Let $\chi_i$ denote the number of correctly received bits by user $i$ (downlink) or from user $i$ (uplink) in a system comprising a user population of $N$ users and $M$ cells. Furthermore, let $\omega$ denote the channel bandwidth and $T$ the time over which the data bits are received. The cell spectral efficiency, $\eta$ is then defined at (1).

$$\eta = \frac{\sum_{i=1}^{N} \chi_i}{T \cdot \omega \cdot M} \quad (1)$$

3.2.2. Peak Data Rate

The peak data rate is the maximum theoretically achievable data rate which can be assigned to a single mobile station assuming error-free conditions when all the available radio resources are utilized for the corresponding link. (In this case, radio resources that are used for physical layer synchronization, reference signals or pilots, guard bands and guard times are excluded.) The minimum requirement for peak data rate is 20 Gbps.

<table>
<thead>
<tr>
<th>Index</th>
<th>Requirement</th>
<th>Brief Definition</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>R8</td>
<td>Energy efficiency</td>
<td>5G radio access technology design should aim for higher energy efficiency against increased device/network energy consumption required on 5G wireless communications.</td>
<td>100 times higher than 4G</td>
</tr>
<tr>
<td>R9</td>
<td>Connectivity</td>
<td>Total number of connected and/or accessible devices per unit area</td>
<td>$10^6$(per km$^2$)</td>
</tr>
<tr>
<td>R10</td>
<td>Positioning</td>
<td>Positioning technology with high resolution is inevitable in establishing an UE centric and UE experiencing services in anytime and anywhere.</td>
<td>[TBD]</td>
</tr>
</tbody>
</table>
3.2.3. User experience data rate

The user data rate is defined as the average user data rate (the number of correctly received bits by users, i.e., the number of bits contained in the SDU delivered to Layer 3, over a certain period of time, measured in bit/s). User experience data rate guarantees the cell edge user data rate that is defined as 5% point of the cumulative distribution function (CDF) of the user data rate.

With $\chi_i$ denoting the number of correctly received bits of user $i$, $T_i$ the active session time for user $i$, the user data rate of user $i$, $\gamma_i$, is defined according to (2).

$$\gamma_i = \frac{\chi_i}{T_i}$$  \hspace{1cm} (2)

3.2.4. Latency

3.2.4.1. Control Plane Latency

Control plane (C-Plane) latency is typically measured as transition time from different connection modes, e.g. from idle to active state. A transition time (excluding downlink paging delay and wireline network signaling delay) of less than or equal to 50 ms shall be achievable from idle state to an active state in such a way that the user plane is established.

3.2.4.2. User Plane Latency

User plane latency (also known as transport delay) is defined as the one-way transit time between an SDU packet being available at the IP layer in the user terminal/base station and the availability of this packet (protocol data unit, PDU) at IP layer in the base station/user terminal. User plane packet delay includes delay introduced by associated protocols and control signaling assuming the user terminal is in the active state. 5G systems shall be able to achieve a user plane latency of less than or equal to 1ms in unloaded conditions (i.e., a single user with a single data stream) for small IP packets (e.g., 0 byte payload + IP header) for both downlink and uplink.
3.2.5. Mobility

The following classes of mobility are defined:

- Stationary: 0 km/h
- Pedestrian: > 0 km/h to 10 km/h
- Vehicular: 10 to 120 km/h
- High speed vehicular: 120 to 350 km/h
- Very high speed train: 350 to 500 km/h

The Table 3.2-2 defines the mobility classes that shall be supported in the respective test environment.

<table>
<thead>
<tr>
<th>Mobility classes supported</th>
<th>Test environments</th>
<th>Base coverage</th>
<th>High speed</th>
<th>Very high speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indoor</td>
<td>Stationary,</td>
<td>Stationary,</td>
<td>High speed vehicular,</td>
<td>Very high speed train</td>
</tr>
<tr>
<td>Microcellular</td>
<td>Pedestrian</td>
<td>Pedestrian,</td>
<td>Vehicular</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vehicular (up to 30km/h)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A mobility class is supported if the traffic channel link can be maintained when the user is moving at the maximum speed in that mobility class.

3.2.6. Handover Interruption Time

The handover interruption time is defined as the time duration during which a user terminal cannot exchange user plane packets with any base station. The handover interruption time includes the time required to execute any radio access network procedure, radio resource control signaling protocol, or other message exchanges between the user equipment and the radio access network. In this case, interactions with the core network are assumed to occur in zero time. It is also assumed that all necessary attributes of the target channel (that is, downlink synchronization is achieved and uplink access procedures, if applicable, are successfully completed) are known at initiation of the handover from the serving channel to the target channel. The 5G system shall be able to support handover interruption times specified in Table 3.2-3.
### Table 3.2-3 The Handover Interruption Times

<table>
<thead>
<tr>
<th>Handover type</th>
<th>Interruption time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-frequency</td>
<td>10</td>
</tr>
<tr>
<td>Inter-frequency</td>
<td></td>
</tr>
<tr>
<td>- Within a spectrum band</td>
<td>10</td>
</tr>
<tr>
<td>- Between spectrum bands</td>
<td>10</td>
</tr>
</tbody>
</table>

#### 3.2.7. Areal Capacity

In order to accommodate the explosive increase of future mobile data traffic, 5G RAN should be able to scale-up system capacity by adding more cells in a target area. If necessary, a metric value in unit of Mbps/m² may be specified.

#### 3.2.8. Energy Efficiency

5G radio access technology design should aim for higher energy efficiency against increased device/network energy consumption required on 5G wireless communications. If necessary, a metric value in unit of bit/Joule may be specified.

#### 3.2.9. Connectivity

Connectivity in 5G is simply not limited to mobile devices. Instead, every single unit mounting a modem function will connect together for any reasons of safety, communication, cozy life, and so on.

#### 3.2.10. Positioning

Positioning technology with high resolution is inevitable in establishing a UE centric and experiencing services in anytime, anywhere, and anything.
References

4.1. 5G Network Architecture

4.1.1. Architecture Overview
The scope of the 5G is not limited in radio technologies, but also includes fixed communication, cloud, and services. Because of the wide scope of the 5G system, single architecture cannot explain every aspect of the 5G system. In this chapter, we provide three models — (1) Logical Model, (2) Reference Model, (3) Operation Model. Each model describes the different architectural aspect of the 5G system.

Logical model is the top level architecture describing softwarization aspect such as virtualization, abstraction, and end-to-end dynamic configuration. SDN/NFV/Cloud related architecture is provided in this model. Related technical issue includes service and virtualization orchestration, service function chaining, slice mapping/setup, network function configuration, resource configuration, and forwarding setup/control.

Reference model is required for describing restructuring of the legacy functions and newly required function. This model describes the network functional level regardless of softwarization. Thus, the network function in this model can be implemented either PNF (Physical Network Function) or VNF (Virtual Network Function).

Operation model is required for describing the physical shape of the network deployment. This model describes the positioning of the function in the real network. Operation model can be considered as deployment reference model for a use case. Different use case can have different operational model.

4.1.2. Logical Model
Figure 4.1-1 illustrates the logical model of 5G network. The model consists of several planes representing Transport, Cloud Infrastructure, Access, COMPA (Control, Orchestration, Management, Policies, Analytics), Business & Service, and etc.

Transport, which consists of X-haul (Front haul and Backhaul) and Optical core, is to provide...
Cloud Infrastructure consists of physical computing/storage/networking resources and virtualization layer middleware (hypervisor or container) which supports the transformation of physical resources into virtual or logical resources needed in case of multi-tenancy environment and network slicing. It contributes to shift the roles of wired and wireless systems in 5G from dumb pipe to distributed cloud data center(s).

RAN and EPC, the traditional functions of mobile network, are restructured completely to be deployed as VNF (Virtual Network Functions) in cloud infrastructure. Access Plane in the model consists of the non-virtualized part of RAN (e.g., RF of C-RAN) and wired access. PNF (Physical Network Functions), corresponding to VNF for cloud environment, can be set for several reasons such as performance enhancement or cost trade-off, and legacy interworking.

COMPA (Control, Orchestration, Management, Policies, and Analytics) has the roles of governing all the planes mentioned above, to control, manage and coordinate them globally. It includes NFV MANO (Management & Orchestration) functions such as Cloud Manager (Open Stack as a typical example) and VNF Manager for Virtual Radio Network Functions and Virtual Core Network Functions, as well as SDN controller for Transport. It also includes Policies for network operation and Data Analytics to convert raw network data into valuable information through the collection and analysis procedures.

The components in Business and Service plane are for; SLA (service level agreement) management for business activities by various 5G network actors, Billing for charging and pricing the network functions and resources used by actors, Service Exposure based on API for providing various network capabilities to external service applications, Brokerage for market mechanism of network resources among actors and service mashup/composition, and Dashboard for visualization of real time network status.
4.1.3. Reference Model

Figure 4.1-2 is a reference model of 5G network which illustrates the domains of Devices, Access network and Core network. It is developed through the analysis on a variety of scenarios for the deployment configurations of 5G network, which eventually resulted in a similar model to 4G network.

Device shown in Figure 4.1-2 could be any of 5G UE types specified as such; advMTC devices, advD2D devices, Moving network devices (e.g. V2X), Infra-less devices and etc. Device can support things or machines in the concept of IoT/MTC as well as human user.

Access Network is comprised of MCNs (Macro Cell Nodes), SCNs (Small Cell Nodes) and possibly WiFi-APs. The SCN can be classified more specifically as several types such as UDN, Hot-zone, and advHeNB etc. MCNs could be 5G or 4G macro-cell node with the feature of dual connectivity working together with SCNs. Even when a Device is situated closer to MCN than SCN, SCN serves it to provide seamless coverage. A spectrum of radio transmission technologies can be adopted in SCNs to improve transmission data rates, such as WiFi and millimeter wave.
Uux interface should support a many-to-many relation between the types of devices and cell nodes, and optimization for the different frequencies and use cases for the types. Two alternatives, single unified radio interface (URI) or multiple radio interface can be foreseen to address the situation; however, the approach needs to be carefully selected based on the thorough research on the flexibility vs. performance, operator’s management, technical complexity, international roaming and economical RoI due to introduction of one of alternatives.

Core Network consists of C-GW (Control Gateway) and D-GW (Data Gateway), providing control plane and data plane respectively, as for the evolution of EPC functions in 4G network. The network entities ‘Provisioning’ and ‘Policy’ in Figure 4.1-2, are the evolution of HSS, PCRF in 4G network respectively, and providing the similar functionality with those previous ones. The nodes (e.g, SCN, MCN, SCN GW) inside the Access Network can be interconnected each other by means of X2x interface which is the evolved X2. For instance, SCN GW can have a role for aggregating multiple SCNs (such as for advanced HeNBs), then X2x is used for the interface.

Access Network interconnects to Core Network by means of the S1x interface. The S1x interface should also support a many-to-many relation between Access Networks and Core Networks. As with the case of Uux, S1x will be either a single unified network interface (UNI) or multiple network interfaces. The technology for S1x needs to be also carefully selected based
on the thorough research on the flexibility vs. performance, operator's management, technical complexity and economical RoI. For a certain case, S1x may not be an external interface, but possibly be internal interface in order to support the low-latency service typ. It will be a deployment issue to be considered by operators.

4.1.4. Operational Model

Figure 4.1-3 shows an operational model of 5G network. In 5G network, the performance of wireline service and wireless service will not be different significantly. The service will move freely in between wireless access and wireline access, and 5G core network covers both wire-line and wireless accesses and traffic. The wired terminals can be HDTV, PC, and home WiFi AP. The wireless access is connected to either macro BS, small cell BS, or Type 2 WLAN. Relay station and Type 1 WLAN are connected to macro BS. Direct connection between small cell and macro BS is not assumed in the figure, but it may be possible. These terminals will be connected to base band unit (BBU) and optical line terminal (OLT) via front haul, backhaul, or passive optical network (PON). The data from BBU and OLT are delivered to 5G core network consists of packet-optical transport network (POTN), gate way, or router. Control plane of the 5G network could be implemented in the virtualized environment such as virtualized radio network function (vRNF) and virtualized core network function (vCNF). Logical GW in the virtualized environment contains the control plane function of the GW. The logical GW controls multiple GW data-plane switches. It is expected to have two new functional blocks in the 5G control plane. First one is Radio Resource Information block. It is required to provide the capability to select the best possible radio access among all available wireless accesses. The typical function of the block includes monitoring the radio resources of multiple RATs, Macro-Relay topology based on channel condition, and etc. Second one is the geometrical location information block. It is required to trace the precise UE location and to identify the best available radio access in the UE location.
4.2. 5G Core Network Enabling Technologies

In this Chapter, the promising candidate technologies for 5G are described in aspects of core and wireless networks.

<table>
<thead>
<tr>
<th>Category</th>
<th>Enabling Technologies</th>
<th>Related 5G Core Network Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F1</td>
</tr>
<tr>
<td>Highly Flexible 5G core Infra</td>
<td>Flexible service chaining for future mobile services</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optimal virtualization of mobile core control functions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dynamic open control protocol for mobile core</td>
<td></td>
</tr>
<tr>
<td>Flat &amp; Distributed Network</td>
<td>Virtualized logical GW with distributed switch</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dynamic mobility anchoring for seamless inter-GW HO</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Signaling mitigation for always-on-app and IoT</td>
<td></td>
</tr>
</tbody>
</table>
In the core network area, three technical areas have been identified as 5G core network enabling technologies. First one is virtualization to achieve highly flexible core network infra. Second one is the distributed architecture to resolve the traffic and signaling explosion. Third one is converged access control transport.

### 4.2.1. Highly Flexible 5G Core Infra

#### 4.2.1.1. General Overview

Role of the core network becomes important in 5G network as it plays a central role not only in accommodating traffic explosion but also in supporting the multiple RAT in a flexible manner [1][2]. Current mobile core network is implemented in a form of tightly coupled software and hardware. Many network functionalities such as network monitoring, IP allocation, QoS, and charging are centralized in a high cost P-GW system. The cost of upscaling the network capacity will be expensive because it will require more and more high
cost P-GWs. In order to resolve the high cost structure, SDN based mobile core network infra is needed. Using SDN enabled software based network can provide various services in the same physical network infrastructure such as multi-tenant network. SDN based mobile core network is being actively studied with this motivation.

Early stage of mobile core network virtualization using VM (Virtual Machine) and OVS (Open vSwitch) is already available from major vendors. Its implementation method is still vendor specific and data plane is tightly coupled with the control plane. To be more flexible, the data plane should be implemented independently from the control plane. Current GW (gateway) system is implemented as a combination of data plane and control plane. In order to achieve the EPC virtualization, a separation of data plane in the GW is required. SDN/OpenFlow based GW implementation is actively studied in the projects such as CellSDN [3], SDMN [4], and SIGMONA [5]. A simplified EPC system using OpenFlow base SDN has been studied in those projects. The control plane of GW is separated from the data plane and implemented in the virtualization environment and the data planes are implemented in the commercial OpenFlow switches. By converting the EPC into the SDN basis, we can obtain many advantages. The details are summarized in the figure and table below.

![Figure 4.2-1 The SDN based EPC Virtualization](image-url)
Table 4.2-2 Comparison of implementation between traditional EPC and SDN based EPC

<table>
<thead>
<tr>
<th></th>
<th>Current : Tightly coupled H/W and S/W EPC</th>
<th>Future: EPC over SDN</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Architecture</td>
<td>Control plane (S/W) and Data Plane (H/W) are implemented in the same system.</td>
<td>Control plane and data plane are separated and implemented as an independent system</td>
</tr>
<tr>
<td>Service Evolution</td>
<td>Difficult to add a new service because system highly depends on H/W</td>
<td>Easy to add variety of services because system is mostly implemented in S/W and less dependent on H/W</td>
</tr>
<tr>
<td>Implementation Cost</td>
<td>When the network is upscale, it requires high cost to purchase expensive network equipment such as P-GW and S-GW</td>
<td>When the network upscale, it only requires less money by adding more data plane switches.</td>
</tr>
<tr>
<td>Network operation</td>
<td>Slow when the network operation when reacting to the real-time network state change and new requirements.</td>
<td>Possible to flexibly operate the network as the network state change.</td>
</tr>
</tbody>
</table>

EPC market was dominated by the global vendors. Only global leading vendors such as ALU, Ericsson, NSN, Huawei, and Samsung dominate the EPC market. However, with rapid progress of SDN technology, EPC can be implemented in a virtualization environment in a form of ‘network in a box’ that can be developed by startups targeting for special purpose network (e.g., network for health care services). There are several startups in this area. They focus on developing a new EPC architecture to eliminate the complexity of existing EPC system rather than virtualize the existing EPC system. Network-in-a-box solution cannot compete with major service providers because it cannot provide the end-to-end solution and complete function sets, but it still secure significant number of customers in the niche market.

4.2.1.2. State of the Art and Research Trend

SDN is a key technology for implementing intelligent open control in the network. Global SDN market is led by three major players. ONF leads OpenFlow protocol standardization, Cisco leads the network equipment, and VMWare leads network virtualization solution. SDN standard related activities are mainly led by ONF (Open Network Foundation) where they standardize separation of control and data plane, centralized control functions, and network programming. ONF proposes SDN based EPC architecture using distributed switches. Combined with OF control, multiple distributed switches handles the GTP based mobile network traffic.
flow by expanding OpenFlow protocol. The architecture of ONF controller and switches are independent from external S1 or SGi interfaces.

In EU, MEVICO [6] is EU FP7 research project studying for future mobile core networking. They proposed new concept of 'OpenFlow controlled GW' where the control functional of S-GW and P-GW is centralized and data plane is designed for balancing the traffic load in the distributed architecture called Ultra Flat Architecture (UFA). Following to the MEVICO project, SIGMONA (SDN Concept in Generalized Mobile Network Architectures) SDN based EPC architecture is being studied more actively [3].

Huawei in China and Bell Labs in US are actively developing EPC solution combined with their technologies. Huawei plans to invest 600M US$ for next 5 years to improve their 4G LTE technology and develop 5G related technologies. They proposed SDMN (Software Defined Mobile Network) to provide maximum level of openness and flexibility to the service providers and terminal vendors by developing new concept of SDN based 5G mobile core solution. SDMN designed the networks in two aspects - OpenFlow (OF) for the fixed transport network and MobileFlow (MF) for the mobile network. MF extracts the signaling message from U-Plane and delivers it to mobile network application such as MME and GW. MF controller let MobileFlow engage in the packet processing. OF implements the IP and Ethernet transport network using OF controller and OF transport engine [4].

CellSDN project is studying splitting the control function and middle box function from GW to realize the mobile core using commodity switches [3]. Using this structure, CellSDN is expected to support more sophisticated policy and more scalable network architecture.

In the industry side, EPC vendors try to upgrade their EPC system using SDN technology and eventually enhance them as solution of 5G core network. Global vendors take different approaches toward EPC over SDN based on their technical strength [7]. The H/W oriented companies who already have high performance dedicated data plane systems are less active than the SW companies who focus on SW based solution and uses COTS system as their data plane system. There are also several startups trying to restructure the existing EPC structure to make it more optimized for SDN based architecture. These small companies rather focus on small and special purpose network than public mobile services. The detailed comparison is provided in the following table.
### Table 4.2-3 Comparison of vendor’s approach toward SDN based EPC

<table>
<thead>
<tr>
<th>Company Characteristics</th>
<th>Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company has its own high performance data plane equipment (e.g., Ericsson, ALU, Huawei, Cisco, Juniper)</td>
<td>Relatively conservative in paradigm shift toward SDN based EPC. Try to slow down the conversing to SDN based mobile network.</td>
</tr>
<tr>
<td>SW based solution provider using COTS data plane equipment (e.g., NSN, ZTE)</td>
<td>Active in changing SDN based EPC. Try to apply their EPC solution to the virtualization environment and SDN.</td>
</tr>
<tr>
<td>Startups developing a new types of EPC solution such as “network-in-a-box” (e.g., Connectem Inc., Polaris Networks, Tecore, WiPro)</td>
<td>Very active toward virtualization and SDN. Take revolutionary approach toward SDN by restructuring the EPC to optimize its performance in the virtualized environment. Focus on small scaled special purpose network such as private mobile health care networks, major corporation, and OTT provider’s own virtual network.</td>
</tr>
</tbody>
</table>

NFV (Network Function Virtualization) working group in ETSI standardizes virtualization to re-arrange the network functionality to the dedicated HW and general purposed servers. While SDN focuses on network abstraction, NFV focuses more on reducing CAPEX/OPEX and saving space and energy. Therefore, SDN and NFV form a mutual supplementary relation in implementing network virtualization. Vendors such as NEC and Telefónica are applying SDN and NFV to develop EPC system combining.

NSN has long term plan for EPC over SDN based on ‘Liquid Network’ roadmap. They proposed an architecture where the mobile application software can be easily added/modified/expanded in the virtualized infra using standard HW based COTS ATCA H/W system environment. Mobile network related S/W such as MSS (Mobile Soft Switch), MGW (Multimedia Gateway), IMS, VoIP server, and HSS can be incrementally installed as network grows.

Ericsson announced ‘Network-enabled Cloud’ which adopted a cloud based network control function where LTE control layer is modified to control the OpenFlow switches and finally achieve integrated network control framework. It also provides the Orchestrated Network and Cloud Management function to create new types of service by adding a Service Exposure layer. They try to provide total SDN infra in service provider’s point of view.
NEC developed vEPC (Virtualized EPC) – an EPC solution on top of Intel server platform. vEPC adopts NEC CGHV (Carrier-Grade HyperVisor) and Intel DPDK (Data Plane Development Kit) to achieve maximum virtualization performance.

Cisco provides ONE (Open Network Environment) – a total solution to support open, programmable, and application-aware network. The ONE has three core elements – SDN controller/Agent, Platform API, and Overlay for virtual network) for both enterprise and service providers.

Korean operators are actively heading toward SDN/NFV based mobile core network. They complete the basic function, performance, and reliability verification on major mobile networking system. They take one example approach to complete the virtualization of mobile network. In first phase, convert the Cloud Network between EPC and Service Network to OpenFlow switch. In second phase, convert EPC to OpenFlow controlled. In third phase, virtualization will be extended to digital units of the base stations, and EPC and Service will be coordinated by Orchestrator.

4.2.1.3. Research Direction

4.2.1.3.1. Open Control Protocol for Mobile Core Network

One of the main challenges in virtualizing the core network is designing mobile core control protocol. Current open protocol (e.g., OpenFlow) only covers fixed network. But in order to apply the SDN in the mobile network, the open control protocol should be upgraded. In order to apply SDN to mobile core, Open Network Control protocol (e.g., OpenFlow) should be extended, because current protocol is designed to support the fixed network. Supporting the core mobile network technology such as GTP (GPRS Tunneling Protocol), handover, and mobile QoS is necessary. Especially, network level synchronization to support seamless handover in SDN environment is another challenge issue. High performance data plane for interworking with the control plane is essential. The control of each switch agent should be tightly organized and synchronized to achieve the seamless handover among the different access points. The data plane configuration should be modified in very high speed.

SDN based 5G core network is studied toward in several directions. The approaches can be divided by three directions. Based on how to allocate the functional entities, the approaches can
be categorized as application based, middle box based, or new protocol based. In the application based approach, all the core network related function is implemented in the control servers and the switch perform the same functions as in the fixed network. In the middle box approach, the 5G core related function is implemented in the separated middle box. Controller controls the chaining of the service flow. Both application based and middle box based approaches require no additional change to the OpenFlow protocol, thus, the existing OpenFlow switch can be used in the network. The new protocol based approach requires protocol and data plane upgrade but can provide more optimal performance than other two approaches. The detail comparison of the three approaches is summarized in the following figure and table. The direction should be determined with careful analysis of existing technical bottleneck.

![Figure 4.2-2 Three approaches toward SDN based 5G core network](image)

**Table 4.2-4 Comparison of the three approaches**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>(a) application based</th>
<th>(b) middle box based</th>
<th>(c) New protocol based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pros</td>
<td>Supporting data plane by expending the virtualized control plane.</td>
<td>Supporting data plane by using dedicated middle box.</td>
<td>Supporting data plane by expanding OpenFlow protocol</td>
</tr>
<tr>
<td>Cons</td>
<td>Minimize the change of OpenFlow data plane.</td>
<td>Minimize the change of OpenFlow data plane.</td>
<td>Can provide optimal performance in delay and handover latency.</td>
</tr>
<tr>
<td>Cons</td>
<td>Big data processing delay By redirecting the packet to the control plane, the performance can be degraded</td>
<td>Big data processing delay The management and equipment cost can be increased by adding dedicated middle box hardware</td>
<td>Requires expanding the control protocols (e.g., OpenFlow) Tighten the relation between controller and switches.</td>
</tr>
<tr>
<td>Example</td>
<td>Huawei (SDMN - Software Defined Mobile Network)</td>
<td>BellLabs (CellSDN)</td>
<td>Cisco (ONE: Open Network Environment)</td>
</tr>
</tbody>
</table>
The research for relation between SDN and NFV technologies is required to apply SDN into 5G network. For example, there are several options mapping the SDN entities into the NFV architectural framework. The positions of the SDN controller, SDN applications, and SDN resources should be identified in the integrated SDN/NFV architecture for 5G network. 5G network should have various wired and wireless cloud resources in order to satisfy the completely different requirements from mobile broadband, massive IoT, and mission critical communications. So multi-domain controller should be provided for multi RAT environment, end-to-end network slicing, and integrated control for multiple administration domains. The research for extension of current SDN technologies is required to satisfy stringent 5G performance requirements such as ultra-low latency and high peak rate. For example, Openflow, the SDN southbound interface, as well as northbound interface for SDN controller should be analyzed and enhanced for carrier-grade 5G network environment.

4.2.1.3.2. Optimal Virtualization of Mobile Core Control Functions
Current network is high cost structure by HW based equipment where the control plane and the data plane are tightly combined. EPC architecture is designed without considering the virtualizations thus is not optimal for future 5G core network. As we described before network-in-a-box solution vendors already try to optimize the mobile core network function by restructuring current EPC architecture to be optimized in virtualization environment. We have to find the more optimal architecture of network control function considering higher flexibility and scalability. The control function should be re-configured dynamically as the network condition is changed. The scope of the core network should be extended to support not only the mobile network but also the fixed network including WLAN. The challenge to optimize the mobile core networking function includes the integration of functional entity and minimization of interface overhead.

NFV MANO (Management and Orchestration) specification in ETSI ISG focuses on the virtualization-specific management and orchestration in the NFV framework. But 5G network should manage not only virtualized network functions and resources but also non-virtualized nodes such as RRU(Remote Radio Unit) of wireless access. So it is required to expand current limitations of NFV MANO functions and technologies.
The scope of virtualization target on 5G network functions should be identified, and the optimization techniques for allocating the VNF to the physical network node should be invented. The relationship between SDN and NFV should be analyzed not only in architectural viewpoint but also in service flow realization aspect at the invocation of service. MEC (Mobile Edge Computing) or fog computing should be introduced into 5G network for low latency requirement of IoT and V2X applications. So the positioning of those technologies in the NFV framework should also be identified.

References


4.2.2. Flat & Distributed Network

4.2.2.1. General Overview

![Diagram showing centralized network architecture](image1)

Current mobile network like EPC is designed in centralized architecture mainly for the convenient management of the entire network. S-GW, P-GW, and MME are implemented as a single system and located in a center node of the network. In the early stage of mobile service, the volume of the mobile service traffic is negligible, and this centralized architecture satisfies the operators need. However, the centralized architecture encounters problems as the volume of the traffic increases. SIPTO (Selected IP Traffic Offload) and LIPA (Local IP Access) is proposed to solve this traffic explosion issue [1][2]. But this approach is for offloading a portion of traffic to Internet bypassing mobile core network. The offloading traffic is treated as a special case, and QoS or HO performance cannot be guaranteed.

In the centralized architecture, the GW becomes bottleneck. All the traffic in the mobile network is delivered to the centralized GW, therefore, the GW becomes the bottle neck of the entire traffic. The nationwide traffic volume of mobile network will exceed multiple Tbps by 2020 [3]. Centralized GW with such high capacity may not be technically feasible or cost effective solution. Therefore, the distributed architecture with multiple GWs will be more reasonable solution to handle the traffic explosion.

![Diagram showing distributed network architecture](image2)
The second problem of the centralized network is transport network scalability. The transport network consists of router/switches or optical connection between nodes. The transport network capacity in the metro and backbone should also be increased in order to deliver all the traffic to the center node. This requires additional transport equipment. Exemplary cost analysis shows that the cost of the transport network is about 2~3 time higher than the EPC system [3]. Therefore reducing cost of the transport network is as important as reducing the cost of core network. With distributed architecture, the traffic of the metro and backbone transport network can be reduced effectively.

Guaranteeing low latency service is another problem in the centralized network. Various content and application servers are located in the server farm behind GW. Future low latency service such as tactile internet service [4] cannot be supported in the centralized network because of the large propagation delay between users and servers. As the technology progressed, the packet processing delay in the gateway and intermediate switches/routers can be reduced, but the propagation delay is determined by the distance of the physical locations. Therefore, the GW should move toward user side to provide the latency services.

4.2.2.2. State of the Art and Research Trend

Need of distributed network architecture is discussed in future mobile network standard body. ITU-R defines the vision of the future mobile network including core network. In [5], it is notified that future core network should be intelligent, flexible, and distributed, and more network edge processing by caching the contents.

UFA (Ultra Flat Architecture) is a network architecture proposed by EURESCOM to accommodate the traffic explosion of the mobile traffic. It changes the hierarchical network architecture to simple distributed network architecture to accommodate the traffic.

In EU, MEVICO project [3] performed on research on future mobile core network. They investigated future direction on mobile core network and proposed a several key technologies to implement the UFA. They identified that core network should be capable of resolving the problems such as explosion of traffic, number of mobile terminal, always-on-app signaling load, heterogeneous wireless access, and terminal of multiple interface. They proposed UFA where the signaling servers and mobile anchoring point are move toward user side. In the
UFA, the GW functions and network control function are distributed to the access side to minimize the traffic load of the backbone network.

To realize the UFA based mobile core network, seamless handover between GW should be supported. In the current mobile network, the P-GW cannot be changed once the terminal is attached to a GW. By converting the network architecture from centralized to distributed model, the handover between GW will occur more frequently. Therefore, the handover mechanism also needs to be distributed. DMM (Distribute Mobility Management) [6] based GW mobility is proposed by IETF DMM working group to overcome the potential problem of existing centralized mobility control scheme. DMM proposes an inter-GW mobility scheme. It overcomes the limitation of fixed anchoring point and provides the capability that the terminal can attach to the multiple GWs dynamically while seamless service is maintained. When the terminal needs GW handover, the terminal is allocated a new IP address. In order to support the seamless service, the terminal should maintain both old and new IP address and session continuity should be maintained in both addresses. DMM is standardized in IETF but is not applied to the EPC based mobile core network because it requires the major functional change in terminal, base station, and control servers. In the 5G network, DMM like distributed mobility control is expected [7].

4.2.2.3. Research Direction

![Figure 4.2-4 Overview of the Fully Distributed architecture](image-url)
4.2.2.3.1. Virtualized Logical GW with Distributed Switch

As explained, hierarchical core network architecture using fixed GW is not scalable for traffic explosion. Distributed data plane architecture is necessary to resolve the traffic explosion. In the virtualization environment, the processing resources can be flexibly allocated anywhere without limitation of its location. Using distributed architecture, the traffic and signaling explosion can be accommodated with logical GW and distributed switches. The data plane should be flat architecture with the GW located more closely toward user side.

It is generally agreed that distributed architecture has advantage in handling traffic. Core network system can be deployed more flexibly as the local traffic demand increases. The higher network utilization can be achieved than the centralized case. Also the cost for transport network can be reduced by reducing the backbone traffic.

However, it may pay additional cost. Maintaining the core network system in multiple sites will requires more operation cost. For example, as many Internet exchange points as the GWs will be required, this will require more equipment and operation cost. Providing reliability in the distributed architecture is another challenge in the distributed network. Since the number of points of failure increases, the maintaining carrier grade reliability will be more difficult. The comparison of centralized and distributed architecture is summarized in the following table.

If we simply deploy more EPC systems to the user side, it cannot resolve the fundamental limitation of EPC. Since current EPC system is designed as standalone system, interworking with other EPC cannot be supported. For example, current EPC cannot support dynamic traffic load balancing with adjacent EPC systems. Simply distributing more EPC system also requires more Internet exchange points that increase the cost of the networks and complicated the management architecture.
### Table 4.2-5 Summary of Centralized vs. Distributed Architecture

<table>
<thead>
<tr>
<th>Feature</th>
<th>Current: Centralized and hierarchical EPC</th>
<th>Objective: Fully Distributed Architecture 5G core</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture Overview</td>
<td>MME/P-GW/S-GW is implemented as a single system and located in the central node of the network.</td>
<td>Control servers are distributed in the access network to be located in closer to the user/terminal.</td>
</tr>
<tr>
<td>GW Flexibility</td>
<td>Fixed P-GW architecture</td>
<td>GW can be flexibly changed to support seamless HO and balance the traffic load among GWs.</td>
</tr>
<tr>
<td>Cost</td>
<td>Require capacity expansion of metro and backbone transport network.</td>
<td>Minimize the cost for upgrade the transport network by reducing the traffic introduced to the metro and backbone transport. Additional cost for installing more Internet exchange points. Potentially more operation cost for handling multiple core network system.</td>
</tr>
<tr>
<td>Delay Performance</td>
<td>Difficult to guarantee the low latency because of the long distance transport to GW.</td>
<td>Possible to support the low latency service by minimizing the propagation delay to the GW.</td>
</tr>
<tr>
<td>Resolution of traffic explosion</td>
<td>Increasing the capacity of GW system Offloading the traffic to reduce the traffic load to the centralized GW</td>
<td>Architecturally resolve the traffic explosion by localizing and minimizing the amount of traffic per GW</td>
</tr>
</tbody>
</table>

#### 4.2.2.3.2. Dynamic Mobility Anchoring for Seamless Inter-GW HO

In the current architecture, there is no mechanism to support the seamless inter-GW handover. Changing the P-GW and IP address of UE is impossible unless the UE perform re-attachment. This fixed service anchoring point architecture leads the inefficient routing in the core network. One of main challenges for the distributed architecture is to resolve the GW dependency. In the distributed network architecture, the mobile core network can be deployed dynamically as
the traffic condition changes. More and more mobile core network can be deployed in
distributed manner as needed. Therefore, seamless inter-GW handover and traffic load
balancing between GWs are required in the distributed architecture.
The figure above depicts the inter-GW HO mechanism in the distributed architecture. The UE
can change its GW without affecting its service. Existing session of old GW should also be
continuously supported when UE moves to the new GW. The idle session in the old GW
should be detected and terminated efficiently. The management and control of multiple IP
addresses in an interface of UE is necessary. The efficient tunneling mechanism between GW
should be supported. The smaller delay performance by optimal routing through shortest path
is possible. By distributing the mobility control servers, it overcomes the network failure that
it may happen in the centralized architecture when a network server failure.
As mobile device has been equipped with multiple radio access technologies and it can be
connected to the networks simultaneously. And the mobile device can transmit IP flows through
the multiple radio access networks. Flow mobility management enables IP flow mobility, with
IP flows belonging a mobile device being moved seamlessly between the multiple radio access
networks. The network operator may provide an improved quality of service with higher
bandwidth available to the user at a low cost by using a Flow mobility management mechanism.
The IETF has standardized the MIPv6 protocols extension to support terminal-based flow
mobility, and PMIPv6 protocol extension to support network-based Flow mobility.

4.2.2.3.3. Signaling Mitigation for Always-on-app and IoT
Together with the traffic explosion, signaling explosion is also needed to be resolved. With
allocating the necessary processing resource as needed, the signaling explosion can be
efficiently resolved. Special consideration is needed for the explosion of the number of mobile
devices. The number of device will be rapidly increased as IoT (Internet of Things) devices
become common. The processing complexity to maintain the terminal and session management
will increase rapidly. Also the various always-on-app in the smart phone triggering the bearer
setup in the core network will increase the control load of the core network. Therefore the
signaling mitigation mechanism should be considered in the 5G core network.
The mobile core signaling should be done in the cloud. The cloud itself needs to be
distributed toward the user side to support quick response to the user’s request. Using the distributed cloud based signaling, signaling architecture should also be distributed. The detection and control mechanism to protect the signaling explosion should be developed. 5GPPP also initiated research on the application-service aware network where the application signaling overhead can be handled more efficiently [8].

4.2.2.3.4. IP Flow -based Packet Management
5G should address the issues of complexity and scalability in GTP-based bearer management in existing 3G/4G mobile networks while maintaining the advantages of bearer management in QoS management, accounting, etc. IP flow management can support not only similar functions as in GTP-based bearer management but also more advanced features such as flow-based mobility which provides mobility based on application recognition and cooperative communication which distributes flows across multiple access networks in order to use limited radio resources more efficiently. However, the IP flow-based packet forwarding should be designed to support short, burst traffic from IoT devices along with the design of more optimized signaling protocols.

4.2.2.3.5. Locator Identifier Separator Protocol
LISP-MN (Locator Identifier Separator Protocol - Mobile Node) is a mobility technology based on LISP, which separates a host identifier (EID, Endpoint Identifier) and its locator (RLOC, Routing Locators). According to LISP-MN, a mobile node is operated as a LISP site. When a mobile node is turned on, it assigns an EID by itself and gets an IP address (RLOC) from a provider. Then, it registers the EID-RLOC mapping to a Map Server. With LISP-MN, mobile nodes can communicate with their corresponding nodes even in the case when they move to other areas because they can continue to use their EID while they should be assigned a new RLOC. LISP-MN is a host-based mobility management technology since mobile nodes should support LISP functionalities like xTR. It also needs network entities such as Map Resolvers and Map Servers. LISP-MN separates control plane and data plane. The architectural feature of LISP-MN that it does not have the entities like Home Agent or Foreign Agent in Mobile IP can prevent triangle routing problems, therefore packets can be transmitted on the shortest path.
References

[1] 3GPP TR 23.829, Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO)
[2] 3GPP TR 23.859 “LIPA mobility and SIPTO at the Local Network (LIMONET)”
4.2.3. Converged Access Control & Transport

4.2.3.1. General Overview

A concept of legacy cellular network is that a mobile user selects a cell, camps on it, receives data services from the cell and changes into another cell by handover. The cellular network determines one cell to provide a mobile user with data services, which is considered as cellular-centric network. The 5G to satisfy various and wide range requirements includes evolved radio access technologies and revolution access technologies. In other words, a 5G network can provide a mobile user with multiple radio accesses simultaneously. In addition, to provide mobile users with optimal radio environment, a 5G network shall decide optimal cells (and radio access) based on user’s feedbacks of available radio status, which is called as a user-centric network. The user-centric network including various radio accesses has functionality that select optimal radio access and provide efficient resource management include an optimal data path. To support a user with massive data services, a 5G network should have a capability to provide multiple radio accesses to a mobile user simultaneously when the multiple radio accesses are on good radio conditions for a user [1]. One of advantages of simultaneous multiple connections (e.g. a macro cell and a micro cell) is separation of user plane and control plane, which can be useful for dual connection in a HetNet radio network. In addition, simultaneous multiple connection on radio accesses shall provide flow-level connection management (carrier aggregation) based on service feature.

Until now, a mobile network architecture has own independent system architecture such as a EPS (Evolved Packet System), WCDMA, mobile WiMAX and Wi-Fi. So new radio access system has own physical infrastructure and an operation & management system, which increases CAPEX and OPEX for network operators. Issues of roaming and handover between the independent radio accesses are solved by interworking GW except between 3GPP family systems. Another issue of independent infrastructure is a non-optimal traffic paths problem when a roaming or handover occurrence without geographical movement between different radio access systems.

Interworking among radio access systems can be classified as a loose-coupled and a tightly coupled structure. Most of current architectures consist of loose-coupled structure, which is
hard to reflect each radio condition of a mobile user. They provide simple inter-radio handover scheme without content information such as user preference, radio condition and service feature. A convergence node has multiple radio accesses that can provide flexible radio resource management and more efficient resource allocation.

4.2.3.2. State of the Art and Research Trend

3GPP standard defines a logical architecture for non-3GPP interworking using Network-based mobility (NBM) (PMIP, S2a, S2b interface) or Host-based mobility (HBM)(MIP, S2c interface). It aims for minimizing the handover latency due to authentication and authorization for network access and can support simultaneous radio transmission capability [2]. The interworking architecture of 3GPP uses an ePDG (or PDG) for untrusted non-3GPP IP access, which may have a non-optimal traffic path due to independent access architectures between a 3GPP mobile system and non-3GPP radio access system such as WLAN with a fixed access network. The Release 12 of 3GPP is trying to solve a WLAN/3GPP radio interworking problem. It aims to provide optimal user experience and unnecessary WLAN scanning [3].

COMBO investigates new integrated approaches for Fixed-Mobile Converged broadband access / aggregation networks for different scenarios [4]. COMBO architecture is based on joint optimization of fixed and mobile access / aggregation networks around the innovative concept of Next Generation Point of Presence (NG-POP). It will suggest a better distribution of all essential functions, equipment and infrastructures of convergent networks. The key objective of COMBO is to define optimized FMC network architectures, which will be quantitatively assessed and compared with respect to Key Performance Indicators such as cost, energy consumption, bitrate, delay, QoS. It provides multi-operator FMC access scenarios to ensure openness and flexibility for network operators and service providers.

The SEMAFOUR project develops a unified self-management system, which enables the network operators to holistically manage and operate their complex heterogeneous mobile networks [5]. The goal is to create a management system that enables an enhanced quality of user experience, improved network performance, improved manageability, and reduced operational costs. One of the objectives is the development of multi-RAT / multi-layer SON functions that provide a closed control loop for the configuration, optimization and failure recovery of the network.
across different RATs (UMTS, LTE, WLAN) and cell layers (macro, micro, pico, and femto). Such coordinated adaptation of radio (resource management) parameters in different RATs and cell layers is imperative for the global optimization of network performance.

4.2.3.3. Research Direction

4.2.3.3.1. Unified Access Control and Resource Management
A 5G network will include various radio accesses such as evolved LTE-Advanced, unlicensed-LTE, Wi-Fi and so on. The 5G network will support both legacy mobile equipment and IoT devices which are many kinds of sensors and actuators. In addition, 5G services have different service features such as delay sensitive services, machine type communication, time tolerant services, vehicle communication and so on. So, it is not reasonable to have a specific access control mechanism per service due to operational cost and management. The 5G network will unify access control for various mobile devices and various services to provide a consistent management and use network resources efficiently. Separated systems which have independent infrastructures are not easy to share their resource information, so they cannot reflect user radio condition into information for resource allocation. The unified access control shall provide resource efficiency by allocate optimal network resources.

4.2.3.3.2. Multi-RAT Flow Aggregation
As mobile equipment having multiple radios are increased, it is possible for a mobile network to have flexibility to select suitable radio resources according to user radio condition, network traffic status, an operator's or user's policy and so on. A 5G Mobile network will be able to use one more radio accesses to transmit data traffic according to service features such as massive streaming data. One of the 5G typical characteristics is an ultra-dense network. An UDN has various range cells such as macro and micro cell, and consists of various kinds of radio accesses. For more efficient data transmission, The 5G network should support multi-RAT flow aggregation which transmits simultaneously data traffic using various kinds of radio access such as LTE, LAA-LTE (Licensed Assisted Access-LTE) and Wi-Fi. A GW should provide the session-level multi-RAT flow aggregation which transmits one session into more
than two RANs. A GW can also schedule data traffic more efficiently if it knows each the radio condition and mobility pattern of a mobile user. The GW receives feedbacks of radio condition, and can determine and allocate network resources to RANs which is available to the mobile user. In addition, the mobility pattern such as speed, and direction can be a candidate information for resource allocation. One of the advantages of Multi-RAT flow aggregation can support seamless handover to moving users.

4.2.3.3. HetNet/Multiple RAT Mobility Control
Heterogeneous networks and multi-Radio Access Technologies (multi-RATs) are also considered as one of the key features of the 5G wireless systems because user wants seamless service in all circumstances. These requirements lead to support mobility across different RATs maintaining the application Quality of Service (QoS) requirements. By taking advantage of multiple RATs, the 5G System will be able to take advantage of the unique characteristics of each RAT and improve the practicality of the system as a whole. For instance, the 4G macro cell is used for exchanging the control messages to maintain the connection, to perform handover. The small cell operating in mmWave unlicensed frequency band would support the gigabit data rate service.

Another key technology to optimize use of multi-RAT technology is virtual RAN. It can seamlessly combine the multi-RATs into a single RAN. It has several advantages such as unified access control and authentication, common mobility management and optimized routing paths to devices for load balancing.

4.2.3.3.4. Wired/wireless Resource Management and Control
Integrated resource management and control of heterogeneous access network enables efficient use of various system resources in a single unified framework by allocating resource based on required QoS and measured signal propagation circumstances. If available resources are low in any one of the access network and it cannot further accept a new user, integrated resource management and control may allocate additional resources borrowed from another access network, or may secure resources for the new user by moving an existing service to a different access network. Integrated resource management and control may select other access network
with less load if the network load exceeds the predetermined threshold value, which may be dynamically modified according to the status of the surrounding network load. If required QoS may not supported through a single access network for a terminal that can use multiple radio interfaces at the same time it is possible to allocate the radio resources for different access networks simultaneously to satisfy the QoS of the user-requested service. It is necessary for the optimization algorithm of the integrated resource management and control to consider a number of factors, such as network status, user's situation.

4.2.3.3.5. Access Condition Aware Content Delivery
Mobile network has structural limitation when deploying and creating new services. The mobile operator tried to support various services using IMS (IP Multimedia Subsystem) but it was not successful. Instead of IMS, 3rd party services become dominant in the mobile network. The mobile network only plays role of simple 'bit pipe' and cannot create additional profit. In this structure, CAPEX and OPEX problem of the mobile operators cannot be solved. Current mobile network also has inefficient architecture when delivering the contents and causes traffic bottleneck and delay. To overcome this problem, access ware content delivery and cloud based mobile networking is needed.

For the mobile cloud networking, efficient resource management such as real-time resource division and allocation is required. Base station and core node should be virtualized to be implemented in cloud. Efficient resource sharing such as RAN sharing, core sharing and resource sharing is required.

For efficient content delivery, condition of access network and user position should be considered when delivering the contents. Access condition aware content delivery (e.g., DASH - Dynamic Adaptive Streaming over HTTP) is needed. To provide best connectivity to the users, LBS based follow-me cloud technology is needed for users to be connected through the best quality radio access.

4.2.3.3.6. Packet Optical Transport Network Switching
With the increase in mobile multimedia services and the spread of internet of things (IoT), the internet data traffic is growing explosively. In the future 5G era, transport networks will
be upgraded to accommodate the real-time and large-capacity traffic in the system which offers ultrahigh-speed optical transmission and high-volume system capacity.

![Figure 4.2-6 The transport network technology roadmap](image)

In the transport network technology-roadmap, the market of circuit transport technologies such as SONET/SDH (including MSPP) are reduced gradually and replaced by the packet transport technologies such as MPLS-TP based PTN with the technology evolution. The transmission speed of the Ethernet/OTN continues to increase and the technology of ROADM/OTN will be continuously developed. With the development of these individual technologies, a new converged optical-circuit-packet (L0-L1-L2) integrated switching system used in POTN (Packet-Optical Transport Network) will be continuously developed to lead the paradigm shift of the transport network.

The step-by-step trends of transmission speed and system capacity are shown in below table.

<table>
<thead>
<tr>
<th></th>
<th>1st Stage (~2017)</th>
<th>2nd Stage (~2020)</th>
<th>3rd Stage (~2025)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmission Speed</td>
<td>~ 100Gbps</td>
<td>~ 400Gbps</td>
<td>~ 1Tbps</td>
</tr>
<tr>
<td>System Capacity</td>
<td>~ Several Tbps</td>
<td>~ Several 10Tbps</td>
<td>~ Several 100Tbps</td>
</tr>
</tbody>
</table>
T-SDN based Optical-Circuit-Packet integrated switching technology is the several 10Tbps scale POTN switching technology integrating 400G Ethernet/MPLS-TP technology, OTN technology and flexible-grid ROADM technology. It provides an open source-based environment that simplifies the management/control of the software-defined transport network through the continuous network evolution.

The below figure is a POTN network based on T-SDN architecture.

Figure 4.2-7 POTN network based on T-SDN architecture

References

4.3 Access Network Enabling Technology

4.3.1. 5G Multi-connectivity

4.3.1.1. General Overview
Multi-RAT connectivity allows a UE to transmit/receive data simultaneously through the radio connections which are established with one or more base stations using different RATs. As shown in the Figure 4.3-1, RAT (Radio Access Technology) of these connections may be different from each other. Multi-RAT connectivity has been emerging as a technique that can provide a high QoE and cost-effective communication leveraging legacy 3G/4G/Wi-Fi networks together with 5G networks.

![Figure 4.3-1 Multi-RAT connectivity](image)

Dual Connectivity is the operation in which a UE utilizes wireless resources allocated by more than two network access points, mostly of the same RAT. Network access points are connected by non-ideal backhaul for the dual connectivity operation. More specifically, the two links of the UE performs wireless communication with base stations in a macrocell and a small cell.
4.3.1.2. State of the Art and Research Trend

In the Dual Connectivity architecture, a Macrocell transmits control messages (Control plane), small cells transmit user data packets (User plane). A base station in the Macrocell controls addition/removal/modification of small cells through RRC connection management, and provides a UE high data rate services using small cells. Detailed discussions and agreements on C-plane and U-plane architectures are described in the following subsections.

4.3.1.2.1. Control plane architecture for dual connectivity

In the dual connectivity operation, there are two options for RRC protocol structure depending on RRC functions of Macrocell and smallcell base stations.

In the option 1, only a Macrocell base station performs the RRC function and it also controls RRC state changes such as addition/removal/modification of small cells. In the option 2, a Macrocell base station performs most of RRC function but smallcell base stations also implement some part of RRC function to assist the operation of the Macrocell base station. The option 1, in which only a Macrocell performs RRC function, has been agreed in the final decision.

Figure 4.3-2 Radio Interface C-plane architecture alternatives for dual connectivity
4.3.1.2.2. **User plane architecture for dual connectivity**

To decide the user plane architecture, this subsection will discuss several options in perspectives of the data transmission and the protocol structure.

The data transmission between a Macrocell base station and a smallcell base station can have options for the bearer split as follows:

- Option 1: S1-U also terminates in SeNB;
- Option 2: S1-U terminates in MeNB, no bearer split in RAN;
- Option 3: S1-U terminates in MeNB, bearer split in RAN.

![Bearer split options](image)

**Figure 4.3-3** Bearer split options

In the case that S1-U terminates at the Macrocell base station, smallcell base station shall support (re-)segmentation. Therefore, the following 4 options can be considered in the protocol perspective.

- Independent PDCPs
- Master-Slave PDCPs
- Independent RLCs
- Master-Slave RLCs

From the bearer split and U-plane protocol options, 9 different options are possible for the U-plane architecture.
In the following, the two alternatives that is agreed in 3GPPP Release 12 Dual Connectivity specification are explained in detail

**Alternative 1A**: S1-U terminates in SeNB + independent PDCPs (no bearer split)

This option defines a U-plane protocol for a bearer at the respective eNB. Therefore, MeNB is not required to buffer or transmit such packets that SeNB is to transmit through EPS bearer. This option has the benefit that the requirements for backhaul between MeNB and SeMB are low and that flow control between the two eNB is not necessary.

![Figure 4.3-4 Alternative 1A](image)

**Alternative 3C**: S1-U terminates in MeNB + bearer split in MeNB + independent RLCs for split bearers

This option assumes that S1-U interface terminates at MeNB and that PDCP layer is implemented at MeNB. Therefore, RLC bearers for the macorcell and the smallcell at UE and eNBs are separated and operates independently. This option has the benefit that there are no security issues and wireless resources of MeBN and SeNB together can be utilized to serve the same bearer.

![Figure 4.3-5 Alternative 3C](image)
The following agreement on the candidate technologies for Dual Connectivity has been achieved in the 3GPP Release 12 specification.

- As a result of discussion on control plane and user plane architecture to support dual connectivity, which is explained in the previous chapter, MeNB RRC protocol structure and 1A/3C architecture has been accepted in the standard specification.
- Small cell handover (addition/removal/modification) procedure by signaling with MeNB in which UE performs the measurement procedure with MeNB based on the RRC configuration information to offload traffic of a data bearer to a small cell base station.
- Security key creation and management procedures for bearers provided by MeNB and SeNB which are to support the bearer establishment/separation/removal operations.
- Power management procedure in which UE transmits data packets using the two uplink radio interfaces (with MeNB and SeNB) simultaneously.

4.3.1.2.3. WALN/3GPP aggregation

Dual connectivity concept is extended to integrate non-LTE radio technology such as WLAN. In 3GPP release 13, WALN/3GPP radio level aggregation WI was developed, and the basic concept of the aggregation is very similar to the dual-connectivity concept. RAN-level aggregation, which simultaneously uses multi-connectivity at radio level, provides QoE improvements, minimizes service interruption, and increases operator control. 3GPP RAN 2 is discussing for WLAN/3GPP radio level integration. WLAN access network becomes transparent to CN in the sense that it should not require WLAN-specific CN nodes and CN interfaces. This provides the operator unified control and management of both 3GPP and WLAN networks as opposed to separately managing them. Solutions should meet the following requirements.

- Improve mobility to/from WLAN while minimizing the core network signaling.
- Improve network control of WLAN offload.
- Improve overall UE throughput by using both cellular and WLAN access.

4.3.1.3. Research Direction

To realize the commercialization and deployment of a smallcell base station in an economical way which supports dual connectivity, researches on system design to reduce implementation
complexity and signaling overhead will be required. The integration of dual connectivity principles with cloudified RAN architecture creates the concept of the multi-connectivity. Without the loss of generality, dual connectivity extends the number of the connected cells with the different coverage and the different radio technologies. There is a discussion on the splitting of the radio protocol to be adopted for cloud-based network architecture as well. The splitting between PDCP and RLC of dual connectivity can be interpreted as the separation of entity requiring real-time processing and non-real-time processing.

Cloudify PDCP and RRC in the central cloud connects multiple cells with macro, micro and small cell coverage. They also connects WLAN AP and even wired access system. UE can transmit and receive the data through any cells which it is connected to according to the required QoS of the data, UE’s mobility and power status etc. Cloud PDCP can determine the optimal transmission path based on the UE’s status, cell load condition and the service type etc. RRC functions can be divided into the central function and the local function, and they collaborate each other to provide the enhanced user experience management. Multi-connectivity is also an easy way for migration to LTE network to 5G-enable network.

The efficient splitting of the protocol functions between the cloud and the BTS are the important research topics for 5G ultra dense network deployment.

Figure 4.3-6 shows the example architecture supporting multi-connectivity with the protocol function split for cloud site and local site.

Figure 4.3-6 5G Multi-connectivity under ultra dense deployment environment.
References


[4] 3GPP TS23.261, IP flow mobility and seamless Wireless Local Area Network (WLAN) offload


[6] 3GPP TS23.861, Network based IP flow mobility

[7] 3GPP TS24.302, Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks


[9] 3GPP TR37.834, Study on WLAN/3GPP Radio Interworking

[10] 3GPP TR37.870, Study on Multi-RAT joint coordination

4.3.2. Radio Access Network Virtualization

4.3.2.1. General Overview
Ideally all baseband processing and radio resource management of a cooperative cell cluster is performed at one central location, with distributed remote radio heads (RRH) left at cell premises. The close proximity of (virtualized) baseband functions ensures fastest coordination of radio resources between attached cell entities and thus best network performance in this cell cluster.

The approach is known as centralized RAN (C-RAN) concept; however there is an add-on price tag due to advanced x-haul transport profiles: attaching “dumb” RRHs requires in general a high-speed/low latency connectivity plane [1].

Centralized RAN approaches are often mentioned in the context of cloud RAN solutions; and in fact C-RAN solutions can bridge the gap to cloudified RAN solutions with virtualized mobile functionalities (deriving from mobile edges) and running on IT platforms: EPC functionality may distribute and in some cases will merge with RAN functions. Generally C-RAN solutions imply higher deployment flexibility, on-demand provisioning of network capacity (elasticity), RAN sharing and RAN “as a Service” opportunities by resource pooling and virtualization of RAN resources. It is expected that 5G will largely rely on such concepts. For the purpose of emphasizing the virtualization architecture, it is rather called as virtualized Radio Access Network (vRAN) instead of C-RAN.

A virtualized Radio Access Network is one in which certain functions are virtualized and implemented on GPP (general purpose processor)-based platforms hosted at the RAN level. It is a step towards a software-based RAN architecture built on NFV, where some previously distributed functions, like baseband processing at each radio site, are centralized, virtualized and pooled, while other functions or applications that may have been more centralized in the past are co-hosted at a pooling node for better network efficiency and enhanced end user QoE. The move from physical network nodes to virtual network nodes on shared hardware makes network functions simpler and less expensive to manage and operate [2].
4.3.2.2. State of the Art and Research Trend

4.3.2.2.1. Cloud RAN with functional splitting

Figure 4.3-7 illustrates the function stack of a LTE eNB: The L1 deals with the baseband processing of the radio signal (i.e. OFDM, MIMO, QAM de/modulator, channel en/decoder), the L2 cares about MAC (incl. HARQ), radio link control (RLC) and packet data convergence (PDCP), whereas the L3 accommodates the radio resource control (RRC) and the radio resource manager (RRM), all supervised by a cross-functional OAM layer. The figure indicates the location of functions for distributed versus centralized base station approaches. The left represents today’s traditional distributed RAN (D-RAN) deployment practice, whereas the right side implies a fully centralized advanced C-RAN hotel solution.

The orange dotted borderline shows the cell site’s RAN interface towards the mobile edge premises, with x-haul requirements increasing from left to right. For traditional D-RAN applications this interface is known as (S1/X2) backhaul, for advanced C-RAN approaches it is referred to as fronthauling (sampled, raw IQ antenna data). The generic term of this RAN interface is called X-haul link (with different profiles).

![Figure 4.3-7 Open RAN Lx Function Splits][1]
The removal of almost all eNB functionality to a central baseband processing location implies that a high amount of unprocessed, raw IQ data needs to be real-time transmitted between connected cell sites to leverage maximal CoMP gains. Real time means RTTs of less than 1ms. IQ transmission rates of C-RAN supporting x-haul links basically depend on the RAT radio channel bandwidth, the sampling rate and bit resolution, but are clearly in the higher gigabit range per sector for today’s configurations (advanced x-haul profile).

Alternatively D-RAN architectures can also realize considerable inter-site CoMP gains by distributing and sharing the channel information in an almost real-time manner (typ. 5ms) among sites via an enhanced x-haul profile misleadingly known as “non-ideal backhaul”. The required extra x-haul bandwidths for eNB cooperation is in the range of 15-30% depending on the CoMP model, but stay clearly below the 1Gbps mark for typical RAT configurations.

Centralized RAN concepts are not always associated with large-scaling digital baseband (DU) units: in suitable HetNet environments site deployment strategies may vary rather following a “best of both” approach where a macro eNB may host the baseband processing of neighboring small cells, referred to as small-scale or macro C-RAN (MC-RAN). Compared to hotel C-RAN solutions (HC-RAN) the x-haul profiles are less demanding w.r.t. link bandwidths and reach (less than 500m). First field roll-outs confirm that in contrast to common D-RAN deployment practices the number of new small cells can be significantly reduced (up to 50%).

Though the interface between the distributed and centralized node is not covered by the 3GPP specification, from the introduction of CoMP, it is already reflected to 3GPP architecture. In the 3GPP 5G RAN workshop held on Sept 2015, the proposal on the flexible functional splitting on the RAN architecture was proposed. Figure 4.3-8 shows the proposed concept, and several function splitting scenarios which could be applied to the various deployment scenarios [3]. According to the fronthaul latency required and the required level of the collaboration between cells, optimal splitting architecture could be chosen by the operators.

Because it is expected that 5G system requires higher bandwidth and low-latency in the fronthaul link, the splitting over L1 (PHY) is more promising. The left most splitting architecture is also considered as an extension of the dual connectivity, that is, multi-connectivity.
4.3.2.2.2. Virtualized RAN (vRAN)

A vRAN evolves eNB processing one step beyond centralization and hoteling by virtualizing most of the eNB function and hosting it on GPP servers. With a centralized RAN, dedicated eNB equipment is stacked in a central site and connected to each radio with fronthaul. Hotel eNB provide limited sharing of stacked eNB resources by interconnecting them to add features like inter-site coordination. But with virtualization, a vRAN runs the eNB function over a pool of virtual machines and offers the possibility to scale the pool dynamically and allocate virtual machines on demand.

The advantages of pooling arise for two reasons. First, not all cells experience peak traffic at the same instant in time. Secondly, a shared virtualized pool can be provisioned on a “just in time” basis, instead of installing extra processing capacity for months or years ahead, as is common practice for distributed eNBs. As a result, vRAN virtualization of load-sensitive baseband functions provides an optimal way to pool basebands for a cluster of cell sites. Figure 4.3-9 shows the example of the resource pooling benefits for varying traffic patterns.
Figure 4.3-9 Radio Resource pooling with the virtualized baseband

A vRAN also allows existing applications or specific parts of them to be hosted at the edge of the network instead of implementing them deeper in the network as is the case with traditional architectures. As a result, better service delivery can be achieved, which improves end user QoE, customer satisfaction and retention. The biggest anticipated gains will come from mobile access to interactive cloud applications, and especially to ultra-low latency service for 5G such as autonomous driving, real-time remote control and tactile internet services. This is known as a mobile-edge computing (MEC), which is one of the key service scenarios for 5G system. Table 4.3-1 shows the benefit of the virtualized RAN architecture.

<table>
<thead>
<tr>
<th>KPIs</th>
<th>Benefits</th>
</tr>
</thead>
</table>
| Network Performance | • Easier scaling, pooling and load balancing across centralized cells  
                     • Better performance using cell coordination features  
                     • Intelligent traffic steering |
| TCO              | • Simplified site acquisition, build and upgrade  
                     • Hardware economies of scale  
                     • Simplified operations  
                     • Lower cost redundancy |
| Revenue          | • Applications implemented at the edge of the network for greater end user QoE  
                     • Fast access to local information for low latency applications |
4.3.2.3. Research Direction

The current research focus on the virtualized RAN research is on the elastic scaling of the RAN functions. Elastic scaling enables operators to deploy just-in-time baseband resources matched to a network’s growing load rather than provisioning capacity to meet each cell’s expected peak demand. The technique will be especially beneficial as we move into the 5G and Internet of Things (IoT) era where future capacity demands are very difficult to predict.

Elastic scaling minimizes the number of servers that need to be deployed and reduces energy consumption. A virtual eNB can be deployed and made operational within just ten minutes to support a new RF resource. There was a demonstration on the cloud-based base station that offers a layered approach with different deployment models to optimize performance by enabling basic capacity to be built near cell sites, and with peak capacity in the data center being redirected to follow traffic demand. The demonstration was on how to enable operators to deploy just-in-time baseband resources, e.g. as network load increases with IoT with the dynamic adjustment of number of instances.

One of the important trend related to the virtualized RAN is that telecom vendors are introducing their own IT solution including servers and switches which can implement both mobile network function and normal IT application with the one HW platform. These solutions accelerate the introduction of the NFV, virtualized RAN etc.

References

[1] NGMN; D9 deliverable; “RAN Evolution – Backhaul and Fronthaul Requirements”; 05/2014
4.3.3. Self-Organizing Network

4.3.3.1. General Overview
The conventional network operating method is based on statistics obtained from OAM and the base station/user device system logs. The 5G network, however, will evolve into the correct operations to solve the network problems by automatically analyzing. The 5G system performs all configuration installation by real-time optimizing the radio environment and network operations to improve the quality of the user service. Moreover, the system stability without service interruption is ensured by performing automatic failure recovery with real-time monitoring. In other words, users can enjoy their 5G service with fully optimized network.

4.3.3.2. State of the Art and Research Trend

4.3.3.2.1. Main features of SON

Self-Configuration
Self-Configuration is a function for setting the network when the first time it started. Plug & Play, the key technology, is a function that enables system operators to configure and inter-work with their existing system automatically when the base station is initially powered on [1]. The allocation of IP address, configuration the interface, product certification, and software download are performed automatically. After that, it obtains default configuration parameters; in addition, PCI configuration and neighboring cell information are collected. Therefore, it can reduce the human error and resource.
Figure 4.3-10 Plug & Play Configuration

The next feature, one of the key technologies, is Automatic Neighbor Relation (ANR). The corresponding function manages a lot of neighbor cells and their complex relationships automatically. Basically, there are two large categories. By detecting the neighbor through the measurement of the device makes. In addition, when the information in the neighbor cell is absent, the support of the OAM makes the neighbor relation.

Figures 4.3-11 ANR process

Self-Optimization

Now, Self-Optimization features will be discussed: Mobility Load Balance (MLB), Mobility Robustness Optimization (MRO) and Coverage and Capacity Optimization (CCO) [2]. The MLB is a function to check the load distribution of each cell based on their RB usage and number
of users and to trigger the users of the high load cell to do handover to a different frequency in order to balance the load of each cell. With this feature, user’s average throughput can be increased while call drops are reduced.

In this figure, the higher-loaded cell is checked, and the users in the cell get handover to the lower-loaded cell to balance the loads between two cells.

Secondly, MRO is a function that aware of the problems such as too-late or too-early handovers and optimizing the handover parameters.
Lastly, CCO is a function to optimize coverage and capacity by changing the parameters based on the UE measurement. Coverage and capacity are analyzed by monitoring MDT data and KPI. If network issues such as weak coverage, lack of capacity are identified, root cause is analyzed automatically. After that, cell coverage and capacity are optimized by changing parameters such as HO, MLB parameter, neighbor cell related parameter and eNB power parameter.

**Figure 4.3-14 Example of CCO operation**

**Self-Healing**

Self-Healing is a function that monitors the hardware problems and repairs it by switching the board automatically without any service interruption [3].

**Figure 4.3-15 Example of Self-healing of Board Fault Function operation**
In the figure, the existing board #1, that has problems, is switched to stand-by board. If the fault is clear, the board will be operated as a stand-by board. If not, the board will be blocked. In addition, cell outage detection, the key feature of the Self-Healing, is possible and compensates this coverage loss.

![Figure 4.3-16 Self-healing of Cell Outage Process]

**4.3.3.2.2. Evolution trend of SON**

The Conventional Self-Organizing Network (SON) is based on network. For co-existence of multi-vendor and multi-RAT (Radio Access Technology), the goal of SON is the self-optimization of network on various environments. Then, SON is developed from Network centric to Scenario and User centric. Adaptive scenarios- urban, suburban, and metro, is considered in scenario centric case. Network is divided and optimized at the grid-level, network is optimized on grid-level. Finally, SON will evolve into user centric case. Big data based analytics will be used to learn user patterns and behaviors. Based on these data, network will be optimized automatically.
The effect of SON is as shown in Fig.4.3-18. CAPEX, OPEX and operation effort will be decreased and customer experience quality will be increased using SON.

References

[1] 3GPP TS32.501, Telecommunication management; Self-configuration of network elements; Concepts and requirements.

[2] 3GPP TS 32.521, Telecommunication management; Self-Organizing Networks (SON) Policy Network Resource Model (NRM) Integration Reference Point (IRP); Requirements.

[3] 3GPP TS 32.541, Telecommunication management; Self-Organizing Networks (SON); Self-healing concepts and requirements.
4.3.4. Xhaul technologies

4.3.4.1. General Overview

Recently, data traffic in access network are rapidly increasing due to various types of smart mobile devices and services. Increasing data traffic require a continuous expansion in access network capacity, and optical access technologies play a critical role. Meanwhile, as mobile broadband services such as long term evolution-advanced (LTE-A) and 5G have been utilized, and many mobile services providers utilize separated type of remote radio head (RRH) and base band unit (BBU) [1]-[8]. The BBU are usually installed in the central office, whereas the RRH are deployed at outside. These centralized radio access network (C-RAN) architecture provides flexible operation of mobile network and reduce CAPEX/OPEX. The physical connection between RRH and BBU, fronthaul, requires guaranteed bandwidth, low latency, and time synchronization.

If the optical access network supporting wireless and wireline services is individually deployed to copy with traffic demand, the CAPEX/OPEX will be very high. To resolve these issues, wired and wireless network resources need to be integrated to achieve efficient deployment of future optical network. Thus, there have extensive researches to share legacy optical infrastructure, resource, and equipment with mobile link with a cost efficient point of view.

The research direction to accommodate 5G mobile traffic in optical access such as radio-over-fiber (RoF) [3-5], radio over Ethernet [6-8], passive optical network (PON), and wireless backhaul is reviewed.

4.3.4.2. State of the Art and Research Trend

4.3.4.2.1. Radio over Fiber (RoF)

Mobile fronthaul based on RoF technology transmits mobile signal with analogy format as shown in Figure 4.3-19. Transmitting baseband signal without digitizing processing could reduce the required capacity of mobile fronthaul. Furthermore, the RoF has no delay arising from digital signal processing due to combining and de-combining of frames of mobile signal, and could utilize optic links used in LTE/LTE-A system.
In mobile fronthaul based on RoF, the baseband signals are mapped onto multiple intermediate frequency carriers and multiplexed. After that, the multiplexed IF carriers are transmitted in mobile fronthaul through optical fiber links. A number of IF carriers could be transmitted on each wavelength without additional bandwidth arising from digitizing processing. Furthermore, it is possible to increase the number of IF carriers by using WDM technology for high-capacity mobile fronthaul.

For the commercialization of RoF technology in mobile fronthaul, there are some required technologies such as 1) Transmission of IF carriers between DU and RU without distortions in analog fiber optic links; 2) Construction of RoF links interconnecting high-capacity DU and compact RU equipment; 3) Control and managing technology between DU and multiple RU. The standardization of mobile fronthaul based on RoF technology have been performed in ITU-T/SG15/Q2 for definition, key technique, applications and so on.

4.3.4.2.2. Radio over Ethernet

In C-RAN, RoE uses Ethernet in fronthaul which is the networks between DU and RU to alleviate the capacity constraint on the fronthaul links in CPRI/OBSAI protocol. In RoE, the DU splits some RF related baseband processing function to RU as in Figure 4.3-20. As a results, the RoE technology provides solutions for capacity constraint on the fronthaul link. The related works on RoE have been studied in IEEE 1904.3 task force (Standard for Radio Over Ethernet Encapsulations and Mappings).

The physical layer could be divided to modulation/demodulation function and channel coding function. When DU splits the modulation/demodulation function to RU, the required capacity could be reduced up to 1/5 compared to CPRI/OBSAI. To implement the RoE technology, there
are some required technologies such as 1) Formatting ethernet frame in RoE; 2) Frame mapping technique compatible with CPRI; 3) timing recovery and synchronization based on Ethernet.

![RoE with split within L1](image)

**Figure 4.3-20 RoE concept**

4.3.4.2.3. Passive optical network (PON)

Passive optical network (PON) is an optical access network technology which can provide 1Gb/s or more high-speed Internet service by applying passive optical distribution devices between the central office and subscribers. PON technology is classified into a time division multiplexing (TDM) scheme and a wavelength division multiplexing (WDM) scheme by the optical signal transmission method. Maintenance and management costs of the network can be reduced effectively because PON technology uses passive components [9]. PON uses a fully validated optical communication technology in metro optical network and long-haul optical network. Thus, PON can ensure price competitiveness with the mass production of related parts for the subscriber services at low cost. PON technology can accommodate subscribers within 20 km distance by utilizing optical fiber cable and PON can configure various network topologies with using a multi-stage passive splitter to satisfy the characteristics of the different regions. Further, PON technology may be configured in the network structure of various forms, such as Ring, Tree, Bus, etc., in which PON can secure 99.999% of network survivability by applying the protection and recovery techniques such as redundancy of the system.

Recently, the low-cost PON technology has been studies actively for using a mobile front holes and backhaul since recent mobile traffic is increased. Timing synchronization function is essential to apply a PON network into the mobile backhaul. This can be implemented using a method such as IEEE 1588v2. International cable TV operators and mobile operators are making a use or planned use of mobile backhaul by using EPON and GPON which are TDM PON technology [10]. Currently, NG-PON2 (Next-generation PON2) technology as 40G class
next generation PON technology is developed for an International standardization in ITU-T SG15. Standardization of NG-EPON technology has been carried out from 2015 second half in the IEEE 802.3 [11, 12]. Standard technologies of the next generation PON presented mobile backhaul / front hall networks as typical applications. In particular, wavelength tunable ptp WDM-PON technology in ITU-T NG-PON2 standard technology, includes main application with the mobile front hall which use the RoF, CPRI and OBSAI protocols.

In TDM-PON, ONUs share a single optical wavelength for signal transmission and use a time division multiplexing method. Therefore, the bandwidth allocated to each TDM-PON ONU is decreasing, therefore, service is delayed when the number of user is increased. IEEE EPON and ITU-T G-PON are typical TDM-PON technologies. Currently, standardization and development of 10 Gb/s TDM-PON system have been completed. Regarding a mobile backhaul based on a TDM-PON technology, Korea and world-wide system manufacturers have developed a LTE backhaul or metro Femto Wi-Fi network by using GPON [13,14]. Considering the transmission capacity and latency of 10 Gb/s TDM-PON, 10 Gb/s TDM-PON would be suitable for the LTE backhaul method of DU-RU integrated base station. However, 5G network has a vision to provide 10-100 times faster speed and 1000 times larger network capacity and requires a high-speed latency. Consequently, TDM-PON technology would not be suitable for 5G fronthaul and 5G backhaul.

![Figure 4.3-21 LTE backhaul network structure using the TDM-PON OLT and ONU](image)

Since the WDM-PON OLT and each WDM-PON ONU communicate using optical signals of dedicated wavelengths WDM-PON can provides always a constant bandwidth to the user irrespective of the number of the ONU that has been operated. Therefore, there is an advantage
of being able to configure the network regardless of the number of channels, an increase in transmission capacity, and the protocol and transmission speed. However, WDM-PON system has an inventory problem caused by possession of different light sources and it present cost burden of WDM-PON system. To solve this problem, the wavelength-independent and wavelength-tunable WDM-PON had been researched and developed. Typical WDM-PON technologies are ITU-T G.698.3 of seed light injected WDM PON technology and ITU-T G.989.2 technology of P-t-P WDM-PON technology. P-t-P WDM-PON provides a mobile backhaul and front haul as the main application fields. WDM-PON can be configured in a ring topology or tree topology, also, by using an optical add-drop multiplexer or an optical power splitter. Since the OLT and the ONUs are directly connected in point-to-point, there is no latency issue so that WDM-PON is applicable to both the mobile fronthaul and the backhaul. CWDM-PON system is inexpensive because a CWDM light source that does not require a temperature controlling function. However, the maximum usable number of wavelengths is limited to sixteen wavelengths. In the future, because of easy increase of operating channels DWDM (dense WDM) system would be preferred than the CWDM system for large capacity 5G fronthaul or backhaul

![Diagram](image.png)

**Figure 4.3-22 Mobile fronthaul structure using the P-t-P WDM-PON**

Figure 4.3-22 is the mobile fronthaul structure by using a P-t-P WDM-PON. WDM downstream signals are transmitted at the DU and demultiplexed after passing through the optical add-drop multiplexers at the nodes. RU uses wavelength tunable transmitter to transmit a single channel upstream signal which is dedicated by the optical add-drop multiplexer. In the case of using optical power splitters at the nodes, RU should apply a wavelength tunable receiver to select a single wavelength signal among the WDM signals.
Figure 4.3-23 Mobile backhaul structure using the P-t-P WDM PON OLT and ONU

Figure 4.3-23 is the mobile backhaul structure by using a P-t-P WDM-PON. WDM downstream signals are transmitted at the OLT and demultiplexed after passing through the optical add-drop multiplexers at the nodes. ONU uses a wavelength tunable transmitter to transmit a single channel upstream signal which is dedicated by the optical add-drop multiplexer. In the case of using optical power splitters at the nodes, ONU should apply a wavelength tunable receiver to select a single wavelength signal among the WDM signals.

References
[7] Suggestions on Potential Solutions to C-RAN by NGMN Alliance V4.0
4.3.5. Vehicular Communication Network (V2X)

4.3.5.1. General Overview

Vehicular communication (V2X: vehicle-to-everything) has many uses, including navigation and driver assistance, travel information, congestion avoidance, fleet management, payment transactions and for traffic control and safety. V2X communication may occur in multiple contexts: vehicle-to-vehicle (V2V) communication, vehicle-to-infrastructure (V2I) communication, vehicle-to-pedestrian (V2P) communication and vehicle to home communication [1]. These uses are referred to as Intelligent Transport Systems (ITS). V2X applications range from personal communication, green transportation, societal mobility and safety to bring more travel convenience, comfort and safety. LTE radio can be efficiently used for vehicular communication because of its inherent radio capabilities such as extensive coverage for urban and rural roads, low latency and higher capacity etc. LTE enhancements improve LTE usage for vehicular communications including D2D communication. [2]

As shown in Figure 4.3-24, 3GPP categorized V2X communication into 3 types of communications. Following is the requirement for such 3 types of communications at the 3GPP study [1].

\textit{Vehicle-to-Vehicle (V2V)}

Network allows such UEs that are in proximity of each other to exchange V2V-related information when permission, authorisation and proximity criteria are fulfilled. The proximity criteria can be configured by the network operators. However, UEs supporting V2V Service can
exchange such information when served by or not served by the network. The UE supporting V2V applications transmits application layer information (e.g., about its location, dynamics, and attributes as part of the V2V Service).

V2V is predominantly broadcast-based; V2V includes the exchange of V2V-related application information between distinct UEs directly and/or, due to the limited direct communication range of V2V, the exchange of V2V-related application information between distinct UEs via infrastructure, e.g., RSU.

*Vehicle-to-Infrastructure (V2I)*

The UE supporting V2I applications sends application layer information to RSU. RSU sends application layer information to a group of UEs or a UE supporting V2I applications.

V2N(Vehicle-to-Network) is also introduced where one party is a UE and the other party is a serving entity, both supporting V2N applications and communicating with each other via LTE network.

*Vehicle-to-Pedestrian (V2P)*

Network allows such UEs that are in proximity of each other to exchange V2P-related information using Network when permission, authorization and proximity criteria are fulfilled. Similar requirement as V2V communication is assumed for V2P communication. The UE supporting V2P applications transmits application layer information. Such information can be transmitted either by a vehicle with UE supporting V2X Service (e.g., warning to pedestrian), or by a pedestrian with UE supporting V2X Service (e.g., warning to vehicle).

V2P includes the exchange of V2P-related application information between distinct UEs (one for vehicle and the other for pedestrian) directly and/or, due to the limited direct communication range of V2P, the exchange of V2P-related application information between distinct UEs via infrastructure, e.g., RSU.

Many companies such as Volvo, Audi, Google, etc. are developing so called “autonomous driving car,” and it aims to help ease the driver as well as to reduce possible traffic accident. V2X technology assists drivers to recognize the speed of surrounding vehicles automatically in such circumstances, and control the car system to stop facing with accident or sudden braking immediately upon grade. In addition, even when close to the object in the blind spot immediately allow to reflect the driving.
4.3.5.2. State of the Art and Research Trend

There are two main communication classes that can support V2X applications. These are cellular-based communication systems (e.g. LTE) and Wi-Fi-based communication (e.g. 802.11p or 802.11n). These systems have different characteristics regarding latency, coverage, reliability, communication costs and data rate. Although the latency of cellular communication systems decreases during the evolution of these systems (from up to several seconds for 2G to under 100 ms for 4G LTE networks), Wi-Fi systems provide a delay of only several milliseconds in most situations. In contrast to that, the coverage of Wi-Fi is significantly smaller (up to 200 m) compared to cellular communication (several kilometers). This is due to the lower transmission power and higher frequency of 802.11p (5.9 GHz). In multi-hop scenarios, the coverage of Wi-Fi systems can be increased, but the drawback is a higher delay. The reliability for both communication classes depends on the environment and on other users in communication range. Typically, cellular system provides higher reliability than Wi-Fi based system, and it also guarantees the QoS (Quality of Service) for the V2X applications comparing Wi-Fi based system. But these systems are operating in an unlicensed spectrum in contrast to cellular communications where the operators have to pay for the frequencies (resulting in higher communication costs). The data rate is similar for both classes. It is significantly influenced by the channel conditions, but the latest extensions of the standards (LTE-A and 802.11ac) enable peak data rates from more than 1 GBit/s. Also hybrid approaches, combining the advantages of both cellular-based and Wi-Fi-based communication systems, are suitable solutions for efficient V2X communications. Since LTE introduced D2D communication link from Release 12, cellular D2D can be used instead of Wi-Fi based system. In addition, 5G communication system aiming less than 1ms latency can overcome the drawback of the high latency of the legacy cellular based V2X service [3]. Table 4.3-2 shows the comparison result on the various candidate communication systems on V2X service.
### Table 4.3-2 Main candidate wireless technologies for on-the-road communications

<table>
<thead>
<tr>
<th>Features</th>
<th>Wi-Fi</th>
<th>802.11p</th>
<th>UMTS</th>
<th>LTE/LTE-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel BW</td>
<td>20MHz</td>
<td>10MHz</td>
<td>5MHz</td>
<td>1.4, 3, 5, 10, 20MHz</td>
</tr>
<tr>
<td>Frequency Band</td>
<td>2.4GHz, 5.2GHz</td>
<td>5.86-5.92GHz</td>
<td>700-2600MHz</td>
<td>700-2690MHz</td>
</tr>
<tr>
<td>Bit Rate</td>
<td>~100m</td>
<td>~1km</td>
<td>~10km</td>
<td>~30km</td>
</tr>
<tr>
<td>Capacity</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>High/Very High</td>
</tr>
<tr>
<td>Coverage</td>
<td>Intermittent</td>
<td>Intermittent</td>
<td>Ubiquitous</td>
<td>Ubiquitous</td>
</tr>
<tr>
<td>Mobility Support</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Very High (~350km/h)</td>
</tr>
<tr>
<td>QoS</td>
<td>EDCA</td>
<td>EDCA</td>
<td>QoS class/bearer selection</td>
<td>QCI and bearer selection</td>
</tr>
<tr>
<td>Broadcast/Multicast</td>
<td>Native broadcast</td>
<td>Native broadcast</td>
<td>MBMS</td>
<td>eMBMS</td>
</tr>
<tr>
<td>V2I support</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>V2V support</td>
<td>Native (ad hoc)</td>
<td>Native (ad hoc)</td>
<td>No</td>
<td>D2D</td>
</tr>
<tr>
<td>Market penetration</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

3GPP approved new study item on “Study on LTE support for V2X services” for Release 14 study item. The objective of this study is to identify use cases and associated potential requirements for LTE support of V2X services taking into account V2X services and parameters defined in other SDOs (e.g. GSMA Connected Car, ETSI ITS (Intelligent Transportation System), US SAE) or related governmental agency (e.g. C-ITS project in Korean Ministry of Land, Transport and Maritime Affairs)\[^5\]. This study was approved at the SA plenary #70, December 2015. Final outcome is TR22.885 v2.0.0, and twenty-seven use cases over three types of V2X (V2V, V2I/N and V2P) covering both safety and non-safety aspects have been studied. Based on the study result, SA approved normative work items at the same meeting \[^6\]. In addition, separate study on V2V communication in RAN1 is on-going, and new work item “Support for V2V services based on LTE sidelink “ was approved at the RAN #70 meeting \[^7\].

### 4.3.5.3. Research Direction

The technical research areas on vehicular communication are V2V, eMBMS for high speed vehicle, low latency communication for high throughput service. The safety service is the major
service for V2X, and the information is categorized as the traffic information generated periodically within wide coverage and the alert information which is available eventually and locally. The former includes the information such as speed limits, optimal speed, traffic condition etc. while the latter includes the information such as construction info, emergency stop and forward collision information. These kind of information can be transmitted both directly by V2V, forwarded by V2I, or by multicast and broadcast based on the network deployment scenarios.

Though D2D is standardized at 3GPP release 12, the standard supports for limited usage scenarios with low mobility and specific application. V2V requires immediate high data rate transmission, fast link setup and low latency transport should be guaranteed.

In addition, priority based fast and reliable scheduling for critical information should be supported. 3GPP is considering to initiate the study on mission critical data application to support the usage scenarios.

eMBMS is one of the important building blocks of V2X service, and wide area information (even local information) which is common to the vehicles in a certain area, can be delivered efficiently via multicast and broadcast manner. Dynamic group management and location-based service are important research area for enhanced MBMS for V2X services. Table 4.3-3 shows the V2X-specific research areas.

### Table 4.3-3 An Example of the main research topics and the considerations to support V2X [4]

<table>
<thead>
<tr>
<th>Enabling features</th>
<th>Key issues</th>
<th>Expected benefits</th>
</tr>
</thead>
</table>
| MBMS             | • Design of lightweight joining/leaving procedures for dynamic groups of vehicles  
|                   | • Backend server role, task and deployment mode definition to support geo-addressing | Efficient CAMs/DENMs dissemination |
| Scheduling       | • Proper mapping of vehicular traffic patterns to existing QCI and/or new QCI definition  
|                   | • Cross-layer scheduling algorithms accounting for mobility and vehicular communication patterns | QoS support and differentiation  
|                   |                                           | No penalization for non-vehicular applications |
| D2D              | • Radio resource management policies to minimize interference in mobility conditions  
|                   | • Mode selection for D2D communication | Localized V2V communications (e.g., CAMs) support eNodeB offloading |
Though V2V could support low latency service, in order to provide the critical information via more reliable communication, fast forwarding through the network (V2I2V: Vehicle-to-infrastructure-to-vehicle) or local distribution based on mobile edge computing (MEC) are considered as a very important deployment options.

V2I2V is an alternative solution of the V2V link with requiring higher reliability, however, it causes longer transmission delay. Dedicated core network functions for vehicular traffic is beneficial to reduce the latency.

MEC is useful to support various V2X services. Based on the local application/contents servers deployed in the local cloud, base station can distribute the traffic, map information to the vehicles. The traffic information and other information reported by vehicles and sensors are gathered and processed at the local MEC server, the information such as road safety, traffic, high resolution map data can be provided to the vehicles. Such information requires high volume of storage, but they are valid only for the limited areas. Thus, it is very efficient to be updated and be processed locally. This approach provides low latency service with higher reliability by integrating eMBMS or single cell point-to-multipoint (SC-PTM) technologies. Figure 4.3-25 shows the example scenarios of V2X service based on MEC. Nokia Networks demonstrates the low latency V2X service based on their own MEC solution [2].

<table>
<thead>
<tr>
<th>Enabling features</th>
<th>Key issues</th>
<th>Expected benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MTC</strong></td>
<td>• Efficient transmission of small amounts of data with minimal network impact</td>
<td>Easier management of some ITS applications (like FCD)</td>
</tr>
<tr>
<td><strong>Enhanced device</strong></td>
<td>• Powered with the vehicle’s battery • Multi-interface platforms</td>
<td>Battery saving • Flexibility offered by multi-technology communications</td>
</tr>
<tr>
<td><strong>Business models</strong></td>
<td>• Incentive-based approaches • Value-added services provisioning</td>
<td>Larger subscribers basin • Higher return-on-investments</td>
</tr>
<tr>
<td><strong>Standardization</strong></td>
<td>• Harmonized MTC and ITS standardization activities • LTE role in ITS reference architecture</td>
<td>Enhanced functionalities/architectures • New use cases and synergic solutions</td>
</tr>
</tbody>
</table>
Figure 4.3-25 MEC-based V2X operation

References

[1] 3GPP TR22.885 v1.0.0 “Study on LTE Support for V2X Services”, September 2015
[2] Nokia Networks, “LTE-Advanced Evolution in Releases 12 - 14 New services to pave the way to 5G”,
[5] SP-150051, “New WID for Study on LTE support for V2X services (FS_V2XLTE)”, SA#67,
    March 2015
[6] SP-150573, “New WID for Study on LTE support for V2X services (V2XLTE)”, SA#70,
    December 2015
    December 2015
4.3.6. Network for IoT

4.3.6.1. General Overview

Internet of Things, IoT, is a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies[1]. The IoT is to provide all kinds of application services with security and privacy requirements satisfied, which uses all the aspects of things through the identification, data gathering, processing, and communication functions. As such IoT technology, advanced sensors and actuators together with Machine-to-Machine communication, M2M, or Machine-type communication, MTC, data-mining, network automation, decision making based on policy, security, privacy, and cloud computing technology are expected to utilized.

Typical network architecture model for the IoT is exemplified in the Figure 4.3-26, which consists of 3GPP network, MTC user equipment, UE, MTC server, and MTC application [2,3]. The MTC server can locate either within 3GPP network or outside depending on the service scenarios.

![Figure 4.3-26 IoT deployment scenarios with Cellular network](image-url)
The MTC device is a kind of UE to be connected to a 3GPP network and to be newly defined to meet the MTC service requirement. In coming 5G era, massive number of IoT devices are expected to connect to network, and thus the MTC device should be designed to reduce its cost considerably, reduce power consumption quite a lot, and enhance its coverage.

There also exists a complementary way to connect the MTC devices to a local access network, so called capillary network, through a MTC gateway device that is a kind of MTC device having 3GPP mobile communication capability. This scenario is useful when the MTC devices are located within building or basement where operations of such MTC devices connecting directly to the 3GPP network are very challenging. The capillary network can provide connection efficiently for the specific local area with the short range communication technologies, such as IEEE 802.15, Zigbee, Bluetooth, etc[4,5].

4.3.6.2. Standardization and Service Requirements

With consensus on that the MTC is an important revenue stream for operators and has a huge potential from the operator perspective, 3GPP has been studying on “provision of low-cost MTC UEs based on LTE” as a work item of Release 12. The Release 12 work item specified a low complexity LTE device for MTC of which bill of material cost is approaching that of an EGPRS modem though a combination of complexity reduction techniques. But, the results of the technical report TR 36.888[6] showed that further complexity reduction can possible. TR 36.888 technical report also showed potential coverage improvement of 15 – 20 dB for both FDD and TDD comparing with normal LTE UE. This can help to support use cases in which MTC UEs are deployed in challenging environment such as inside building by compensating gain loss with various complexity reduction techniques. However the work item “Low cost & enhanced coverage MTC UE for LTE”, even though it progress significantly, was removed from Release 12 scope and continued in Release 13 due to time limitation.

Regarding MTC, the work item objectives in Release 13 are as follows[7].

- Specify a new low complexity UE category/type for MTC operation in any LTE duplex mode (full duplex FDD, half duplex FDD, TDD)
- Reduced UE bandwidth of 1.4 MHz in downlink and uplink
- Reduced maximum transmit power
- Reduced support for downlink transmission modes
- Further UE processing relaxations can also be considered within this work item: Reduced maximum transport block size for unicast and/or broadcast signaling, Reduced support for simultaneous reception of multiple transmissions, Relaxed transmit and/or receive EVM requirement including restricted modulation scheme, Reduced physical data channel processing (e.g. relaxed downlink HARQ time line or reduced number of HARQ processes), Reduced support for CQI/CSI reporting modes

• Target a relative LTE coverage improvement – corresponding to 15 dB for FDD – for the UE category/type defined above and other UEs operating delay tolerant MTC applications with respect to their respective nominal coverage
  - Subframe bundling techniques with HARQ for physical data channels (PDSCH, PUSCH)
  - Elimination of use of control channels (e.g. PCFICH, PDCCH)
  - Repetition techniques for control channels (e.g. PBCH, PRACH, (E)PDCCH)
  - Either elimination or repetition techniques (e.g. PBCH, PHICH, PUCCH)
  - Uplink PSD boosting with smaller granularity than 1 PRB
  - Resource allocation using EPDCCH with cross-subframe scheduling and repetition (EPDCCH-less operation can also be considered)
  - New physical channel formats with repetition for SIB/RAR/Paging
  - A new SIB for bandwidth reduced and/or coverage enhanced UEs
  - Increased reference symbol density and frequency hopping techniques
  - Relaxed “probability of missed detection” for PRACH and initial UE system acquisition time for PSS/SSS/PBCH/SIBs can be considered as long as the UE power consumption impact can be kept on a reasonable level.
  - The amount of coverage enhancement should be configurable per cell and/or per UE and/or per channel and/or group of channels. Relevant UE measurements and reporting to support this functionality should be defined.

• Provide power consumption reduction for the UE category/type defined above, both in normal coverage and enhanced coverage, to target ultra-long battery life:
  - When defining the detailed solutions for the Release 13 low complexity UEs and the solutions for coverage enhanced UEs, strive to reduce active transmit/receive time. (e.g.,
minimizing the required number of repetitions by minimizing sizes of control messages).
- Modification, including redesign, addition or removal, of signals/channels can be considered if this can achieve significant power consumption reduction.
- Reduction of measurement time, measurement reporting, feedback signalling, system information acquisition, and synchronization acquisition time etc., can be considered if this can achieve significant power consumption reduction.

• Provide power consumption reduction for the UE category/type defined above, both in normal coverage and enhanced coverage, to target ultra-long battery life: Half duplex FDD, full duplex FDD, and TDD should be supported but since half duplex operation is particularly beneficial from device complexity and power consumption point of view, the solutions specified within this work item should be optimized for half duplex FDD and TDD.
• Reduced mobility support can be considered if this is needed to fulfil the objectives.
• The agreements and working assumptions made during the initial work carried out during the corresponding Release 12 work item should be used as a starting point when applicable.

As an one of work item for enhancement for MTC in Release 12, MTC capillary network where a MTC device acts as a gateway for other devices [4,5]. Figure 4.3-25 illustrates such MTC gateway in capillary network. The MTC gateway device is a kind of MTC UE that has a functionality to communicate with 3GPP mobile network. The local access devices are to be connected to short range wireless communication such as IEEE 802.15, Zigbee, Bluetooth, etc. The MTC gateway device acts as an agent for the local access devices in the capillary network and hence the local devices are not visible by the operator network directly. The MTC Gateway Device performs procedures such as authentication, authorization, registration, management and provisioning for the Local-Access Devices connected to it using local connectivity mechanisms. Depending on use case scenario, the MTC gateway can also connect with MTC devices that have mobile communication capability as well as local devices that do not have 3GPP mobile communication capability as shown in Figure 4.3-27. The MTC server can be located within or outside operator network depending on use case scenarios.
The service requirements for MTC are released in TS 22.368 as a stage 1 specification document in Release 13 [8]. The service optimized for MTC is different from that optimized for human-to-human communication. Compared to human-to-human communication, the MTC has differences with the followings:

- Different market scenarios
- Data communications
- Lower costs and effort
- A potentially very large number of communicating terminals with
- To a large extent, little traffic per terminal

Since MTC applications have quite different characteristics from each other, the system optimization is quite differently designed for every MTC application. Thus, MTC Features are defined to provide structure for the different system optimization possibilities that can be invoked. MTC Features can be individually activated. MTC Features have been defined as follows:
• Low Mobility;
• Time Controlled;
• Small Data Transmissions;
• Infrequent Mobile Terminated;
• MTC Monitoring;
• Secure Connection;
• Group Based MTC Features: (Group Based Policing; Group Based Addressing)

References

[1] ITU-T Y.2060 (06/2012), Overview of the Internet of things,
[2] 3GPP TR 23.888 v11.0.0 (2012-09) System improvements for Machine-Type Communications
[4] 3GPP TR 22.888 v12.0.0 (2013-03) Study on enhancements for Machine-Type Communications
[6] 3GPP TR36.888 v12.0.0, Study on provision of low cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE (Release 12), June 2013
[7] 3GPP web-site, “Evolution of LTE in Release 13” February 18,205, See also RP-141865
[8] 3GPP TS 22.368 v13.1.0 (2014-12), Service requirements for Machine-Type Communications (MTC) Stage 1 (release 13)
4.3.7. Network Technologies for Massive connectivity

4.3.7.1. General Overview
As a future service oriented network, besides today's human centric communication, 5G network is expected to connect machine-type devices. When a car sensor or medical device that sends information to a database in real time, the data is analyzed to improve fuel efficiencies or enhance patient care. Image numerous sensors or actuators weaved in your clothes, worn around your body, distributed in your immediately environment such as car and room, and in further distance such as next block, the city or even the earth. The connection, transportation, storage and processing of all sensor data open up opportunities to create completely new business model. This is the new IoT (Internet of Things) reality.
Cisco, Nokia, Ericsson and Huawei each believe that the connected Things will rise sharply over the next 5-10 years, meaning there could be as many as 100 billion connected Things. In a similar vein, GSMA predicts the total figure for connected Things could grow to 24 billion by 2020 whilst Gartner forecasts that number at 30 billion. Those connections will span wearable, automotive, intelligent building, metering, smart city, healthcare and consumer electronics applications; the exact number is not easy to determine, but all analysts agree it will be huge. IoT is expected to create added value of about $ 1.9 trillion. The 5G network should continue evolution towards more efficient mobile broadband support; whilst also providing the connectivity that the IoT needs. In the final evolutionary stage, the previously unconnected things, processes, humans and date generated converge by their nature. This is the heart of IoT. Through relevant analytics, the data generated from humans and things can be turned into data-driven applications and services to create richer experiences, and incredible economic opportunity.

4.3.7.2. State of the Art and Research Trend
The traffic considered by massive IoT has distinct features from human centric traffic. Firstly, IoT applications, such as alarm, protection, and real-time control, require much stringent network access latency, which could be less than a few milliseconds. Secondly, the IoT traffic packets are typically small in size, sporadically, and uplink dominated. Thirdly, some IOT devices behave in a group manner, i.e. a large amount of devices will access the network
simultaneously. This asymmetric nature of traffic in conjunction with sporadic patterns of IoT applications motivates studies of an efficient random access scheme with the massive number of devices. Finally, the IoT devices are usually constraint in size, battery, memory capacity, processing capability etc. This adjure a multiple access scheme to support a low-duty cycle operation to prolong the battery lifetime and lightweight security and management mechanism.

More specific, the requirements for the IoT connectivity are as follows:

• Support a very large number of low-cost IoT devices per cell that can be regarded as disposable;
• Support very low power consumption allowing operation for a decade or more on small batteries;
• Support deep coverage of low-rate services into highly-shadowed locations;
• Support stringent latency and reliability requirement, in particular for safety-critical applications and industrial control applications;
• Optimized for small payloads, with low signaling and control overhead;
• Deployment friendly with MBB services to provide global coverage and allow controlled QoS.
• Support for navigation and location tracking;

Various solutions have been proposed to connect billions IoT devices wirelessly to the cloud. They can be generally classified as the one-hop solution and the multi-hop solution through relay or gateway nodes in the link for the purpose of improving coverage. The network architecture is shown in the Figure 4.3-29.
The one-hop solutions can be further divided into three classes: licensed-spectrum based, unlicensed spectrum based, and white space based. The licensed spectrum guarantees controlled interference management and allows operators to provide good QoS services. And as a proofed technology, licensed spectrum based solution provide continuous coverage of a large area.

To solve the congestion induced by burst IoT traffic, ACB (Access Class Barring) only allows part of devices to initiate the random access. High priority services are guaranteed by the 16 different access classes. EAB (Extended Access Barring) is a step forward by not permitting devices with delay tolerant applications to access network totally. A new UE category 0 is introduced for low cost MTC (LC-MTC) in 3GPP to achieve 49%~52% cost saving and at least 15dB coverage improvement based on techniques of:

- Reduction of peak rate with maximal bandwidth of 1.4MHz, Max TBS =1000 for unicast traffic, and Max TBS=2216 for broadcast traffic;
- A single receive antenna;
- Half Duplex-FDD;

3GPP RAN2 makes more effort on issues about LC-MTC UE capability indication, enhancements to cell reselection/handover, parallel reception, and HD-FDD impacts. UE Power Consumption Optimizations and Signaling Overhead Reduction were considered. The benefit of LC-MTC lies in the efficient utilization of existing LTE spectrum resources.

In 3GPP GERAN, a clean-slate IoT solution is under study to save more power consumption.
and cost for IoT devices, and more than 20dB coverage improvement without hardware impact on the base station and possible reusing existing core network. The key features are as follows:

- A narrow 200 kHz bandwidth operating in the refarmed GSM spectrum or the guard bands of LTE carrier;
- A single carrier modulation with frequency and time division multiple access for downlink (15 kHz carrier spacing) and uplink (5 kHz). Uplink channel bonding is used to provide higher data rates.

Some proprietary solutions over unlicensed spectrum below 1 GHz also adopt cellular network architecture. An ETSI LTN (Low Throughput Network) ISM was created to conduct standardization in 2012, including SigFox, Semtech, Actility, HP, Covéa Technologie as founding members. The terminal transmits ultra narrow band signal (from about 100Hz several KHz) to the eNB. Due to the limited signal bandwidth, the receiver sensitivity is very good because of small noise bandwidth. Combining a very large spreading factor, good coverage can be obtained. However, due to the low data rate, high accuracy TCXO is required for to achieve large MCL (Maximum Coupling Loss). Though the unlicensed spectrum is free of usage, the severe interference makes it hard for QoS guarantee. Most enterprise networks work over unlicensed spectrum for the obvious economic reason with interference tolerance technologies. It is similar for the white space spectrum solutions, which has different permitted transmit power density from the unlicensed solution. Though larger transmit power is allowed in white space, the interference increases correspondingly.

Typical multi-hop solutions are the two-hop ones. The IoT device is connected to network via a relay or gateway. The first hop can adopt ZigBee (IEEE 802.15.4) or WiFi (IEEE 802.11 b/ah) technologies in unlicensed spectrum and D2D or relay technologies in licensed spectrum. Zigbee and WiFi work in below 1 GHz, such as the 868MHz/915MHz ISM bands. The receiver sensitivity of IEEE 802.11ah is 10 dB lower than that of legacy WiFi thanks for narrower bandwidth (mandatory 1 MHz and 2 MHz) and lower data rate. The IEEE 802.11ah adopts low PAPR waveform, long sleeping duration and header compression etc for low power consumption. The second hop can be either cellular or wired connection such as ADSL.

The coverage of IoT devices can be improved through D2D (device-to-device) communication by relaying in the uplink direction. In 3GPP, there are two main design directions the
proximity services based on D2D. In network-assisted D2D, the network performs all the decisions in regards to resource sharing mode selection, power control, scheduling, selection of transmission format. While in D2D with minimal network assistance, the network provides only synchronization signals to the devices. However, the D2D are not optimized for IoT device, particularly when the wearable IoT devices are considered. Wearable are focusing on fitness and wellness, healthcare and medical, infotainment application and markets. In the long term future, wearable are expected to be a completely new product category that is trying to incorporate sensors and computing on an individual's body, like televisions and feature phones. The same as IoT devices, the wearable are usually constraint in size, battery, memory capacity, processing capability. They are difference in that wearable are usually accompanied with smart phone in up to 85% time. The rest are stand-alone time is usually inconvenient to carry a cell phone or unavailable, for example during swimming, jogging and running. Wearable are more suitable for children and older than cell phone. Compared with other IoT device, wearable are particularly sensitive to security and privacy, and power consumption. The extreme in size and volume of wearable put a hard limitation on battery capacity, which is usually hundreds mAh, compared with thousands mAh battery for today's smart phone.

4.3.7.3. Candidate technologies
As mentioned in the previous section, massive IoT access is quite different from the classical mobile broadband design which mainly focuses on increasing data rate and system capacity. The key features of 5G IoT network can be concluded into four aspects: ubiquitous coverage, low terminal cost, extremely low terminal power consumption, and ultra-reliability and short latency. Usually massive connection IoT services do not require large channel bandwidth. It is nature to consider a narrow band system design. The narrow band device are well known for low complexity and low cost. Applying transmission power for within a narrow band, network coverage and connection capability can be improved. The new filtered waveform technologies, e.g. F-OFDM (filtered-OFDM) and FBMC (filter bank multi-carrier) can lower out-of-band interferences, and utilize non-contiguous and fragmented bands efficiently, as well as decouple technical solutions between different subband. This feature also helps to accommodate different categories of IoT services and even the human centric service. By supporting multiuser
superposition transmission, multiple access technologies, such as SCMA (sparse coding multiple access), allows grant-free transmission. The benefits include simplified signaling process, reduced power consumption, and short access latency. Moreover, the coding, modulation, multiple access, and signaling process in different subbands can be configured independently. A low cost and low power D2D communications is envisioned to facilitate the large scale deployment of IoT services. By borrowing battery and resource from the relaying smart phone, the wearable device can be extremely simple and small. All the network management, control security and even charge can go to the accompanied smart phone.

For low-latency high-reliability IoT services, new waveforms supporting shorter frame length and TTI (Transmission Time Interval) needs to be considered. The short TTI may cause larger CP overhead, the supporting non-CP and CP sharing between multiple symbols can be utilized. Again, the grant-free transmission of SCMA can achieve “zero” waiting time for uplink traffic. The network forwarding delay can be reduced through sinking some core network functions to access network and localizing service contents. To improve transmission reliability, technologies such as advanced coding and space/time/frequency diversity, and multiple connections can be utilized. Ultra-dense network (UDN) deployment with various small cells in 5G will have beneficial effect on alleviating congestion and overloading of the massive IoT connections.

References

In this Chapter, key enabling technologies for 5G are described in aspects of wireless networks. Table 3.2-1 describes how the technologies contribute to the 5G requirements that are introduced in Chapter 3.

<table>
<thead>
<tr>
<th>Category</th>
<th>Enabling Technologies</th>
<th>5G RAN Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wide and Flexible Bandwidth Technology</td>
<td>Millimeter-wave Band</td>
<td>✓  ✓  ✓</td>
</tr>
<tr>
<td></td>
<td>Heterogeneous Multi-RAT Integration</td>
<td>✓  ✓  ✓</td>
</tr>
<tr>
<td></td>
<td>Cognitive Radio and Spectrum Sharing</td>
<td>✓  ✓</td>
</tr>
<tr>
<td>Advanced Modulation and Coding</td>
<td>Advanced Modulation: FQAM</td>
<td>✓  ✓</td>
</tr>
<tr>
<td></td>
<td>Advanced Channel Coding</td>
<td>✓  ✓  ✓</td>
</tr>
<tr>
<td>Duplexing</td>
<td>In-band Full Duplexing</td>
<td>✓  ✓  ✓</td>
</tr>
<tr>
<td></td>
<td>Dynamic TDD</td>
<td>✓  ✓  ✓</td>
</tr>
<tr>
<td>Multiple Access and Waveform</td>
<td>Multiple Access: NOMA</td>
<td>✓  ✓</td>
</tr>
<tr>
<td></td>
<td>SCMA</td>
<td>✓  ✓</td>
</tr>
<tr>
<td></td>
<td>New Waveform</td>
<td>✓  ✓  ✓</td>
</tr>
<tr>
<td>Large Scale Antenna</td>
<td>Large-scale Antenna below 6GHz</td>
<td>✓  ✓  ✓</td>
</tr>
<tr>
<td></td>
<td>Large-scale Antenna above 6GHz</td>
<td>✓  ✓  ✓</td>
</tr>
<tr>
<td>Advanced Interference Management</td>
<td>Advanced Receiver for Simultaneous Non-Unique Decoding (SND)</td>
<td>✓  ✓</td>
</tr>
<tr>
<td></td>
<td>Sliding Window Superposition Coding (SWSC)</td>
<td>✓  ✓</td>
</tr>
<tr>
<td>Access Architecture-related Radio</td>
<td>Enhanced Wireless Backhaul</td>
<td>✓  ✓</td>
</tr>
<tr>
<td>Technologies</td>
<td>Moving Network</td>
<td>✓  ✓</td>
</tr>
<tr>
<td></td>
<td>Device-to-Device (D2D) Communication</td>
<td>✓  ✓</td>
</tr>
<tr>
<td></td>
<td>Edgeless Cellular Network</td>
<td>✓  ✓  ✓</td>
</tr>
</tbody>
</table>
5.1. Wide and Flexible Bandwidth Technology

5.1.1. Millimeter-wave Band

5.1.1.1. General Overview

As one of the most innovative and effective solutions to realize the 5G vision and requirements, the use of large chunks of underutilized spectrum in the very high frequencies such as the millimeter-wave (mmWave) bands has recently gained significant interest. Traditionally, due to the high propagation loss and lack of cost effective components, among other reasons, these frequencies have mostly been utilized for outdoor point-to-point backhaul links or for carrying high resolution multimedia streams for indoor applications, but not for cellular access links. In order to refarm these underutilized spectrums for future outdoor cellular applications, two key hurdles must be overcome, i.e., sufficiently large geographical coverage and support for mobility even in Non-Line-of-Sight (NLOS) environments where the direct communications path between the transmitter and the receiver is blocked by obstacles.

5.1.1.2. State of the Art and Research Trend

There have been concerns with utilizing mmWave frequency bands for mobile cellular communications. Some of these concerns regarding the propagation characteristics at higher frequencies such as higher penetration, precipitation and foliage losses are legitimate even though the actual amounts of additional propagation losses vary depending on the material of the building, the strength of rain, or the thickness of foliage. The most common misunderstanding, however, on the propagation characteristics at higher frequencies is that they always incur a
much higher propagation loss even in free space compared with lower frequencies, and thus not adequate for long range communications. To clarify this misunderstanding, let us start with the Friis transmission equation,

$$P_r = P_t + G_t + G_r + 20 \log\left(\frac{c}{4\pi RF}\right) \text{ [dB]}$$

where $P_r$ is the receive power in unobstructed free space, $P_t$ is the transmit power, $G_t$ and $G_r$ are the transmit and receive antenna gains, respectively, $R$ is the distance between the transmitter and receiver in meters, $f$ is the carrier frequency, and $c$ is the speed of light. The received power can be easily seen as inversely proportional to the frequency squared when an ideal isotropic radiator ($G_t = 1$) and an ideal isotropic receiver ($G_r = 1$) are used at each ends. In reality, however, antennas or an array of antennas with antenna gains of $G_t$ and $G_r$ greater than unity are typically employed at both ends, and the antenna gains are proportional to the frequency squared given a fixed physical aperture size [8]. Given the same physical aperture size, therefore, transmit and receive antennas at higher frequencies, in fact, send and receive more energy through narrower directed beams, which is not commonly recognized [9]. In order to verify this, measurements have been conducted in an anechoic chamber using two antennas supporting 3 GHz and 30 GHz, respectively, as shown in Figure 5.1-1.

*Figure 5.1-1 Results of verification measurements of propagation loss predicted by Friis equation*
A patch antenna at 3 GHz and an array antenna at 30 GHz of the same physical size were designed for this measurement and placed within an anechoic chamber at each communication ends. As expected from the Friis equation and the argument above, results in Figure 5.1-1 show the same amount of propagation loss regardless of the operating frequency when an array antenna of the same physical aperture size is used at the 30 GHz receiving end. In addition, when array antennas are used at both transmitting and receiving ends at 30 GHz, the measured receive power is 20 dB higher than that of the 3 GHz patch antenna case. Along with the aforementioned lab measurements, there have been recent studies regarding the outdoor channel propagation characteristics which have shown the potential for utilizing higher frequency bands for cellular communications [10-14]. In [10] and [11], outdoor channel measurements were carried out at 38 GHz and 28 GHz, respectively, on the campus of University of Texas at Austin. Another channel measurement campaign was conducted at 28 GHz to produce measurement data for a sub-urban environment at Samsung Electronics site in Suwon, Korea [3][12]. In addition, investigation of the channel characteristics in a dense urban environment was made in Manhattan, New York [13-14]. And measurements and analysis of the channel characteristics in urban micro environments was made in Daejeon, Korea [16]. In [17], thanks to ray-tracing simulations, analysis of the channel characteristics in urban micro environments (Daejeon urban area and Manhattan downtown) was made. All these channel measurements were carried out at 28 GHz instead of 60 GHz and E-Band with many aspects considered including regional regulatory status and availability of significant amount of licensed spectrum etc.

In [16], the measurable path loss is around 140 dB and 11 different channels are sampled to omni-directional PDPs and to calculate the pathloss in Daejeon street canyon, UMi scenario. In UMi channel characteristics, the LOS path loss in the bands of interest appears to follow Friis' free space path loss model quite well. Just as in lower bands, a higher path loss slope (or path loss exponent) is observed in NLOS conditions. The shadow fading in the measurements appears to be similar to lower frequency bands, while ray-tracing results show a much higher shadow fading (>10 dB) than measurements, due to the larger dynamic range allowed in some ray tracing experiments.

Recently, joint work of 5G channel model up to 100 GHz collaborating with industry and academia was published to whitepaper [18]. Following the channel model and characteristics
in [18], there are measurements of the delay spread above 6 GHz which indicate somewhat smaller ranges as the frequency increases, and some measurements show the millimeter wave omnidirectional channel to be highly directional in nature.

The 5G channel model in [18] is extended to mmWave channel up to 100 GHz based on the current 3D-SCM framework [15] including additional features considered for mmWave spectrums. In the double-directional channel model, the multipath components are described by the delays and the directions of departure and the direction of arrival. Each multipath component is scaled with a complex amplitude gain. Then the double-directional channel impulse response is composed of the sum of the generated double-directional multipath components. The double-directional channel model provides a complete omni-directional statistical spatial channel model (SSCM) for both LOS and NLOS scenarios in UMi environment. These results are currently analyzed based on the ray-tracing results, which is compared with the measurement campaign done in the same urban area. The final results will be derived from both the measurement and ray-tracing results.

For fast-fading modeling, ray-tracing based method is useful to extend the sparse empirical datasets and to analyze the channel characteristics in both outdoor and indoor environments. Ray-tracing results provide full information in spatio-temporal domain, which can be extracted to parameterize the double-directional channel model. Current preliminary large-scale parameters for small-scale fading model in UMi are analyzed using ray-tracing results on UMi street-canyon area in Daejeon, Korea shown in Figure 5.1-2, which models the same area conducted the measurement campaign [16].

Figure 5.1-2 Daejeon street canyon environments conducted measurement and the ray-tracing
The channel parameters in delay and angular domains are extracted from the ray tracing simulation. The angular spread of the mmWave channel is calculated by the elevation and azimuth angle spreads. The values of delay spread in the 28 GHz band are smaller than the values of delay spread in the conventional cellular band. This is mainly caused by the propagation characteristic of the mmWave band in a lower-scattering environment, where paths that involve effects like multiple diffractions and penetrations are more strongly attenuated. The large-scale parameters are extracted and verified that the excess delays at both 28 GHz are exponentially distributed, and azimuth angle of departure (AoD) and azimuth angle of arrival (AoA) follow a Laplacian distribution as reported in [17].

These study results in [18] reveal that the key parameters characterizing the propagation properties of the mmWave bands such as the path loss exponent, delay spreads and angle spreads are comparable to those of typical cellular frequency bands when transmit and receive antennas are used to produce beamforming gains. For instance, transmission links were found to be established for a distance of up to 10-950 meters with path loss exponents in the range of 2.89-4.14 for NLoS and 1.85-1.98 for LoS environments in urban micro environment, which are similar to those measured in the traditional cellular bands. Path loss exponents below 2 are frequently observed due to constructive addition of the reflected paths and the direct paths in street corridors or tunnels in LoS environments. While more extensive measurement campaigns are currently being carried out by the authors of [19] in Asia, U.S. and Europe to build a comprehensive statistical mmWave channel propagation model, it is evident that the mmWave bands have strong potentials as a candidate band for next generation cellular services. After the verification of the channel feasibility, the next step is to develop underlying core technologies to most efficiently utilize the abundant spectrum in the mmWave bands and to prove commercial viability.

5.1.1.3. Research Direction

5.1.1.3.1 mmWave Beamforming Algorithm

An appropriate beamforming scheme to focus the transmitted and/or the received signal into a desired direction in order to overcome the unfavorable path loss is one of the key enablers
for cellular communications at mmWave frequency bands. The small wavelengths of mmWave frequencies facilitate the use of a large number of antenna elements in a compact form factor to synthesize highly directional beams corresponding to large array gains.

Depending on the beamforming architecture, the beamforming weights required to form the directive beam could be applied in the digital domain or in the analog domain. Digital beamforming is done in the form of digital precoding that multiplies a particular coefficient to the modulated baseband signal per RF chain. For analog beamforming, on the other hand, complex coefficients are applied to manipulate the RF signals by means of controlling phase shifters and/or Variable Gain Amplifiers (VGAs). When combined with an OFDM system, digital beamforming is carried out on a subcarrier basis before the Inverse Fast Fourier Transform (IFFT) operation at the transmitter and after the FFT operation at the receiver, whereas analog beamforming is performed in time domain after the IFFT operation at the transmitter and before the FFT operation at the receiver.

In general, digital beamforming provides a higher degree of freedom and offers better performance at the expense of increased complexity and cost due to the fact that separate FFT/IFFT blocks (for OFDM systems), Digital-to-Analog Converters (DACs) and Analog-to-Digital Converters (ADCs) are required per each RF chain. Analog beamforming, on the other hand, is a simple and effective method of generating high beamforming gains from a large number of antennas but less flexible compared to digital beamforming. It is this trade-off between flexibility/performance and simplicity that drives the need for hybrid beamforming architectures, especially when a multitude of antennas is required as in the mmWave bands.

The Figure 5.1-3 illustrates a hybrid beamforming architecture applied both at the transmitter and the receiver. In this architecture, the sharp beams formed with analog beamforming (phase shifters) compensate for the large path loss at mmWave bands and digital beamforming provides the necessary flexibility to perform advanced multi-antenna techniques such as multi-beam MIMO.
The simulated performance of the hybrid beamforming architecture in mmWave bands are presented in [5], where link and system level simulation results are provided with various numbers of transmit/receive antennas and RF chains. Using a 500 MHz bandwidth at 28 GHz, [5] presents some notable results for the hybrid beamforming system including an 8 dB gain over the conventional spatial multiplexing scheme and an 8 Gbps average sector throughput with 16 antennas with 4 RF chains at the base station and 8 antennas with a single RF chain at the mobile station.

### 5.1.1.3.2. mmWave Beamforming Prototype

The main purposes of the mmWave prototype are to check the feasibilities of mmWave bands for sufficiently large geographical coverage for cellular services and the support for mobility even in NLoS environments. In this section, as an example, we present a detailed description of the mmWave beamforming prototype developed and tested at the DMC R&D Center, Samsung Electronics, Korea.

An mmWave adaptive beamforming prototype was developed including RF units, array antennas, baseband modems, and Diagnostic Monitor (DM) as shown in Figure 5.1-4.
Both transmit and receive array antennas have two channels and each comprises 32 antenna elements arranged in a form of Uniform Planar Array (UPA) with 8 horizontal and 4 vertical elements, confined within an area of 60 mm by 30 mm. This small footprint was made possible due to the short wavelength of the carrier frequency at 27.925 GHz. Two channels at transmit and receive array antennas are designed to support various multi-antenna schemes such as MIMO and diversity. The array antenna is connected to the RF unit that contains a set of phase shifters, mixers and related RF circuitry. The set of phase shifters control the phases of the signals sent to the antennas to form a desired beam pattern. Therefore, by setting the phase shifter values to a particular set, transmit and receive array antennas are capable of forming a sharp beam pattern in the intended horizontal (azimuth) and vertical (elevation) angles.

In order to reduce the hardware complexity, a sub-array architecture was employed to group 8 antennas into a sub-array, thus requiring only 4 RF units per channel instead of 32. The reduction in the number of RF paths results in a reduction of antenna gain at the desired angle (except antenna boresight), a reduction of beam scanning ranges and an increase in the side lobe levels, but still meets the overall beamforming requirements. The resulting Full Width at Half Maximum (FWHM) of the beam at the antenna boresight is approximately 10°.
horizontally and 20° vertically with the overall beamforming gain of 18 dBi. In addition, a set of beam patterns is predefined to reduce the feedback overhead required for the adaptive beamforming operation between the transmitter and the receiver, where the overlapped beam patterns cover the intended service area with a unique beam identifier (ID) for each beam. These beam IDs are used by the baseband modem to control the phase shifter weights and to feedback the preferred transmission beam information back to the transmitter.

The baseband modem shown in Figure 5.1-4 was designed and implemented for real time operation with commercial off-the-shelf signal processing units including Xilinx Virtex-6 Field Programmable Gate Arrays (FPGAs), and ADC and DAC each with up to 1 Giga-samples/sec (Gsps) conversion rate. The analog signal ports of the modem Analog Front-End (AFE) are connected to the RF/antenna input (output) port to transmit (receive) the complex analog baseband signal. Furthermore, the baseband modem is linked to a DM program developed to visualize the operational status of the system and to collect system statistics including data throughput, packet error rates, transmit/receive beam IDs, received signal constellations and signal strengths etc.

Two sets of mmWave beamforming prototype as specified above were built, playing the roles of a base station and a mobile station, and various lab and field tests in both indoor and outdoor environments were performed. For the downlink transmission, the base station periodically transmits a sequence of beam measurement signals in predefined beams so that the mobile station can carry out, also in predefined receive beams, channel quality measurements of the transmit-receive beam pairs and thus select the best beam pair for data transmissions. The selected base station transmit beam ID is fed back to the base station for the subsequent downlink transmission until the next update incident. In this fashion, the base and the mobile stations quickly establish the wireless communications link and adaptively sustain the link even in high mobility condition. The communications link setup for the uplink is done in an analogous way where the roles of the base station and the mobile stations are interchanged. The developed mmWave beamforming prototype was designed to complete the search for the best transmit and receive beam pair within 45 msec.
5.1.1.3.3. **mmWave Test Results**

Using the mmWave adaptive beamforming prototype described in the previous section, rather comprehensive indoor and outdoor field tests were carried.

An aggregated peak data rate of 1.056 Gbps was achieved in the lab with negligible packet error using two channels at the base station supporting two stationary mobile stations with 528 Mbps each. In an outdoor range test in an LoS environment, the communication range with negligible errors (Block Error Rates (BLER) of less than 10\(^{-6}\) with the block size of 672 bits) was verified up to 1.7 km with transmission power headroom of 10dB still left over [20]. The 1.7 km limit was due to spectrum license issues and the authors are confident that much longer ranges are in fact possible.

In addition, outdoor coverage tests were conducted to demonstrate the service availability in a typical outdoor environment for both LoS and NLoS sites. The tests were performed at sites surrounded by tall buildings where various channel propagation effects such as reflection, diffraction, or penetration are expected to take place, as shown in Figure 5.1-5.

![Figure 5.1-5 Outdoor coverage test results of mmWave beamforming prototype](image)

As can be seen from the test results in Figure 5.1-5, satisfactory communications links were discovered even in NLoS sites more than 200 meters away, mostly due to reflections off neighboring buildings. On the other hand, there were a few locations where a proper link
could not be established, i.e., coverage holes, which necessitate solutions for coverage improvement such as optimized cell deployment, inter-cell co-ordination, relays, or repeaters. Considering one of the important operation scenarios in practical cellular networks, communication between an outdoor base station and an indoor mobile station was also investigated. The test results, shown in Figure 5.1-6, presents link qualities between an outdoor base station to an indoor mobile station placed inside one of typical modern office buildings with heavily tinted glass with more than 150 meter separation. These types of buildings are representative of presenting highly unfavorable propagation (penetration) conditions even for current cellular frequency bands below 6 GHz.

![Figure 5.1-6 Outdoor to Indoor penetration test results of mmWave beamforming prototype](image)

As can be seen in Figure 5.1-6, surprisingly amicable indoor coverage results were obtained with only the totally obstructed, farthest side of the building resulting in lost connections. While the spots showing block error rates around 10~20 % can be improved with conventional error correction schemes such as Hybrid Automatic Repeat reQuest (HARQ) and modulation/coding adaptation schemes, remaining coverage holes would need to be covered with other alternative schemes such as repeaters and indoor femto cells as in traditional cellular systems. Lastly, mobility support was also tested in NLoS setup where the direct path was blocked by a tall building and the mobile station was moving at the speed of 8 km/h in random directions.
The test results were extremely encouraging and resulted in error free transmission at 264 Mbps and less than 1% of block error rate at 528 Mbps transmissions due to the fast adaptive beamforming algorithm running at both communications ends. As mentioned in the previous section, the design capability of the adaptive joint beam searching and switching algorithms implemented in our prototype could easily support mobility higher than 8 km/h. The verification of higher mobility support and hybrid beamforming schemes are currently under way and the results will be published upon completion. We believe that these and ensuing results will provide a firm ground for the development of mmWave beamforming based 5G cellular networks.

References


[18] Aalto University, BUPT, CMCC, Nokia, NTT DOCOMO, New York University, Ericsson, Qualcomm, Huawei, Samsung Electronics, Intel, University of Bristol, KT Corporation, University of Southern California, “5G Channel Model for bands up to100 GHz”, Dec. 6, 2015. (available : http://www.5gworkshops.com/5GCM.html).


5.1.2. Heterogeneous Multi-RAT Integration

5.1.2.1. General Overview
The 5G mobile communication system aims at reaching a data rate comparable to the rate which is provided by the high-speed wired-line. As one of the key technologies for achieving the goal, the carrier aggregation (CA), which is adopted in the 3GPP LTE-Advanced, is expected to be more enhanced in the 5G. The CA will meet new bandwidth-hog services with high spectrum efficiency.

The future network is expected to be heterogeneous: the more small cells will be deployed in a macro cell; and in particular the cellular will be evolved into ‘the heterogeneous radio access networks (h-RATs)’ where the cellular will be trying to make the most of the adjacent RATs such as Wi-Fi, LAA-LTE, and other licensed networks [1]. It is also expected that mobile equipment with the cellular/Wi-Fi dual mode will be ever-increasing and furthermore they will be equipped with more various wireless network interfaces. Although the offloading by Wi-Fi is being used at the present time, the 5G is expected to be evolved into the directions such as the CA between h-RATS and the seamless inter-RAT handover which are able to reinforce the user’s QoS/QoE. That makes heterogeneous multi-RAT integration more pivotal in 5G, particularly for enhanced mobile broadband services such as rich video or mobile cloud.

5.1.2.2. State of the Art and Research Trend
3GPP LTE-Advanced Release 10 allows the carrier aggregation (CA) of up to 5 component carriers (CCs) of which each CC may occupy a maximum bandwidth of 20 MHz, which results in a total aggregated bandwidth of up to 100 MHz. The aggregated carriers can be either in the same band (intra-band CA) or in different bands (inter-band CA). Some examples of world-wide network having shown the willingness to service a CA network are summarized in Table 5.1-1. Release 10 CA has been designed for intra-site CA (for co-located cells), while Release 11 CA also supports multiple uplink timing advances and other enhancements to support inter-site CA (for non-collocated cells). Release 12 introduces aggregation of both TDD and FDD carriers [2-3].
Table 5.1-1 LTE Aggregation Bands [2]

<table>
<thead>
<tr>
<th>Country</th>
<th>Operator</th>
<th>Bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korea</td>
<td>SKT</td>
<td>B1, B3, B5</td>
</tr>
<tr>
<td></td>
<td>KT</td>
<td>B1, B3, B8</td>
</tr>
<tr>
<td></td>
<td>LGU+</td>
<td>B1, B5, B7</td>
</tr>
<tr>
<td>China</td>
<td>CMCC</td>
<td>B39, B40, B41</td>
</tr>
<tr>
<td>Japan</td>
<td>KDDI</td>
<td>B1, B18, B26</td>
</tr>
<tr>
<td></td>
<td>Softbank</td>
<td>B1, B8, B11</td>
</tr>
<tr>
<td></td>
<td>NTT</td>
<td>B1, B19, B21</td>
</tr>
<tr>
<td>USA</td>
<td>AT&amp;T</td>
<td>B2, B4, B13, B17</td>
</tr>
<tr>
<td></td>
<td>Verizon</td>
<td>B2, B4, B13</td>
</tr>
<tr>
<td></td>
<td>Sprint</td>
<td>B25, B41</td>
</tr>
<tr>
<td>Europe</td>
<td>All operators</td>
<td>B3, B7, B20</td>
</tr>
</tbody>
</table>

Cooperation between other wireless access technologies has become the center of special attention as an efficient method to increase the network capacity. The capacity offloading to IEEE 802.11 specification-based wireless local area network (WLAN, so called Wi-Fi) among non-cellular wireless technologies is considered to be an efficient scheme with easy deployment and cost-efficiency [4-6]. Both Release 11 and Release 12 include the offloading policy framework by which they support inter-system mobility and inter-system routing [4]. The policies for inter-system mobility address how the access network is selected by a user equipment (UE) which is able to be connected with only a single network. The inter-system routing policies, instead, describe a way for the UE to route traffic through other wireless interfaces which are able to be accessed by the UE. The above policies are determined by network-basis, in other words, they are given by the operator’s network for UE to follow. In addition, policies are not dependent on the UE profile, but these are specified by the specific area, time zone, and type of contents. In order to investigate the efficiency of 3GPP offloading policy framework called access network discovery and selection function (ANDSF), [5] defines practical, dynamic offloading strategy to utilize limited knowledge which an operator possesses about other networks and executes the performance analysis. In [6], quantitative results of research for performance of offloading the cellular mobile data into Wi-Fi network are shown.
In connection with the above offloading, various researches supporting carrier aggregation between Wi-Fi and LTE link are in progress, and in particular, some efforts to integrate LTE with Wi-Fi were on the move through integrated femtocell and Wi-Fi (IFW) [7-12]. 3GPP standard provides three approaches to offload LTE traffic to Wi-Fi network in IFW network. In local IP access (LIPA) [7], a single UE is allowed to access regional indoor and business network by femtocell without passing the operator’s mobile core network. The second approach is the selected IP traffic offloading (SIPTO) [7] where UE traffic is allowed to directly enter into the internet from femtocell by detouring the operator’s core network. The above LIPA and SIPTO schemes are transparent to UEs and both of them are designed to avoid congestion in the operator’s core network. The third approach is IP flow mobility (IFOM) [8] which is transparent to RAN side instead of UE, unlike LIPA and SIPTO. In this case, UE is able to make a determination of initiating simultaneous connections with both cellular and Wi-Fi AP. In [12], under the expectation that a number of Wi-Fi APs are already deployed and they will be more increasing, it is being highlighted that the replacing APs with Wi-Fi-supporting LTE small cells has the benefits of supporting both Wi-Fi users and cellular users while exploiting the existing Wi-Fi Hotspot backhaul. With the goal of enhancing throughput per user by the Wi-Fi-integrated LTE small cell, it provides dynamic switching and carrier aggregation between licensed bandwidth and unlicensed bandwidth, and it carries out performance analysis by developing an optimization model. Meanwhile, close-integration between LTE and Wi-Fi is scheduled as one of Release 12 research topics [13], and in addition 3GPP starts to see LTE usage in unlicensed spectrum (also known as LTE-U) [14]. Both operators and users are able to decide a preferred network among the possible candidates because the coverage areas of those networks are partly or fully overlapped in HetNet. This determination depends on the availability of technologies in the terminal as well as the operator network construction. The traffic steering scheme which directs the user connection into a specific network in HetNet can improve user QoS while increasing the HetNet capacity and [15] investigates the traffic steering schemes by static adjustment of mobility parameters in HetNet and tries to achieve the traffic steering goal by inter-RAT handover. For the purpose, it controls threshold values of B2 and 3A event which are control parameters of inter-RAT handover defined in 3GPP standards. In [16], to deal with the load balancing between UMTS
network and WLAN network, the dynamic auto-tuning process is performed by using fuzzy theory framework. According to the evolution of mobile devices, a single device is able to be connected with multiple wireless technologies. It is, in addition, expected that there is a practical opportunity to build up efficient multi-homed data exchange with the raised possibility to transmit/receive simultaneously over other access networks. In particular, it is possible to provide a number of devices with multi-homed connection through 3GPP access network such as LTE and LTE-A and non-3GPP access network represented by Wi-Fi, and it is expected to accommodate the explosive ever-increasing mobile traffic by improving network capacity in 5G mobile communication environments. Although multi-path TCP (MPTCP) [17], mobile SCTP [18], and host identity protocol (HIP) [19] are proposed as solutions for the multi-homed devices, these are mainly concerned with end-to-end signaling, thus requiring the correspondent nodes to be extended with specific functionality related to the respective protocol [20].

In Q3 2014, SKT (one of ISPs in South Korea) announces the achievement of 1 Gbps data rate through TCP level aggregation between LTE-A and Wi-Fi. In general, the data aggregation can be performed on the application layer, transport layer such as MPTCP, or lower layers like RLC/PDCP layer, and the implementation of the aggregations on layers above IP is easier and frequently demonstrated. Although the aggregation on the lower layer is able to provide best performance since it is easier to adapt to the varying wireless channel and traffic loading conditions, the lower layer aggregation requires more changes to wireless access stratum infrastructure, which is a major impediment for the implementation [21].

Hotspot 2.0 (HS2.0), which is a working group in Wi-Fi Alliance (WFA), is aimed to enable seamless roaming among Wi-Fi networks and between Wi-Fi and cellular networks [1], [22]. In particular, HS2.0 specifies some Wi-Fi functionalities facilitating the Wi-Fi/Cellular-integration. HS2.0 certification program is called Wi-Fi Certified Passpoint which certifies the selected functions from the IEEE standards based on IEEE 802.11u, 802.11i, and 802.1X. While functions in Passpoint Release1 are complementary to the 3GPP ANDSF, the Release 2 is competing with the ANDSF in that UE is informed of indications related to its network selection. 3GPP is in progress of aligning functions from ANDSF and HS2.0 as a work item of its Release 12 specification.
With the goal of spectrum overlay technology development, SOLDER (Spectrum OverLay through aggregation of heterogeneous DispERsed bands) project was launched in November of the year 2013 with a duration of 3 years [23-24]. The main objective of the SOLDER is to provide the aggregation of such HetBands enhancing thereby the overall composite capacity and QoS at the UE. For the purpose, the SOLDER project has the following detailed objectives: 1) to design and develop PHY/MAC techniques for efficient CA over the HetNets and h-RATs; and 2) to develop radio resource management exploiting the full potential of heterogeneous carriers for providing seamless and enhanced service delivery [24].

Recent literature sets extensive focus on context-aware optimal RAT selection in a Multi-RAT environment [25-29]. The work in [25] deeply analyzes the state-of-the art in the field of multi-RAT selection and proposes a unifying model for representing the solutions. In order to support efficiently multi-RAT selection, context-aware schemes are introduced [25-27]. The context-awareness is considered as a necessary element for self-organizing network (SON) since its parameters are able to be automatically changed by continuously monitoring the varying environments. Reference [28] devises a distributed per cell approach for cell/RAT assignment using load and signal strength based adaptive learning mechanisms. On the other hand, fully distributed UE-centric techniques for inter-RAT network selection are considered in [29], where extensive theoretical characterization of load-aware UE-centric network selection is presented.

Recently, 3GPP has approved and studied a Release 13 Work Item on ‘LTE-WLAN radio level integration and interworking enhancement (LWA)’ which allows for scheduling decisions for each link at a packet level based on real-time channel and load-aware radio resource management across WLAN and LTE to provide significant capacity and QoE improvements. The LWA does consider only WLAN networks deployed and controlled by an operator and its partners. In LWA, 3GPP considers two basic deployment scenarios: 1) in collocated eNB and WLAN scenario, the integrated eNB and WLAN AP is considered or the eNB has an ideal backhaul to a WLAN logical entity which is connected to the APs; 2) The non-collocated eNB and WLAN scenario is when such backhaul is not ideal. The eNB, as an anchor point for both data and control planes in LWA, connects to the core network via regular S1 interfaces (S1-C and S1-U). For traverse of data packets and control signaling between eNB and WLAN, a new
interface is being standardized by 3GPP and will enable aggregation operation and feedback from WLAN on channel and user conditions to assist LWA scheduling operation.

5.1.2.3. Research Direction
There are many opportunities to carrier aggregation in different types of spectrum, including licensed spectrum, unlicensed spectrum (e.g., Wi-Fi), and new emerging forms of spectrum, such as TV white space (TVWS) and spectrum freed by concepts such as authorized/licensed shared access (ASA/LSA). The multi-RAT CA should overcome the problem of aggregation of different access technologies in h-RATs.

5.1.2.3.1. 3GPP licensed – 3GPP unlicensed CA
3GPP is willing to deploy LTE system in unlicensed spectrum so called LTE-U, which enables operators to increase their network capacity and allows UEs to achieve higher QoS/QoE with a comparatively cheap cost. Therefore a mixture of heterogeneous CCs from licensed LTE and unlicensed LTE issues some challenges as follows:

- The various CCs in licensed and unlicensed spectrum are heterogeneous, which makes the lower layer CAs in PHY/MAC/RLC/PDCP a real challenging problem.
- Due to the necessity of multiplexing CCs in different spectrum bands, UE goes through transmission power problems, which may prohibit their aggregation.

5.1.2.3.2. 3GPP – non-3GPP CA
The CA between 3GPP system (LTE) and non-3GPP (Wi-Fi) is the most challenging issue in h-RATs because there are obviously many differences between licensed LTE and unlicensed Wi-Fi. LTE transmissions are based on OFDMA/SC-FDMA managed by a centralized scheduler whereas Wi-Fi transmissions are based on CSMA/CA procedures without centralized control. In particular, Wi-Fi is susceptible to interference since the unlicensed band is shared with other users. A major concern with simultaneous operation of LTE and Wi-Fi bands is interference due to the UE coexistence of multi-radios operating in adjacent or overlapping radio spectrum bands [23]. The CA of LTE and Wi-Fi could be performed at different levels, such as above IP layer and below IP layer. The upper layer CA like MPTCP needs more
research on the optimization process for commercial usage whereas a tight integration of LTE and Wi-Fi at the lowest layer such as the MAC layer is required to be researched with aims to increase significantly the overall data throughput and to enhance user QoS/QoE.

The typical application scenarios for the aggregation of Wi-Fi and LTE are small cells with a large number of users and limited spectrum availability in the licensed band. If the small cell supports the CA of LTE and Wi-Fi, it can use Wi-Fi as a supplementary carrier which may be based on either 6 GHz technologies such as 802.11ad and 802.11ax or 60 GHz 802.11ad. The CA scenario would bring forth some enhancements more than the scenario where LTE and Wi-Fi are used separately since the base station could balance the traffic according to the needs of the users and the capabilities of both technologies [23].

If the tight integration of LTE and Wi-Fi is possible, the CA with other systems such as TVWS spectrum system and WPAN system can be also expected to be possible.

5.1.2.3.3. RRM-related Challenges

The major RRM issues for h-RATs are the dynamic resource allocation, interference management, and traffic load balancing between the heterogeneous aggregated resources. In order to perform the functions, the RRM should utilize measurement data from both the base station and the users. However, the h-RATs may provide a different sort of measurements which make the RRM very complex. The media independent handover functions, which are defined in IEEE 802.21 [30] and provide an abstraction layer to the measurements from different RATs, may be utilized in order to deal with the RRM-related challenges. Thereby a set of common measurement enables the RRM to efficiently and dynamically aggregate traffic over the h-RATs.

The Main RRM issues to be addressed under h-RATs CA application could be summarized as follows.

- Hybrid scheduling mechanisms and dynamic resource allocation for application-specific approach, various UE types, and different channel conditions over various bands
- Procedures for CC activation/deactivation at MAC layer which allow for more degrees of freedom to allocate resources
- Optimization algorithms such as load balancing and handover optimization in terms of aggregated h-RAT resources usage
5.1.2.3.4. **Smart Signaling for Vertical Soft Handover between Heterogeneous RATs**

5G wireless systems are expected to support various UEs and extensive services with diverse and distinct requirements. To allocate adaptively the heterogeneous network resources to the above devices and services can efficiently make the network performance be higher. For that purpose, decision making can be optimized by utilizing various context information which is able to be gathered from user environment, devices, and services. In h-RATs environment where multiple wireless accesses are available at the same location, user mobility should be efficiently supported by exploiting simultaneous usage of multi-RATs to the maximum, in order to improve the network performances such as the user experience and the network capacity. Therefore, it is required to develop the h-RATs handover decision/execution mechanisms which allow to reduce the required signaling exchange during handover and to optimally place a UE’s session(s) on RATs based on the user preferences and mobility pattern information, the network availability as well as the operators’ policies.

**References**


5.1.3. Cognitive Radio and Spectrum Sharing

5.1.3.1. General Overview

Owing to the prodigious interest in the use of emerging 5G communications including small cells, device-to-device (D2D) communications, Internet-of-Things (IoT), and so on [1], the spectrum bands required to support all types of wireless services have increased. Thus far, the spectrum assignment was static to a specific licensed service and its users, causing spectrum scarcity because spectral resources are limited. It is well known that spectrum scarcity occurs due to the inefficient spectrum allocation, rather than the actual physical shortage. According to real measurements, nearly 70% of the allocated spectrum is either unutilized or idle [2, 3]. To mitigate this problem, new spectrum assignment concepts, cognitive radio (CR), dynamic spectrum access (DSA), and spectrum sharing have been considered.

According to ITU-R [11], spectrum sharing is classified into vertical sharing and horizontal sharing. In vertical sharing, secondary users (SUs) should be able to identify the unutilized or idle licensed (and unlicensed) spectrum which is originally licensed for primary users (PU). To ensure communication of PUs, spectrum databases which contain geo-location information about the utilization of PUs should be deployed to aid SUs to identify the idle spectrum easily while protecting the licensed PUs. Such examples are TV white space (TVWS), Spectrum Access Sharing (SAS) [10] and Licensed Shared Access (LSA) [6-8]. Currently, several standards such as IEEE 802.11af [9] are working on accessing the idle TVWS through the spectrum database. SAS and LSA allow access to 3.5GHz and 2.3GHz in the USA and EU, respectively, as long as existing radar and satellite systems are protected.

In horizontal sharing, different systems access the same spectrum band with no priority. Such an example is LTE-U and licensed assisted access (LAA) where LTE systems access unlicensed (ISM) bands where they co-exist with Wi-Fi and other unlicensed devices. Horizontal sharing is also possible together with vertical sharing; e.g., different systems co-exist with the aid of spectrum database which informs them of the spectrum usage.
5.1.3.2. State of the Art and Research Trend

5.1.3.2.1 Spectrum database

Although spectrum sensing was the leading scheme to identify whether a channel is idle or busy, it puts high computational cost on devices and accuracy issues on low SNR environments because sensing may result in “False Positives” or “False Negatives”. Therefore, starting from the year 2010, the idea of using geo-location database that contains information about available bands has been adopted in most standards. The geo-location DB (or TVDB) stores information such as protected areas, propagation model, allowed transmission power and the duration of availability for each channel. This will reduce the computation associated with the device and provide accurate information about PUs with SUs. PUs at TV bands also include devices such as wireless microphones, talkback devices or in-ear-monitor, which in general are called Program Making and Special Events (PMSE) devices. TVDB should also contain information about PMSE devices [4]. These unutilized TV bands are known as TVWS as a potential spectrum of most CR standards. Compared to bands that are currently used for wireless communication, TVWS offers better channel characteristics in terms of the coverage range and penetration of obstacles.

![Diagram of Cognitive radio network based on TVDB](image-url)
The architecture of database based cognitive radio is depicted in Figure 5.1-7 CR users can set up a network in ad hoc or infrastructure mode. Networks in the ad hoc mode can access the TVDB through the Internet while networks in the infrastructure mode can either use a dedicated wired network or use the Internet. Regulatory bodies monitor how the band is being used fairly and also states which information the TVDB should include. PMSE booking system provides information about PMSE device activity. TV broadcasting associations monitor if the SUs interfere with current broadcasting program. The definition and information in TVDB depend on the requirement of regulatory bodies, which is discussed in [9].

An example of TVDB is found in the design and architecture of Google TVDB that already deployed a working TVDB for the USA through which one can easily search idle TV channels by giving the longitude and the latitude of a location. Figure 5.1-8 shows the available channels in the case of latitude = 37 and longitude = -95.5. The detailed architecture and the contributors of the database is given in [5].
5.1.3.2.2. LSA
LSA provides a CR-based architecture to obtain more frequency bands around 2.3 GHz which have been assigned to radar systems in the EU. LSA defines three zones: exclusion zone, restriction zone, and protection zone. The exclusion zone means a geographical area within which LSA Licensees are not allowed to have active radio transmitters. The restriction zone means a geographical area within which LSA Licensees are allowed to operate radio transmitters, under certain restrictive conditions (e.g. maximum EIRP limits and/or constraints on antenna parameters). The protection zone means a geographical area within which Incumbent receivers will not be subject to harmful interference caused by LSA Licensees’ transmissions.

![Figure 5.1-9 The architecture of LSA systems](image)

LSA Repository (LR) supports the entry and storage of information describing Incumbent’s usage and protection requirements and conveys availability information to authorised LSA Controllers. LSA Controller (LC) is located within the LSA Licensee’s domain and enables the LSA Licensee to obtain spectrum resource availability information from the LR. The architecture of LSA is shown in Figure 5.1-9.

5.1.3.2.3. SAS
SAS is a CR-based system that enables SUs to access a spectrum band around 3.5GHz in the USA which have been originally assigned to radar and satellite systems. In SAS, there are three
types of accesses: Incumbent Access (IA), Priority Access License (PAL), and General Authorized Access (GAA) as shown in Figure 5.1-10. IA users would be entitled to full protection for their operations within their deployed areas. PAL users would have interference protection from other PALS/GAAs users but should expect interference from and avoid interference to Incumbent users. GAA users would not be entitled to interference protection and should accept interference from other users. The main use cases of PAL and GAA are expected to be TD-LTE small cell and IEEE 802.11ac, respectively, at 3.5 GHz.

5.1.3.3. Research Direction

5.1.3.3.1. Carrier aggregation using more spectrum bands
As shown in Figure 5.1-11, Qualcomm has proposed a carrier aggregation technique based on licensed and unlicensed bands together which can increase LTE spectrum with no additional cost. This technique, called LTE-U and LAA, is said to improve the throughput of existing Wi-Fi systems. Similarly, using LSA and SAS, more spectrum bands at 2.3 GHz and 3.5 GHz can be aggregated with existing cellular carriers.
5.1.3.3.2. Small cells using spectrum sharing
Spectrum sharing is already being used for small cells: e.g., Almost Blank Subframe (ABS) provides time domain multiplexing for small cells and macrocells in HetNets. Most spectrum sharing techniques such as LTE-U, LAA, and SAS consider small cells as the first use case, because small cells are relatively free of interference issues, owing to the small transmission range, and they should be able to deal with much data transmission.

5.1.3.3.3. Dynamic radio configuration for self-organizing network (SON)
SON includes self-configuration, self-optimization, and self-healing for efficient network operations. Especially, for self-configuration, dynamic radio configuration is needed and it may adopt the concept of rendezvous which have been studied for dynamic spectrum access to find a common channel for network initialization.

5.1.3.3.4. Inter-cell interference management using spectrum sharing
For 4G mobile systems, inter-cell interference problems have been resolved using Inter-Cell Interference Coordination (ICIC), eICIC, and Coordinated Multipoint (CoMP). As different kinds of cells or radio access technologies (RATs) are deployed together, existing inter-cell interference solutions may use spectrum sharing techniques.

5.1.3.3.5. Coexistence
As the geolocation database largely protects the co-existence issue [14], adoption of CR techniques can deal with the basic problems of coexistence. Inter-network communication can
lead to the solution of coexistence (communication between TVWS database and WLAN APs). SAS to manage the service that incorporates a dynamic database [15]. Still, it is a challenging issue to develop internetwork communication technology, potential interference mitigation technique, and a solution for SAS dynamic database to manage the use of TVWS.
References


5.2. Advanced Modulation and Coding

5.2.1. Advanced Modulation (FQAM)

5.2.1.1. General Overview
One of key requirements for 5G is enhancement of a cell-edge performance, which means every user can enjoy experience of Gigabit anywhere. Conventional approaches are interference managements (e.g., interference cancellation, interference avoidance). These approaches deal the interference as the Gaussian. However, it was proven that the worst case about distribution of additive noise in a wireless network is the Gaussian distribution in a sense of channel capacity [1]. Therefore, interference design, which means making inter-cell interference (ICI) as non-Gaussian, is required instead of interference managements. The distribution of ICI depends on a modulation of a interfering signal. An active interference design can be achieved by using a novel modulation.

5.2.1.2. State of the Art and Research Trend
A frequency and quadrature amplitude modulation (FQAM), which is a combination of frequency shift keying (FSK) and quadrature amplitude modulation (QAM) can be used as an active interference design scheme. Several results of research are available in literatures to deal with the performance of systems with FQAM [2-7]. However, most of the prior results of research focus on an uncoded error performance or a trellis coded modulation (TCM). FQAM was firstly proposed in [2], where the performance of a TCM system utilizing the proposed modulation scheme was shown to outperform uncoded QAM for a given bit rate under additive white Gaussian noise (AWGN) channels. Thereafter, in [3], the study was further extended to the multiple TCM based systems. Also, in [4-5], a modulation scheme that conjoins a multi-tone FSK and QAM was proposed and the uncoded error performance of the proposed system was analyzed. Recently, in [6], the performance of Reed Solomon coded orthogonal frequency division multiplexing (OFDM) systems using FQAM was analyzed under AWGN and Rayleigh fading channels. Even with a sub-optimum receiver employing noncoherent detection for detecting active frequency tones, the error performance at a given
bit rate was shown to be superior to that of the QAM based systems. These results are motivated to analyze the error performance and throughput of an optimum receiver for FQAM with bit interleaved coded modulation (BICM) and coded modulation (CM) [7].

5.2.1.3. Research Direction
Firstly, we analyze the performance of the FQAM-based BICM and the CM systems, respectively, under AWGN channels when a coherent detection is used. The derived probability density function (pdf) of correlator’s outputs are used to compute the BICM and the CM channel capacity as well as corresponding normalized throughputs for the considered systems. Furthermore, maximum likelihood (ML) soft decision decoding metrics for a binary and a non-binary Turbo coded systems with FQAM are derived for practical performance evaluations. Numerical results show that similar to FSK-based systems, the FQAM-based CM systems significantly outperform the FQAM-based BICM systems. The results also show that the bandwidth efficiency of FQAM is dramatically enhanced compared to that of FSK, while the error performance of FQAM is maintained similar to that of FSK. Moreover, the FQAM is shown to approach a theoretical channel capacity limit when combined with CM systems, for modest code rates in low signal-to-noise ratio (SNR) regions where the channel capacity limit cannot be achieved with either QAM or FSK.

Next, we analyze the performance of the downlink cellular OFDMA networks (as specified in the 3GPP LTE standardization) using FQAM instead of QAM, especially for cell-edge users. Numerical results show that for the case with FQAM, the statistical distribution of ICI observed at the cell-edge highly deviates from the Gaussian distribution and has a heavier tail than the Gaussian pdf, while that with QAM does not. The results also show that a transmission rate of a cell-edge user with FQAM significantly improves compared to that with QAM. For example, for 3-cell structure at a frame error rate (FER) of 0.01, the transmission rate of a cell-edge user increases approximately 3-times by replacing 4-ary QAM with 32-ary FQAM. In addition, it should be required to evaluate the system level performance. Moreover, the reduction of complexity (especially, the decoding) for non-binary channel coding must be considered.
Figure 5.2-1 is the 16-ary FQAM signal constellation that is a combination of 4-ary FSK and 4-ary QAM, and the simulation results are shown in Figure 5.2-2 to Figure 5.2-5.

Figure 5.2-2 Minimum required SNR versus the BICM and the CM channel capacity (achievable code rate) for FQAM, FSK and QAM-based systems under AWGN channels.
Figure 5.2-3 Binary and M-ary Turbo coded FER versus $E_b/N_0$ for FQAM, FSK and QAM-based systems under AWGN channels

Figure 5.2-4 Sample histogram of ICI for downlink LTE systems with 32-ary FQAM and 4-ary QAM, ITU Pedestrian B, 3km/h
Figure 5.2-5 FER performance versus information bit rate of downlink LTE systems with FQAM and QAM where NF =240 bits, ITU Pedestrian B (3 km/h) II

References


5.2.2. Advanced high-speed Channel Coding

In IMT-2020 vision [5], one of key requirements in fifth generation mobile networks (5G) is that peak data rate requires up to 50Gbps within 1ms latency, which is 50 times higher data rate within 1/10 times shorter latency than requirement of 4G LTE. In the future, many kinds of various high quality of service scenarios, such as multimedia UHD streaming and AR/VR services, will be expected and then required peak data rate per user will grows up very highly. On the other hand, these kinds of services also are delay-sensitive and so requires data transmission based on shorter feedback delay and lower latency as much as possible. Based on these requirements, 5G channel coding technology should support the faster processing speed in the limited latency, that is, decoder throughput should be so much higher than that of 4G systems. Then, more power consumption and more hardware complexity will be required to support higher decoder throughput. But the increment of power consumption and complexity rises in proportion to the increment of decoder throughput. So a couple of Gbps data transmission can give us serious impact on the practical complexity and power consumption. As a result, to realize the fifth generation mobile network (5G), the consideration on the efficient channel coding technology is very important.

5.2.2.1. Low-Density Parity Check (LDPC) Codes

5.2.2.1.1. State of the Art and Research Trend

In 1991, Turbo code, invented by Berrou et al., shows performance nearby Shannon limit for large packet transmission and adopted in 3GPP standardization. Until now, it is well-proven commercial technology for mobile network including WCDMA, HSDPA, LTE, and so on. Recently, LTE system required more data rate per UE and so up to 1 Gbps Turbo decoder was developed [6]. It was possible by using the parallel algorithm such as QPP interleaver and high density CMOS technology for commercialization. On the other hand, LDPC (Low Density Parity Check), revised by Mackey in 1996, also shows performance nearby Shannon limit based on graph theory. As Quasi-cyclic type LDPC [2,3] was developed, the efficient parallel decoding architecture was realized and became to have strong advantage for high performance and high decoder throughput. In 2003, LDPC was adapted firstly in the European broadcasting
standardization DVB-S2 and widely spread in the broadcasting system and connectivity system such as IEEE802.11n and 11ac and so on.

As mentioned in the previous section, channel coding for 5G system required the efficiency in terms of power consumption and complexity for Gbps data transmission. In the most of pre-5G system have target to 5 Gbps as a data rate per UE. Figure 5.2-6 shows comparison of power consumption for decoder between Turbo codes and LDPC codes. In comparison, 3 types of Turbo code with the different parallel factor of QPP interleaver-8, 16, 64 is shown. In these cases, the Turbo code with 64 parallelism require the low power consumption, but it still needs 10 times more power than the worst case of LDPC. Furthermore, it requires 439 mW in the latest CMOS technology and it is very significant impact on the practical modem design.

![Required Power for 5 Gbps](image)

**Figure 5.2-6 Comparison of power consumption between Turbo code and LDPC code**

Figure 5.2-7 shows comparison of chip area for decoder between Turbo codes and LDPC codes. In comparison, higher parallelism for turbo codes, then higher chipset area. Hence, it means that there is some tradeoff between power consumption and chip area depending on parallelism of Turbo code. Finally, to reduce the power consumption of Turbo code, higher parallelism should be required, then it gives significant impact on chipset area.
Therefore, although Turbo code was well-proven technology in mobile communication until now, some serious consideration of usage is needed for the 5G mobile system. Considering the practical power consumption and complexity, LDPC code for 5G mobile system is preferred.

5.2.2.1.2. Research Direction

In this section, required research direction for LDPC is described. The length and rate compatibility of LDPC is an important property in the mobile communication. First, the transport block size is variable based on modulation order, coding rate and allocated resource blocks for each user [1]. To support different sizes of transport blocks efficiently, we propose the length compatible Quasi-cyclic LDPC (QC-LDPC) codes. The QC-LDPC code [2][3] is defined by the parity-check matrix which consists of $L \times L$ blocks which are zero matrix or circulant permutation matrices. The $mL \times nL$ parity check matrix $H$ of QC-LDPC codes is defined by

$$H = \begin{bmatrix}
P^{a_{11}} & P^{a_{12}} & \cdots & P^{a_{1(n-1)}} & P^{a_{1n}} \\
P^{a_{21}} & P^{a_{22}} & \cdots & P^{a_{2(n-1)}} & P^{a_{2n}} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
P^{a_{m1}} & P^{a_{m2}} & \cdots & P^{a_{m(n-1)}} & P^{a_{mn}}
\end{bmatrix}.$$
where \( a_{ij} \in \{0, 1, \cdots, L-1, \infty\} \) and \( P_i \) is the circulant permutation matrix which shifts the identity matrix to the right by \( i \) \((0 \leq i < L)\) times. The length of the LDPC code, the number of parity check bits and the number of information bits is \( nL, mL \) and \( (n-m)L \), respectively. Therefore, the length of information bits and coded bits are variable depending on circulant permutation matrix size \( L \) which is predetermined integer numbers. If the transport block size is greater than maximum information size, transport block is segmented, coded and transmitted. Moreover, the LDPC information size is adjusted to transport block by selecting the suitable circulant permutation matrix size \( L \) and shortening method. Second, we also propose rate compatible QC-LDPC codes. Rate compatible property is important to achieve high throughput in HARQ (Hybrid Automatic Repeat Request) and AMC (Adopted Modulation and Coding) systems [4]. The rate compatible is supported by puncturing and repetition based on one or more QC-LDPC codes designed by taking into account the rate compatibility.

5.2.2.2. Polar Codes

5.2.2.2.1. State of the Art and Research Trend

Polar code is the first code family that has been theoretically proven to achieve the Shannon Limit [12]. Since the polar coding was invented, a lot of work has been done on decoding algorithm, rate-compatible coding scheme, and hardware implementation.

Polar code has both low encoding and decoding complexity. It is reported in [13] that a throughput of 237Gbps can be implemented on a FPGA.
Figure 5.2-8 Polar code has a significant improvement against turbo code

When concatenated with CRC bits (of CB/TB in LTE), the performance of polar codes under list decoder can achieve a significantly better performance than turbo codes, especially for short code length [14-16]. Figure 5.2-8 shows comparison against turbo code. Under a list decoder, Polar code can provide a better and better performance with increasing the list size.

Figure 5.2-9 Frame error rate test on FPGA of (1024, 512) polar code
Another favorite property is polar code has no error floor. Figure 5.2-9 shows the FER curve of (1024, 512) polar code over AWGN channel. No error floor was observed as FER goes down to 1e-9.

Figure 5.2-10 The code rate and code length of polar codes can be flexibly adjusted.

Besides, polar codes can support flexible coding length and coding rate [17]. Under a specific code length $N$, the rate of polar code can be simply adjusted by increasing or reducing the number of the sub-channels which are selected to carry the information bits. The code length $N$ can be reduced from a base code length $M = 2^{\lceil \log_2 N \rceil}$, which is some power of 2, through a bit-reversal puncturing.

Based on the rate and length compatible coding scheme, polar coded hybrid ARQ scheme is proven to be capacity-achieving and shows the better performance than turbo-coded scheme in LTE [18]. In each re-transmission, the less reliable information bits in previous transmissions
are collected and generate a new codeword. The re-transmitting information bits are determined based on the predefined sub-channel reliabilities, which make all the transmissions to share the same residual coding rates. Figure 5.2-11 gives a simple example of four transmissions with length $N = 16$. The initial is code rate is $3/4$. After four transmissions, all the four transmissions have a residual code rate is $3/16$. An FER comparison against turbo codes with maximum four transmissions is given in Figure 5.2-12. Polar codes can achieve the better performance than turbo in every transmission.

![Figure 5.2-12 FER performance of (2304, 1224) polar code under 16QAM with four transmissions](image)

5.2.2.2. Research Direction

In this section, future research direction for polar code is described. The decoding algorithms for polar code, including successive cancellation decoding and list decoding, require rather different hardware design. The published hardware decoders are mostly on FPGAs. Other implementation schemes such as DSP and ASIC are to be further explored.
References


5.3. Duplexing

5.3.1 Full Duplex Radio

5.3.1.1 General Overview
In-band full-duplex (IBFD) technology allows simultaneous uplink and downlink service in a same frequency band. This functionality theoretically provides two times higher spectral efficiency compared to time division duplexing (TDD) or frequency division duplexing (FDD) [1, 2]. Figure 5.3-1 shows the achievable sum rate performance in IBFD system. It can be explained that IBFD system achieves more gains than conventional half duplex (HD) system under an ideal assumption [1,2].

![Achievable sum rate vs SNR](image)

**Figure 5.3-1 Performance Comparison of IBFD system with HD system**
The basic advantages offered by the features of IBFD transmission can be summarized as follows [1]:

- Can double ergodic capacity: Full utilization of time and frequency resources makes it theoretically possible for IBFD transmission to double the link capacity, as compared to HD transmission.

- Can reduce feedback delay: Reception of feedback signaling (such as control information, channel state information (CSI) feedback, acknowledge/no-acknowledge (ACK/NACK) signals, resource allocation information, etc.) during data signal transmission enables shorter air interface latency in feedback information.

- Can reduce end-to-end delay: In relay systems, relay nodes with IBFD transmission can reduce end-to-end delay because the relay node simultaneously receives data from a source node and transmits data to a destination node.

- Can improve network secrecy: The use of simultaneous transmission at two nodes means that eavesdroppers receive mixed signals that are hard for the eavesdropper to decode due to interference signals.

- Can improve the efficiency of ad-hoc network protocols: Because all nodes are transmitting, IBFD transmission can solve the ‘hidden node’ problem in ad-hoc networks. Furthermore, the fact that simultaneous listening and sensing is being performed on a frequency band while the signals are being transmitted means that each node can decide whether or not the other nodes are transmitting signaling and thus, prevent collisions.

- Can increase spectrum usage flexibility: By retaining the option to use one frequency band (IBFD transmission) or two different frequency bands (HD transmission) for uplink and downlink, each transceiver can select either the IBFD or the HD transmission mode.

Despite this attractive list of advantages with IBFD transmission, however, IBFD device faces an inherited strong interference that is originated from its own transmitted signal. It is called as self-interference (SI). The level of SI signal is extremely higher than a desired one. For example, a UE at the edge of a small-cell can incur path losses of 110 dB. The maximum allowable transmit powers of the eNB and UE are 24 dBm and 21 dBm, respectively, and the power of the received signal at the eNB can be up to $21 - 111 = -90$ dBm. If we assume
15 dB isolation between the eNB’s transmit and receive signal paths, then the SI can be up to $24 - 15 - (-90) = 99$ dB stronger than the desired received signal [3].

This strong SI seems very difficult to cancel out perfectly but the amount of self-interference can be effectively reduced since we know the characteristics of interfering signal originated from device itself. In practical environment, however, there should be residual SI due to the imperfect self-interference cancellation (Self-IC) at IBFD device. The performance of IBFD system with residual SI due to imperfect channel estimation of self-channel is investigated is in [4]. Therefore, we can conclude that the powerful Self-IC techniques become an indispensable component for feasible IBFD system.

In addition to this SI, IBFD transmission also has some other disadvantages, which can be summarized as follows [1]:

- **Self-interference (SI):** Simultaneous transmission and reception in a single frequency band can cause the transmitted signals to loop back to their receive antennas. In practical FD systems, the desired signal from the pair node propagates much longer compared to SI from itself. The power difference between the desired signal and the self-interference increases exponentially as the distance lengthens. The SI power can be up to 99 dB stronger than the desired received signal and the statistical model of residual SI is limited according to the performance of various self-IC schemes. Along with direct SI, furthermore, reflected interference signals which are partially obstructed by nearby obstacles also exist.

- **Imperfect interference cancellation:** In practical environments, the SI cannot be perfectly canceled for a variety of reasons, such as the non-linearity of hardware components in the RF chain (the SI power is beyond the feasible range), estimation errors on the self-channel and the received SI signal (various reflected interferences), and incompleteness of various cancellation techniques.

- **Increased inter-user interference:** Because all adjacent nodes are transmitting simultaneously, the number of inter-user interferers increases by almost a factor of two and the aggregate interference at a node increases as well.

- **Increased consumed power and complexity:** Canceling out both SI and inter-user interference at each node will require each node to have additional components for interference cancellation, thus consuming more operating power and resources.
In this white paper, we address the state-of-art research trend of Self-IC techniques in antenna, analog, and digital domains.

5.3.1.2. State of the Art and Research Trend

In this section, we will confine ourselves to briefly addressing the cancellation requirement and the basic Self-IC techniques for feasible IBFD system.

5.3.1.2.1. Cancellation Requirements

When a node simultaneously transmits and receives with IBFD transmission, it causes strong SI to its receiver RF chains. This strong SI can cause errors in hardware components at receiver, such as low-noise amplifier (LNA), analog-to-digital converter (ADC). This SI must be effectively cancelled out for feasible IBFD system. Table 5-3 1 shows that the target Self-IC requirement (RSelf-IC) for IBFD system in the worst case. $P_{t,\text{max}}$ is the maximum transmit power and NF is noise figure at receiver which is defined in 3GPP and $NL(BW)=-174\text{dB}+10\log_{10}(BW)+NF$ is receiver noise level and $R_{\text{Self-IC}} = P_{t,\text{max}}-NL$ is Self-IC target according to the node types [3]. The first step is to perform an antenna Self-IC scheme using a variety of antenna techniques. After that, we can sequentially employ analog and digital Self-IC to cancel any residual SI [2].

<table>
<thead>
<tr>
<th>Node Type</th>
<th>$P_{t,\text{max}}$</th>
<th>NF</th>
<th>$NL(BW)$</th>
<th>$R_{\text{Self-IC}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macro eNB</td>
<td>46dBm</td>
<td></td>
<td></td>
<td>142dB</td>
</tr>
<tr>
<td>Pico eNB</td>
<td>30dBm</td>
<td>5dB</td>
<td>-96dBm</td>
<td>126dB</td>
</tr>
<tr>
<td>Small eNB</td>
<td>23dBm</td>
<td></td>
<td>-92dBm</td>
<td>119dB</td>
</tr>
<tr>
<td>Cellular UE</td>
<td>23dBm</td>
<td>9dB</td>
<td>-92dBm</td>
<td>115dB</td>
</tr>
</tbody>
</table>

5.3.1.2.2. Antenna Self-IC

Antenna Self-IC techniques suppress the SI before it enters the receive RF chain [5,6,7]. The most intuitive approach to antenna cancellation is using path-loss. The propagation path-loss depends on the placement of antennas and the distance between Tx and Rx antennas.
Furthermore, circulator [8,9] and cross-pol antennas [10] also can be utilized for antenna Self-IC.

5.3.1.2.3. Analog Self-IC

After antenna Self-IC, it is necessary to suppress SI on the analog circuit before it enters the LNA and ADC. To create a replica signal in analog domain, the signals from transmit RF circuits are sequentially modified and aggregated with multiple parallel lines of varying delays and tunable attenuators [8,9]. We can then cancel the SI signal in analog domain using the replica signal before LNA.

5.3.1.2.4. Digital Self-IC

Digital Self-IC is the last step in the Self-IC process and suppresses any residual SI after the ADC. Basically, the estimated residual SI is subtracted from the received signal in digital domain [4, 5]. For this, it is necessary for estimating effective channel which suffers the distortions (e.g., wireless channels, active circuits in Tx/Rx chains, and antenna/analog Self-IC).

5.3.1.2.5. Hardware Implementation of IBFD System

SI power can be dramatically reduced using a combination of the above self-IC techniques in the antenna, analog, and digital domains. Before we can properly evaluate the performance of the self-IC techniques, however, the IBFD system must first be implemented. The authors in [6-19] implemented IBFD systems with WARP platforms to verify the feasibility of IBFD systems in small-sized devices such as laptop computers. The authors of [14] used the Universal Software Radio Peripheral (USRP) platform. The authors of [15] and [16] designed their hardware platforms for evaluating the IBFD system with larger system bandwidth or with different center frequencies.
Table 5.3-2 Performance comparison of Self-IC schemes for IBFD systems.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Band</th>
<th>BW</th>
<th>App.</th>
<th>Equipment</th>
<th>Antenna Type</th>
<th>Num. of Ant. (Tx,Rx)</th>
<th>Ant. Configuration</th>
<th>Cancellation Performance Antenna</th>
<th>Cancellation Performance Analog</th>
<th>Cancellation Performance Digital</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>[6]</td>
<td>2.4GHz</td>
<td>200MHz</td>
<td>All App.</td>
<td>VNA Agilent E8363A</td>
<td>Directional Antenna</td>
<td>(2,2)</td>
<td>Separated</td>
<td>51dB</td>
<td>None</td>
<td>None</td>
<td>51dB</td>
</tr>
<tr>
<td>[7]</td>
<td>5GHz</td>
<td>10MHz</td>
<td>WiMAX</td>
<td>WARP Platform</td>
<td>Dual-polarized Antenna</td>
<td>(1,2)</td>
<td>Separated</td>
<td>55dB</td>
<td>None</td>
<td>None</td>
<td>55dB</td>
</tr>
<tr>
<td>[8]</td>
<td>2.4 GHz</td>
<td>625 kHz</td>
<td>Bluetooth/Wi-Fi</td>
<td>WARP Platform</td>
<td>Typical Wi-Fi Antenna</td>
<td>(1,1)</td>
<td>Separated</td>
<td>45dB</td>
<td>35dB</td>
<td>80dB</td>
<td></td>
</tr>
<tr>
<td>[11]</td>
<td>2.4GHz</td>
<td>10MHz</td>
<td>802.11n</td>
<td>WARP Platform</td>
<td>3dBi Desktop Antenna</td>
<td>(1,1)</td>
<td>Separated</td>
<td>None</td>
<td>45dB</td>
<td>28dB</td>
<td>73dB</td>
</tr>
<tr>
<td>[12]</td>
<td>2.4GHz</td>
<td>40MHz</td>
<td>802.11</td>
<td>WARP Platform</td>
<td>7dBi Desktop Antenna</td>
<td>(1,1)</td>
<td>Separated</td>
<td>44dB</td>
<td>30dB</td>
<td>74dB</td>
<td></td>
</tr>
<tr>
<td>[13]</td>
<td>2.4GHz</td>
<td>10MHz</td>
<td>802.11n</td>
<td>WARP Platform</td>
<td>7dBi Desktop Antenna</td>
<td>(1,1)</td>
<td>Separated</td>
<td>57dB</td>
<td>24dB</td>
<td>None</td>
<td>81dB</td>
</tr>
<tr>
<td>[14]</td>
<td>2.4GHz</td>
<td>5MHz</td>
<td>825.15-4</td>
<td>USRPv1</td>
<td>Omni-directional Antenna</td>
<td>(2,1)</td>
<td>Separated</td>
<td>33dB</td>
<td>20dB</td>
<td>53dB</td>
<td></td>
</tr>
<tr>
<td>[15]</td>
<td>2.4GHz</td>
<td>20MHz</td>
<td>802.11</td>
<td>WARP Platform</td>
<td>7dBi Desktop Antenna</td>
<td>(2,1)</td>
<td>Separated</td>
<td>65dB</td>
<td>20dB</td>
<td>85dB</td>
<td></td>
</tr>
<tr>
<td>[16]</td>
<td>370MHz</td>
<td>100kHz</td>
<td>Relay</td>
<td>PENTEK board</td>
<td>Omni-directional Antenna</td>
<td>(7,3)</td>
<td>Separated</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>60dB</td>
</tr>
<tr>
<td>[17]</td>
<td>2.4GHz</td>
<td>20MHz</td>
<td>802.11</td>
<td>WARP Platform</td>
<td>Dual-polarized Antenna</td>
<td>(1,1)</td>
<td>Separated</td>
<td>72dB</td>
<td>23dB</td>
<td>95dB</td>
<td></td>
</tr>
<tr>
<td>[18]</td>
<td>915MHz</td>
<td>30MHz</td>
<td>ISM Band</td>
<td>Circulator/Spectrum</td>
<td>Circularly Polarized</td>
<td>(1,1)</td>
<td>Shared</td>
<td>None</td>
<td>40dB</td>
<td>None</td>
<td>40dB</td>
</tr>
<tr>
<td>[19]</td>
<td>2.4GHz</td>
<td>10MHz</td>
<td>802.11</td>
<td>Circulator/Spectrum</td>
<td>Omni-directional Antenna</td>
<td>(1,1)</td>
<td>Shared</td>
<td>None</td>
<td>70dB</td>
<td>None</td>
<td>70dB</td>
</tr>
<tr>
<td>[9]</td>
<td>2.4GHz</td>
<td>80MHz/20MHz</td>
<td>802.11ac</td>
<td>WARP Platform/Rohde/Schwarz</td>
<td>Off-the-shelf</td>
<td>(1,1)</td>
<td>Shared</td>
<td>None</td>
<td>62dB</td>
<td>38dB</td>
<td>100dB</td>
</tr>
<tr>
<td>[10]</td>
<td>2.4GHz</td>
<td>20MHz</td>
<td>802.11n</td>
<td>WARP Platform</td>
<td>Off-the-shelf</td>
<td>(3,3)</td>
<td>Shared</td>
<td>None</td>
<td>70dB</td>
<td>38dB</td>
<td>108dB</td>
</tr>
</tbody>
</table>

5.3.1.3. Research Direction

As mentioned above, cascaded Self-IC techniques are fundamental components to enable IBFD system. In practical environment, there still remain many challenges for Self-IC technique. In this section, we address the key elements to enhance the performance of Self-IC techniques for feasible IBFD system.

5.3.1.3.1. SI Channel Characteristics

As mentioned above, effective channel (after antenna Self-IC and analog Self-IC) should be measured for analog and digital Self-IC. Due to the strong power level of SI signal and the distortion in SI signal after each Self-IC steps, it is hard to accurately figure out the
characteristics of SI channel (the distortion and non-linear characteristic due to imperfect hardware components at RF chains, non-uniform channel response due to multi-path channel, time varying channel, etc.). Therefore, a more powerful channel estimation scheme is required to accurately estimate effective channel of SI signal. Furthermore, aggregated SI feature of IBFD system with multiple antennas would differ from that of IBFD system with a single antenna. Therefore, a more powerful Self-IC technique is required to make MIMO FDR system feasible.

5.3.1.3.2. Adaptive Self-IC Algorithm
Several studies in the literature assume that the SI channel remains static during Self-IC. When the IBFD system is applied to mobile devices, such as small cell BS or cellular mobile phones or laptop computers, however, the SI channel would normally vary in a realistic environment. Therefore, it will be important to investigate adaptive analog and digital Self-IC techniques that are robust to the time-varying SI channel. Furthermore, a fast and low-complex tuning algorithm should be designed for guaranteeing the reliable performance of Self-IC technique.

5.3.1.3.3. Small Size Self-IC Circuit
As the device size becomes smaller for portable items such as cellular mobile phones and laptop computers, systems will require smaller-size Self-IC components for IBFD transmission due to space and power limitations.

5.3.1.3.4. Advanced Self-IC for Infrastructure-based System
As the transmit power increases, the power of SI signal also increases, whereas the performance of Self-IC is dramatically reduced due to the limited dynamic ranges of LNA and ADC. Therefore, making an IBFD system feasible in infrastructure-based system will require the advanced Self-IC technique to deliver more than 142dB (See Table 5-4).

5.3.1.4. FDR Application: UE-specific Dynamic TDD
Current mobile networks like 3GPP Long Term Evolution (LTE)/LTE-A operate time division duplex (TDD) and frequency division duplex (FDD) system based on synchronous transmission
in different cells to eliminate the base station (BS)-to-BS and user equipment (UE)-to-UE interference. Especially, TDD systems have the advantages of accommodating asymmetric downlink (DL) and uplink (UL) traffics by varying the TDD DL-UL configurations [20]. To handle cell-specific asymmetric data traffic, time-division LTE (TD-LTE) system allows different UL-DL configuration in different cells. The main technical reason for such restricted operation of the TD-LTE system is to prevent strong BS-to-BS and UE-to-UE interference due to the usage of opposite transmission directions in different cells.

However, this synchronous operation with a common UL-DL configuration in a cell for all users may not match the instantaneous traffic situation in a particular user. The amount of data traffic for DL and UL varies with time and between UEs. To efficiently address these diverse traffic demands, UE-specific DL-UL configuration emerges as one of the main techniques in future wireless communication systems.

References


5.3.2. Dynamic TDD

5.3.2.1. General Overview
The key characteristics of 5G networks will be ultra-dense small cell deployments, and very diverse applications with equally diverse requirements ranging from ultra-low latency to Gigabit data rates. For small cell deployments and full support of diverse application requirement, it is beneficial to employ a TDD-based air-interface with the option to flexibly assign transmit resources to uplink and downlink communication in each frame. This type of flexible TDD switching between uplink and downlink is also known as dynamic TDD. In case of a small cell with only a single active user in it, it is quite obvious that data rates experienced by the user could be improved by scheduling the TDD direction depending on the user’s traffic pattern. For this reason, several 5G candidate technologies for small-cell deployments feature very flexible dynamic TDD schemes, cf. e.g. [1]. New frame structures capable of full dynamic TDD are proposed, and it is accepted as one of the main key features of 5G system.

In order to introduce the concept to the existing LTE system, 3GPP release 12 introduced support for more rapid TDD configuration [2,3]; there it is called “enhanced Interference Mitigation and Traffic Adaptation” (eIMTA). However, eIMTA doesn’t have full flexibility of the switching because its legacy LTE co-existence could generate serious performance degradation from the interference.

5.3.2.2. State of the Art and Research Trend

5.3.2.2.1. eIMTA in LTE Flexible TDD
In TDD in general, each Transmission Time Interval (TTI) the transmission is pre-configured for downlink, for uplink or for special subframe. LTE supports 7 different TDD patterns which
offer uplink / downlink ratios from approximately 6:4 to 1:9 within a system-frame consisting of 10 successive TTIs. For UEs supporting eIMTA implementations flexible subframes are introduced. These can be configured dynamically either for uplink or for downlink. Legacy UEs will be configured with an uplink-heavy TDD configuration and are thus limited to uplink transmission in the flexible subframe. The eNB will not schedule a legacy UE with an uplink grant in case it wants to use the subframe for an eIMTA UE in the downlink.

Cell-specific TDD pattern selection introduces the problem of cross-link interference between neighboring cells and with this potentially severe variations in interference power / SINR between successive TTI even when the radio channel itself does not change. Without any specific action, this can seriously deteriorate the performance of the link adaptation and hence the user experience.

This scheme allows for flexible utilization of the available radio resources by means of dynamic adaptation of the TDD uplink / downlink subframe configuration thereby greatly improving user experience especially in low to medium load. Some important aspects of eIMTA, namely flexible TDD pattern selection, separate CSI measurements, enhanced uplink power control and clustering, were discussed and evaluated by means of system-level simulations. Packet throughput gains for the small cell scenarios as reported by 3GPP in [3] are confirmed. It was observed that great care must be taken in designing the eIMTA system in order to benefit from it and to have a good balance between uplink and downlink performance and to ensure fairness among UEs in the system.

**5.3.2.2. Dynamic TDD in 5G dense small cell**

Dynamic TDD was studied as one of the key promising techniques to provide higher throughput, and low latency in METIS project [4].

Key technical building blocks on dynamic TDD includes the efficient design of frame structure and control signaling, interference coordination between DL-UL signals and the advanced receiver design etc.

Figure 5.3-2 shows the example of the subframe structure for 5G new radio interface which is proposed at the METIS project. Each subframe has timeslot for DL and UL control, DL or UL data, so network can schedule DL or UL subframes flexible manner. This frame structure
is also one of enablers to achieve the low latency of 1ms RTT (Round-Trip-Time) which is the key capability of 5G system, while it provides the capability of dynamic TDD.

Aimed at accomplishing the 5G requirements and based on the analysis described previously regarding the main deficiencies of existing technologies, a TDD optimized radio has been devised. The TDD optimized radio has a flexible TDD frame structure that enables: (i) fast TDD access and fully flexible UL-DL switching, (ii) cost-efficient pipeline processing, (iii) low power consumption, (iv) clean HARQ, (v) support for advanced receivers, (vi) support for frequency coordination and reuse.

The tight latency requirements of 5G system lead to short RTT with fast control signaling, fast TDD switching periodicity and support for flexible UL/DL ratio. In order to enable robust and fast control plane, control signaling is embedded to each subframe and time separated from the data plane. Control symbols are located before data symbols in order to allow fast and cost efficient pipeline processing at the receiver. This way the UE can process the control information while transmitting/receiving in the data part, which type of processing reduces the latency. Also, this time separation of control and data part is aimed at reducing the power consumption at the UE too. If the UE does not transmit/receive in the data part nor receives any command during the control part regarding the data part, it can turn-off its receiver. In order to allow fully flexible allocation of different control and data part patterns for consecutive frames, Tx and Rx control parts need to be separated from each other and from the data symbols with a guard period (GP).

![Figure 5.3-2 Example of radio subframe for 5G radio: METIS proposal](image)

The conventional method to indicate TDD frame configuration is to broadcast the information to the entire cell. LTE TDD signals the information on system information block (e.g. SIB1). In case of eIMTA, though the information is signaled on PDCCH, it is not dedicated to a UE,
and is transmitted to every UE is capable of decoding the eIMTA-RNTI. Different from legacy TDD system, the dynamic TDD system does not utilize any explicit signaling to indicate the TDD configuration, instead it signals it implicitly by using user dedicated scheduling. Figure 5.3-3 shows the concept of the implicit signaling of dynamic TDD.

One of the key enablers of dynamic TDD is the scheduling scheme. Several schemes have been studied in many works on dynamic TDD [5-7]. In [5], user throughput has been evaluated at a system level for a macro based hexagon layout. Simulation results in a network with remote radio units and a central baseband unit performing flexible UL/DL switching in a clustered manner were shown in [6]. In clustered dynamic TDD, remote radio units in a group keep the same time slot allocation pattern that can be adapted over time based on interference conditions and instantaneous traffic demand. In [7], the so called phantom cell concept is investigated in combination with a dynamic TDD scheme and a simplified system level analysis is carried out. METIS [8] also investigated two resource allocation; a completely decentralized scheme which makes independent scheduling decisions in each cell, and uses only interference knowledge from the past and a completely centralized scheme which makes globally optimal scheduling decisions based on a central network entity. Many resource allocation variants are known to exist in between, such as implicit coordination schemes [9, 10], where cells do not exchange information but base their decisions on interference measurements, or explicit coordination schemes based on the exchange of information among cells, but still with decentralized decisions [11].

**Standalone decentralized:**

The standalone decentralized scheme is a fully decentralized scheme wherein no signaling takes place between small cell access points. In this scheme, a small cell access point makes resource allocation (scheduling) as well as uplink downlink switching decisions for its users in a selfish
manner. To efficiently perform scheduling in scheduling slot \( \# t \), it is assumed that the users of a small cell feedback the signal to interference noise ratio (SINR) per resource block based on the estimate from time slot \( \# t-1 \). Thus, the standalone decentralized scheme is interference-aware based on the SINR feedback.

**Fully centralized scheme:**

In the fully centralized scheme, the scheduling and uplink downlink switching decisions are collectively made for a coordination group of neighboring cells by a central entity. It is assumed that the same entity possesses instantaneous interference information received from all the other small cells which are not in the coordination group.

Based on system level evaluation result of METIS, dynamic TDD provides 1.67 times higher throughput gain to LTE system [12], and upto 60% of latency reduction gain according to the system models [8]. In addition to latency gain, 20-60% of gains on the resource utilization can be provided. Figure 5.3-4 shows the deployment scenario to utilize the evaluation. Figure 5.3-5-Figure 5.3-6 shows the comparisons of the performance on the delay and the throughput between the fixed and flexible DL/UL allocation with the different scheduling scheme by system simulation [8]. Simulation results shows that only with flexible DL/UL switching higher performance gain is achieved. Though centralized scheduling gives more performance gain, it implies more complexity. In order to exploit centralized scheduling with low complexity, clustering with duplicated areas are proposed [6].

![Figure 5.3-4 Indoor deployment with high and low isolation (AP’s isolated by walls)[8].](image-url)
Figure 5.3-5 Exemplary DL and UL packet delay cumulative distribution for packet size 128 kB and high (dashed) and low (solid) cell isolation scenarios and four scheduling options [8].
5.3.2.3. Research Direction

The TDD optimized radio has a flexible TDD frame structure that enables support for non-conventional type of communications such as D2D and self-backhauling. Dynamic TDD can be considered with D2D communication, and DL, UL and D2D link could
be scheduled dynamically based on the traffic demand and the channel condition [13]. D2D is one of the key enablers of 5G system, and the D2D in the existing LTE is overlaid under macro LTE network with limited condition. However, in the ultra dense network in 5G system, D2D will become more dominant usage scenario, and more flexible allocation of D2D opportunity will be required. Dynamic TDD with D2D links can provide amount of gain in the point of latency as well as the throughput.

Another useful scenario of dynamic TDD is a self-backhauling [14]. Dynamic TDD also enables dynamic sharing of access and backhaul resource for 5G system. In fact, self-backhauling can be implemented with full duplexing technology in mmWave link, which is highly isolated by directional beam pattern. This flexibility enables the duplexing pattern to be dynamically optimized to current traffic loading and channel conditions. In addition, the duplexing can be adapted to local topological constraints. This adaptation is particularly valuable since the number of hops and their capacity are likely to vary significantly due to different cell sizes, propagation obstacles and availability and quality of wired backhaul.
References


5.4. Multiple Access and Waveform

5.4.1. Multiple Access

5.4.1.1. NOMA

5.4.1.1.1. General Overview

Multiple access schemes allow a large number of mobile terminals to share the allocated radio resources such as frequency, time, and code in most efficient manner. Orthogonal multiple access, in which radio resource allocated to a given terminal is not repeatedly allocated to another one, can minimize interference between simultaneously serviced terminals.

NOMA is an approach to support multi-user with common resource in power domain while orthogonal multiple access uses specific resource for a user in time domain or frequency domain[2-4]. According to QoS of data for supported terminals, a base station decides whether to share an allocated resource using NOMA scheme or not.

In NOMA scheme, each user in a set for a transmission has different power level and modulation order by their channel status. Here, each user in a set with the same resource affects the other user in the same set as interference. To enhance system performance, NOMA requires interference canceller in the receiver. Normally, a faraway terminal from the base station in a set serviced with high transmission power is operated without interference cancellation, on the other hand the other terminal near to the based station in the set serviced with low transmission power decodes data after interference cancellation using the data to the faraway terminal. This sharing resource processing can improve the system capacity up to about 30% compared to the OFDMA [4].

Figure 5.4-1 Resource allocation using OFDMA vs. NOMA
5.4.1.1.2. State of the Art and Research Trend

5.4.1.1.2.1. Downlink NOMA

NOMA is also known as a superposition coding scheme, hierarchical modulation, or layered modulation which have been studied in the broadcasting communication system. It allows for transmitting user-1 signal $x_1$ and user-2 signal $x_2$ over the same resource by superposition. In other words, the output signal of FFT which sums two signals, each with its own individual power, is given by the superposed signal $x = \sqrt{P_1} x_1 + \sqrt{P_2} x_2$ for NOMA, the received signals of user 1 and user 2 are given as

$$y_1 = h_1 x + n_1$$
$$y_2 = h_2 x + n_2$$

where $n_1$ and $n_2$ are the receiver Gaussian noises and $h_1$ and $h_2$ are the channel coefficients. Let us assume that user 1 has the better channel condition than user 2, i.e., $|h_1| > |h_2|$. Allocating more power to user 2, i.e., $P_2 > P_1$, user-2 signal $x_2$, corresponding to interference in user 1, can be detected successfully and cancelled from $y_1$, which allows for detecting user-1 signal $x_1$. The successive interference cancellation leads to the following signal-to-noise ratios (SNR) for user 1:

$$SNR_1 = \frac{P_1 |h_1|^2}{N_{0.1}}$$

where $P_2 > P_1$ and $N_{0.1}$ is the noise power density of $n_1$. Meanwhile, SNR for user 2 is given as

$$SNR_2 = \frac{P_2 |h_2|^2}{P_1 |h_1|^2 + N_{0.2}}$$

where $N_{0.2}$ is the noise power density of $n_2$. The capacities of user 1 and user 2 are represented respectively as $R_1 = \log_2(1 + SNR_1)$ and $R_2 = \log_2(1 + SNR_2)$.

Assuming the received SNR of 20 dB and 0 dB for two users, respectively, Figure 5.4-2 shows a sum rate of two users as varying the allocated power ratio. In this figure, it is clear that NOMA provides more capacity than OFDMA.
At the transmitter, the data for a set of the same resource is merged with allocated power rate for each user after the mapper, and the data of multi-user is transmitted as the form of \( N_{HP} + N_{LP} \) order constellation where \( N_{HP} \) is the constellation order of a high power allocated user and \( N_{LP} \) is the constellation order of a low power allocated user. At the receiver, the user with high power allocation ratio decodes the received signal which is contained other signal as noise in a set and the user with low power allocation ratio decodes own information after successive interference cancellation using the data of user with high power allocation ratio.

Each user in a set has the different order constellation and code rate depending on channel status, so various combinations of constellation and code rate are possible. NOMA performance comparing to OFDMA is shown in the Figure 5.4-4.
Figure 5.4-4 BLER and throughput curve depending on the different code rates of 64QAM, HP QPSK and LP 16QAM in AWGN channel.

NOMA could be one solution to avoid severe interference problem in the overlapped area between cells or beams because the interference can be used to potentiate the receiving power of the user which is in the low SINR zone. In two beams in a cell, one beam affects from the other in the middle of two beams. In this case, co-operating with two beams is possible in the overlapping area.

Figure 5.4-5 An example of cooperation in NOMA

Assuming 3 users received from two beams in a sector, we consider a weak user (UEA) with low signal-to-interference ratio (SINR) in a overlapped area by two beams is grouping with a strong user (UEC or UED) with high SINR in the center of a beam. In the case of two beams,
the optimum power ratio of strong user (UEC or UED) is the cross point of the curves of
the achievable throughput rate against OFDMA of UEA and UEC & UED according to the
power allocation ratio of the strong user. The greater of SNR gap between UEA and UEC (or
UED), the achievable data rate of increase against OFDMA scheme is getting higher.

Figure 5.4-6 Performance variation by power rate and SNR of UEs with two beams.

In NOMA, the grouping of users to maximize the efficiency of cell throughput and the power
allocation ratio in a group are important. Therefore, a base station has to take exact feedback
signals about each channel state from terminals to reduce the performance loss. In this
processing, signaling overhead between a base station and users increases and the signals are frequently occurred. Furthermore, many reference signals are used for high quality channel measurement of users in downlink.

A receiver for NOMA including SIC structure results in increase of hardware complexity and processing delay to re-decode after interference cancellation. For the simple receiver without SIC, candidate schemes for superposition transmission have been studied [9].

5.4.1.2.2. Uplink NOMA

As the demand of mobile internet traffic volume is increasing, the improvement of uplink efficiency is required as well as the downlink. Moreover, massive connectivity and cell edge throughput are important key performance indices for the next generation mobile communication systems. It is shown that the average throughput and cell edge throughput are significantly improved by adopting NOMA in the uplink [10-12]. Larger number of users can be supported simultaneously as well with uplink NOMA. Additional potential advantages of the uplink NOMA are low latency and signaling cost; By adopting the uplink NOMA, a grant-free transmission, that does not require a grant from the base station before it transmits, can be implemented [13].

Figure 5.4-7 shows the channel capacity comparison of uplink OMA and NOMA with certain AWGN channels where the number of simultaneously transmitting users is two. The capacity region of OMA can be obtained by

\[ R_1 < \alpha \log\left(1 + \frac{P_1}{\alpha N_0}\right) \]
\[ R_2 < (1 - \alpha) \log\left(1 + \frac{P_2}{(1 - \alpha)N_0}\right) \]

where \( \alpha \) is varying between 0 and 1. The capacity region of uplink NOMA system is given by

\[ R_1 < \log\left(1 + \frac{P_1}{N_0}\right) \]
\[ R_2 < \log\left(1 + \frac{P_2}{N_0}\right) \]
\[ R_1 + R_2 < \log\left(1 + \frac{P_1 + P_2}{N_0}\right) \]
From Figure 5.4-7, we can see that the capacity region of OMA is suboptimal. The maximum sum rate can be achieved by using OMA at certain operating point. However, the fairness is poor at the point where the maximum sum rate is achieved by OMA. From the figure, we can see that the fairness can be improved by adopting NOMA. It is shown that the user fairness can be enhanced by using NOMA in cellular uplink [10].

A simple example of an uplink NOMA system is described as follows: The received signal is given by \( \sum_{i=1}^{K} h_i x_i + n \), where \( h_i \) denotes the channel gain between the user \( i \) and the base station, \( x_i \) denotes the signal transmitted by user \( i \), \( K \) denotes the number of simultaneously transmitting users, and \( n \) denotes white Gaussian noise. Note that the time indices are omitted in the above equation. The received signal constellation is a superposition of multiple constellations as in the downlink NOMA. The receiver at the base station can apply a successive interference cancellation (SIC) technique to detect the messages from different users.
The procedure of SIC is described as follows: The receiver decodes the data for one user with largest channel gain. The reconstructed signal is then subtracted from the received signal. The receiver decodes the data for another user with the second largest channel gain using the remaining signal. The receiver continues this operation until the data for all users are decoded. Scheduling [12], power control [14], and user pairing [15] are important issues in uplink NOMA system design. For an uplink NOMA scheme with OFDM, a subcarrier allocation algorithm is proposed [16]. An uplink NOMA scheme with multiple antenna is investigated in [14].

When the two users connected to the same base station are communicating with each other, a PNC technique [17] can be used in the uplink NOMA. In PNC, the base station does not need to detect all messages from the users individually. Instead, it is enough for the base station to detect the XORed message, \( m = m_1 \oplus m_2 \), where \( m_k \) denotes a discrete message transmitted by user \( k \). The message \( m \) can be broadcasted so that the users can decode the messages from the other users by applying an XOR between the message it has transmitted and the one it has received.

### 5.4.1.1.3. Research Direction

To optimize NOMA scheme in a system, scheduling issues of how to group users in a cell as a set using the same resource, how many users are in a set, and how to allocate power for each user in a set are to be stressed. These are based on the received SNRs from each users, so it is necessary to obtain accurate channel estimation. Robust reference signals are allocated for the fine channel estimation. The measured channel information of the users is transmitted to base station by feedback. A base station has to handle the feedback information with heavy use of resource in uplink which degrades the cell throughput. Also, the scheduler should consider ACK/NACK feedback from the users of each set and decide how to process re-transmission. When the transmission is failed for a while, the scheduler judges whether to maintain the set or rearrange the sets in a cell or release.

In the multi-user system, successive interference cancellation has obvious advantage to obtain clear signals using re-decoding without the strong interference in the multi-user receiver, but the processing delay by sequential decoding is existed [7]. The complexity and latency of the
receiver are important for system feasibility. Modulation scheme using superposition constellation for multi-user which does not require the SIC structure could be one of solution to reduce the burden of hardware and time in the receiver.

In 3GPP RAN meeting, NOMA has been discussed in the study item called “MUST” since RAN1#80bis meeting in April, 2015. “MUST” means Multiuser Superposition Transmission for LTE and the group investigates the gain of simultaneous transmission with superposition coding considering realistic deployment scenarios, traffic model and trade-offs between system-level gain, UE complexity, signaling overhead [18].

References


5.4.1.2. Sparse Code Multiple Access (SCMA)

5.4.1.2.1. General Overview
Air interface design has always been the crown jewel and the foundation of every generation of mobile communications, while the design of waveform and multiple access technology is the very first step on its way of standardization. In current 4G systems, OFDMA techniques are used, which is an orthogonal multiple access scheme. In particular, radio resource is divided into two-dimensional orthogonal time-frequency grids in OFDMA and each grid can only be used by one user at a time. It is obvious that the number of simultaneously accessible user is strictly proportional to the number of available orthogonal resources, and is thus limited. Facing the 5G requirement of massive capacity and massive connectivity, non-orthogonal multiple access solutions have become the research focus of 5G air interface design [1]. Moreover, the requirements for IoT services are diverse and of extreme ends. On one hand, applications like smart sensors and text-based messaging are examples of extremely high volume applications that will require very low data rates and will not be sensitive to latency. On the other hand, ultra-low latency and extremely high reliability will be essential requirements for services in mobile industrial automation, vehicular connectivity, and other IoT applications. If we continue to use the current multiple access protocols in LTE, scheduling such large amount of devices mainly with small packets will consume long time and the scheduling overhead will eat up more than half of the radio resources, so it is not able to meet neither the spectrum efficiency or latency requirement. In light of this, the novel multiple access schemes in 5G are needed not only to enhance the overall uplink and downlink spectrum efficiency, but also boost the number of connections especially in uplink, and enable low-overhead, ultra-low latency, and energy efficient access protocols as well. In short, motivated by the 5G requirement to support massive capacity and massive connectivity, the following features are desired for 5G novel multiple access schemes [2]:

- Relax of the orthogonality requirement of LTE
- Resistance to multi-user interference
- Boost in overall UL and DL system SE and connectivity
- Enabler of grant-free transmission with low overhead and low latency
• Simplified scheduling and more flexible user pairing, even for high mobility users
• Enabler for the building of UE-centric no-cell systems

5.4.1.2.2. State of the Art and Research Trend

5.4.1.2.2.1. Technology and Features
SCMA [3,4] is a novel non-orthogonal multiple-access technique introduced for future 5G wireless networks. In SCMA, the coded bits from each layer is directly mapped to a multi-dimensional complex codeword, which is selected from predefined layer specific codebook. Each layer's data is spread over a block of resources in a sparse manner. Thus the sparse codewords of multiple layers are overlaid in code and power domains and carried over shared time-frequency resources. The predefined codebooks can be assigned flexibly to one user or multiple users, depending on user's rate requirement.

Figure 5.4-8 Illustration of a multi-user SCMA system

Figure 5.4-8 gives an illustration of an uplink multi-user SCMA system. Such system has the following features.

• **Non-orthogonal** Multiple modulation symbols from different users are superposed together on each RE. For example, on subcarrier 1 in Figure 5.4-8, symbols from UE1, 3, and 5 are overlapped with each other.
• **Overloading** Due to the non-orthogonal nature, the system may become overloaded if for
each SCMA block, the number of overlaid layers is larger than the length of spreading. In other words, an overloaded SCMA system can accommodate more data layers than the number of orthogonal resources. For example, in Figure 5.4-8, 6 data layers can be supporting by length 4 spreading, resulting in an overloading rate 150%. Increasing the spreading length or changing the codebook structure, we can obtain overloaded system with even higher overloading rate such as 300% or even higher, implying that 3 times more users can be packed into the same amount of resources.

• **Sparse Spreading** Different from the traditional CDMA spreading, SCMA is a new frequency domain spreading with low density. The spreading enhances the robustness of link adaptation and is good for coverage improvement, while the sparse feature helps to limit the total interference on each RE and at the same time limit the complexity of the receiver.

• **Multi-dimension mapping** Each SCMA codeword is a multi-dimensional constellation vector selected from the pre-designed layer specific SCMA codebooks. In general the non-zero spread modulation symbols from the same data layer are different but their dependency is optimized to provide a large average Euclidian and product distance between any two points on the constellation, and thus providing extra shaping gain and coding gain over low density spreading but with the simple repetition of a traditional QAM constellation on the non-zero tones. The latter way of multiple access is referred to as low density spreading (LDS) [5].

• **Compatible with f-OFDM and MIMO** SCMA can be used over OFDM or f-OFDM, so it can not only inherit the benefit of OFDM but also enjoy the flexibility offered by f-OFDM in employing different numerologies in different sub-bands. Moreover, SCMA can be easily combined with all MIMO modes defined in LTE in the way that the superposed SCMA signal is taken as one MIMO layer.

**Codebook Design**

Codebook design is the key feature that distinguishes SCMA from other non-orthogonal multiple access schemes [4]. SCMA codebooks are layer-specific, which defines the mapping rules from information bits to multi-dimensional modulation symbols for each SCMA layer.
Figure 5.4-9 shows an example of a set of 6-by-4 SCMA codebooks, one for each layer so in total 6 codebooks each with size 4-by-8, meaning each codebook consists of 8 codewords (which corresponds to 8 points constellation, may not necessarily look the same as 8PSK), and the length of each codeword is 4 (the same as spreading length). For each layer, the codeword is selected based on the input bits. Different from the traditional modulation procedure, which maps \( m \) bits to one of the \( 2^m \) modulation symbol, SCMA modulation and codebook mapping maps \( m \) bits to one of the \( 2^m \) multi-dimensional codewords. In the figure, since there are 8 codewords per codebook, 3 bits will be mapped to one codeword each time. The codewords from different layers are overlaid with each other in the air.

![Figure 5.4-9 SCMA codebook illustration: bit-to-codeword mapping.](image)

The spreading structure of SCMA can also be expressed by tanner graph. Figure 5.4-10 shows the example of the spreading structure for the 6-by-4 SCMA codebooks in Figure 5.4-9. For one SCMA block, each data layer is abstracted as a variable node (VN), and each resource element is abstracted as a function node (FN). The edges connecting VN \( i \) and FN \( j \) means that VN \( i \) has spread onto FN \( j \). Again, we can see in this example, each VN spreads onto 2 FNs and there are 3 data symbols superposition at each FN. The symbols on the two out-going branches of each VN are generated by the multi-dimensional constellation design. Iterative decoder can be applied on the graph to get all data layers decoded at the same time.
It has been shown that with multi-dimensional constellation employed in the SCMA codebook design, shaping and coding gain can be achieved to provide robust overloaded transmission and higher link quality. For instance, Figure 5.4-11 shows the multi-dimensional modulation gain that SCMA has over LDS. The left-hand-side figure shows single user SCMA and LDS with 2, 4, 6 layers respectively, 4 codeword per each codebook so equivalently 4 point modulation is employed with rate 1/2 turbo coding. We can see that SCMA outperforms LDS when there is inter-layer interference, and the larger the interference, the higher the gain (about 0.5 dB gain for 4 layers and 1 dB gain for 6 layers). This is due to the power variation for the symbol mapped on the two non-zero tones of the codebook, which helps MPA receiver to operate more efficiently to cancel inter-layer interference. On the other hand, the right-hand-side figure illustrates another advantage of SCMA over LDS, which is the shaping gain due to the fact that SCMA enjoys additional degrees of freedom in the multi-dimensional constellations design. About 2.5dB gain can be observed for 2 layers (no inter-layer interference) of 16 point constellation.
Low Complexity Receiver

As introduced above, compared with orthogonal multiple access, non-orthogonal multiple access schemes can accommodate more users and thus increase the system throughput at the cost of using much more complex joint detection algorithm at the receiver end. Low complexity receiver design is thus an important topic to implement SCMA in practical systems. Thanks to the sparseness in the codebook and the multi-dimensional constellation design, the complexity of SCMA detection can be greatly reduced and well controlled through the following two factors.

- The sparseness level of the SCMA codebook, namely the number of zero components in each SCMA codeword. With such sparseness, low complexity message passing algorithm (MPA) can be employed as the joint detector to achieve near optimal maximal likelihood detection performance [6]. The MPA detector can be further combined with channel decoder in an iterative manner to boost link quality [7].

- The low projection of SCMA multi-dimensional constellation per dimension [4]. An example of low projection codebook and the corresponding constellation mapping is shown in Figure 5.4-12. In the example, a codeword of length 4 is used, corresponding to 4-point multi-dimensional constellation. However, with the low projection technique, the constellation points on each dimension (each non-zero tone) can be reduced from 4 points to 3 points. Similar design can be used to project higher modulation codebooks (e.g., 16-point codeword to 9-point on each tone), resulting in even larger complexity reduction.
One thing needs mentioning is that, in the above example, each codeword (e.g. the codeword corresponds to coded bits “00”) has non-zero component only in one tone. Such codebook with only one non-zero component is a zero-PAPR codebook if the number of subcarriers used equals to the spreading length of SCMA codebook which is especially suitable for narrow bandwidth IoT devices.

5.4.1.2.3. Application Scenarios
As shown in Figure 5.4-13, SCMA can be applied in varied scenarios with different performance demand, including but not limited to

- Uplink overloaded access to support large number of devices and heavy traffic
- Uplink grant-free access to support large number of devices with sporadic small packets and ultra-low delay requirement
- Downlink multi-user transmission with flexible user pairing for system throughput and user experience enhancement
- Downlink multi-point coordinated transmission or interference cancellation
Robust Overloaded Access

The overloading ratio of an SCMA system is defined as the ratio between supported data layers and the spreading length. Increasing the spreading length $n$ can increase the supported number of layers in the manner of $\binom{n}{2}$ (selecting any two non-zero components in $n$ resource elements).  Keeping the link quality close to the non-overloading systems is the key requirement for the design of a robust overloaded system. Thanks to joint optimization of sparse spreading and multi-dimensional constellation in the codebook design, SCMA can support robust overloading. As we can see from Figure 5.4-14, SCMA works well in the high overloading scenario, and the link-level performance can approach the single user bound even for 300% overloading. As we can see from Figure 5.4-14, SCMA works well in the high overloading scenario, and the link-level performance can approach the single user bound even for 300% overloading. Moreover, given the same bandwidth and the same per user power constraint, to support the same number of users and the same spectrum efficiency requirement per user, SCMA has much better performance than OFDMA.
Blind Detection and Grant-free Access

Grant-free multiple access is a mechanism that eliminates the dynamic request and grant signaling overhead. Without dynamic scheduling, the devices in grant-free mode can enjoy “arrive-to-go” service and transmit at any instance. Collision is possible to happen but thanks to the nature of SCMA codebook design, SCMA based grant-free access is robust to codebook collisions [9]. Figure 5.4-15 shows the performance of SCMA with codebook collision. We can see that the performance of every two user’s codebooks collide is only less than 0.5dB worse than that of the no collision case. This is an attractive solution for small packets transmission with sporadic traffic, such as in most of the IoT service scenarios. Furthermore, a blind multi-device reception technique [10] can be applied at the SCMA receiver to detect device activities and the information carried by them simultaneously. Due to these benefits, SCMA can support massive connectivity, reduce transmission latency, save signaling overhead, and provide energy saving.
Flexible User Pairing and Open-Loop Transmission

User multiplexing over the same time-frequency resources improves the overall throughput of a downlink wireless network. Downlink multi-user SCMA is an open-loop multiplexing scheme where different code domain layers are assigned to different users without the need of full CSI knowledge of the co-paired users [8]. With a very limited need for channel knowledge such as CQI, a TP simply pairs users together with appropriate power allocation among multiplexed layers. Compared to closed-loop MU-MIMO, open-loop SCMA system is more robust against dynamic channel variations in high speed scenarios, therefore, more suitable for moving network scenarios with high data rate and mobility-robust requirements.

Open-Loop CoMP to Enable UE-Centric No-Cell Systems

One of the solutions for interference coordination in wireless networks is cooperation among TPs which is known as CoMP. Most proposed CoMP schemes are closed-loop and are based on short-term CSI feedback from users to cooperating TPs. CSI feedback can be more challenging in the future networks due to excessive number of users and TPs especially for UDN where a UE is seen by a large number of TPs. SCMA CoMP, with inter-TP layer assignment through a central scheduler, can provide an open-loop CoMP solution without
knowledge of short-term multi-TP CSI [11]. In particular, different SCMA codebook sets are assigned to different TP antennas, and terminals jointly detect the signals from multiple TPs within their CoMP coordination cluster. Such open-loop CoMP scheme will not only reduce the overhead needed for multi-TP CSI feedback, but also significant increase of the robustness to channel aging. Moreover, the open-loop nature makes TP collaboration much easier to happen which can further enable UE centric system design with dynamic and flexible TP collaboration. Figure 5.4-16 shows example application scenarios of SCMA CoMP in vehicle to infrastructure (V2X) and ultra-dense networks (UDN).

![Figure 5.4-16 Examples scenarios for SCMA based open-loop CoMP](image)

**Future Work**

As 5G standardization is going to start, SCMA has been identified as important candidate for future 5G multiple access schemes [12, 13]. Currently, the basic scenario of using SCMA for uplink transmission with 300% overloading and multi-user SCMA multiplexing have been verified in the field trial [14]. The SCMA based grant-free transmission and open-loop CoMP prototypes are soon to be built and tested. However, there is still some research work needs to be done in the near future. For instance, the constellation or codebook constrained capacity analysis is not yet solved [15, 16], and different ways to further improve codebook design [17-20] and receiver optimization [21-23] are encouraged. Moreover, protocol, pilot and control channel design for SCMA based grant-free transmission or open-loop CoMP need more standardization oriented investigation [24-26].
References

5.4.2. Waveform design

5.4.2.1. General Overview

Why new waveforms?

Orthogonal frequency division multiplexing (OFDM) has been widely researched and deployed for broadband wireless communications such as 3GPP LTE/LTE-A, IEEE 802.11, IEEE 802.16, and etc. OFDM offers a number of advantages: orthogonality amongst subcarriers, ease to implementation via fast Fourier transform (FFT) and inverse FFT (IFFT), one-tap equalization with low computational complexity, and simple extensionality to support frequency division multiple access, and multiple-input multiple-output (MIMO).

In 5G scenarios, however, traditional OFDM has several drawbacks due to its own structural characteristics [1]. First, because of the presence of cyclic prefix (CP), OFDM is not suitable for ultra-low latency systems. To achieve ultra-low latency under 1ms, the frame duration must be reduced extremely. However, to combat with the multipath channels, CP should be inserted and it causes numerous loss in spectral efficiency when the frame duration is very low. Next, the use of rectangular impulse responses in OFDM is not appropriate for spectral and fragmentation. For 5G scenarios targeting a unified frame for different types of traffics such as MTC, cognitive radio (CR), coordinated multipoint (CoMP), and etc., the very good spectral localization is required. The rectangular pulse waveform of traditional OFDM leads to undesirable magnitude responses that suffer from large side lobes in the frequency domain. The large out-of-band radiation causes severe interference when heterogeneous traffic types are simultaneously served by sharing the unified resource grid. Further, to avoid the interference OFDM waveform imposes generous guard bands and it severely deteriorate spectral efficiency. Finally, to prevent inter-carrier interference (ICI), OFDM signals should be tightly synchronized. However, dealing with synchronization is not trivial task in many communication scenarios such as uplink multiple access, machine type communication (MTC), device-to-device (D2D) communications, CoMP, and etc. Sporadic traffic generated by internet of things (IoT) should not be forced to be integrated into the bulky synchronization procedure. Consequently, OFDM has fundamental limits to support asynchronous communication systems.

To overcome these shortcomings of OFDM, it is highly expected to introduce a new waveform
technology to satisfy 5G requirements. For designing new waveform, it is desirable to satisfy following properties: robustness against ISI/ICI and synchronization error, well-localized pulse responses in time/frequency domain, suitability for short burst transmission, low out-of-band power radiation, and etc.

5.4.2.2. State of the Art and Research Trend

5.4.2.2.1. Waveform research trend

New waveform, as one of the key enabling technologies in 5G physical layer, attracts growing research attentions in recent years. In order to meet the 5G transmission requirements on ubiquitous access and explosive traffic growth, new waveform research is undergoing a paradigm shift from orthogonal to non-orthogonal design approaches. The European Union initiated an integrated project in the seventh framework program (FP7) including Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS), physical layer for dynamic spectrum access and cognitive radio (PHYDYAS), and etc., which aims to lay the foundation of 5G mobile and wireless communications system. The U.S. national science foundation (NSF) has also decided to grant NYU-Wireless and Auburn university to 'gain a deep understanding' of 5G radio communications. Meanwhile, in the industrial area, both vendors and operators are motivated to investigate their internal research programs and actively joining collaborative research projects for 5G wireless communications, including Ericsson, Huawei, T-Mobile and other partners. The research on new waveform can be classified into two groups—modified OFDM based new waveform and filter bank multicarrier (FBMC) based new waveform. The research on FBMC is initiated from PHYDYAS project of FP4 in Europe. FBMC receives many attentions because of its various benefits over OFDM such as well-localized responses in time/frequency domain, low out-of-band radiation, higher spectral efficiency, and etc. However, FBMC also has several drawbacks. Since FBMC should be combined with offset quadrature amplitude modulation (OQAM), FBMC cannot be directly applied to several conventional techniques such as channel estimation, space-time code, MIMO detection, and etc. In addition, since FBMC has long transition time, FBMC is not suitable for short burst transmission. There are various modified FBMC structures which overcome several drawbacks of conventional FBMC systems. Windowed
cyclic prefix circular OQAM (WCP-OQAM) is proposed for block transmission of FBMC signals [2] and FBMC-QAM is proposed to transmit QAM signals through FBMC waveform. By using Generalized FDM (GFDM), short burst transmission and QAM transmission are possible. On the while, as a different approach from FBMC, several modified OFDM techniques have been studied in [3][4]. To make OFDM symbols having low out-of-band radiation, Universal-filtered multicarrier (UFMC) and filtered OFDM (f-OFDM) are proposed in [3] and [4], respectively. By means of very good frequency localization property, UFMC and f-OFDM are adequate for supporting 5G unified frame for different types of traffic.

5.4.2.2.2. Advanced OFDM

5.4.2.2.1. UFMC: Universal-Filtered Multi-Carrier

Universal filtered multi-carrier (UFMC) is a recently introduced technique generalizing filtered OFDM where the filtering is done over the frequency sub-band [3][5]. The conventional OFDM requires CP to get the robustness against inter-symbol interference (ISI) from multi-path fading and temporal synchronization error. In UFMC, instead of CP, the filtering is performed per sub-band comprising multiple sub-carriers, e.g., per physical resource block (PRB).

Figure 5.4-17 depicts the transmission and reception processing of UFMC. Total M sub-carriers are divided into B sub-band groups. Then the UFMC filtering is performed to the temporal signals after passing through $M$-point IFFT. Here, the filter response for each sub-band is obtained from the prototype filter applying phase rotation. After filtering, time domain signal has ramp-up and ramp-down portion at the both ends, thus, it leads robustness against ISI in itself without additional CP insertion. However, since temporal filtering is performed by linear convolution operation, the length of a UFMC symbol increase according to the length of filter. For example, we get $M+L-1$ symbol length of UFMC after $L$-length filtering. Usually, the length of filter $L$ is equal to the length of CP in conventional OFDM. Therefore, it is difficult to expect higher spectral efficiency for UFMC compared to OFDM.
The time domain filtering is equivalent to the frequency domain windowing per sub-band. As shown in Figure 5.4-18, the frequency domain windowing suppresses the spectral side-lobe levels and thus the ICI between different resource blocks stemming from, say, lack of synchronism or carrier frequency offsets (CFOs). Therefore, UFMC has many benefits to support various 5G network scenarios such as asynchronous signaling in the uplink random access channel (RACH). Further, by means of low out-of-band radiation, the guard band of UFMC can be more shrunk than that of OFDM and consequently spectral efficiency can be improved. By means of a very similar waveform structure like conventional OFDM, UFMC ensures backward compatibility for other significant communication techniques such as MIMO transmission without additional processing.

As for the drawback of UFMC, the computational complexity for signal processing linearly increases with the number of sub-band groups. In OFDM, IFFT is performed only one time...
for entire sub-carriers, on the contrary, in UFMC, IFFT as well as filtering is performed per sub-band. As a result, at the both transmitter and receiver sides, complicated signal processing blocks are required.

5.4.2.2.2. f-OFDM: Filtered OFDM

As shown in Figure 5.4-19, filtered OFDM (Ff-OFDM) is a similar technique with UFMC that performs per-subband filtering to the OFDM signal [4]. For UFMC, in all existing paper or technical report, the subband bandwidth is one PRB (i.e., 180KHz). While for f-OFDM, subband bandwidth could be any PRB (>=1 PRB). The subband filtering leads low out-of-band radiation. Therefore, f-OFDM is much less sensitive to synchronization than OFDM, thus, f-OFDM is more suitable for asynchronous transmission.

Although the basic concept of waveform design is similar with UFMC, the way of symbol transmission of f-OFDM is different to that of UFMC. In UFMC, the symbols are transmitted with specific time interval maintaining symbol isolation. On the contrary, in f-OFDM, the symbols are transmitted consecutively and the tails of the symbols are overlapped each other. In other words, theoretically f-OFDM has non-orthogonal transmission structure including intrinsic interference. To deal with ISI caused by the intrinsic interference and multipath fading channel, f-OFDM requires CP like conventional OFDM. However, the ISI has very small power and has negligible impact on detection performance, which could be reflected on the EVM of f-OFDM. Actually, the EVM loss of f-OFDM is almost the same as CP-OFDM with a given spectrum mask and ACLR requirement. In this sense, f-OFDM is actually an orthogonal waveform.
f-OFDM, with the introduction of a subband filter on top of CP-OFDM, could inherits all the CP-OFDM transmission schemes, including pilot, MIMO, equalization etc. and overcomes the shortcomings of OOB and strict synchronization.

Only for very narrow sub-band, the time domain mainlobe of filter in the overlapping region will broaden, which means intrinsic interference becomes larger, it is required to use expanded CP as shown in Figure 5.4-20, or, a portion of original CP is used to avoid the ISI due to the filter tail without extending CP length. However, this case only happens in 1RB subband filtering for 256QAM case, which almost does not exist in the practical system.

![Figure 5.4-20 F-OFDM symbol block](image)

Since f-OFDM has very low out-of-band radiation, the guard band can be reduced thus spectrum resources can be more efficiently used for non-orthogonal and sporadic traffic signals. Therefore, f-OFDM also has appropriate structure for supporting 5G networks having unified frame structure controlled by and advanced 5G RACH.

5.4.2.2.3. Zero-Tail DFT spread OFDM

In both OFDM and SC-FDM (also known as DFT-s-OFDM), CP is essential to mitigate the multi-path fading. But insertion of CP would cause some limitation as mentioned earlier, the restriction in a specific system design. CP larger than the length needed to overcome the delay spread increases unnecessary overhead and loss of throughput, by contrast, if the CP is not sufficiently long to deal with the real delay spread observed in the system, a higher block error rate (BLER) would be measured. Thus, CP length should be designed to fit both the estimated delay spread (which depends on the particular radio environment and the related coverage objective) and the frame duration and has been a hard-coded value. It is natural that the use of an adaptive CP with length set with fine granularity to fit different channel responses is
unfeasible in practical scheduled systems due to the constraint of fixed frame duration [6]. However, the current networks are heterogeneous and so will be the future ones. Consequently, each of the composing systems may use different numerologies (for example LTE, which in several occasions throughout the document is used as a baseline, uses OFDM in DL and SC-FDM in UL and has defined short and extended CP depending on the coverage scenario). The mismatch between different numerologies and in particular, different CP lengths, will cause mutual asynchronous interference even if the neighboring nodes are synchronized on a frame basis, see Figure 5.4-21 [6]. While the advanced receiver solutions such as IRC and SIC can successfully cancel mutual synchronous interference, but are unable to cope with asynchronous one, which results in poor link performance.

The zero-tail DFT-s-OFDM signals are proposed to overcome the limitations imposed by a fixed-length CP. Specifically, to decouple the radio numerology from the radio channel characteristics by replacing the fixed length CP with a set of very low power samples (named zero-tail), which are part of the inverse fast Fourier transform (IFFT) output. The devised solution is aimed at preserving the orthogonality of the data sub-carriers at the receiver.

![Diagram](image)

**Figure 5.4-21 (a) CP-based vs. (b) zero-tailed DFT-s-OFDM signals (with different tail lengths)[6]**

The zeros at the tail of an OFDM/DFT-s-OFDM symbol, which substitute the CP, are obtained as a natural input of the IFFT instead of blanking the last samples of IFFT. This preserves the orthogonality of the subcarriers, enabling simple one-tap equalization at the receiver. The duration of the zero part $T_{so}$ fulfills the requirement imposed on CP (see section on frame numerology), so that the energy of an OFDM/DFT-s-OFDM symbol does not split over its
adjacent symbol, which maintains the cyclic property of the signal at the receiver. Such a symbol has the properties as follows [6]:

- Dynamic adaptation to the expected delay spread and propagation delay without modifying the system numerology (adaptive overhead vs hard-coded overhead).

This allows to avoid the previously mentioned limitations (throughput loss or increased BLER). The delay spread can be estimated in various ways, e.g., from periodical pilot sequences [1], and the Tso can be periodically set according to the estimations. In other words, the hard-coded overhead needed to cope with a multipath channel with an estimated fading profile is substituted by an overhead adjusted to the radio environment. In particular, UEs located near a serving BS can use lower overhead (avoid the use of large zero part) and thus will benefit of a larger throughput compared to a case when UEs located near their serving BS use the same CP as those UEs located at the cell edge (where the CP is adjusted to fit certain BLER). Likewise, if the hardcoded CP is shorter than the delay spread as seen by the edge UEs, their BLER will be higher than the one of edge UEs, which can dynamically (based on delay estimates reflecting measured data) set the zero part of the OFDM/DFT-s-OFDM symbol.

- Supporting the coexistence of systems using different numerologies – enables the possibility of achieving time alignment among systems operating over channels with different dispersion characteristics.

Since the zero samples are part of the OFDM/DFT-s-OFDM symbol itself, systems operating over different types of channels can use the same numerology [6]. When the systems are synchronized on a symbol and frame level, the interference can be managed either through advanced receivers such as IRC (to cancel mutual synchronous interference when simultaneously transmitting) or by coordinating their transmissions (to avoid interference). In the context of 5G systems, an indoor 5G cell can be set to operate in proximity with an outdoor 5G cell, and the two cells can coordinate their transmissions such as mutual interference is not generated. Further, hybrid solutions that use a short CP and zero-tail in case of longer propagation delay and delay spread, can enable coexistence with the current radio standards using CP-based solutions [6].
In summary, the usage of zero-tail signals allows decoupling of the radio numerology from the expected channel characteristics, which brings corresponding benefits in terms of overhead/BLER as well as in terms of attaining system coexistence. The zero-tail signals can be generated with a modified form of the traditional DFT-s-OFDM chain as illustrated in Figure 5.4-22 [6].

![Figure 5.4-22 Zero-tail DFT-s-OFDM signal generation [6]](image)

It is worth to note that a limited number of zeros is also appended at the head of the signal to prevent power regrowth at the tail due to the cyclicity of the IFFT [6]. While the zero-tail duration has to be adjusted to the radio environment, the zero-head represents a pure system overhead. The zero-head can be set to be extremely short (e.g., 2 samples out of 1200) without affecting the link performance.

![Figure 5.4-23 A zero-tail DFT-s-OFDM signal [6]](image)
The zero-tail DFT-s-OFDM has link performance similar to that of the traditional DFT-s-OFDM. Interestingly, zero-tail DFT-s-OFDM has also the advantage of a much better spectral containment, due to the usage of both a low power head and a low power tail, which smoothens the transition between adjacent time symbols [6]. Since the pre-DFT insertion of zeros is a trivial operation, the complexity of zero-tail DFT-s-OFDM is the same as the one of a traditional LTE uplink transceiver. Further, zero-tail DFT-s-OFDM preserves the same benefits of traditional DFT-s-OFDM in terms of robustness to hardware inaccuracies. Further details are reported in [6]. The main limitation is again the impossibility of accessing the frequency domain for frequency selective link and rank adaptation.

An analysis of the characteristics of the transmit signals is carried out first. Figure 5.4-19 shows the Complementary Cumulative Distribution Function (CCDF) of the Peak-to-Average Power Ratio (PAPR) of zero-tail DFT-s-OFDM, assuming 16QAM modulation. The performance of OFDM and DFT-s-OFDM is also included for the sake of comparison. It is well known from literature that DFT-s-OFDM exhibits lower PAPR than OFDM due to its quasi-single carrier nature [7]. This allows the transmit power amplifier to work with a lower back-off, with remarkable advantages in terms of power efficiency. Zero-tail DFT-s-OFDM introduces a PAPR penalty of around 0.5 dB due to the presence of the low power samples in the tail. However, a considerable performance margin over OFDM is preserved.

Figure 5.4-24 displays the Out-Of-Band (OOB) emissions of zero-tail DFT-s-OFDM, assuming 1200 subcarriers configuration. where $\delta$ represents the acceptable offset of power regrowth with respect to the minimum of the oscillating part of the envelope function. When the zerohead is not added, zero-tail DFT-s-OFDM has approximately the same OOB emissions of OFDM/DFT-s-OFDM. However, the presence of the zero-head leads to significantly lower OOB emissions. This is due to the smooth transition between adjacent time symbols ensured by the low power samples at both the head and the tail of the signals.

PRB-specific zero-tail DFT-s-OFDM fulfills nominally all requirements and its potential in 5G is currently under evaluation.
5.4.2.2.4 UW-OFDM

The OFDM-based communication systems require a guard interval to prevent inter-symbol interference (ISI). A cyclic prefix (CP) has been widely used in wireless communication systems adopting OFDM. However, the spectral efficiency becomes lower because the received signal in CP period is not used at the receiver.

To overcome this problem, a novel technique called UW-OFDM has been recently proposed [9]. In this technique, tails of DFT samples of each OFDM symbols have the same sequence called unique word (UW) in time domain. Thus, an explicit guard interval such as CP is not required.
UW is used as a guard interval and it can also be used for channel estimation and synchronization purpose. The correlations among the subcarriers in generating UW can be used by the receiver to improve the bit error rate (BER) as well [11]. It is shown that the UW-OFDM-based systems outperform the CP-OFDM systems in terms of the BER [10,11]. Moreover, the UW-OFDM system has much lower out-of-band (OOB) radiation compared to the CP-OFDM system [11,14].

Figure 5.4-25 shows the transmit data structures of symbols using CPs and UWs. Unlike the data structures using CPs, the data structure using UW includes the time period of UW inside the DFT interval. By including the UW time period inside the DFT interval, the explicit guard interval is not necessary while the output of the DFT at the receiver can be equalized by a one-tap frequency domain equalizer (FEQ) for each subcarrier.

Since the UW is defined in time domain, which is the output of the inverse discrete Fourier transform (IDFT) at the transmitter, it is not straightforward to generate the OFDM symbols containing the UW. There are two categories for UW generation [12].

The first category is a systematic coding. There are two approaches in the systematic coding category. The first one is a two-step approach where the time domain signal with zero UW is first generated and the UW is added to it. This approach is described by $X' = [x_p]$ and $X' + [0]_{x_u}$, where $x_p$ denotes payload and $x_u$ denotes unique word. The inputs of the IDFT...
for generating \( xp \) are subcarriers that are one of three types: data subcarrier, redundant subcarrier, or zero subcarrier. The permutation matrix should be properly designed to obtain the input of the IDFT from the concatenation of data subcarriers and redundant subcarriers. Redundant subcarrier vector is obtained by a linear combination of the data subcarrier vector. The second approach is a direct approach where the payload and the unique word are directly generated at the output of the IDFT. Here, the redundant subcarrier vector is obtained by a sum of linear combination of UW and a linear combination of data subcarrier vector.

The second category is nonsystematic coding, where the input of the IDFT is made of linear combination of the data subcarrier vector by a generator matrix. It is shown that the UW-OFDM system designed with nonsystematic coding has lower BER compared to that with systematic coding [11].

There is much research activity in both design and analysis of the UW-OFDM system. Regarding the transmitter design, the redundant carrier position selection and the optimization of the generator matrix for non-systematic coding are important design issues [11,13]. Regarding the receiver design, several linear estimators and the sphere decoding (SD) were investigated [15,16]. It was shown that the performance gain of the SD is significant over linear estimators [16]. There are several research results on power spectrum characterization of UW-OFDM system [11,14]. It was shown that the UW-OFDM system has much lower OOB radiation compared to the CP-OFDM system [11,14].

References


5.4.2.2.3. FBMC: Filter Bank Multi-Carrier System

5.4.2.2.3.1. FBMC-OQAM: Filter Bank Multi-Carrier – OQAM

Research on FBMC was started a half century ago, and its several variants devised so far can be categorized into three schemes [1]: staggered modulated multitone (SMT), cosine-modulated multi-tone (CMT), and filtered multi-tone (FMT). SMT transmits a sequence of quadrature amplitude modulation (QAM) symbols on double side-band while CMT transmits that of pulse amplitude modulation (PAM) symbols on vestigial sideband, but those have the same spectral efficiency (which will be verified using symbol density below). FMT transmits that of QAM symbols, but it employs guard bands between subcarriers, so there is loss in spectral efficiency as much as a guard band ratio. Accordingly, SMT and CMT are more spectrally efficient than FMT and OFDM. Hereafter, SMT will be focused on since it has been more extensively studied than the other two types of FBMC due to its more similarity with OFDM.

The block diagrams of FBMC and OFDM are depicted in Figure 5.4-26. In FBMC, the filtering is done per subcarrier at transmit side and the symbols are detected by matched filtering at receive side. In FBMC, the filter should be well localized in time-frequency domain to near perfect reconstruction for offset QAM (OQAM) signals. Further, according to the filter characteristic, FBMC has very low out-of-band radiation and robustness for doubly dispersive channels.

![Figure 5.4-27 Block diagrams of OFDM and FBMC](image)

Figure 5.4-27 shows an example of prototype filter response of FBMC. By means of well localized characteristic of FBMC, in time domain the CP is unnecessary anymore and in frequency domain higher spectral efficiency can be achieved by minimizing guard band. Also,
FBMC has robustness against synchronization error, thus, FBMC is very suitable for asynchronous communication environment.

In FBMC, the filtering can be done in frequency or time domain [6]. In frequency domain, the filter is implemented in the manner of frequency spreading approach. The frequency-domain signals are $K$ times oversampled. Then oversampled frequency-domain signals pass through the frequency-domain filter by using convolution operation. Here, the length of filter becomes $2K - 1$ and it leads a twofold increase in the IFFT size. Therefore, the computational complexity for IFFT operation increases. The length of FBMC symbol also increases $K$ times compared to the conventional OFDM. In time domain, FBMC can be implemented by polyphaser network (PPN) structure. The time-domain signals obtained by IFFT are multiplied with the time-domain response of filter. By using PPN structure, the computational complexity for IFFT operation can be reduced.

In FBMC, there exists transition time because of its overlap and sum structure as shown in Figure 5.4-28 FBMC symbols are transmitted after overlapping and summation with $M/2$-sample interval. Then total symbol length becomes $NM + KM + M/2$ where $K$ is overlapping factor, $M$ is the number of subcarriers, and $N$ is the number of FBMC symbols. On the other hand, when $N$ OFDM symbols are transmitted with $L$ length of CP, total length of OFDM symbols becomes $N(M + L)$. As a result, the CP overhead in OFDM increases with the number of transmit symbols, while, the transition time of OFDM is constant regardless of the
number of transmit symbols. To ensure the orthogonality among overlapped signals, the filter of FBMC should be designed to satisfy orthogonality condition. However, since the orthogonality condition can be satisfied only for real signals, the intrinsic interference occurs in imaginary parts. Because of this intrinsic interference, it is desirable to use OQAM for FBMC waveform.

![Overlap and sum structure of FBMC](image)

**Figure 5.4-29 Overlap and sum structure of FBMC**

However, FBMC has several drawbacks. First, it is very difficult to directly apply the conventional OFDM techniques or MIMO techniques using complex signals to FBMC systems. Therefore, backward compatibility of FBMC is the most important challenging issues which should be solved. Second, because of overlap and sum structure of FBMC, transition time is required. For short transmit symbols, the overhead caused by transition time can be larger than overhead of CP. Therefore, FBMC is not adequate waveform for short burst traffic signals.

### 5.4.2.3.2. Burst transmission for FBMC-OQAM

FBMC-OQAM is not suitable for short burst transmission. To overcome this drawback, various solutions are proposed in many prior works [2][7]. In [7], a modified FBMC-OQAM system is proposed such that the transition time is truncated. Since the more truncation is performed, the more signal-to-interference ratio (SIR) is degraded, there exists trade-off between SIR performance and transition time. Figure 5.4-29 shows the burst signal of FBMC and shortened burst signal of FBMC after truncation, respectively. It is shown that $M(K-1)$ truncation is possible without severe SIR distortion [7]. However, since $1/2M$ samples transition time still exists, we cannot achieve maximum spectral efficiency.
In [2], windowed cyclic prefix circular OQAM (WCP-COQAM) is proposed. In WCP-COQAM, transition time is eliminated by performing circular convolution operation for filtering. The details of WCP-COQAM are introduced in the next subsection.

Figure 5.4-30 An example of burst signal of FBMC

5.4.2.3.3. WCP-COQAM: Windowed Cyclic Prefix Circular Offset-QAM

Windowed cyclic prefix circular OQAM (WCP-COQAM) waveform is obtained after windowing and adding CP to FBMC-COQAM signals [2]. Figure 5.4-30 shows the transmitter and receiver structures of FBMC-COQAM. In FBMC-COQAM, a linear convolution operation is replaced by a circular convolution for filtering. Figure 5.4-31 depicts overlap and sum structures of FBMC-OQAM, FBMC-COQAM, and WCP-COQAM, respectively. In FBMC-COQAM structure, transition time is not occurred for transmitting KM symbols because of the circular convolution operation. However, the filtering based on circular convolution injures the well-localized property of the prototype filter. As a result, the most important benefits of FBMC which are robustness against ISI and low out-of-band radiation are lost in FBMC-COQAM. To compensate such a performance degradation, WCP-COQAM is proposed by adding CP and windowing to FBMC-COQAM. By inserting CP, the robustness against ISI and ICI is ensured. The low out-of-band radiation is obtained by using windowing operation. Further, since
WCP-COQAM also does not have transition time, WCP-CQOQAM can support short burst transmission. However, additional CP and windowing causes loss in spectral efficiency.

![WCP-COQAM transceiver block diagram](image_url)

Figure 5.4-31 WCP-COQAM transceiver block diagram

![Overlap & sum structures of FBMC-OQAM, FBMC-COQAM, and WCP-COQAM](image_url)

Figure 5.4-32 Overlap & sum structures of (a) FBMC-OQAM, (b) FBMC-COQAM, and (c) WCP-COQAM

5.4.2.2.3.4. GFDM: Generalized Frequency Division Multiplexing

Generalized frequency-division multiplexing (GFDM) is a recent physical layer scheme designed to overcome the major broadband and real-time challenges for 5G systems. GFDM is a flexible general multicarrier modulation scheme, i.e., the conventional OFDM and singe-carrier frequency domain equalization (SC-FDE) are special cases of GFDM [8].
The transceiver structure of GFDM is represented in Figure 5.4-32. In GFDM, circular convolution based filtering is performed like WCP-COQAM. Each subcarrier is filtered within a GFDM block, and filter impulse response plays an important role in the system. Figure 5.4-33 shows the frame structure and an example of filter impulse responses. The transmit filter impulse response is not restricted to be rectangular and the circular convolution is performed for filtering. As a result, burst block transmission is also possible in GFDM systems. However, the localization property of prototype filter is degraded due to circular convolution filtering, thus, CP and windowing are required. Consequently, spectrum resources can be efficiently utilized for sporadic traffic signals by means of low out-of-band radiation while the overhead of CP causes rate loss in time domain symbols. Generally, since the orthogonal condition is not considered for GFDM filter design, there is intrinsic interference. Therefore, appropriate receiving filter is required such as matched filter, zero-forcing (ZF), minimum mean square error (MMSE), and etc.
Figure 5.4-34 View of GFDM frame structure:

(a) impulse responses for transmit filter, (b) transmit matrix where it is possible to observe the time slot blocks with $K$ subcarriers.

Since GFDM can use complex QAM signals for transmit symbols, it is easy to directly combine to conventional communication techniques. Also, GFDM is suitable to supporting ultra-low latency and short burst traffic due to its block transmission structure.

5.4.2.2.3.5. FBMC – QAM system

The conventional FBM systems adopt OQAM since the orthogonality conditions are satisfied only in the real field [6]. Also, the intrinsic interference is the main obstacle for the FBMC systems. Therefore, the conventional FBMC systems cannot be used together with the conventional OFDM and MIMO techniques such as channel estimation, space-time code, detection algorithm, and etc [6]. To mitigate the intrinsic interference various approaches have been proposed [9], [10]. In [9], a MIMO maximum likelihood detection (MLD) was proposed based on the transformation of two-dimensional inter-symbol interference into one-dimensional inter-symbol interference. However, the proposed scheme in [9] cannot perfectly eliminate the intrinsic interference. In [10], a FFT-FBMC system was proposed in order to avoid the intrinsic interference. However, the FFT-FBMC system has the loss of spectral efficiency due to the addition of CP and therefore has also trade-off between the BER performances and the CP lengths compared to the OFDM systems.
For the other approach, a new FBMC-QAM system with two prototype filters is proposed in [11] as shown in Figure 5.4-35. The proposed FBMC-QAM uses two prototype filters. One prototype filter is used for the even-numbered sub-carrier symbols and the other prototype filter is used for the odd-numbered subcarrier symbols. This individual filtering method can enable FBMC to use QAM symbols without the intrinsic interference. However, the OOB of the proposed FBMC-QAM in [11] is even worse than CP-OFDM. It is not possible to get a modulation scheme holding at the same time the following properties: orthogonality in the complex field, time/frequency localization, and symbol density. The proposed FBMC-QAM gives up the frequency energy confinement among these filter properties in order to transmit QAM symbols without the CP (see Figure 5.4-36). FBMC-QAM in [11] is improved in [12] by considering trade-off between frequency localization and intrinsic interference. Figure 5.4-37 depicts the FBMC-QAM system using multiple prototype filters in [12]. But the intrinsic interference between subcarriers is worsen although the OOB performance is improved, which means there exists obvious performance gap for medium and high MCS between FBMC-QAM in [12] and CP-OFDM for data transmission. This kind of intrinsic interference due to non-orthogonal WF also makes equalization, channel estimation, MIMO and non-orthogonal multiple access rather complex.
5.4.2.3. Research Direction

5.4.2.3.1. Comparisons (Cons and Pros, Backward compatibility)
Comparisons between OFDM and new waveforms are summarized in Table 5.4-1.
First, let us compare OFDM with new waveforms in terms of spectral efficiency. The loss in spectral efficiency comes from guard band in frequency domain and overhead in time domain in conventional OFDM. All new waveforms are proposed to improve the out-of-band emission (OOBE) in frequency domain, but they have various OOBE performance, hence leads to different improvement in spectral efficiency. f-OFDM and FBMC-QAM/OQAM outperforms
other WFs and enable a rather high spectrum utilization (e.g. 1% guard band needed for a 20MHz system bandwidth in the evaluation). GFDM is much worse than other new WFs due to the circular filtering operation, although it is slightly better than OFDM. UFMC, due to its CP length filter constraint, needs more guard band than f-OFDM and FBMC-QAM/OQAM WFs (additional 5% in the evaluation) in order to meet the spectrum mask requirements.

In time domain, FBMC-OQAM and FBMC-QAM do not require CP, so they have no CP overhead in time domain. While the problem for FBMC family waveforms is that they have long filter tail (usually 4 symbols duration), which will introduce extra time domain tail overhead between TTIs, therefore, guard interval has to be reserved in order to accommodate the tail per TTI. Moreover, the GI per TTI makes frame structure design a little bit complex. For UFMC, there is no filter tail overhead due to its CP length filter design. And for f-OFDM, the filter overhead only exists in the DL/UL switching point for TDD frame structure, and could be absorbed into GP. Even though GFDM and WCP-COQAM use the CP, the CP is inserted with different way compared to OFDM. For OFDM, the CP is added to every OFDM symbol. On the other hand, the CP of GFDM (or WCP-COQAM) is inserted only one time for M overlapped symbol. Therefore, GFDM and WCP-COQAM have lower CP overhead than OFDM. However, the cyclic convolution operation in GFDM and WCP-COQAM will worsen the OOB, some extra time domain window has to be added in order for OOB improvement, which again introduces extra time domain overhead which depends on the guard band overhead requirements. Thus, the time domain overhead for all new WFs has to be carefully
studied, which includes the CP, filter tail overhead and time domain window overhead together, rather than only CP overhead. Furthermore, the impact of time domain overhead on various frame structure design has to be considered.

EVM performance should also be evaluated in order to compare the orthogonality of various WFs, given the spectrum mask and ACLR constraints. In the table, the EVM for different WFs are compared given a 20MHz system bandwidth and 99% spectrum utilization, with 3GPP spectrum mask and EVM requirements. The OFDM, f-OFDM, FBMC-OQAM, UFMC and windowed-OFDM has negligible EVM loss, and could be treated as orthogonal WFs (The orthogonality of FBMC-OQAM only happens in real domain). For FBMC-QAM in [11] and [12], and GFDM, the EVM due to the WF non-orthogonality far exceeds 3GPP specification on EVM tolerance. And advanced non-linear receiver has to be used to suppress the severe self-inference. Furthermore, the kind of intrinsic interference due to the WF design also makes the pilot design, MIMO application and non-orthogonal multiple access application very complex.

<table>
<thead>
<tr>
<th>Waveform</th>
<th>LTE(OFDM)</th>
<th>f-OFDM</th>
<th>FBMC-OQAM</th>
<th>FBMC-QAM1</th>
<th>FBMC-QAM2</th>
<th>GFDM</th>
<th>UFMC</th>
<th>W-OFDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVM (64QAM)</td>
<td>0.57%</td>
<td>0.85%</td>
<td>0.65%</td>
<td>28.24%</td>
<td>10.07%</td>
<td>24.8%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

In terms of short burst transmission, it is not desirable to use FBMC. Because of the overlap and sum structure, FBMC has long transition time. Therefore, the spectral efficiency of FBMC is significantly degraded for the short burst traffic environment.

Basically, the computational complexity of new waveform techniques is much higher than conventional OFDM. For FBMC, GFDM, and WCP-COQAM, the IFFT/FFT size increases by overlapping factor K. Although it is possible to reduce the IFFT/FFT size by using PPN structure, the complexity is still higher than OFDM. Further, the complexity of UFMC and f-OFDM linearly increases with the number of sub-band groups. f-OFDM has lower complexity than UFMC for wideband application due to the fact that UFMC requires per PRB filtering. For backward compatibility, only FBMC-OQAM has a obvious problem. Since all new waveforms except for FBMC-OQAM can transmit QAM signals, the conventional communication
techniques including channel estimation, space-time code, MIMO detection, and etc., can be combined easier compared with FBMC-OQAM. However, for current FBMC-QAM WFs, some advanced non-linear receiver has to be used to combat the intrinsic interference. In addition, the intrinsic interference also makes the pilot design and MIMO detection more complex than OFDM. For FBMC-OQAM, it is required to design a new MIMO transmission techniques which can be operated by using OQAM.

As stated previously, the time overhead should include CP and the GI due to filter & window tail. For FBMC-QAM/OQAM, and WCP-COQAM WFs, some extra guard interval (GI) has to be reserved per TTI in order to accommodate their filter tail or window overhead, despite of CP overhead saving. Therefore, the time overhead has to be calculated considering the CP, filter tail and window overhead. It’s not correct to say OFDM has high time overhead, since OFDM has no tail or window overhead.

### Table 5.4-1 Summary of Wireless Network Requirements

<table>
<thead>
<tr>
<th></th>
<th>Frequency Localization</th>
<th>Time Overhead</th>
<th>Out-of-band Radiation</th>
<th>Short Burst Transmission</th>
<th>Complexity</th>
<th>Backward Compatibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>OFDM</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Yes</td>
<td>Low</td>
<td>Yes</td>
</tr>
<tr>
<td>UFMC</td>
<td>Moderate</td>
<td>Variable</td>
<td>Reduced</td>
<td>Yes</td>
<td>High</td>
<td>Yes</td>
</tr>
<tr>
<td>F-OFDM</td>
<td>Very high</td>
<td>Variable</td>
<td>Negligible</td>
<td>Yes</td>
<td>High</td>
<td>Yes</td>
</tr>
<tr>
<td>FBMC-OQAM</td>
<td>Very high</td>
<td>Low1)</td>
<td>Negligible</td>
<td>No</td>
<td>High</td>
<td>No</td>
</tr>
<tr>
<td>GFDM</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
</tr>
<tr>
<td>WCP-COQAM</td>
<td>Moderate</td>
<td>Low</td>
<td>Reduced</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
</tr>
<tr>
<td>FBMC-QAM</td>
<td>Low or moderate</td>
<td>Low1)</td>
<td>Reduced</td>
<td>No</td>
<td>High</td>
<td>No</td>
</tr>
</tbody>
</table>

1) Especially when the number of consecutive transmitted symbols is extremely large, the time overhead could be negligible

### 5.4.2.3.2. Research Direction

Designing new multicarrier waveform becomes a key physical layer technology components for 5G communication systems. Through many research project groups, e.g. PHYDYAS, METIS, 5GNOW, and etc., a number of research outputs have been developed related to new
waveform techniques such as new filter design, numerous link level performance evaluation, and combining with other communication techniques. For practical implementation, it is very important to evaluate the performance of new waveform through extensive system level simulations. Also, the efforts of standardization for new waveform techniques are required such as new physical layer frame structure, control signaling, reference signal design, modification in modulation and coding scheme, and etc.

One of the most important issue in new waveform is related to multiple access technique. Combining multiple access technique, new waveform can have many benefits compared to OFDM such as low complexity and low overhead [13]. As a part of study on multiple access, combing new wave form with non-orthogonal multiple access (NOMA), e.g. interleaved division multiple access (IDMA), has been studied to achieve massive connectivity and high reliability [14]. Also, by using the well-localization characteristic of new waveform, enhanced CR technique should be studied to more efficiently use spectrum resources [15]. According to the low out-of-band emission, the interference constraint can be more relaxed in new waveform systems and it can change various conventional CR techniques in terms of spectrum sensing, spectrum sharing, and dynamic access.

It is already discussed that asynchronous transmission is required for sporadic cellular IoT environment [16]. The robustness against synchronization of new waveforms enable to implement asynchronous networks. The communication among IoT nodes such as D2D, M2M, and MTC arise in the manner of random access like wireless sensor networks. Usually, these nodes are controlled in the manner of distributed algorithm. By using new waveforms, synchronism-insensitive networks can be achieved.

To achieve ultra-low latency networks, block transmission of new waveforms should be more discussed. Current block transmission schemes such as WCP-COQM have several drawbacks like enormous computational complexity and low energy confinement. Also, implementing ultra-low latency networks requires cross-layer optimization between new waveform design and MAC protocol.
References


5.5 Large Scale Antenna

5.5.1. Large Scale Antenna Below 6GHz

5.5.1.1. General Overview
Large scale antenna system has attracted significant attention in the wireless industry and academia in the past few years as one of candidate technology for the next generation evolution towards 5th generation (5G) cellular systems. It can extend a large number of antennas placed in a two-dimensional (2-D) antenna array panel for realizing spatially separated transmission links to a large number of mobile stations (MSs). The arrangement of these antennas on a 2-D panel can possibly allow the extension of spatial separation to the elevation domain as well as the traditional azimuth domain in MIMO systems. This section discusses the features and observation of performance benefits on a large-scale antenna system along with the ongoing efforts in standard organization (e.g. 3GPP) to incorporate large-scale antenna features in the next evolution of cellular systems. Furthermore, a design of a 2-D antenna array which plays one of the key roles in the implementation of a large scale antenna system for below 6 GHz frequency band is also discussed.

5.5.1.1.1. Trends
In response to the increase in wireless data traffic, wireless technology has focused much of its efforts in providing cutting-edge techniques to improve spectral efficiency and user experience. Among such techniques are multi-input-multi-output (MIMO), coordinated multi-point transmission/reception (CoMP), and carrier aggregation (CA). CoMP relies on coordination between multiple transmission and reception points to enhance UE performance at cell edges but requires a very capable backhaul connection for inter-site coordination. CA simultaneously utilizes multiple frequency bands to enhance peak data rate and network’s load balancing capability but requires the use of large frequency resources. Although each of these techniques represents a major step forward in improving system performance, further developments of new technologies are required to meet an exponentially growing demand for wireless data traffic.
Large scale antenna system is one of the key technologies currently studied in the industry and academia for the next generation wireless system. By incorporating large-scale antenna system into current network system (e.g. LTE systems), it is expected that system throughput will be drastically improved beyond what it is possible in conventional systems. Compared to CoMP and CA, large scale antenna system is capable of enhancing system performance without requiring a very capable backhaul or large frequency resources. This section provides details of large-scale antenna system below 6 GHz in terms of possible deployment scenarios, practical 2-D antenna array implementation, and possible enhancements to the current standards.

5.5.1.1.2. Concept of Large Scale Antenna Below 6 GHz

The key source of performance enhancement by utilizing large-scale antenna system is its ability to handle high-order MU-MIMO. Compared to conventional systems where the maximum number of MU-MIMO co-scheduled UEs is limited to two or four with 4 or 8 antennas, respectively. Large-scale antenna system is capable of supporting a significantly larger number of MU-MIMO UEs with a larger number of antennas. Consider a system with $N_T$ transmit antennas at eNB, $K$ co-scheduled UEs, and downlink transmission power of $P$. With channel conjugate precoding, the received signal for the $k$th UE can be derived as

$$y_k = \sqrt{\frac{P}{N_T K}} h_k h_k^* x_k + \left( \sqrt{\frac{P}{N_T K}} \sum_{l \neq k} h_l h_l^* x_l + n_k \right)$$

where $x_k$ is the transmitted signals for the $k$th UE, $h_k$ is the downlink channel for the $k$th UE and $n_k$ is the Gaussian noise at the $k$th UE’s receiver. Theoretically, as the number of antennas increases, the cross correlation of two random channel realizations converges to zero as shown in [1]:

$$\lim_{N_T \to \infty} \frac{h_k h_l^*}{N_T} = \delta_{kl}$$

where $\delta_{kl} = 1$ if $k = l$ and $\delta_{kl} = 0$ otherwise. As a result, assuming a large $N_t$ and $K$ ($K < N_t$), the average signal-to-interference plus noise ratio (SINR) for each UE can be approximated as
for the case where inter-user interference is significantly larger than the noise variance. The above analysis is for the downlink but a similar analysis can be applied to the uplink multiple access channel [1]. Although the simple analysis above is based on an ideal signal model, important insights can be obtained. From the equation above, it can be observed that the SINR at each UE linearly increases as a function of the number of antennas. Additionally, if the number of antennas increases at the same rate as the number of co-scheduled UEs, the same SINR can be maintained. In other words, if the number of transmit antennas increases by a factor of $G$, the number of UEs that can be co-scheduled using the same wireless resource can also increase by a factor of $G$ without any sacrifice on SINR. For example, if the number of antennas increases from 10 to 100 while the number of UE's increases from 2 to 20, a tenfold system capacity increase can be achieved. Note that the above analysis assumes an uncorrelated channel at the transmitter and a very large number of transmit antennas. However, actual wireless channels have some degree of correlation depending on the environment or antenna implementation and an eNB has a finite number of antennas.

Figure 5.5-1 Concept of large-scale antenna system below 6GHz
Based on the analysis, large-scale antenna system utilizes multiple antennas placed in a 2-D antenna array panel to realize high order MU-MIMO transmissions. High order MU-MIMO refers to the use of a large number of antennas at the base station to transmit or receive spatially multiplexed signals to or from a large number of terminals. Figure 5.5-1 depicts an enhanced base station with large-scale antenna capability transmitting simultaneously to multiple user equipment. The antennas at the eNB are placed on a 2-D antenna array panel where every antenna is an active element. These active antenna elements allow dynamic and adaptive precoding to be performed jointly across all antennas. As a result of such precoding, a base station can realize more directed transmissions in the azimuth and elevation domains simultaneously to a larger number of UEs.

Compared to MIMO transmissions of conventional LTE systems, large-scale antenna system has two important differentiating factors. Firstly, the number of antennas has been increased beyond what is supported in conventional systems (up to 8 antennas). As a result, beamforming and spatial user multiplexing capability can be significantly improved. Secondly, the antennas are no longer assumed to be placed in a one-dimensional linear array but in a 2-D planar array. The main motivation of the planar placement is to reduce the form factor of the antennas to be more practical. For example, supporting 64 antennas at 2.5 GHz in an 8-by-8 planar array with 0.5 λ spacing would result in a form factor of 50 cm by 50 cm. However, if the antennas are placed in a linear array, the array would be 4 m wide making it unpractical. While a planar array does reduce the effective spacing between different antenna elements compared to a linear array, it provides the benefit of being able to extend spatial separation to the elevation domain as well as the traditional azimuth domain. More details on the design of the 2-D antenna array and its impact on the system performance are presented in the following sections.

5.5.1.1.3. Possible Deployment Scenarios for Large Scale Antenna Below 6 GHz

In order to fully exploit the enhanced beamforming and spatial user multiplexing of large-scale antenna system, a base station with large-scale antenna capability should be deployed in scenarios where such characteristics can provide system performance enhancement. Figure 5.5-2 shows some examples of such system deployment in urban micro, urban macro, high...
rise, and high population density scenarios.

In practical situations, most UEs in urban locations are indoors on different floors. Having the capability to control the beam direction in the elevation domain as well as the azimuth domain presents new opportunities to enhance the system performance in such scenarios. One important scenario is the urban outdoor to indoor scenario between an outdoor eNB and indoor UEs on different floors. Transmissions originating from the outside of the building to UEs located on different floors can be better separated using beamforming in the elevation direction.

Another scenario that is of importance is the high population density scenario where a large number of UEs are closely located with one another in a hot zone. Examples of such high population density scenario are:

- Shopping malls
- Stadiums or concert halls
- Transportation hubs such as major airports or train stations

A key characteristic of the high population area scenario is that a large number of people are located in a limited area generating high traffic demand simultaneously. Typically, in such scenarios, hundreds or even thousands of UEs in a hotzone can simultaneously try to access the cellular system leading to severe QoS instability. MU-MIMO transmission from the 2-D antenna array can be made simultaneously for multiple UEs in such scenarios taking advantage of the additional beam directivity. For example, in a shopping mall with a high ceiling, the 2-D antenna array can be positioned on the ceiling facing downwards to provide high order MU-MIMO transmission.
5.5.1.4. Two Dimensional Active Antenna Array Below 6 GHz

In order to realize the benefit of large-scale antenna in a practical shape of a system, an efficient implementation of a 2-D antenna array is a key requirement. A 2-D antenna array should be designed such that active antenna elements in horizontal and vertical placements can provide adaptive precoding in both azimuth and elevation domains. In doing so, the 2-D array should have a form factor that is adequate for actual deployment while providing sufficiently efficient radio frequency characteristics.

An actual functioning example of a large-scale antenna array configuration is shown in Figure 5.5-3. The array comprises four stacked panels, each having eight sub-arrays arranged in 8-by-1 (8-horizontal by 1-vertical) configuration. The spacing between two adjacent sub-arrays is $d_H = 0.5\lambda$ in the horizontal direction and $d_V = 2\lambda$ in the vertical direction (between centers of adjacent sub-arrays). Each sub-array is composed of four patch antenna elements arranged in a 1-by-4 configuration and fed with equal magnitude and phase by a single feed port. Thus, the example antenna array has a total of 32 feed ports (32 channels) and a form factor of approximately 1 m by 50 cm.

One of the key features of this antenna array configuration is that the patch antenna elements
are disposed in the $\phi = \pm 45^\circ$ directions which results in dual-linear polarization on the two diagonal planes ($\phi = \pm 45^\circ$ with reference to the x-y coordinate system shown in Figure 5.5-3). Due to this configuration, the $\pm 45^\circ$ and $-45^\circ$ polarized sub-arrays have the same beamwidths in the elevation ($\phi = 0^\circ$) and azimuth ($\phi = 90^\circ$) planes and are affected more alike by the channel characteristics than a $0^\circ$ and $90^\circ$ dual-polarized array version. Notice also that the $\pm 45^\circ$ and $-45^\circ$ sub-arrays are interlaced along both horizontal and vertical directions so as to increase isolation between adjacent sub-arrays (since they are orthogonally polarized). Thus, scanning the array beam in the azimuthal plane, for example, could involve sub-arrays from two adjacent panels (e.g., the four $\pm 45^\circ$ from the top panel and the four $\pm 45^\circ$ from the panel below it), if the $d_H = 0.5\lambda$ spacing is to be maintained.

Figure 5.5-3 Active array and feed network example
The patch elements of each sub-array are fed through a corporate microstrip line feed network printed on the bottom layer of the feed board. Energy is coupled to the patches through rectangular slot cutouts on the ground plane, on the other side of the feed board. This feeding technique provides better bandwidth, higher isolation between adjacent patch elements, and also more flexibility in adjusting the air-gap between the antenna and feed board (see board stack-up detail in Figure 5.5-3), than the conventional probe feeding. The air gap between the antenna board and the feed board (ground plane) is tuned so as to maximize the bandwidth and achieve the specified gain. Finally, a low-loss radome covers the antenna and system enclosure and protects it from the elements.

For consistent radiation, it is important that all sub-arrays are phase matched, i.e. the difference of electrical lengths from the feed ports to the patch antennas should not exceed more than $1^\circ$-$2^\circ$. To ensure phase matching, all sub-arrays have the same feed network adjusted to fit both $\pm45^\circ$ rotated patches by merely mirroring the microstrip sections, as seen in Figure 5.5-3 (mirroring sections in our case does not change the electrical length). Further, within each 4-element patch sub-array, all microstrip sections were phase matched to better than $1.4^\circ$, from the common feed port to the patch antenna.

The antenna panel is designed for one candidate band to support LTE TDD#41 (2.496 – 2.69 GHz) and is designed to achieve a gain of about 10 dBi per sub-array with beamwidth of 25$^\circ$ and 65$^\circ$ in the azimuth and elevation domains, respectively. Figure 5.5-4 to Figure 5.5-7 show a picture of a fabricated active array panel and measurement results in an anechoic chamber environment. Each sub-array has lower than -15 dB reflection coefficient in the target band while the mutual coupling between adjacent sub-arrays is below -20 dB. Measured co-phase and cross-pol radiation patterns for one of the eight sub-arrays on the azimuth and elevation planes at 2.6 GHz are also shown in Figure 5.5-7. The sub-array co-pol gain is about 10 dBi at $\theta = 0^\circ$ and the cross-pol gain about 10 dB below that. The beamwidths at azimuth and elevation planes are about 65$^\circ$ and 24$^\circ$, respectively.
Figure 5.5-4 Antenna calibration in an anechoic chamber

Figure 5.5-5 Magnitude of reflection coefficients of 8 sub-arrays
5.5.1.2. Progress in Standard Organization (3GPP)

Full Dimension MIMO (FD-MIMO) [2] is one of large-scale antenna technologies currently studied in the 3GPP for the next generation long-term-evolution (LTE) systems. As a first step,
a study item [3] has been initiated to study a new channel model under which future evaluation of the antenna technologies will be performed. Follow-up 3GPP study and work items on FD-MIMO are expected in early 2014 focusing on providing specification support for the technology. The focus of the study and work items will be to identify key areas in the LTE specification that needs to be enhanced in supporting up to 64 antenna ports placed in a two-dimensional (2-D) array. By incorporating FD-MIMO into LTE systems, it is expected that system throughput will be drastically improved beyond what it is possible in conventional LTE systems. Compared to CoMP and CA, FD-MIMO is capable of enhancing system performance without requiring a very capable backhaul or large frequency resources.

Geometry-based stochastic channel models have been developed and refined over years by a number of research groups such as 3GPP, 3GPP2, ITU and the WINNER initiative [5-6]. Spatial channel model (SCM) [4], an example of a geometry-based stochastic channel model [6], is widely used in the 3GPP community to evaluate performance of different wireless technologies. Traditional SCM used in the 3GPP community is a 2-D channel model, where an elevation angle of each signal path is always assumed to be zero. While such an approach is acceptable for evaluating performance of systems with horizontally placed linear antenna arrays, modeling of elevation angles is necessary when evaluating a FD-MIMO technology utilizing a 2-D antenna array.

In a three-dimensional (3-D) SCM, each signal path has to be modeled with an elevation angle as well as an azimuth angle. A 3-D spatial channel model takes into account the wireless channel propagation in the elevation direction as well as the azimuth direction. One of the main challenges is to model the correlation of large scale parameters as well as the statistical distribution of elevation angles. The large scale channel parameters, such as ASD (azimuth spread at departure), ASA (azimuth spread at arrival), ESD (elevation spread at departure), ESA (elevation spread at arrival), shadow fading, Rician K-factor, and delay spread, have been shown to be correlated. As a result, for a terminal, the cross-correlations of these parameters must be measured and modeled. In addition, the large scale parameters are correlated for different terminals as well if the terminals are closely located. These cross-correlations have not been as extensively measured and reported in the literature. Nevertheless, in some references (e.g. WINNER+), the elevation spread is assumed to have the same spatial correlation as the
azimuth spread; such approximation is considered reasonable, since the azimuth and elevation spreads originate from the same clusters, and their autocorrelations may behave in a similar manner. Further study and measurements are needed to confirm this assumption and determine whether or not the same approach could be taken for the other elevation parameters.

In some references (e.g., WINNER II [7]), the distribution of the elevation spread is assumed to be wrapped Gaussian, which is symmetric around the mean. However, in other measurements, it is observed that the distribution of elevation angles is asymmetric, and thus the distributions that reflect the asymmetric nature of elevation angles are considered. This aspect is reflected in the WINNER+. A double exponential (or Laplace) distribution is proposed in WINNER+ which models such skewness by using two standard deviations (left and right). A proper modeling of the mean and variance is another key challenge of extending a 2-D channel model to a 3-D channel model.

Currently, in 3GPP, a study item has been finalized the details of the three-dimensional (3-D) channel model. The following key topics are included in the new channel model; generic steps of channel generation, height dependent pathloss generation, correlation of large scale parameters, and statistical distribution of elevation components. The outcome of the study item will be used in the future evaluations for the standardization of not only FD-MIMO but also other advanced antenna technologies that utilize 2-D antenna arrays.

5.5.1.3. Research Direction

FD-MIMO technology can take the benefit of reduced form factor for higher frequency spectrum. Taking into account that WRC-15 is expected to allocate new licensed spectrums up to 6 GHz, it would be necessary to investigate further evolution of FD-MIMO with more number of antennas, e.g., around 100 or even more, for near-term evolution of the current LTE-Advanced system. Further evolution of FD-MIMO concept for using a few hundred antennas toward introduction of even higher frequency spectrum such as mmWave bands in 5G mobile communication systems would be a longer term research direction.
References


[6] 3GPP TR 25.996 V10.0.0, Spatial channel model for Multiple Input Multiple Output (MIMO) simulations.

5.5.2. Large Scale Antenna System Above 6 GHz

5.5.2.1. General Overview

To cope with the increasing traffic demands in wireless mobile communication systems, the millimeter wave (mmWave) has been paid attention as a candidate frequency band. However, unlike the concept of large scale antenna system in below 6 GHz, the large scale antenna system based on mmWave focuses on the sharp beamforming for coverage, which highlights the antenna calibration issue [1]. In addition, the rank deficiency to a single UE, enabling the data multiplexing, leads MU-MIMO technology more feasible when the wireless channel conditions in mmWave band are considered. Until now, a few types of beamforming structures called hybrid beamforming have been proposed [2], where the two different types of antenna configuration are assumed. One is singular typed patch array antenna, and the other is modular typed patch array antenna.

Table 5.5-1 Parameters for link budget calculation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center freq.</td>
<td>20 GHz</td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>1 GHz</td>
<td></td>
</tr>
<tr>
<td>Rainfall</td>
<td>47 mm/h</td>
<td>99.99%</td>
</tr>
<tr>
<td>5G-BS output power</td>
<td>27 dBm</td>
<td>Avg. output power/Beam (ref. 3GPP medium range BS Power class)</td>
</tr>
<tr>
<td>UE output power</td>
<td>23 dBm</td>
<td>3GPP UE class 3</td>
</tr>
<tr>
<td>Noise Figure</td>
<td>8 dB</td>
<td></td>
</tr>
<tr>
<td>Required SNR</td>
<td></td>
<td>code rate 2/3</td>
</tr>
<tr>
<td></td>
<td>4 dB (QPSK)</td>
<td>(ref. BLER = 0.1)</td>
</tr>
<tr>
<td></td>
<td>10 dB (16QAM)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 dB (64QAM)</td>
<td></td>
</tr>
<tr>
<td>Antenna gain</td>
<td></td>
<td>5G BS (H: 22.5°, V: 15°)</td>
</tr>
<tr>
<td></td>
<td>17.8 dB</td>
<td>UE H: 45°, V: 30°</td>
</tr>
<tr>
<td>Path loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2 dB/Km</td>
<td></td>
</tr>
<tr>
<td>Rainfall loss</td>
<td></td>
<td>57 mm/h (ref. V ray)</td>
</tr>
<tr>
<td></td>
<td>6.59 dB/Km</td>
<td></td>
</tr>
<tr>
<td>System margin</td>
<td>6 dB</td>
<td>LOS</td>
</tr>
</tbody>
</table>
Referring to Table 5-6 and Figure 5.5-8, the beamforming gains at Tx and Rx sides can achieve the cell coverage in mmWave based wireless mobile communication systems. For line-of-sight case the maximum communication range is restricted to about 300 m, where 64-QAM 2/3 code rate and 16-QAM 2/3 code rate are assumed for downlink and uplink, respectively. After all, without beamforming technology, its cell coverage is bounded to ultra pico cell range of several tens of meters.

5.5.2.2. Concept of Large Scale Antenna Above 6 GHz

Compared to the concept of large scale antenna below 6 GHz, a greater number of antenna elements can be easily accommodated in the given size of antenna aperture at mmWave frequency band. It is noted that a greater number of antenna elements can make it possible to generate sharp beamforming to the desired direction and prevent the loss of power, and effective to handle the interference management.

Classification of Antenna Structure

- Modular Typed Antenna Structure
- Singular Typed Antenna Structure

Candidates of mmWave based 3D Beamforming Technology

- Fixed Beamforming
• Analog Fixed Beamforming
• Adaptive Beamforming
  - Analog Adaptive Beamforming
  - Digital Adaptive Beamforming
  - Hybrid Adaptive Beamforming

Depending on the configuration of a patch array antenna, there are two different types of patch array antennas. Figure 5.5-9 shows a modular typed antenna structure while a singular typed antenna is in Figure 5.5-10. In Table 5-7, the feature of each structure is summarized. Since each antenna type has advantages and disadvantages itself, the antenna type should be carefully selected according to the use cases of the large scale antennas.

![Figure 5.5-9 Modular typed antenna structure](image)

**Figure 5.5-9 Modular typed antenna structure**

![Figure 5.5-10 Singular typed antenna structure](image)

**Figure 5.5-10 Singular typed antenna structure**
### Table 5.5-2 Comparison of singular and modular typed antennas

<table>
<thead>
<tr>
<th>Good</th>
<th>Modular type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular type</td>
<td>Burden of electrical steering can be reduced through mechanical tilting</td>
</tr>
<tr>
<td>• Beam design with various</td>
<td>✓ Obtaining beams that are even in all angles is easy</td>
</tr>
<tr>
<td>shape and performance is</td>
<td>✓ Simple beam-forming is possible</td>
</tr>
<tr>
<td>possible through electrical</td>
<td>• Low burden for antenna calibration due to wide</td>
</tr>
<tr>
<td>steering</td>
<td>beam-width</td>
</tr>
<tr>
<td>✓ Pin-point beamforming is</td>
<td>• Low hardware complexity (# of phase shifters)</td>
</tr>
<tr>
<td>possible</td>
<td>(= (\text{# of antenna groups}) \times (\text{# of antennas/group}))</td>
</tr>
<tr>
<td>✓ Fine steering in the region</td>
<td>(\times (\text{# of antennas}))</td>
</tr>
<tr>
<td>where there is no</td>
<td>((\text{# of adders}) = 0)</td>
</tr>
<tr>
<td>side-lobe issue</td>
<td></td>
</tr>
<tr>
<td>✓ Interference management</td>
<td>• Managing interference among neighboring beams</td>
</tr>
<tr>
<td>among neighboring beams is</td>
<td>easy</td>
</tr>
<tr>
<td>easy</td>
<td>• Fine steering is not possible</td>
</tr>
<tr>
<td>• Burden of electrical</td>
<td></td>
</tr>
<tr>
<td>steering can be reduced</td>
<td></td>
</tr>
<tr>
<td>through mechanical tilting</td>
<td></td>
</tr>
<tr>
<td>• Obtaining beams that are</td>
<td></td>
</tr>
<tr>
<td>even in all angles is easy</td>
<td></td>
</tr>
<tr>
<td>• Simple beam-forming is</td>
<td></td>
</tr>
<tr>
<td>possible</td>
<td></td>
</tr>
<tr>
<td>• Low burden for antenna</td>
<td></td>
</tr>
<tr>
<td>calibration due to wide</td>
<td></td>
</tr>
<tr>
<td>beam-width</td>
<td></td>
</tr>
<tr>
<td>• Low hardware complexity</td>
<td></td>
</tr>
<tr>
<td>(# of phase shifters)</td>
<td></td>
</tr>
<tr>
<td>= (# of antenna groups)</td>
<td></td>
</tr>
<tr>
<td>(\times (\text{# of antennas/group}))</td>
<td></td>
</tr>
<tr>
<td>(\times (\text{# of antennas}))</td>
<td></td>
</tr>
<tr>
<td>((\text{# of adders}) = 0)</td>
<td></td>
</tr>
<tr>
<td>Bad</td>
<td></td>
</tr>
<tr>
<td>• Antenna calibration issue</td>
<td>• Managing interference among neighboring beams</td>
</tr>
<tr>
<td>arises due to narrow</td>
<td>is difficult</td>
</tr>
<tr>
<td>beam-width and fine steering</td>
<td>• Fine steering is not possible</td>
</tr>
<tr>
<td>• Coverage is limited (Great</td>
<td></td>
</tr>
<tr>
<td>side-lobe arises when a</td>
<td></td>
</tr>
<tr>
<td>beam deviates from the bore-</td>
<td></td>
</tr>
<tr>
<td>sight. Reasonable coverage</td>
<td></td>
</tr>
<tr>
<td>is within (\pm 40^\circ) of</td>
<td></td>
</tr>
<tr>
<td>bore-site)</td>
<td></td>
</tr>
<tr>
<td>• Beam that covers the whole</td>
<td></td>
</tr>
<tr>
<td>coverage and a beam that</td>
<td></td>
</tr>
<tr>
<td>supports a specific user</td>
<td></td>
</tr>
<tr>
<td>might be different</td>
<td></td>
</tr>
<tr>
<td>- Beam operation should be</td>
<td></td>
</tr>
<tr>
<td>alternating according to TTI</td>
<td></td>
</tr>
<tr>
<td>• High hardware complexity</td>
<td></td>
</tr>
<tr>
<td>(# of phase shifters)</td>
<td></td>
</tr>
<tr>
<td>= (# of beams) \times (\text{# of antennas})</td>
<td></td>
</tr>
<tr>
<td>((\text{# of adders}) = (\text{# of antennas}))</td>
<td></td>
</tr>
</tbody>
</table>

Every beamforming technology should keep the balance between the performance and the cost, where Figure 5.5-11 depicts the operations of a transmitter and a receiver in each beamforming structure, being described hereafter. Compared to beamforming in Sub-6 GHz, beamforming in mmWave is tightly coupled with the coverage issue, demanding a high cost for beamforming due to antenna calibration. The candidates of beamforming technologies are introduced as follows.

- Analog Fixed Beamforming

The Analog Fixed Beamforming is easy to implement compared to other 3D beamforming technologies. A transmit side prepares for a fixed beamforming weight vector set and the fixed
3D beams are controlled by a UE feedback or a UE transmit signal measurement at eNB. The significant advantage is that it does not require an antenna calibration and a phase shifter.

- **Analog Adaptive Beamforming**
  The Analog Adaptive Beamforming is well applicable in the interference limited environments. It is advantageous that the number of RF transceivers is proportional to the number of independent symbols to be transmitted, but the calibration in analog domain is needed to work.

- **Digital Adaptive Beamforming**
  This has been widely applied in wireless mobile communication systems. It processes a beamforming weight vector in digital domain. Its advantage is in the possibility of a fine control beamforming, but it demands the number of RF transceivers equal to the number of antenna elements, and antenna calibration is a key factor for beamforming with strong directivity.

- **Hybrid Adaptive Beamforming**
  For Hybrid Adaptive Beamforming, there are two beamforming weights \((\mathbf{F}_{RF}, \mathbf{F}_{BB})\) each corresponding to the analog and digital domain precoder. In a given wireless channel, the combination of the precoders should approximate the optimum precoder, \(\mathbf{F}_{opt}\). The following equation shows the optimization problem that finds the best digital and analog domain precoders. The optimal precoders \((\mathbf{F}_{RF}, \mathbf{F}_{BB})\) are obtained through minimizing the difference between optimal precoder \(\mathbf{F}_{opt}\) and their combination, \(\mathbf{F}_{RF} \times \mathbf{F}_{BB}\), under an unitary constraint to the analog precoder and a power constraint.

Similarly, to Analog Adaptive Beamforming and Digital Adaptive Beamforming, the antenna calibration in Hybrid Adaptive Beamforming is also unavoidable.

\[
(\mathbf{F}_{RF}^{opt}, \mathbf{F}_{BB}^{opt}) = \arg\min_{\mathbf{F}_{RF}, \mathbf{F}_{BB}} \left\| \mathbf{F}_{opt} - \mathbf{F}_{RF} \mathbf{F}_{BB} \right\|_F
\]

\[
\text{s.t. } \mathbf{F}_{RF} \in \mathcal{F}_{RF}
\]

\[
\left\| \mathbf{F}_{RF} \mathbf{F}_{BB} \right\|_F^2 = N_s
\]
From the simulation result in Figure 5.5-12, the performances of Analog, Digital, Hybrid Beamforming are compared. And in a coded domain it can be seen that the performance of Analog Beamforming is not much degraded compared to the cases of Hybrid and Digital Beamforming under the given mmWave wireless channel condition referring to [3].
5.5.2.3. Research Direction

Figure 5.5-13 shows one of possible scenarios for 5G where the conventional 4G network accommodates mmWave based 5G network, and the small cells can be allocated in Stand-alone or Macro-controlled with the frequency below and above 6 GHz. The 3D beamforming with a large scale antenna system above 6 GHz at 28 GHz is described in consideration of mmWave-based Stand-alone small cells.

In the scenario 3D multiple beams are transmitted and there are multiple Beam Spots formed by 3D multiple beams as shown in Figure 5.5-14. For the purpose of cell capacity and interference control 3D multiple beams are regarded as small cells called Beam Spots and its control is performed at eNB in a centralized way. The available antenna configuration for 3D beamforming could be a modular-typed patch array antenna, and its physical design is shown in Figure 5.5-15.
Figure 5.5-13 Prospect on 5G network configuration

Figure 5.5-14 Multiple 3D Beam Spots
In a given scenario the 3D beams could work in transmission modes such as Single Beam Transmission (SBT), Coordinated Multiple Beam Transmission (CoMBT-SM, CoMBT-CoMP), and Beam Coverage Enhancement (BCE) as shown in Figure 5.5-16, Figure 5.5-17, and Figure 5.5-18.
In addition, for energy saving and inter-beam interference control Discontinuous Beam Transmission (DBT) and Dynamic Beam Switching (DBS) could also be applied.
References


5.6. Advanced Interference Management

5.6.1. Advanced Receiver for Simultaneous Non-Unique Decoding

5.6.1.1 General Overview
As 4G cellular systems densify their cell deployment, co-channel interference becomes a major source of obstacles to cell throughput improvement. In addition, cell edge users suffer more from co-channel interference, which may govern end users’ experiences. Although some network-side solutions for co-channel interference management have been introduced in current 4G standards, it turns out that most of those solutions yield only meager gains in realistic cellular environments. Among all the limitations of 4G systems, high co-channel interference in the downlink1 is one of the biggest concerns, as the densely overlaid heterogeneous network deployment with full resource reuse is deemed inevitable in future cellular systems. As a matter of fact, co-channel interference management has driven recent LTE releases, and important features, such as inter-cell interference coordination (ICIC) and coordinated multipoint (CoMP) communication, have been introduced. Basically, the interference management in 4G LTE is mostly a network-side operation and transparent to receivers. Network-side interference management is beneficial to ensure backward compatibility with legacy users and easy to deploy by extending the legacy network. The user equipment (UE)-side interference management can be a solution to alleviate the issues relevant to network-side interference management.

5.6.1.2. Advanced Interference Management in Network Information Theory
The advantages of advanced receiver and joint scheduling techniques can be supported by the information-theoretical analysis of the achievable rate region.
To take a glance at the impact of advanced receivers, let us assume that the two transmit points are not coordinated, but simply transmit i.i.d. Gaussian random codes, which are known to achieve the capacity of the point-to-point AWGN channel with a proper choice of covariance matrices. Suppose that UE 1 is a conventional receiver that only tries to correctly decode the desired signal and treats the interference signal as noise. This strategy is referred to as interference-as-noise (IAN) decoding, and the relationship between R1 (data rate from transmit point 1) and R2 (data rate from
transmit point 2) at UE 1 is represented in Figure 4.48. Note that the achievable rate of IAN decoding depends only on the interference power, not on the data rate of the interference signal. Now, suppose that UE 1 insists on jointly and correctly decoding both desired and interference signals, which is referred to as the simultaneous decoding (SD) strategy. Then the channel configuration reduces to a Gaussian multiple access channel (MAC), and the achievable region of the rate pair \((R_1, R_2)\) is represented in Figure 5.6-1. Unlike IAN decoding, SD utilizes the codebook structure of the interference signal. In a mild interference condition, correct decoding of the interference signal provides perfect interference mitigation and raises the achievable rate \(R_1\) to the interference-free rate. However, in a harsher interference condition, where \(R_2\) is too large, correct decoding of the interference signal is an excessive requirement, and the achievable rate \(R_1\) falls to zero. Note that neither of the two regions in Figure 5.6-1 includes the other; that is, neither IAN nor SD provides a universally best interference management. Another decoding strategy, simultaneous non-unique decoding (SND) [1], tries to decode the interference signal jointly with the desired signals as SD does. However, unlike SD, SND does not care about any decoding errors in interference decoding. It is known that SND outperforms both IAN and SD, and can achieve the union of the IAN and SD regions in Figure 5.6-1. In fact, SND is the optimal decoding strategy for the interference channel employing point-to-point channel codes, and it forms the theoretical basis of the UE-side AIM (i.e., the advanced receiver).

Figure 5.6-1 Achievable rate regions: a) treating interference as noise; b) simultaneous decoding; c) simultaneous non-unique decoding (advanced receivers); d) jointly achievable rate region of two advanced receivers.
Returning to the discussion on the two-cell-two-receiver model, the rate region that is jointly achievable by UEs 1 and 2, when both UEs are advanced receivers, is determined as the intersection of the achievable regions of the UEs. In Figure 5.6-1, an example of achievable region of advanced receivers is shown. To emphasize the benefit of advanced receivers, the achievable rate region of conventional receivers is depicted in the same figure. Given the achievable rate region, joint scheduling can be identified as a process of selecting a service point, that is, a rate pair (R1, R2) within the achievable rate region, such that some network utility function is optimized. More commonly, a scheduling process encompasses selecting a UE to be served at a given resource, such as time and frequency, when there are multiple candidate UEs in the cell coverage. However, in the above single-UE example, we can omit this procedure and focus on the rate selection. Therefore, scheduling for the two cells cannot be separated; coordination between them is required.

5.6.1.3. Summary and future works

High-level requirements and practical aspects related to these technologies, such as the transmitter coordination strategy and receiver architecture, as well as the theoretical basis and their implication on 5G system design, are discussed. A performance evaluation based on the current LTE system quantified the benefit of these technologies; the combination of advanced receivers and joint scheduling provides over 50 percent improvement in cell edge throughput without sacrificing the cell average throughput. This gain demonstrates that if 5G networks incorporate advanced interference management, they will provide virtually edgeless end-user experiences.

References

5.6.2. Advanced Interference Management via Sliding-Window Superposition Coding (SWSC)

5.6.2.1. General Overview
IMT-2020 vision states as one of key requirements in the fifth generation mobile networks (5G) that cell edge performance requires 100 Mbps, which is 10 times higher than the requirement 10 Mbps (20 MHz x 5 carriers) of LTE-A [1]. As small cells become densely deployed in 5G, performance degradation due to co-channel interference will be a more serious problem and hard to meet the cell edge requirement of 5G. As a first step dealing with the co-channel interference problem on the receiver side, conventional linear receivers were improved by considering statistical information of interference channels and thus developed into interference rejection combining (LMMSE-IRC) receivers [2, 3], which are now widely and commercially used in industry. Network assisted interference cancellation and suppression (NAICS) was first studied for 3rd Generation Partnership Project (3GPP) LTE-A in Release 12, in which several kinds of interference-aware receivers and control signaling methods for them were discussed. Symbol-level interference-aware receivers can now be used in cellular networks after the standardization of NAICS [4]. It has been verified in information theory that interference-aware techniques always increase performance and more importantly that maximum-likelihood (ML) decoding performance can be achieved by simultaneous non-unique decoding (SND), when p2p coding techniques are imposed [5, 6, 7]. Conventional receivers, however, have large performance gap from the SND decoding performance. Moreover, SND must use some form of multi-user sequence detection, which can't be implemented in its current form without sacrificing one or the other, and thus it's been very difficult to achieve the SND performance in practice. Therefore, unprecedented technology is required to both break through the current limitation and satisfy the 5G requirement. Such key enabling technique for 5G would have to achieve the SND performance and be implementable at low complexity.

5.6.2.2. State of the Art and Research Trend
Sliding-window superposition coding (SWSC) has been proposed as a new information-theoretic coding scheme to mitigate co-channel interference by achieving the SND performance
at low complexity [8]. SWSC transmitter sends one message over multiple blocks rather than a single block via superposition coding without rate splitting (see Figure 5.6-2 for the example of two signal layers at sender 1 and one signal layer at sender 2). SWSC receiver recovers a desired message by decoding an interference signal and a desired signal sequentially based on received signals within a given sliding window. More specifically, the SWSC transmitter consists of the following: 1) Known information is used at the beginning and the end of the transmission, 2) One message is sent over multiple blocks, and 3) A transmitted signal is generated by the superposition of multiple codewords each carrying a different message in the symbol-level (for example, the superposition of two QPSK symbols results in a 16-QAM symbol). The SWSC receiver cancels out known information from a received signal and decodes an interference signal and a desired signal sequentially. For instance, assuming that a current sliding window covers blocks 1 and 2, it cancels out known information from a received signal in block 1 using the SWSC transmission structure. Since a desired signal is not as yet known, it decodes an interference signal first by considering the desired signal as noise and cancels out the decoded interference signal from the received signal in block 1. Now it recovers a desired signal from received signals in both blocks 1 and 2 by treating unknown signals as noise, and cancels out the decoded desired signal from the received signal in block 2. It then decodes an interference signal in block 2, and cancels it out from the received signal in block 2. In this way from block 1 to b, it performs sliding-window decoding and successive cancellation decoding over all blocks. SWSC increases the data rate of an interference signal by handling co-channel interference via utilizing known information at every decoding step to decode interference better than the conventional approaches, which, in turn, helps increase the rate of a desired signal recovered at another receiver. Thus the sum-rate increases up to the SND decoding performance. The detailed analysis can be found in [8, 9].

5.6.2.3. Research Direction

Given the theoretical development [8, 9], an important research direction is to realize the theoretical concept of SWSC into practical transmission coding schemes. In [10, 11, 12], SWSC has been implemented via off-the-shelf codes such as LTE turbo codes. As promised in theory, these implementations track the SND performance very closely in Gaussian interference channels.
with signal-to-interference ratio (SIR) = +0.2/-0.2 dB shown in Figure 5.6-3, and has a large gap to the conventional receivers including treating interference as noise (IAN), interference-aware detection (IAD) [13], and interference-aware successive decoding (IASD) [14]. The basic implementation in [11] has been further extended to an adaptive transceiver design [12], which tracks the SND performance even under a broad range of interference channel conditions over time and frequency like LTE fading channels. In particular, the proposed design controls the number of layers and applies adaptive bit mapping schemes between layers for this purpose. This design also utilizes various decoding methods using the SWSC transmission structure to combat outdated channel estimation errors caused by changing channel conditions. The SWSC receiver can perform robustly against channel estimation errors by changing the decoding order adaptively according to instantaneous SIR. When evaluated in Ped-B interference channels with average INR = 10 dB, the proposed adaptive transceiver design performs better than IAN by 9.5 dB, IAD (NAICS receiver) by 7.5 dB, IASD (the most advanced interference-aware receiver in practice in prior to SWSC) by 3.3 dB, and basic SWSC by 2.7 dB at BLER = 0.1 (see Figure 5.6-4). SWSC requires network assistance to mitigate co-channel interference in similar way like NAICS receivers. For instance, cooperation between eNodeBs is required for joint resource allocation and signaling for code rates of interference signals and user-specific RNTI information. In addition, system level simulation should be evaluated for further verification of the SWSC performance.

Another interesting research direction is to investigate how similar ideas can be extended to mitigate more than one interference signal. Currently, even theoretical SWSC has not been fully developed for interference channels for three or more users.

![Figure 5.6-2 The encoding and decoding operation of the SWSC scheme for b blocks.](image-url)
Figure 5.6-3 Sum rates of SWSC, IASD, IAD, and IAN for the Gaussian interference channel with SIR = -0.2/0.2dB.

Figure 5.6-4 BLER for the 2/2-layer SWSC scheme with bit mapping rule optimization and iterative soft decoding, the 2/1-layer SWSC scheme with successive hard decoding (basic SWSC), IASD, IAD, and IAN in the SISO symmetric Ped-B interference channel with average received INR = 10 dB. Sum-rate = 1.2 bits/D. Modulation: 4-PAM.
References


5.7. Access Architecture-related Radio Technologies

5.7.1. Advanced Small Cell

5.7.1.1. General Overview
Since it is anticipated that future mobile traffic growth is tremendously explosive due to broadband hyper-realistic media transmission and massive connected devices in years of 2020~2030, wireless network densification using small cells will be a key solution of 5G to accommodate the required capacity. High-traffic areas include urban hotspot areas, subway/train stations, bus stops, venues, stadiums, shopping malls, and high-rise buildings. Mobile traffic demands can be fluctuated and non-uniform according to time and place, so that typical traffic profile can be estimated and properly utilized.

Small cells have lower transmit power and smaller coverage than those of macrocells and are deployed as different kinds of microcells, picocells, femtocells, in-building cells, relays, and personal cells according to their usage and outdoor/indoor deployments. Various small cells offer different capabilities in terms of size, DU/RU integration, installation place/type, transmit power, coverage, RAT, frequency bands, simultaneous connections, and backhaul links. In 5G, ultra-dense small cells over the macrocells, i.e., heterogeneous network will be more installed, particularly in urban areas.

In frequency perspective, more various frequencies such as unlicensed band (e.g., 5 GHz) and millimeter-wave band (e.g., 3~30 GHz) will be used in small cells due to relatively cheap frequency cost and broad bandwidth. In particular, more refined interference management schemes will be considered for unlicensed 5G access between unplanned cells with different RAT and with different operators.

The millimeter-wave small cells can be favorably operated by using compact-sized massive MIMO base station and precise beamforming functionality that can effectively mitigate interference to other UEs.

Therefore, ultra-dense heterogeneous network, where many small cells are densely deployed on macrocells or independently, will play a key role in designing low-cost and high-capacity 5G network, and optimized operation of heterogeneous network will be important for enhancing...
capacity and mobility performances.
However, the profit of mobile operators has been decreasing and expenditure for deploying and operating high-capacity network has been increasing. So, high-capacity wireless network at a very-low cost (CAPEX and OPEX) is also a challenging goal for 5G.

Figure 5.7-1 illustrates an example of the small cell-based heterogeneous network deployment in 5G network. 5G RAN can be deployed regardless of legacy LTE/3G RAN and controls 5G macrocells as well as 5G small cells. Table 5-8 shows the classification of various small cells.

![Figure 5.7-1 Ultra-dense heterogeneous network deployment with various small cells in 5G](image)

**Table 5.7-1 Classification of various small cells**
*(A typical deployment scenario is assumed)*

<table>
<thead>
<tr>
<th></th>
<th>UE density per cell</th>
<th>Cell density</th>
<th>Cell size</th>
<th>Data rate per UE</th>
<th>Degree of interference</th>
<th>Examples of deployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picocell</td>
<td>High</td>
<td>Medium</td>
<td>~300m</td>
<td>Medium</td>
<td>Medium</td>
<td>Outdoor hotspot</td>
</tr>
<tr>
<td>Femtocell</td>
<td>Low</td>
<td>High</td>
<td>~50m</td>
<td>High</td>
<td>Low</td>
<td>Home, cafe</td>
</tr>
<tr>
<td>In-building cell</td>
<td>Medium/High</td>
<td>Medium</td>
<td>~50m</td>
<td>High</td>
<td>Medium</td>
<td>Office, shopping mall</td>
</tr>
<tr>
<td>Personal cell</td>
<td>Low</td>
<td>High</td>
<td>~10m</td>
<td>Low</td>
<td>High</td>
<td>D2D</td>
</tr>
<tr>
<td>Stadium cell</td>
<td>Very high</td>
<td>High</td>
<td>~10m</td>
<td>Medium</td>
<td>High</td>
<td>Stadium, exhibition</td>
</tr>
</tbody>
</table>
5.7.1.2. State of the Art and Research Trend

5.7.1.2.1. Literature Review
Current state of the art in heterogeneous network researches and 3GPP standards including various interference management techniques was described in [1]. Network densification schemes over space and frequency using neighborhood small cells and advanced IC receivers were proposed in [2]. Resource-aware cell association and power control methods for challenging interference management in 5G multi-tier networks were studied in [3].

The main challenge of 5G hyper-dense small cell deployment will be the random deployment, dynamic on-off, flexible connection to core networks, and flat system architecture. Cooperative distributed radio resource management algorithms for time synchronization, carrier selection, and power control for solving these problems were described in [4].

In [5] and [6], technical combinations of small cells, coordinated multipoint transmission, and beamforming based on massive MIMO to enhance the spectral efficiency with affordable complexity were studied.

5.7.1.2.2. Standard Trend
3GPP standardization works on small cell enhancements are briefly reviewed based on Release 12 and 13 to outlook 5G standard activity.

Since the interference problems in heterogeneous network deployment and cell densification may significantly degrade the overall performance, proper interference management schemes such as coordinated multipoint (CoMP) transmission/reception and enhanced intercell interference coordination (eICIC) were standardized in Releases 10 and 11, respectively.

The eICIC schemes can be classified into time-domain, frequency-domain, and power control techniques. First, in time-domain eICIC scheme using almost blank subframe (ABS), transmissions of the victim users are scheduled in time-domain resources where the interference from other cells is mitigated. Secondly, in frequency-domain eICIC scheme, control channels and physical signals of different cells are scheduled in reduced bandwidths in order to have totally orthogonal transmission of these signals at different cells. Thirdly, in power control scheme, cell range expansion (CRE) offloads more traffic to small cell from macrocell.
Also, reduced power subframe (RPS) scheme can mitigate capacity degradation due to ABS. In joint processing (JP) CoMP schemes, joint transmission (JT) and dynamic point selection (DPS) schemes are used. In coordinated scheduling/beamforming (CS/CB) CoMP schemes, the scheduling and beamforming decisions at each base station are made jointly via coordination. Note that the coordination interval in CoMP is much faster than eICIC, and CoMP gives significant gains over eICIC; however, requires fast backhauling and coordination capability. Higher-order modulation like DL 256QAM and reduction of UE-specific reference signal to exploit the favorable RF condition and UE’s low-mobility in small cells were also discussed in Release 12.

5.7.1.3. Research Direction

The major directions of ultra-dense heterogeneous network include the increase in capacity and cell-edge data rate, operational efficiency and cost reduction, energy saving, support of ideal and non-ideal backhauls, and various deployment scenarios.

5.7.1.3.1. Cloud RAN for Ultra-dense Heterogeneous Network

Since very large number of small cells will be deployed in 5G, we need to significantly reduce costs due to base station units, backhauls, and supporting facilities. Therefore, ultra-dense heterogeneous network based on centralized and cloud RAN (C-RAN) network architecture is a promising approach. However, C-RAN with ideal and non-ideal backhauls should be considered for accommodating various deployment scenarios.

In addition, network function virtualization (NFV) can be applied for low-cost, large-scale pooling of base-band processing units.

Required backhaul and fronthaul capacity will also be increased much due to the increase in data traffic and number of cells, requiring low-cost and high-capacity links. From a deployment perspective, most of operators will install 5G RAN systems of multiple vendors (in addition, the vendor of 5G RAN can be different from 4G’s). Therefore, open, standardized interfaces for backhaul and fronthaul (e.g., ORI) will be highly desirable for cost-efficient inter-operation.

Fast and automated RAN configuration, optimization, and recovery tools based on self-organizing network (SON) algorithms will be important in ultra-dense heterogeneous network because
they help the operator to handle the massive and complicated tasks of 5G network optimization due to the dense cell deployments and multi-band operation. Especially, interference management and load balancing features should be incorporated in 5G SON system. In addition, RAN analytics using big-data will be helpful for operating intelligent and robust 5G network.

5.7.1.3.2. Advanced Inter-Cell Coordination and Interference Management
As cells in 5G denser and more heterogeneous than LTE, advanced inter-cell coordination and interference management techniques for ultra-dense heterogeneous network will be more essential to fully exploit the gains of cell densification.

5.7.1.3.2.1. Advanced Cell Coordination and Cooperation
Coordinated transmission from multiple cell nodes with multi-band is very effective in mitigating the interference due to the deployment of heterogeneous network. Multi-connectivity where the UE is simultaneously connected to two or more cell nodes will be feasible with dynamic and tight cooperation between the macrocell and small cells, thus requires ideal backhaul connections. Therefore, the cell coordination and cooperation schemes can provide multiple benefits in addition to throughput aggregation of multiple nodes. In addition, downlink/uplink separation, where the UE can associate to the macrocell in the downlink and to the small cell in the uplink and vice versa, is advantageous in a heterogeneous network deployment because of the big difference on coverage and load between the macrocell and small cell. As cells are denser, we also need to reduce energy consumption due to huge number of small cells. Energy efficiency as well as throughput performance through dynamic turn-off of small cells at low or no load (e.g., office in-building cells in the nighttime) can be improved. More robust interference avoidance from cell reference signal (CRS) can be achieved by using the new carrier type (NCT) transmission in the small cells, where the cell node does not transmit control signal at no traffic in the cell and does transmit CRS more sparsely.

5.7.1.3.2.2. UE-side Interference Cancellation Receiver
Since CRS interference can significantly degrade ICIC performance of heterogeneous network, advanced UE-side interference management is required in 5G.
Advanced interference cancellation receiver in UE can mitigate strong CRS interference by using linear and nonlinear interference cancellation techniques. Linear IC (Interference Cancellation) is spatial minimum mean squared error (MMSE) processing receiver, while non-linear IC can estimate and reconstruct the interference signal, and then remove the interference. Furthermore, proactive interference management like interference alignment techniques can be introduced to neutralize interferences in ultra-dense heterogeneous network.

5.7.1.3.2.3. Advanced Load Balancing and Mobility Management

Data traffic and UEs will be more unevenly distributed across geographical cells and multiple frequency bands in ultra-dense heterogeneous network, so that the performance of cell-edge users can be significantly degraded by the load imbalance problem. Thus, dynamic load balancing techniques are required to fully utilize multi-band and system resources of cells.

The main purpose of small cells is to offload traffic from macrocell to small cells to properly balance cell load in connected or idle mode. More delicate interference management techniques combined with CRE will be required for achieving this.

Very frequent handovers or ping-pongs can occur for the fast-moving UEs in small-sized dense heterogeneous network area. Heterogeneous network based on enhanced local area concept splits data and control planes, so that large-sized macrocell is responsible for coverage and mobility and small cell handles data transmission. Another promising solution is to build a virtual, single cell by combining small cells and macrocell, which leads to edge-less heterogeneous network.

References


5.7.2. Enhanced Fixed Wireless Backhaul

5.7.2.1. General Overview

The rising popularity of connected devices drives the rapid growing need for mobile broadband capacity and coverage, and additional enhanced backhaul requirements. There are several ways to support rising mobile traffic demands: by deploying more advanced wireless technologies; by securing additional spectrum; by implementing techniques that are more spectrum efficient; and by densifying the macro layer. Continued enhancement and densification may not always be the most cost-efficient ways to boost capacity at hotspots, indoors and cell edges. But, small and low-powered nodes can provide additional capacity in those situations by off-loading the macro layers. Adding radio network capacity and coverage through the deployment of small cells as part of a heterogeneous network requires a higher degree of backhaul scalability and flexibility. Backhaul availability depends on the feasible placements of cells, installation costs, and time needed for site acquisition and installation.

Traditionally, the backhaul should not limit the radio access network and should have sufficient end-to-end performance to meet the desired user quality of experience (QoE) in everywhere. This is equally important for backhaul in both the macro and small cells of a heterogeneous network. To enable high QoE maintaining coordination between radio nodes and layers, high performance end-to-end backhaul in terms of speed, delay, and delay variation is recommended. There is a wide variety of transmission characteristics with different backhaul options that can affect the levels of coordination in a heterogeneous networks [1]. To deliver maximum QoE across networks, operators should aim to build backhaul links for
high performance such as LOS microwave and optical links whenever possible. Where no fixed links are available, wireless is the natural choice, as it enables short time-to-market and a higher degree of freedom in site placement. For an even higher degree of freedom, NLOS wireless links may be used.

This section introduces enhanced wireless backhaul requirements and technology issues related to backhaul environments and spectrum.

### 5.7.2.2. Standardization and Market Requirements

Since 3GPP LTE Release 10, network densification using small cells has been an important evolution direction in 3GPP to provide the necessary means to accommodate the anticipated huge traffic growth, especially for hotspot areas. In the RAN Release 12 Workshop in June 2012, 3GPP identified future requirements and candidate technologies including key areas of capacity increase, energy savings, cost efficiency, support for diverse application and traffic types, higher user experience, and backhaul enhancement [2]. As a potential technology to meet these requirements, small cell enhancements has been started as a Release 12 items, and the backhaul is considered as another important aspect for enhanced small cells, especially when considering the potentially large number of small cell nodes to be deployed. 3GPP has decided both the ideal backhaul (i.e., very high throughput and very low latency backhaul, e.g., dedicated point-to-point connection using optical fiber or line-of-sight (LOS) microwave) and non-ideal backhaul (e.g., typical backhaul widely deployed today, e.g., xDSL, non-LOS (NLOS) microwave) to be considered. Table 5-9 listed examples of non-ideal backhaul.

<table>
<thead>
<tr>
<th>Backhaul technology</th>
<th>Latency (one way)</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiber</td>
<td>10 – 30 ms</td>
<td>10 Mb/s – 10 Gb/s</td>
</tr>
<tr>
<td>DSL</td>
<td>15 – 60 ms</td>
<td>10 – 100 Mb/s</td>
</tr>
<tr>
<td>Wireless backhaul (typically NLOS)</td>
<td>5 – 35 ms</td>
<td>10 – 100 Mb/s, up to 1 Gb/s</td>
</tr>
</tbody>
</table>

NGMN also identified in their white paper that deploying large number of small cells near to the consumers would help solve the capacity problem for the radio access network, but create a new one for backhaul, which must provide connectivity at sufficient capacity and quality of
service [3]. They also provided several considerations for wireless backhaul solutions that can be differentiated by the propagation between sites, the spectrum used and the topologies formed by the resulting network. They see NLOS is generally only practical with carrier frequencies below 6 GHz, and not at ‘microwave’ frequencies in the 10's of GHz where penetration losses are significantly larger. It was noted that the limited low frequency spectrum suitable for NLOS propagation is sought after for mobile access, more spectral bandwidth is available at microwave frequencies, giving the potential for higher capacity. NLOS solutions potentially offer better coverage in dense urban environments, provided the links support sufficient throughput to be usable.

Senza Fili Consulting confirmed trends that backhaul is still a key element of small cell deployment and TCO (Total Cost of Ownership), and small cell backhaul requirements are different from macro cell backhaul [4]. In [4], it was claimed that small cell backhaul has a more stringent requirement set, which makes it a more challenging solution than macro cell backhaul in terms of several factors: low cost; small form factor; high capacity; low power; fast and easy installation that can be done by semiskilled employees; tolerances of sway and ability to operate from precarious locations; ability to cope with changes in the environment that can be unforeseen and are not under the operator's control; scalability to accommodate the addition of new small cells within the same footprint; low latency, to support LTE's X1 and S2 interfaces. In [4], it was insisted that fiber is the best solution for small cell backhaul, so where it is available and cost effective, it typically wins over wireless solutions. But, it is not always available, and when it is, often it is not cost effective, because either the installation costs (mostly due to trenching) or the operating costs (i.e., leasing costs) are too high. If the operator selects wireless backhaul, the choice among technologies becomes more complex, because it depends on a larger number of factors, including spectrum and LOS availability, and capacity requirements. If other things being equal, operators prefer NLOS solutions in licensed spectrum. But, these solutions typically have less capacity and higher latency, and this makes them unsuitable for high-capacity small cells or for backhaul links that support multiple cells. In [4], it was also claimed that there are emerging trends that variety in small cell deployments requires flexibility in the backhaul. In sub-6 GHz bands where both LOS and NLOS are possible, NLOS is the dominant architecture because it gives more flexibility.
Although it has been believed that NLOS works in sub-6 GHz, and it can be spectrally inefficient and lead to severe limitations in capacity that make many sub-6 GHz solutions not fit for LTE and Wi-Fi small cells, backhaul sharing, or multi-hop backhaul. But, at the same time, there is a promising exploration of using high-frequency bands for NLOS and near-line-of-sight scenarios, leveraging the signal reflection, diffraction and penetration in the environment. This approach may lead to innovative high-capacity backhaul solutions that use spectrum that is less expensive and more widely available than sub-6 GHz. In [5], it was also commented that NLOS capability is typically possible in frequency ranges below 6 GHz, while with NLOS, operators do not have to align antennas as they do for point-to-point LOS radios, which can ease installation and reduce deployment times.

5.7.2.3. Research Direction

5.7.2.3.1. Topology
Several references provided wireless backhaul network topology for small cells [3][5][6]. With wireless backhaul, operators have options for network topology depending on how they need or want to connect the small cells to a macro cell site or other aggregation point. The connections can be point-to-point, which requires LOS, or point-to-multipoint, whereby a hub site connects multiple small cells in NLOS or LOS conditions and backhaul capacity is shared among the small cells. In [3], it is considered that usage of traditional LOS technology may restrict the coverage of wireless backhaul to small cells and potentially drive up costs if alternative paths need to be used. Near- or even non-line-of-sight options are under consideration, and a schematic overview of different application of wireless technologies are suggested in [3]. There might be cases where specific small cells base station cannot be directly connected to the macro cell site via a single wireless link because of physical obstructions, but can be reached via another small cell node. In these cases, more complex topologies like chains and trees could be used. Such topologies would require the small cell backhaul solutions to support multiple wireless links as well as a traffic aggregation functionality. Connecting small cell node via chains or even trees may be an appropriate topology when they are installed e.g., on lamp-posts or any other place few meters above street level. In those cases, it is sufficient
that only one of the small cell node is connected to the backhaul network (e.g., via macro cell site) and further connectivity is provided among the small cell base station themselves. The small cell layer might have connections to two different macro base station sites for resiliency reasons, i.e., if one of the connections goes down the other one could take over – but both are not used simultaneously. Whilst multi-hop topologies can provide extensive connectivity, they may do so at the cost of capacity and latency performance. Requirements for these aspects need to be defined.

There are the challenges posed by locations to deploy microwave backhaul without a clear line of sight. Traditionally, several established methods can be used to overcome the challenges. For example, in mountainous terrain, passive reflectors and repeaters are sometimes deployed. However, this approach is less desirable for cost-sensitive small cell backhaul, as it increases the number of sites. Daisy chaining in urban areas can be considered to reach sites in tricky locations that is also effective for small cell backhaul [6].

5.7.2.3.2. Line-of-sight Backhaul

According to the fact that millimeter-wave wireless technologies (e.g., 60 GHz or 70 - 80 GHz E-band) are actively considered for wireless backhaul networking, considering LOS situations is required because millimeter-wave links are especially good for LOS. Even in ITU-R 1411-7, only LOS propagation models are proposed for millimeter-wave links, i.e., a free-space propagation model with the path-loss coefficient of 2.2.

In LOS backhaul networking, two technical features are fundamentally required: (i) multi-hop relaying/routing and (ii) self-organization.

**Multi-hop relaying/routing:** Suppose that multiple backhaul devices are deployed in backhaul networks and two of them may want to construct a wireless backhaul link even if they are not in LOS situations. Then, relay devices are required to make an end-to-end communication link with multiple LOS links. Therefore, advanced multi-hop routing and relaying schemes are required to maintain LOS links in backhaul networking.

**Self-organization:** In LOS backhaul networking, intermediate relay devices can be a malfunction, then the other intermediate relay devices should take the role of the malfunctioned relay device. In addition, when backhaul devices are deployed in a random manner, the deployed
backhaul devices should be able to construct backhaul links among the backhaul devices in a self-organization manner.

From one example implementation of modular antenna arrays (MAA) by Intel (16 elements in one module) for the 60 GHz or E-band (70 - 80 GHz) millimeter-wave LOS backhaul wireless networking, four main advantages can be observed, i.e., (1) more flexibility in terms of antenna module placement, (2) feedline power loss reduction, (3) extensible architecture to achieve wide range of antenna gain and aperture to achieve regulatory EIRP limits, and (4) capable of realizing true Massive MIMO with independently phase-controlled antenna elements. Also, MAA-based millimeter-wave LOS backhaul shows 1 Gbps data rates up to 440 meters when WiGig/IEEE 802.11ad-based modulation and coding scheme (MCS) basedband/MAC is considered.

5.7.2.3.3 Non-line-of-sight Microwave Backhaul for Small Cells

Deploying high-performance microwave backhaul in places where there is no direct line of sight brings new challenges. It is traditional belief in the telecommunication industry that sub-6 GHz bands are required to ensure performance for such environments. In [6], it was demonstrated that high frequency systems using above 20 GHz can outperform those using sub-6 GHz bands even in locations with no direct line of sight. In order to determine the importance of NLOS-system properties (diffraction, reflection and penetration), measurement test was carried out on two commercially available microwave backhaul systems in different frequency bands, 5.8 GHz and 28 GHz [6].

According to the results in [6], contrary to common belief, microwave backhaul in bands above 20 GHz will outperform sub-6 GHz systems under most NLOS conditions. The key system parameter enabling the use of high frequency bands is the much higher antenna gain for the same antenna size. So, it is possible to plan NLOS backhaul deployments that provide high network performance. And so, in the vast amount of dedicated spectrum available above 20 GHz, microwave backhaul is not only capable of providing fiber-like multi-gigabit capacity, but also supports high performance backhaul for small cells, even in locations where there is no direct line of sight.

Reference [7] investigated the feasibility of using microwave frequencies for fixed non-line-of-sight wireless backhauling connecting small cell radio base stations with an aggregation node in an outdoor urban environment, i.e., a typical heterogeneous network.
scenario. It was found that the higher frequencies offer not only larger bandwidths but also higher antenna gains which would ideally work to their advantage. These advantages may be lost when taking antenna alignment errors and rain into account.

In [7], it was concluded that the higher frequencies showed best performance in ideal scenarios with no rain or antenna alignment errors, while the lower frequencies (2.3 GHz and to some extent 10 GHz) did not perform well.

5.7.2.3.4. Spectrum Options for Wireless Backhaul

For small cell backhaul, simplicity and licensing cost are important issues. Light licensing or technology-neutral block licensing are attractive alternatives to other approaches such as link licensing, as they provide flexibility. Using unlicensed frequency bands can be a tempting option, but may result in unpredictable interference and degraded network performance. Spectrum availability and licensing conditions vary from country to country. In [5], key characteristics of spectrum for small cell backhaul was summarized as follows:

Sub-6 GHz: There is licensed and unlicensed spectrum in the sub-6 GHz frequency bands that can be used for small cell backhaul. In the licensed category, much of the spectrum is allocated and already used for mobile broadband access, such as 1 GHz and 2 GHz bands. But, there are some fragments of unpaired frequencies (time-division duplex, TDD) that many operators have, as well as some frequencies that operators do not yet use for access, such as the bands above 3 GHz. In unlicensed sub-6 GHz bands, such as 2.4 GHz and 5 GHz, small cell backhaul would share frequencies with Wi-Fi, Bluetooth and many other applications. The bands are uncoordinated and potentially crowded, which makes interference avoidance a challenge. The benefit of this spectrum is that it is free to use and doesn’t involve a lengthy license application process.

Microwave 6 GHz - 60 GHz: These are traditional microwave frequencies that are used extensively for macrocell backhaul and other types of wireless transport networks. Licenses are allocated on a per-link or area basis. Radios can be configured in point-to-point or point-to-multipoint configurations. Backhaul installations in these bands require LOS and antenna alignment. However, research from [6] showed that operating between 23 GHz and 60 GHz under NLOS conditions performed better than systems in the sub-6 GHz bands.

Millimeter-wave 60 GHz: The spectrum is unlicensed or lightly licensed in at least 22 countries,
which makes it low cost in some significant markets. There is a large chunk of spectrum available in the band—the band between 4 GHz and 9 GHz of contiguous spectrum. Also, it can support high capacity links from 100 Mbps to 1 Gbps over distances up to 1 kilometer, as well as compact antenna designs.

Millimeter-wave 70 GHz – 80 GHz: Spectrum in the 71-76 GHz and 81-86 GHz frequency bands (also referred to as E-band) is considered suitable for rooftop aggregation of small cell traffic from the street level. This is because it supports very high capacity of more than 1 Gbps over distances up to 3 kilometers.

References

[1] Ericsson White Paper, It all comes back to backhaul, February 2012
5.7.3. Moving Network

5.7.3.1. General overview

It is expected that future 5G wireless access networks will enable the hyper-connectivity society in which humans can be connected with anything anywhere anytime through intelligent mobile devices. In order to make such a large number of connections, there is a need to reduce the environmental constraints of mobile communication systems by intelligently deploying new wireless nodes so that users are able to utilize ubiquitous networks.

A significant number of users who access the future broadband wireless networks will be on board in public transportation vehicles such as buses, trams, and trains or in personal cars. Therefore, one of promising approaches is to deploy one or several relay node(s) mounted on the vehicles, so called mobile relay nodes (MRNs) which build their own cell(s) inside the vehicle to serve vehicle users. In other words, the MRN constitutes wireless hotspots. However, the communications through the MRNs suffer from severe quality degradation due to high vehicular penetration loss (VPL), and thus it is necessary to place antennas in order to reduce or even eliminate the VPLs. In addition, owing to no constraints on size and power consumption, the MRNs can enjoy freedom to exploit various advanced antenna technologies and smart signal processing schemes, compared to the regular user equipment (UE) connected directly to a macro base station [1].

3GPP has incorporated relays in the Release 10 standard (LTE-Advanced) for the purpose of cell coverage extension and capacity increase [2-5]. 3GPP TR 36.806 [2] provides four architectures and has selected the proxy S1/X2 architecture as the baseline to support a fixed relay node for Release 10 since the proxy S1/X2 has more benefits in comparison with other alternatives. However, new configurations like multi-hop and moving relays are opening up a range of possible architecture solutions [3-4]. Taking into account that mobile relay support has additional considerations such as the group mobility, persistent IP connectivity during handover and so on., 3GPP TR 36.836 [5] identifies scenario and requirements for high speed train moving at speed reaching 350 km/h, then identifies the key properties of MRNs, and assesses the benefits of MRNs over existing solutions in fast-moving environments. Moreover, it also defines mobility procedures over various architectures and has selected both of the
architectures for future work on mobile relays. As described above, basic research is currently in progress including case study and requirements research.

In order to support moving cells/networks, there is a need for smart broadband mobile wireless backhaul which should potentially be fully integrated in the resource allocation and interference coordination in the mobile communication system. In other words, integrating moving cells/networks in wireless access networks can thus enable full control of the Quality of Service (QoS), and at the same time improve the capacity and energy efficiency of the wireless access networks. In addition, it is required to raise the capacity of backhaul link which may be the point of bottleneck in total system capacity, and to be able to maintain high spectral efficiency at high moving speed.

Meanwhile, the above MRNs can potentially expand their capability to provide moving small cells within the macro network by serving users outside the vehicles. Thus, future vehicles and transportation systems may play a crucial role in wireless networks by providing additional communicating capabilities and becoming part of the communications infrastructure to improve capacity and coverage of the mobile communications system. However, there are also challenges in using MRNs, such as smart broadband backhauling, design of efficient resource allocation and interference management techniques, as well as proper mobility management schemes to exploit the benefit of group handover for vehicular UE devices attached to the same MRN.

The following subsection gives an overview of state of the art and research trend on moving networks and elaborate on the visions related to moving networks.

5.7.3.2. State of the Art and Research Trend

The 3GPP standard group introduced relay nodes in order to improve the coverage with high data rate, provide enhanced cell-edge user data rate, and easily deploy a temporary network [2-10]. The fixed relay nodes (FRNs), which have been defined by 3GPP, have two types: type 1 RNs are non-transparent RNs that can form their own cells, and terminate all layer 2 and layer 3 communication protocols; type 2 RNs are transparent RNs that replicate the cell ID of their donor eNB (DeNB) [6]. In recent 3GPP studies, a dedicated MRN deployment has been considered as a cost-effective solution to serve data-hungry UE inside public transportation vehicles [5].
In 3GPP TR 36.806 [2], four architecture alternatives such as Alt1, Alt2, Alt3, and Alt4 were considered for a case of FRN. In the architecture Alt1 called full-L3 relay, the S-GW/P-GW (Serving Gateway/Packet Data Network Gateway) entity serving the relay is separated from the DeNB and the relay is transparent for DeNB. In the architecture Alt2 called Proxy S1/X2, the S-GW/P-GW functionality serving the relay is embedded in the DeNB and the relay S1/X2 interfaces in the architecture are terminated in the DeNB. The distinguished feature of the architecture Alt3 is that the relay bearers are terminated in the DeNB while Alt4 is that the user plane of the S1 interface is terminated in the DeNB. Resultantly, the Proxy S1/X2 architecture was selected as the baseline architecture for Rel-10 since the architecture in comparison with other three alternatives has more benefits in terms of RN/DeNB complexity, other Nodes impact, deployment complexity, UE mobility and QoS, and standardization effort and complexity. However, taking into consideration that mobile relay support has additional requirements such as the group mobility, persistent IP connectivity during handover, etc., there is a need to re-evaluate the relay architecture aiming to clarify the basic differences and distinguish among the alternatives in the context of mobility support for MRNs.

In order to efficiently serve the data-intensive UEs on board vehicles which move at high speed as well as with high VPL, a dedicated MRN deployment has been considered in 3GPP TR 36.836 [5]. TR 36.836 has defined scenario and requirements for high speed train moving at speed reaching 350 km/h, and then identified the key properties of MRNs. Moreover, it has defined mobility procedures for total six architecture alternatives such as Alt1 and Alt2 in TR 36.806, three Alt2 enhancements (Alt2-1, Alt2-2, Alt2-3), and new Alt4 (different from that of TR 36.806), as some candidate alternatives for deploying MRNs. Solutions based on using the existing system elements, such as optimizing the deployment of macro cells, using layer-1 repeaters, or LTE as backhaul and Wi-Fi as access on board, have also been discussed in [5]. Resultantly, the dedicated MRN deployment has been considered as a cost-effective solution and both Alt1 and Alt2 have been selected as the baseline architecture to support mobile relays. In case of 3GPP standardization, as described above, basic researches are currently in progress, including case study and requirements research for providing passengers service and group mobility by deploying MRNs in a limited environment such as high-speed train.

In [3], the author has reviewed the basic 3GPP fixed relay architecture alternatives presented.
in [2] for enabling mobility of the relay node. In particular, mobile relay scenarios based on Alt1 and Alt2 alternatives were analyzed and the corresponding handover procedures were described for comparative purposes. The author insists that for mobile relay changing its backhaul link from Source-DeNB to Target-DeNB the full-L3 relay has more benefits than Proxy S1/X2. One of such reason is that the Alt1 supposes the stable IP anchor point supporting continuous IP connectivity for the MRN and thus it does not require the time-consuming relay attach procedure when backhaul link is re-established. As a result of this study, the author has concluded that the Alt1 is more suitable for mobile relay handling than other architecture alternatives and can be selected as baseline architecture in a mobile relay networks.

The work in [7] was an initial study of the benefit of deploying an MRN in a single cell noise limited scenario. In this paper, the end-to-end outage probability (OP) was considered as the metric to evaluate the performance. Both FRN assisted transmission and direct transmission were used as references. An FRN performance lower bound was derived when the UE position was known. In order to achieve a fair comparison, the optimal FRN position that minimized the average OP was obtained numerically when only knowing the UE position distribution. From the results we can see that when the VPL is moderate to high, the MRN can significantly lower the end-to-end OP at the vehicular UE. In other work [8] of the same authors as those in [7], the end-to-end OP performance of an MRN was investigated in the presence of CCI when only considering the effect of path-loss and small scale fading. A two cell setup was considered, and the FRN assisted transmission and direct transmission were used as reference cases. The authors did not consider any ICIC schemes, and hence the results could be seen as a worst case study. Better performance could be possible in case of different ICIC schemes. In this study, the authors also employed more practical channel models for path-loss and small scale fading, and the optimal FRN position that minimized the average OP was obtained numerically. The results showed that MRN assisted transmission had similar performance to direct transmission when the VPL is 10 dB and as we increased the VPL, MRN assisted transmission greatly outperformed direct transmission as well as FRN assisted transmission. Furthermore, the MRN was operated at a much lower transmit power than the FRN. Hence, the use of MRN also has the potential to improve the energy efficiency of the network.

The authors in [9] suggested that in case of train several backhaul antennas can be
interconnected to form a cooperative and coordinated relay system which can strengthen the backhaul link by using antenna selection techniques. Also, the authors proposed that by a proper placement of indoor and outdoor antennas, an MRN can circumvent the VPL caused by a well isolated vehicle. In [10], taking into considerations that communications of high-speed railway systems suffer from problems such as Doppler spread, radio condition abrupt change and handover failure, the authors proposed and evaluated mobile relay solution with group mobility by a series of system simulation which witnesses an improvement in train user throughput as well as system throughput, and higher handover success ratio with a decrease in radio link failure ratio.

As the high speed trains have been widely deployed across countries, the demand of using mobiles in such an environment in E-UTRA is growing faster. Therefore, it is necessary to investigate potential issues and scenarios, and enhance the performance including RRM (Radio Resource Management), downlink/uplink performance in the high speed scenarios. In order to meet those goals, a study item for performance enhancement in high speed scenarios had been processed until RAN4 #77. The study identified new scenarios and investigated uplink/downlink performance under the existing/new scenarios. For the existing scenarios new requirements were identified, and for the new scenarios the issues on RRM and demodulation were identified. The study captured the proposal for the candidate solutions in the Technical Report [11]. Some outputs of the study items are as follows:

- No new cell identification and RSRP/RSRQ requirements in non-DRX are required in existing high speed scenarios.
- It is feasible to specify the new PDSCH demodulation requirements under the ETU600 to verify UE performance.
- It is agreed to specify the new PUSCH requirements under ETU600.

Based on the results of the study item, a new work item for performance enhancement in high speed scenarios will be proposed.

An inherent issue of this LTE-A based MRN deployment is that the MRN is likely to become a bottleneck. The reason is two-fold: 1) although the LTE-A is capable of providing Giga bit-per-second (Gbps) data throughput, the data throughputs of the high mobility users are much lower than those of low mobility users since a high-order modulation like 64 QAM
(Quadrature Amplitude Modulation) and the advanced antenna technologies with MIMO are not applicable in such high speed environments (when it comes to very high-speed trains, above 350 km/h, available data throughputs between the base station and vehicle are barely over 20 Mbps); 2) It is difficult to meet the ever increasing demands of higher data rate communication with the limited frequency resources of below 3 GHz.

An effort to deal with the bottleneck is underway in a project named Mobile Hotspot Network (MHN). The research goal is to develop a mobile wireless backhaul solution exploiting millimeter waves (mmWaves). The first prototype will provide 1-Gbps data throughput per train by using 250-MHz bandwidth at speed of 400 km/h as of 2015 [12-15]. The MHN system adopts Wi-Fi as user access links on board and delivers mmWave-based mobile wireless backhaul between train cars and radio units along tracks. The backhaul links are composed of MHN Digital Unit (mDU), MHN Radio Unit (mRU) and MHN Terminal Equipment (mTE). Each mDU, connected to mRU via optical fibers, serves baseband signal processing and higher layer protocol. Beamforming technologies are necessary to compensate the high loss of the mmWave. A fixed beamforming has been applied to the mRU and mTE. In the MHN, end users are the passengers inside the train cars. Their aggregate data are pipelined through mTE and transmitted to mDU via mRU. The basic bandwidth for the MHN system is 125 MHz, and the sampling rate is 184.32 MHz which is 6 times higher than that of LTE. The system bandwidth can be extended up to 500 MHz by carrier aggregation method in the current version. Figure 5.7-2 shows an example of the MHN system architecture.

Figure 5.7-2 Architecture of mobile hotspot network for the trains
5.7.3.3. Research Status and Direction of MHN

A huge number of people are using subway trains every day around the globe. For instance, about 8 million passengers are using Seoul subway trains in Korea each day. Many of them surf the Internet, watch videos, check E-mails, and chat online with the smart phones. Some passengers try to access 3G/4G networks directly according to availability and affordability, and some others access Wi-Fi on board, if available. The quality of data services that they experience on the vehicles change greatly depending on how many users are there at that time whether the access is direct on 3G/4G or indirect via Wi-Fi. Since the capacity of wireless backhaul between trains and base stations is limited up to about 10 Mbps as of 2015, the Wi-Fi users on the subway trains might experience slow connections or even frequently failures. Therefore, the backhaul capacity needs to be largely improved sooner for better service quality.

A research project on the mobile wireless backhaul for high speed trains named MHN, introduced in the previous subsection, utilizes millimeter wave bands to improve the backhaul capacity. The approach used in the MHN project can keep spectral efficiency of 4 bps/Hz per train at vehicle speed of 400 km/h, so that it could have 2-Gbps capacity with 500-MHz band in carrier aggregation manner [16]. Theoretically, it could achieve 4-Gbps data throughput with 1-GHz band, and expand to 8-Gbps capacity by using a polarization MIMO technology. If the MHN architecture, currently developed, is applied in the near future, a subway train will be equipped with 1-Gbps capable backhaul, which can be further expanded up to 8-Gbps capacity in the next phase.

Cell coverage is an important topic in the practical and economical aspect. Since the MHN stands on higher frequency carriers, and free space path loss is proportional to frequency, the cell coverage is expected to be rather short. Table 5-10 shows an example of link budget calculation assuming that a transmitter (TX) is 1 km away from a receiver (RX). TX power of 20 dBm and beamforming gain of 22 dBi for both TX and RX were assumed. In addition to the free space loss at 30-GHz radio frequency and connection loss, the atmospheric and rain loss were considered as well. The resultant received power was -65 dBm.
Table 5.7-3 Link budget calculation in an MHN system

<table>
<thead>
<tr>
<th>TX power</th>
<th>TX antenna gain</th>
<th>Loss</th>
<th>RX antenna gain</th>
<th>RX power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Connection</td>
<td>Free space</td>
<td>Atmospheric</td>
</tr>
<tr>
<td>20dBm</td>
<td>22dBi</td>
<td>2dB</td>
<td>124dB</td>
<td>0.2dB</td>
</tr>
</tbody>
</table>

Since a receiver sensitivity is associated with a receiver algorithm, the system margin corresponding to -65 dBm of received power may change depending on the receiver implementation and channel environments. Trackside antennas should be deployed along train routes with keeping appropriate distances between neighbored sites to confirm the overall cell coverage by using coverage related information gathered from the link budget calculation and measurements on propagation. Handover from a serving cell to a target cell around the cell edge is challenging in the MHN system where the transmitter and receiver are beamformed as narrowly as 8 degrees of 3-dB beam width. Figure 5.7-3 shows beam radiation patterns and a 3-dimensional beam image. In case of using such narrow beams and applying fixed beamforming, the receiver inside (or on top of) the train reports rapid decreases of a home cell power because it faces side lobes through back lobes while passing by a trackside antenna. As the train runs faster, time duration for handover gets shorter. Since the home cell power drops so fast, e.g., about 20-dB drop within 400 ms under 100 km/h of speed, the handover should be performed successfully within the short time duration. The handover requirements posed on the MHN system should be addressed to keep average data throughput high over all the paths.

![Beam radiation pattern](image-url)

Figure 5.7-3 Beam radiation pattern (inset: 3-dimensional image)
The feasibility of wireless links between a TX and RX, developed for the MHN system, had been demonstrated in June of 2015. Figure 5.7-4(a) shows two vehicles loaded with a TX and an RX, respectively. They ran on a test road, illustrated in Figure 5.7-4(b), with the relative speed of maximum 100 km/h. The test resulted in 500 Mbps of data throughput as shown in Figure 5.7-4(c). The maximum distance was larger than 1 km during the test.

The MHN system is envisioned to be applicable in the railroad communications in the near future. Subsequently, next step must include some field trials in the real subway environments. Besides, further researches on performance enhancements need to be conducted. They should cover topics such as robust handover under very high speed, adaptive beamforming for
coverage enhancement, and polarization MIMO for capacity improvement. According to the 5G wireless network requirements, the maximum mobile speed of 500 km/h should be supported and peak data rate of 20 Gbps and user experience data rate of 100 Mbps should be supported. The above mentioned MHN system, of which the provisions are limited to the maximum mobile speed of 400 km/h and peak data rate of 8 Gbps, is required to advance to achieve the maximum speed of 500 km/h and the peak data rate of 20 Gbps. Moreover, there are still issues to be discussed and solved in order to realize mmWave-based smart backhauling: 1) the resource allocation and interference in the mobile relay when trains are moving from opposite directions; 2) the deployment and handover decisions of users in a platform where there are usually more than one train arrive or depart, stop or pass by; 3) the different decision for real time traffic and non-real time traffic passengers; 4) the efficient group mobility for users on board very high-speed vehicles; 5) the deployment of moving networks in various vehicle environments which will be more complex than the train environment with the limited number of trains on fixed route railways, and 6) cooperative communication schemes for moving networks. In particular, in case that we deploy moving networks for vehicles on comparatively free routes by utilizing mmWave-based backhauls, the mmWave backhauls may be blocked by various obstacles and thus we need an ad-hoc networking which can provide detour paths for blocked backhauls. This ad-hoc networking with multi-hop mobile wireless backhauls between vehicles will enable cooperative communication schemes for moving networks.
References


5.7.4. Device-to-Device (D2D) Communication

5.7.4.1. Overview
Device-to-Device (D2D) transmission is a wireless technology where devices can communicate directly with each other without routing the data through a network infrastructure. Recently, a great attention has been paid to D2D technology in 3GPP to implement what is appeared to embark on a new chapter. However, the current accomplishments in 3GPP only stress the importance of public safety based on D2D communication or small commercial services via D2D discovery which may promote the growth of the current mobile market. In this article, we first investigate the state of the art in D2D technology which has been worked and accomplished in 3GPP then give some small talk to give rise to discussions about provisions for future D2D technology in terms of 5G perspectives.

5.7.4.2. State of the Art in D2D Technology
From a spectrum point of view, part of LTE carrier or public safety (PS) specific carrier is now allocated for D2D applications. More specifically, parts of physical resource blocks (PRBs) in some UL subframes of LTE can be scheduled for D2D by eNB or pre-configuration. Also, such frequency bands as Korea’s 20 MHz FDD spectrum (718-728, 773-783 MHz) and Band 14 of USA First Responder Network Authority (DL: 758-768, UL: 788-798 MHz) will be used for D2D communications to provide PS services.
Up until now, D2D services are classified into PS and non-PS (commercial).
From the technology point of view, D2D transmission can be distinguished between synchronization, discovery, and communication. While discovery is for in-coverage non-PS services, D2D communication mainly focuses on both out-of-coverage and in-coverage push-to-talk (PTT) service for PS.

5.7.4.2.1. Synchronization
The choice between a synchronous and an asynchronous D2D operation be determined in terms of the performance, the control overhead and the power consumption of each device, etc.
Roughly speaking, the throughput performance of synchronous operation is twice of that of asynchronous and we need to keep in line with the legacy LTE system using OFDMA in DL and SC-FDMA in UL. Keep in mind that synchronization between devices is mandatory, not optional to use OFDMA or SC-FDMA. For these reasons, the state of the art in D2D technology has adopted synchronous mechanism between devices.

The main features of synchronization are divided into: 1) resource allocation, 2) signal and channel design, and 3) procedures for synchronization.

5.7.4.2.1.1. Resource allocation for synchronization

As always do, the first thing we have to consider to do something is to secure necessary resources. Current LTE D2D technology prepares two periodic time-frequency resources to support synchronization. The number of synchronization resources is 1 for in-coverage and 2 for out-of-coverage. One resource among the two for out-of-coverage is the same as that of in-coverage.

The reason we need multiple (two) periodic resources for synchronization is to resolve half-duplex problem in synchronization propagation. In the perspective of current D2D specification, simultaneous D2D transmission and D2D reception is prohibited for many reasons. However, the synchronization propagation is still supported through multiple hops. While only one hop extension is permitted through UE-to-Network relay in D2D communication, the number of hops does not matter as far as synchronization propagation is concerned. Devices just start and stop synchronization transmissions when the required conditions are met.

Since D2D devices operate in half duplex mode, they can either transmit or receive at any given time but cannot transmit and receive simultaneously. Therefore, there are cases where a device (device 1) deriving synchronization using received synchronization signals and channels from another device (device 2) using synchronization resource 1 and relays it to its neighbor devices by transmitting its synchronization signal and channel using synchronization resource 2 as shown in Figure 5.7-5.
5.7.4.2.1.2. Synchronization signal and channel design

Upon detection of synchronization signal, devices acquire sequence indices as well as time and frequency synchronization.

D2D synchronization signal is composed of primary and secondary sequences. Two kinds of length-62 primary sequences reusing Release 8 PSS sequences except different root indices are defined. In order not to use Release 8 PSS root indices (25, 29, 34), D2D primary synchronization signal uses root index 26 for in-coverage and 37 for out-of-coverage. Since the D2D secondary signal reuses Release 8 SSS sequences as it is, there are 168 synchronization sequences for each in-coverage and out-of-coverage.

D2D synchronization channel conveys essential system information including 14-bit D2D frame number (DFN) which consists of 10-bit SFN and 4-bit offset indication to identify radio frame and subframe, respectively. Other information included are TDD configuration, in-coverage indicator, and system bandwidth. TBCC is used for D2D synchronization channel. The index of synchronization signal sequence is needed to decode this channel. Scrambling applied to data and all the unknowns as to DMRS are to be known from the synchronization signal sequence.

The periodicity of D2D synchronization resource is 40 ms and one subframe is allocated at every period as shown in Figure 5.7-6.
5.7.4.2.1.3. Synchronization procedures

The up to date D2D technology allows any device to transmit or stop synchronization transmission of signal and channel when the required conditions are met. Since there are many scenarios (in-coverage, out-of-coverage, and partial coverage) and cases under a given scenario, there cannot be a single procedure which controls all the synchronization transmission. Also, we need to define a priority when a device selects synchronization reference among the received synchronization signals and channels. When a device becomes a synchronization source which transmits synchronization with reference to received ones, in-coverage source is prioritized over out-of-coverage.

Following is an example of the synchronization procedure: when an out-of-coverage device wants to become a synchronization source for itself without any reference, it randomly selects one sequence from the 168 sequences for out-of-coverage and set in-coverage indicator to 0.

5.7.4.2.2. Discovery

From a physical layer point of view, discovery is just a broadcasting of repeated small data. The main application of discovery is an advertisement which may promote penetration in mobile market. Discovery has nothing to do with communication except Device-to-Network relay application as in Figure 5.7-7. In order to establish Device-to-Network connection, discovery is a prerequisite condition for selecting relay device. Once a relay device is selected, the data passes through the relay device from the network to the remote device and vice versa, subject to the specified D2D communication.
The main features of discovery are divided into: 1) resource allocation, 2) transmission format, and 3) procedures for discovery.

5.7.4.2.1.1. Resource allocation for discovery

Discovery resource is configured by the eNB. In Figure 5.7-8, offset indicates the start of the first discovery period with respect to SFN#0 of the serving cell. In the event of a non-integer number of discovery periods within the SFN cycle, the last discovery period extends into the next SFN cycle. All the subframes and PRBs consisting of discovery period are not allowed to be allocated to a discovery transmission. The eNB configures some subframes by using a bitmap of length $N$ and some PRBs for D2D discovery. Where $N$ is the number of subframes and specified depending on the duplex mode and TDD configuration.

Figure 5.7-8 Discovery transmission and resource allocation.
5.7.4.2.2. Discovery transmission format

The sizes of the discovery message and the discovery resource are fixed. The size of discovery resource spans over two contiguous PRBs in the frequency domain and one subframe in the time domain.

As per eNB configuration, the number of repeated discovery transmissions per period can be 0, 1, 2, or 3: if configured to 2, RV0, RV2, and RV3 of turbo coded discovery message are transmitted every period. Each redundancy version of discovery message occupies one discovery resource and every transmitted redundancy versions within a period uses contiguous subframes in the time domain.

Although the transmit format for decoding is known to a receiving device in the perspective of physical layer, discovery message contains 32-bit message integrity check (MIC) and 4-bits time-calibration information.

5.7.4.2.2.3. Procedures for discovery

Two types of discovery transmissions are supported: One (type 1) is a device autonomous scheduling where a device having a discovery message to transmit randomly selects discovery resource and transmits. The other (type 2B) is eNB scheduling where the eNB allocates discovery resource for $j$th transmission at discovery period $i$.

In type 1 discovery, a device randomly selects one discovery resource and transmits RV0 every discovery period with the eNB-configured probability of $p$. If retransmissions are configured, the next discovery resource position in the same period is derived from the previous position. Retransmissions in the same period use contiguous subframes in the time domain. But hopping is applied for the transmissions of different redundancy versions.

In type 2B discovery, the eNB instructs discovery resource position for an initial transmission of discovery message in the first period. All the other discovery resource positions in the next periods including retransmissions are derived from this eNB instruction.

Since the devices near the transmitter can decode initial transmission, they don’t have to wait to collect other redundancy versions. However, the devices far from the transmitter have to wait to collect and combine all the redundancy versions.
5.7.4.2.3. Communication

As in discovery, D2D communication is also a broadcasting of repeated data in the physical layer point of view.

With respect to SFN#0/DFN#0, multiple communication periods are multiplexed into an SFN cycle in the time domain as shown in Figure 5.7-9. Each period consists of a control pool and a data pool. Multiple fixed size control resources are multiplexed into a control pool in the time and frequency domain and each control resource conveys scheduling message indicating resource allocation and transmit format of the data to be transmitted in the following data pool of the same period.

![Figure 5.7-9 D2D communication period, control and data pool.](image)

5.7.4.2.3.1. Resource allocation for D2D communication

As previously stated, some of the LTE UL subframes can be configured for D2D communication. To indicate D2D subframes, a bitmap of length \( N \) is used. While only one length-\( N \) bitmap is mapped to control pool, multiple bitmaps are multiplexed into the data pool. Among the LTE time-frequency resources in the pool, subframes indicated by ‘1’ in the bitmap and PRBs configured by eNB are used by D2D communication.

5.7.4.2.3.2. Channel design for D2D communication

Two channels are defined for D2D communication: control and data channel. Control channel conveys scheduling information such as time-frequency resource allocation, hopping, time resource pattern, MCS, group destination ID required to decode data channel.

All the information required to decode control channel is known to every devices. A device having data in the transmit buffer has to select control channel index, which corresponds to two control resources in the control pool. For out-of-coverage, a device randomly selects...
control channel index, but in-coverage it is scheduled and informed to a device by eNB. Finally, the device transmits RV0 and RV0 on the two control channel resources, respectively. A device wanting to receive necessary data has to decode all the control channels transmitted in the control pool. If group destination ID in a control channel matches to itself, the device tries to decode corresponding data channel according to the previously received control channel indications: resource allocation, hopping, and MCS. If there is no relevance to group destination ID, the device doesn’t have to monitor data pool.

HARQ is supported for data channel without feedback, which means all PDUs are transmitted 4 times in the order of RV0, RV2, RV3, and RV1.

All the subframes configured for D2D communication in the data pool are collected and then partitioned by grouping $K$ subframes of each. For a specific device point of view, $k$ subframes among the $K$ are used for D2D transmission. The available positions of $k$ subframes among the $K$ are represented by ‘Time Resource Pattern’.

Figure 5.7-10 shows an example where the values of $N$, $K$, and $k$ are 40, 8, and 2, respectively. Here, the bitmap representation of time resource pattern is ‘10010000’. After the transmission of PDU0 of redundancy versions of RV0, RV3, RV2, and RV1, PDU1 is transmitted. In this way, multiple PDUs are transmitted until the end of the data pool.

From an individual device’s point of view, both the available D2D subframe positions and the number of contiguous PRBs per subframe are known by ‘time resource pattern’ and ‘time-frequency resource allocation’ field in the control channel, respectively. Even though the number of PRBs per subframe is the same over the period, hopping is supported to obtain frequency diversity. Since, the size of time-frequency resource, hopping, and MCS are configured per period basis, dynamic scheduling is supported for D2D communication.

Figure 5.7-10 HARQ transmission of D2D communication.
5.7.4.2.3.3. Procedures for D2D communication

There are two scheduling modes in D2D communication: an eNB scheduling and an autonomous scheduling.

For out-of-coverage case, an autonomous scheduling is applied: The transmit device randomly selects control resource index without carrier sensing or any other techniques to avoid collision and then transmits control channel on the selected control resources. Of course, the device constitutes the contents of the control channel of its own decision. Therefore, collisions may occur.

For in-coverage case, both autonomous and eNB scheduling can be applied. Unlike the autonomous scheduling, there is no need to worry about a collision since the eNB schedules time-frequency resources of all the D2D communication transmissions. Most of the contents of the control channel are originated from the eNB via LTE downlink control channels, PDCCH or EPDCCH.

5.7.4.3. D2D on 5G perspectives

We had seen the advent of new generation of mobile communication systems is always based on the maturity of new technology regarding multiple access or waveform. In the same context, we understand WCDMA and OFDMA as 3G and 4G, respectively. But at the time we are discussing 5G, there is no deserved promising technology commonly recognized. However, from a user's point of view, what is important is not a technology but a service which can be distinguished from previous generations. IoT applications using wireless technology are popped up services recently coloring 5G features. In that sense, we should consider D2D as candidate technology suitable for providing IoT.

Spectrum

In order to accommodate massive connectivity and various kinds of IoT data including both narrow and wide bands, we need to consider unlicensed as well as licensed bands. Since making a charge for every IoT connection is not an easy problem, some IoT applications on unlicensed bands should be considered as a manufacturing point of view.

Unicast

Unlike public safety (PS) services based on PTT, other IoT applications regarding medical and private security require unicast connection. Critical time as well as tolerable delay is also
considered. Taking the varieties of D2D applications into account, D2D could be a gateway to introduce a new waveform and/or a multiple access technology viewed as 5G.

Multi-hop ad-hoc

Self-organization is essential for sensor network for IoT applications. Traditionally, multi-hop ad-hoc has been the key technology for this purpose but still there is no satisfactory result. For coverage extension, current state of art takes advantage of repeated transmissions. Multi-hop may contribute to extreme coverage extension.

Others

Vehicle-to-Vehicle (V2V) is the first D2D commercial application in 3GPP. The number of cars would be approaching the volume of human population in the near future. Therefore, industries gaze at smart cars equipped with V2V technology. Considering the fact that industries’ de facto standard has strong impact on this field, current D2D technology should be enhanced in terms of efficient resource allocation, channel access, and transmission technology.

References

5.7.5. Edgeless Cellular Network

5.7.5.1. General Overview
In the past, mobile operators have focused on how to increase the number of subscribers and how to provide mobile broadband service. Recently, the need for improved user experience has raised since various user requirements and services available. To cope with such a trend, the target user experience data rate for 5G communication systems was set in the range of 100 Mbs to 1 Gbps which is 10 to 100 times greater than that of 4G. User experience data rate can be thought as cell-edge user throughput since cell-edge user usually experiences the lowest throughput in the cell. ITU has defined cell-edge user throughput as the 5% point of CDF of the user throughput normalized with the overall cell bandwidth [1]. Under conventional cellular networks, throughput difference between a cell edge user and a peak user is in the order of tens. Therefore, without special measures to improve the cell-edge user throughput, the peak data rate should be boosted in the order of hundreds or more to satisfy the target user experience data rate of 5G - which is impossible. There are several well-known technologies to improve the cell-edge performance such as coordinated multipoint transmission/reception (CoMP) in LTE. Moreover, several companies have proposed several edge-user performance enhancing solutions. More recently, user-centric network architecture has emerged as a new candidate to provide high consistent user throughput. Specifically, ‘pCell technology’ has demonstrated that it can provide high consistent data rate over the coverage. In the rest of this chapter, we will briefly summarize various technologies toward edgeless cellular network.

5.7.5.2. Research Trends
In LTE releases 10 and 11, a coordination between transmission points was discussed under CoMP. Downlink CoMP can be divided into two categories [2]:

- Coordinate Scheduling/Beamforming (CS/CB): In CS/CB, the user data is not shared among the transmission points. The data is transmitted from a specific point. Instead, the scheduling and the link adaptation can be coordinated among the transmission points.
- Joint Processing (JP): The transmission to a user can be carried out from multiple transmission points. The transmission from different transmission points can be simultaneous
(joint transmission) or alternate (dynamic point selection).

CoMP can improve the cell-edge user throughput, but the gain is shown to be limited due to the limited channel state information feedback and the limited interface capability among the transmission points.

CA Omni technology is an advanced carrier aggregation (CA) scheme applicable to all deployment scenarios [3]. Extending CA capability from normal intra-site CA to inter-site and HetNet CA to help provide a truly seamless user experience. With CA Omni solution, spectrum utilization can be maximized to achieve edgeless network by enhancing CA scheme with centralized resource coordinator.

A “Smart Scheduler” has been proposed to increase cell-edge performance [4]. As the name implies, Smart Scheduler is a scheduling scheme that achieves performance gain with legacy LTE specification. Smart Scheduler has main features such as frequency selective scheduling, interference shaping, QoS differentiation, intra-frequency load balancing, uplink power control, and multi-cell scheduling. Smart Scheduler is shown to improve cell edge throughput by approximately 110% with non-ideal backhaul.

A prototype communication system adopting user-centric architecture (‘pCell technology’ meaning ‘personal cell’) has been demonstrated in 2015 [5]. In the demonstration, they have shown that with off-the-shelf LTE devices, their technology can yield 9x spectral efficiency gain and such gain can be secured over the entire coverage. It is quite interesting in two perspectives in that high gain can be obtained throughout the entire coverage and such a gain does not require specification change. To provide high space selectivity, number of RRHs are deployed in the coverage and one centralized base-band unit controls the RRHs. Basically, the central base-band unit and RRHs constitute a C-RAN architecture. The pCell technology is user-centric in that serving RRHs for each user are selected based on the location and transmission range of each user. As the user moves or the channel changes, serving RRHs can be dynamically changed. The pCell technology adopts zero-forcing precoding to provide interference free signal. As the users are perfectly decoupled by precoding in spatial domain, the full time-frequency resource can be used to each user and thus, each user can presume itself as a single user in a cell (hence the name ‘pCell’). The pCell gain increases linearly with the number of users until the number of users does not exceed that of pCell antennas. The
high data rate can be achieved everywhere in the coverage area. pCell technology has its meaning in that it is first known genuine user-centric architecture and implemented a sophisticated RF circuit calibration algorithm to estimate channel based on channel reciprocity. Often pCell technology is compared with CoMP because multiple transmit points cooperate to increase throughput of users. However, apparent differences exist between pCell technology and CoMP. The differences are summarized in Table 5.7-4 in terms of LTE specification.

**Table 5.7-4 Comparison of CoMP and pCell technology**

<table>
<thead>
<tr>
<th></th>
<th>CoMP</th>
<th>pCell technology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transmission Mode</strong></td>
<td>TM10</td>
<td>TM7</td>
</tr>
<tr>
<td><strong>DM-RS (UE-specific)</strong></td>
<td>Multiple DM-RS ports required (Orthogonality by code, frequency, and time)</td>
<td>Single DM-RS port for all users (Orthogonality by spatial precoding)</td>
</tr>
<tr>
<td><strong>CSI Acquisition</strong></td>
<td>UE feedback (Primary) Channel reciprocity (open)</td>
<td>Channel reciprocity (Primary) UE feedback (open)</td>
</tr>
<tr>
<td><strong>Data Transmission</strong></td>
<td>Non-coherent/coherent</td>
<td>Coherent (zero-forcing)</td>
</tr>
<tr>
<td><strong>Number of Transmission Points</strong></td>
<td>Limited (≤ 3)</td>
<td>Unlimited</td>
</tr>
<tr>
<td><strong>Applicable Area</strong></td>
<td>Cell-edge</td>
<td>Entire coverage</td>
</tr>
</tbody>
</table>

### 5.7.5.3. Challenges in edgeless cellular network

In this sub-section, we will briefly introduce challenges of edgeless cellular networks. Specifically, we will address challenges with user-centric architecture.

One of the most challenging problems in user-centric architecture is acquiring channel state information. In TDD, channel reciprocity can be utilized to obtain the CSI where sophisticated antenna and RF circuit calibration are huge problems. The CSI acquiring problem becomes even worse when it comes to FDD. Since uplink and downlink are working in different frequency bands, channel reciprocity does not hold anymore and a user feedback is required. As in the example of pCell, serving RRHs are selected dynamically. The number of serving RRHs, index of RRHs, and channel state changes dramatically according to user movement and time progresses, and to achieve high promising gain in user-centric architecture, user should feed back accurate CSI. However due to its limited capacity for feedback channel, performance degradation seems inevitable.
While the conventional communication systems are primarily designed for single-user (users have different time/frequency resources), user-centric architecture is inherently multi-user system (users share the same time/frequency). As a number of users share the same time, frequency resources with user separation done in spatial domain, user selection became a new challenge. To guarantee high and consistent data rate to all users, the aggregated channel composed by aggregating channels to each user should be in ‘good condition’. Otherwise, some users cannot acquire high data rate and consistency breaks.

User centric architecture is meaningful when a user can utilize many transmission points for communication. If there are enough RRHs, users can choose or disregard RRHs for its benefit. However, if there is not enough number of RRHs, such a selection is impossible and a user should be connected to a few neighboring RRHs. A typical example of such a system is legacy communication systems where one high-rise base-station covers the area and user should attach to one base-station. Once there are enough number of transmit points, user can select with its preference and user centric architecture becomes meaningful. Hence, network densification is essential for user centric architecture to be meaningful. With dense network, there are numerous challenging problems in deployment and operation.

References


5.8. Application Specific Radio Technologies

5.8.1. Massive Connectivity

5.8.1.1. NB-IOT

5.8.1.1.1. General Overview
As one of usage scenarios for 5G, Massive Machine-type communications (mMTC) are characterized by a very large number of connected devices transmitting a relatively low volume of non-delay-sensitive data. Also, MTC devices are required to be low cost, and have coverage enhancement and a very long battery life [1].
Diverse technologies and use cases have been developed within and beyond the scope of cellular MTC. In the unlicensed spectrum, Weightless™ and SIGFOX are already commercialized for low power tracking use cases. In the licensed spectrum, 3GPP initiated the work on eMTC and Cellular Internet of Things (IoT) to provide the worldwide connectivity needed for most IoT use cases. Especially, NarrowBand IoT (NB-IoT) as a single solution for Cellular IoT will be completed in June 2016.
However, as more and more things get connected, the industry needs to avoid fragmentation. These connected things range from low complexity devices to high quality devices such as smart phones, sensors, actuators and vehicles. In addition, the number of connected devices is expected to amount up to $10^6$ devices per km$^2$ with various performance and operational requirements [2]. As a result, 5G technologies can be based on the unified air interface or flexible air interface converged into the unified framework to provide massive connectivity for a variety of devices.

5.8.1.1.2. State of the Art and Research Trend
NB-IoT is a radio access based on a non-backward compatible variant of E-UTRA so that the massive number of IoT devices can transmit small amount of data with relaxed delay characteristics. It is expected that the NB-IoT device density per cell is up to $52,547$[3], which is 10 times less than that of 5G mMTC. In addition, NB-IoT devices are required to improve coverage up to 164 dB MCL since a large number of IoT devices are deployed in a basement,
or in deep inside building. NB-IoT devices are required to reduce complexity and cost so that they can be deployed on a mass scale or in a disposable manner. Furthermore, NB-IoT devices are required to reduce the power consumption so that they can have a ten years battery life. NB-IoT should support three different modes of operation. Stand-alone operation utilizes the spectrum currently being used by GERAN systems as GSM re-farming carriers. Guard-band operation utilizes the unused resource blocks within a LTE carrier's guard-band. In-band operation utilizes resource blocks within a LTE carrier. NB-IoT UE RF for both downlink and uplink has 180 kHz bandwidth operating HD-FDD with type B guard period.

![Figure 5.8-1 Three different modes of operation for NB-IoT](image1)

![Figure 5.8-2 Collision Avoidance with legacy LTE signals for in-band operation](image2)
For the downlink, OFDMA with 15 kHz subcarrier spacing is used [4]. For in-band operation, physical channels and signals shall be designed to avoid collision with legacy LTE signals, i.e. Cell-specific Reference Signal (CRS), PDCCH and Channel State Information Reference Signal (CSI-RS) as shown in Figure 5.8-2. A single synchronization signal is newly designed for the different modes of operation so that the synchronization signal is spanned up to 11 OFDM symbols within one subframe for normal CP and punctured by LTE CRS if a collision occurs. The synchronization signal of NB-IoT consists of primary synchronization signal (PSS) and secondary synchronization signal (SSS) which indicates one of 504 physical cell identities (PCID).

Common data and control channel frameworks are designed for the different modes of operation. For all downlink channels, the maximum size of a transport block (TB) is no less than 520 bits, which is applicable to Tail-biting Convolutional Coding (TBCC).

NB-IoT supports physical downlink channel carrying Master Information Block (MIB), which contains information required to acquire the rest of the system Information that is grouped into different System Information Blocks (SIB). Scheduling of different SIBs is PDCCH-less with different periodicity. Physical downlink control channel for NB-IoT is supported for scheduling downlink and uplink grants.

For the uplink, SC-FDMA with 15 kHz subcarrier spacing is used for multi-tone transmission which can be occupied by 3, 6, or 12 subcarriers. In addition, two uplink numerologies, i.e. 3.75 kHz and 15 kHz, are configurable for single-tone transmission as a special case of SC-FDMA.

Single-tone transmission in SC-FDMA achieves high PA efficiency regardless of modulation scheme reducing the overall development cost and avoiding market fragmentation. However, multi-tone transmission in SC-FDMA requires peak-to-average power ratio (PAPR) reduction mechanism, e.g. PSK modulation and PAPR reduction precoding.

In all three modes of operation, one preamble-based PRACH scheme is used for all MCL cases. NB-PRACH is transmitted in a single tone of 3.75 kHz subcarrier spacing with frequency hopping within 12 subcarriers.

MAC, RLC, PDCP and RRC procedures are based on existing LTE procedures and protocols, and optimizations are required to support small data transfer based on the new design of
physical layer. Furthermore, S1 interface to CN and related radio protocols is enhanced to support signaling reduction for small data transmission [5].

References

5.8.1.2. Compressed Sensing

The massive connectivity of Internet-of-things (IoT) devices is characterized by the requirement of a large number of devices with low activity that can quickly access a base station (BS) whenever necessary. Herein, the low activity indicates that a device transmits or receives its data quite rarely. Note that supporting high activity devices is more related to a capacity enhancement issue, rather than such massive connectivity. In this sense, this subsection mainly focuses on an efficient uplink access for a large number of IoT devices characterized by low activity.

Connection with only a small number of devices among a large number of devices is closely related to the problem that a concept of compressed sensing (CS) can be employed to solve. The CS provides us with the unique solution of underdetermined linear equations where the number of variables is more than the number of linearly independent equations, when there are only nonzero solutions much less than the number of variables. In this regard, this subsection reviews uplink access methods based on this concept.

5.8.1.2.1. Basics of Compressed Sensing

Prior to reviewing previous works on efficient uplink access using the CS, the basic principles for the CS are introduced. Let $\mathbf{x}$, $\mathbf{y}$, and $\Phi$ denote $N \times 1$ vector including $K$ nonzero elements, $M \times 1$ vector, and $M \times N$ matrix, respectively. Linear equation $\mathbf{y} = \Phi \mathbf{x}$ is considered, where $M < N$ and the rank of $\Phi$ is $M$. In general, a unique solution cannot be determined, as there are more than one solutions. That is, the estimation of $N$ variables from $M$ measurements is impossible. However, if the number of nonzero unknowns is given, the CS enables the unique solution to be derived. If $\mathbf{x}$ has $K$ nonzero elements (this is called $K$-sparse), by appropriately designing $\Phi$ for measurement, $\mathbf{x}$ with $N$ unknowns can be uniquely determined only from $M$ measurements, with an overwhelming probability. In general, compressive sensing is a signal processing approach behind this concept, i.e., efficiently acquiring and recovering a signal by solving the underdetermined linear systems, based on the principle that a signal can be recovered by its sparsity with much fewer samples than required by the Shannon sampling theorem.

In CS, a sensing matrix $\Phi$ should meet the restricted isometry property (RIP). This property can be interpreted as determination of the unique solution is to distinguish the points in $N$. 
dimensional space, and if two points can still be distinguished after the projection into the dimensions less than \( N \), one can determine the unique solution in this reduced dimension. A representative example of sensing matrix to meet this RIP is a matrix whose elements are independently taken from a random distribution, e.g., Gaussian distribution. When using this Gaussian distribution-based sensing matrix to determine the unique \( N \times 1 \) vector \( x \), which has \( K \) nonzero elements, the amount of measurement, \( M \), is given by \( M = O\left( K \log\left( \frac{N}{K} \right) \right) \).

![Figure 5.8-3 Straightforward application of compressed sensing for multiple access](image)

An \( N \times 1 \) unknown vector \( x \) can be recovered through minimizing \( l_1 \)-norm that sums the absolute value of each element, which can formulated as the following convex optimization problem:

\[
\min_{x} \| x \|_1 \quad \text{subject to} \quad y = \Phi x
\]

The above method is still complex; thus, the methods with less complexity, i.e., greedy algorithms such as orthogonal matching pursuit (OMP) have been studied and they can be used for the recovery of \( x \). In contrast, in practical systems, the sparsity information, i.e., \( K \), is not typically known a priori. Moreover, it may not be easy to implement the sensing matrix consisting of Gaussian random variables. In this regard, the focus of recent studies has been moving on to issues such as the design of sensing matrix for unknown \( K \) and more implementation-friendly sensing matrix.
5.8.1.2.2. On-Off Random Access based on Compressed Sensing

Let \( x \) denote the uplink transmission data for mobile station (MS) 1 to MS \( N \). In this case, unknown vector \( x \) of linear equation, \( y = \Phi x \), consists of the information about the position and value of each element. Herein, the position corresponds to the MS identifier while the value means the data value of each MS. That is, the nonzero element can be considered as the data that an MS sends while the zero element is equivalent to the event where an MS does not transmit any data. Figure 5.8-3 illustrates the multiple access situation in which \( s = [s_1, s_2, \ldots, s_N] \) represents the vector including the transmission symbol for each MS, and \( h_n \) denotes the wireless channel gain between the \( n \)-th transmitter and receiver.

![Figure 5.8-3 Multiple access situation](image)

Since there is no interference between wireless links, a channel matrix \( \mathbf{H} \) is given by a diagonal matrix whose diagonal elements are \( h_n \)'s. The transmission code is assumed to be assigned to each MS a priori by its serving base station (i.e., the receiver), and the spreading code assigned to the \( n \)-th MS is denoted by \( \phi_n \). As a result, the received signal without background noise can be expressed as follows:

\[
y = \sum_{i=1}^{N} h_n \phi_i s_i
\]

In the above equation, \( N \), \( K \), and \( M \) denote the total number of MSs, the number of active MSs who have transmission data, the length of the code assigned to an MS, respectively, where we assume \( K < M \ll N \). In practical wireless communications, received signals are affected by thermal noise, so \( \omega \) denotes \( M \times 1 \) additive noise vector; thus, \( y = \Phi x + \omega \), where the \( n \)-th column of sensing matrix \( \Phi \) is given by \( h_n \phi_n \). The receiving performance depends on the signal to noise ratio (SNR) as well as the value of \( M \). Thus, the effect of SNR may lead to the near-far problem. The work of [1] assume that the uplink transmissions are time
synchronized while the work of [2] studies the asynchronous transmissions for CS-based multiple access.

5.8.1.2.3. Random data traffic transmission via compressed sensing

When an MS transmits its data to its serving BS without any scheduling by the BS, the MS needs to inform the BS of the MS ID and its data. The CS-based access introduced in Section 5.8.1.2.2 allows for BS both to identify the transmitter while decoding the received data. In [3], the CS-based medium access control method other than that in [1] is proposed. The scheme proposed in [3] supports a variable frame structure.

Figure 5.8-4 presents the variable frame structure proposed in [3] for CS-based random traffic transmission. A BS transmits the signal indicating the beginning of a frame for distributed access, and each MS spreads its data using the spreading code assigned to itself if it needs to transmit its data. At this time, the MS transmits a part of its spread data sequentially until the BS indicates the end of a variable frame rather than transmitting all of its spread data. Thus, within a variable frame, each MS always can transmit one unit of data at most. Once a BS can successfully decode any data unit, the BS informs MSs of the termination of a variable frame. Here, the BS determines if it receives a data successfully by checking the same decoding results on consecutive time slots within a variable frame. This is based on the principles of sequential compressed sensing [4]; that is, the CS can recover the original data better by increasing or adjusting the size of the sensing matrix. As a result, the work of [3] is to apply the principle of sequential CS to medium access control protocol straightforwardly. In [3], it is shown that its proposed method is superior to carrier sensing multiple access (CSMA) in terms of the mean and variance of packet transmission delay as long as traffic load is low, through simulations. However, this method has several drawbacks: e.g., (i) communication fails if the BS signals informing the beginning and ending of a variable frame is corrupted; (ii) the spreading code based on CS is still based on random variables, which is not implementation friendly.

A similar framework can be applicable to MAC protocol for wireless LAN [5], the CS-based access is executed for the purpose of uplink resource request prior to transmitting its data. First, the access point (AP) broadcasts the request solicitation indicating that MSs can now
transmit uplink resource request. Then, an MS or a host transmits the code designed to inform the AP of the need of uplink resources. Through the principles of CS discussed in the above, the AP can realize which nodes request uplink resources and it designates which nodes transmit without any collision by broadcasting a bitmap for multiple grants. The results in [5] exhibits that the CS-based approach has better performances compared with the conventional distributed coordination function (DCF).

5.8.1.2.4. Compressive sensing based one-shot random access for machine type communication

The stringent requirements of the fifth generation network would not only include support for massive connectivity, but also quick channel access for delay-intolerant applications. Some of the MTC applications require a sub-millisecond channel access which cannot be achieved by the random access-based request and grant procedure in the current LTE specification. Techniques based on permanent assignment of spreading codes as in Section 5.8.1.2.2-4 are also not scalable with the foreseen number of MTC devices under a base station. Moreover, MTC is also characterized by small packets which make the channel access procedure of LTE inefficient with high overhead.

In [6], a random access scheme is presented which allows MTC devices to access the channel in one shot, i.e., immediately upon request. All users associated the same base station are assumed to be synchronized and randomly select a set of sequences to spread their signals. The main issue in serving a massive number of MTC devices is the overload in random access channel (RACH). For example, in an LTE system where there are 64 preambles to request channel access with, the collision rate exceeds 10% when 8 devices are active at each random access opportunity. To mitigate this problem, we need to have a way for users to contend for channel with less collision. In [6], a one-shot random access is proposed with the objective of alleviating the RACH congestion problem where users spread their signals with multiple sequences. It is the multi-sequence spreading random access (MSRA) scheme, in which active MTC users select a set of multiple sequences (sequence set) to spread their signals. Applying a set of multiple sequence to spread one’s signal has an advantage of twofold:

1. As a large pool of sequence sets can be generated from fewer unique sequences. For example, generating 1000 sequence sets each composed of 4 sequences from 128 unique
sequences. This reduces the collision rate which is the main bottleneck in the massive random access.

2. A diversity gain may also be achieved as a larger Euclidean distance can be achieved by employing multiple sequences than a single sequence. The advantage of MSRA can be discussed with illustration for single-sequence spreading random access (SSRA) and MSRA schemes in Figure 5.8-5(a) and Figure 5.8-5(b), respectively. In these figures, each frame is divided into smaller groups so as to reduce detection complexity, in which a part of symbols of transmission bursts from active users is facilitated. In this illustration, there are three active users, each selecting its own spreading sequence once in each group for the single-sequence spreading random access scheme, or each selecting its own spreading sequence for each symbol for the multi-sequence spreading random access scheme. Since Euclidean distance between the multi-user signals can be increased by employing the different spreading sequence for each symbol over a group, i.e., incurring much less correlation among the multi-user signals (e.g., all symbols in the different colors in Figure 5.8-5(b)), MSRA tends to outperform SSRA, in which the multi-user signals would be confused easily at the receiver due to higher correlation among them (e.g., light vs. dark brown sequences in Figure 5.8-5(a)).

In [7], a greedy algorithm which exploits the uniform sparse structure in CS-based channel access techniques, known as iterative order recursive least square (IORLS) estimation, is also presented. When users are either active or inactive for the entire frame, then received signal can be expressed in a matrix form as
where the index $i$ denote the symbols in the frame. Then, user symbols $x_i$'s have the same support structure which can be exploited for better accuracy in the CS estimation. In [7], it has been shown that for highly delay-intolerant applications where retransmission is not allowed, the MSRA system with IORLS can support 12x more users than the current LTE system.

### 5.8.1.2.5. Compressive channel and multi-user detection for post-LTE systems

The CS framework can be applied in three areas in communication networks: sparse data decoding, sparse channel estimation, and sparse users' activity. Most natural signals are sparse in some basis. If the data can be sensed and taken in that domain, CS can be applied to decode the signal with a lesser rate than the required Nyquist rate. This phenomenon is called sparse data decoding. On the other hand, channels are usually sparse with most of the channel gain components zero or near to zero. Consider a channel gain of $h = [h_1, h_2, \cdots, h_k]$ with each path of $l_k$ taps. If the nonzero (significant) components of $h$ are much lesser than $l_k$, then $h$ is said to be sparse [8,9]. Sparse users' activity deals with the situation in which only few MTC users from a total set of users under a BS are active at a time. This is results from the sporadic and uncorrelated nature of MTC communication. In [8,9], sparse channel and users' activity have been exploited to allows for transmitting both control and data information in the same time, as a new means of a random access procedure. A power ratio $\alpha$ is set for the portion of power allocated to control signal (preamble) transmission and hence, $(1 - \alpha)$ is the power ratio for data transmission. The received signal can then be represented as

$$y = D(p)h + C(h)x + \omega$$

where $D(p)$ and $C(h)$ are the measurement matrices for the sparse channel $h$ and sparse users' vector $h$, respectively. Then, successive decoding can be applied to first estimate the channel and then the data.

In [10], the sporadic data for machine-to-machine (M2M) is overlaid on the constant human-to-human (H2H) communication. In the work, the power allocated to M2M and H2H traffic with a power ratio $\alpha$, as in [8,9]. The BS first treats the H2H data as noise and decodes the M2M data using the IORLS algorithm. Then, the data for H2H is decoded by canceling the decoded M2M signal from the received signal.
References


5.8.2. V2X Communications

5.8.2.1. General Overview

ICT capability is considered a focal point for competitiveness of next generation vehicles on top of traditional exterior/interior design and mechanical performances. Recently, 3GPP has initiated a standardization program for LTE-based Vehicle to Everything (V2X) technologies composed of Vehicle to Vehicle (V2V), Vehicle to Pedestrian (V2P) and Vehicle to Infrastructure/Network (V2I/N), whose specification is expected to be available around 2016~2017. V2X communication passes the information from a vehicle to any entity that may affect the vehicle, and vice versa.

Leveraging the spectrally-efficient LTE-based air interface and the cost-effective network deployment, the next generation V2X technologies can substantially improve to enhance safety, convenience and inforainment including driver assistance, vehicle safety, travel information, and traffic operations, even in the high mobility and densely populated vehicle environment. Moreover, V2X communication is expected to further evolve to support highly mission-critical connected car services such as autonomous driving and platooning as well as broadband inforainment services in conjunction with 5G new air interface design. Therefore, there is significant societal benefit and commercial value to delivering safety, mobility and convenience applications that rely on V2X. This section gives an overview of state of the art and applications on V2X communications, and elaborates on key technical solutions related to V2X.

5.8.2.2. State of the art and trend

5.8.2.2.1. Progress in 3GPP

Recently, 3GPP activities have included use case, service requirements, network architecture enhancements, and radio level enhancements to support LTE V2X services. In 3GPP TSG SA meeting #69, Sep. 2015, a technical report for Release 14 study item on LTE support for V2X services was approved. The study item was targeted to investigate the essential use cases and requirements for the following

• V2P (Vehicle-to-Pedestrian): covering LTE-based communication between a vehicle and a device carried by an individual (e.g., handheld terminal carried by a pedestrian, cyclist, driver or passenger).

• V2I/N (Vehicle-to-Infrastructure/Network): covering LTE-based communication between a vehicle and a roadside unit/network. A Roadside Unit (RSU) is a transportation infrastructure entity (e.g., an entity transmitting speed notifications) implemented in an eNB or a stationary UE.

All the above services are defined from the end-to-end view point. For V2V, it can be implemented either by directed communication over PC5 interface between the UEs (sidelink), or the eNB over the conventional Uu interface (uplink and downlink).

In 3GPP TSG RAN meeting #68, June 2015, a study item of feasibility study on LTE-based V2X services was approved. For support of PC5 interface based V2V, the study item is targeted to identify necessary enhancements to the resource allocation mechanism (e.g., mitigate impact of half duplex constraint, reduce resource collision, enhance pool structure, enhance resource patterns, sidelink control information transmitted in the same subframe as the associated data) to meet identified requirements for robustness, latency, overhead and capacity. Also, they plan to identify any necessary PC5 enhancements for high Doppler case (e.g., up to 280 km/h up to 6 GHz) and also synchronization based on GNSS, at least for out-of-coverage operation.

For the support of Uu interface-based V2V, and PC5/Uu interface-based V2I/N and V2P, the study item in 3GPP TSG RAN is planned to evaluate the feasibility of Uu transport for V2V and V2P in terms of meeting latency requirements, network coordination required, resource efficiency, and energy efficiency of UE, and to identify and evaluate enhancements required to support each of eNB-type and UE-type RSU. According to the current SA status, solutions for UE-to-UE relaying based on a new architecture for UE-type RSU will not be studied. Also 3GPP has a plan to identify and evaluate the necessity of enhancements to multi-cell multicast/broadcast for reduced latency and improved efficiency. Support for PC5 transport for V2V services is of highest priority in 3GPP until Dec. 2015.

5.8.2.2.2. V2X use cases

In the literature and standardization bodies, there has been much interest in studying V2X use cases.
cases made up with V2V, V2P (pedestrian or vulnerable road use), V2I/N. The range of V2X application can be extended to personal communication, social benefit to bring more comfort, convenience as well as road safety services. Recently, by several major automotive brands, V2X has specified as a key technology to improve driver experience and comfort, or to eliminate the driving side of motoring altogether with autonomous driving systems. Some V2X use cases are summarized as follows [1].

- **Forward collision warning**

  This use case is intended to warn the driver of the host vehicle in case of an impending rear-end collision with a remote vehicle ahead in traffic in the same lane and direction of travel. This helps drivers in avoiding or mitigating rear-end vehicle collisions in the forward path of travel.

- **Pedestrian warning**

  This V2P application provides warning information to vulnerable road users, e.g. pedestrian or cyclist, of the presence of moving vehicles in case of dangerous situation. Thus, by noticing imminent danger of a collision to a pedestrian, the pedestrian warning can be effective way to reduce accidents between car and pedestrians especially at intersections.

![Figure 5.8-6 Pedestrian warning](image)

- **Road safety and traffic control via infrastructure**

  In this use case, infrastructure nodes such as road side units and traffic safety servers for road safety generate and distribute traffic safety-related messages.
• Platooning
This use case allows a vehicle with V2V capability to join and leave a group of corporative-adaptive-cruise-control vehicles. Platooning provides convenience and safety benefits to participating vehicles, and improves road congestion and fuel efficiency.

![Figure 5.8-7 Platooning](http://www.sartre-project.eu)

• Autonomous driving
Autonomous driving enables vehicles to share their sensor data with other vehicles around them. By detecting hidden-from-view threats, especially in urban areas, it can extend the horizon of awareness beyond what the driver can see.

![Figure 5.8-8 Hidden vehicle problem](https://www.youtube.com/watch?v=QoD0m0-4nm8)
• High rate in-vehicle infotainment

V2X can be applied to support multiple high-end devices (displays) in vehicles for personalized services, e.g., video streaming, online gaming, etc. As illustrated in Figure 5.8-9, the concept car exhibited at CES 2015 by Mercedes-Benz had multiple UHD displays to support high quality in-vehicle infotainment services.

![Figure 5.8-9 High rate in-vehicle infotainment](image_url)

5.8.2.3. Research direction

5.8.2.3.1. Technical challenges

Technical challenges for V2X result from the environmental characteristics of vehicles, and tight requirements based on the mission critical applications related to safety. The main issues could be summarized as follows.

• Supporting high vehicle mobility is a crucial challenge for V2X, since handling high Doppler issue should be resolved especially for PC5-based V2V which has the “dual mobility nature”. Also, it accompanies needs for the frequent cell change and for adaptation to the fast change of channel state.

• Heavy vehicle density can be a challenge for V2X. More specifically, for example, the number of vehicles in a unit area increases drastically in urban cases or in traffic congestions.

• Requirement for extremely high data rate is a challenging issue since some applications which require heavy data transmission could be requested from multiple in-vehicle devices simultaneously.
• Low latency and high reliability are technically most demanding issues for V2X communications, especially for ITS safety use cases.

5.8.2.3.2. Candidate Technologies

To tackle some of the biggest challenges mentioned in previous section, basic research is currently in progress extensively to find key enablers for V2X. In this subsection, several core candidate technologies are listed as follows.

5.8.2.3.2.1. Enhanced D2D

In 3GPP release 12/13, D2D specification has been designed with the requirements of public safety and commercial consumer applications. There are several challenges on direct usage of release 12/13 D2D for V2X. Release 12/13 D2D communication targeted public safety services. The public safety services are typically operated in a low frequency band, e.g., 700 MHz. Also VOIP is the main service for public safety services. Low UE mobility was assumed (maximum 60 km/h). Simple random resource allocation is used under the assumption that the UE density is low: 3GPP assumed 3 ~ 7 transmitting UEs/km2. Release 12/13 D2D discovery targeted specific applications. The message size is fixed at physical layer.

In V2X services, much higher UE mobility than public safety services should be assumed and heavy UE density should be taken into account. Therefore, release 12/13 D2D should be enhanced to support V2X services.

For supporting high UE mobility, a new physical format is necessary to handle extremely fast channel variation caused by high Doppler and high frequency offset since D2D has the nature of dual mobility and the relative speed and residual synchronization error can be doubled. Also, high frequency band (e.g. 5.9 GHz) is under consideration for V2X services. High UE mobility causes frequent topology change. Thus, distributed local synchronization in Release 12/13 D2D may not be suitable. A common synchronization reference such as GNSS can be helpful for V2X. However, the D2D signal based synchronization method can be useful to cover the case where GNSS is not available.

For supporting heavy UE density, enhanced resource allocation is necessary to alleviate inter-device interference. Two types of interference are considered: resource collision and
power leakage to unused frequency resources. Battery consumption should be taken into consideration for V2P services in the enhanced resource allocation. “Always channel monitoring” is not a suitable design. The requirements on latency and reliability should be met in any case. There may exist inverse correlation between UE mobility and UE density. Adaptation to the traffic situation (e.g., by a network entity in charge of radio parameter control) can be helpful.

5.8.2.3.2.2. Multi-cell coordinated broadcast

For optimized support of Uu based V2X operation, it is important to identify necessary enhancements to multi-cell multicast/broadcast for reduced latency and improved efficiency. Local downlink multicast/broadcast mechanisms can be firstly considered for Uu based V2X operation where target broadcast area of a V2X message is not large. For V2V, the maximum target range is about 300 m. So it is enough to send a message only in adjacent cells. Coordination between adjacent cells can be utilized to improve efficiency. CoMP-like operations such as joint transmission and cell switching can be considered. To improve latency, local routing can be helpful since there is no need to travel the messages via the core network.

![Figure 5.8-10 Multi-cell coordinated broadcast](image)
5.8.2.3.2.3. Vehicle MIMO

In order to fulfill the needs for high data rate and reliability for the future V2X services, multiple vehicle antennas can be utilized. Since implementation of the large size antenna array in the outside of vehicle would be a drawback in terms of vehicle design and aerodynamics, in-vehicular distributed antenna system could be beneficial where multiple small antenna units are distributed on the different locations of vehicles. Thanks to a variety of spaces for vehicle antennas, these antenna units can be hidden at each corners of the car where various types of antennas are attached currently as shown in Figure 5.8-11. Since antenna units are quite far from each other, wireless channel characteristics of them would be quite different. Therefore, in-vehicle distributed MIMO has a potential to provide improved spatial diversity gain as well as spatial multiplexing gain.

References

Chapter 6 Concluding Remarks

As opposed to the previous generation of mobile communication standards in which peak data rate has been steadily increased with wider channel bandwidth, 5G is more than just enhancement in data rate. It would not be all about the smart phone, but also all other types of devices that can provide new user experiences with higher capabilities, and integrates physical and digital worlds. In fact, 5G is all about the end-to-end ecosystem with deeper innovation for the fully mobile and connected society in the future. 5G usage scenarios have been embodied in three different directions: enhanced mobile broadband for augmented user experience, massive connectivity for Internet of Things, and ultra-reliable and low latency communication for mission-critical applications. 5G will be enabled by a set of key capabilities for radio access, including peak data rate, user experienced data rate, area traffic capacity, mobility, connection density, and latency.

An nearly consensus view of 5G radio access technology is a set of tightly-coupled existing and new radio access technology, i.e., evolved LTE, WLAN, and potential new RATs on below 6GHz and above 6GHz bands. We have identified the enabling wireless technologies for 5G in various approaches, e.g., wider bandwidth technology with millimeter-wave band and multi-RAT integration, advanced modulation & coding, full radio duplexing, non-orthogonal multiple access and the underlying new waveforms, large-scale antenna technology, and advanced interference management. Furthermore, we also have investigated access architecture-related and application-specific radio technologies, such as advanced small cell, wireless backhaul, moving cell, and D2D communication, as another tier of essential features in 5G. Meanwhile, highly flexible 5G core infrastructure with flat and distributed network would be sought for more efficient and scalable deployment under a framework of network virtualization.

Much of our efforts go toward sharing information, encouraging universal standardization, and accelerating the technology innovations by cooperating with all participants in the global 5G ecosystem. One particular activity would be a global cooperation on developing an interim standard for pre-5G demonstration, e.g., targeting at Pyeong Chang Winter Olympic Games in 2018.
### Appendix 1: Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D-SCM</td>
<td>3-Dimension Spatial Channel Model</td>
</tr>
<tr>
<td>3GPP</td>
<td>3rd Generation Partnership Project</td>
</tr>
<tr>
<td>5GPPP</td>
<td>5G Infrastructure Public Private Partnership</td>
</tr>
<tr>
<td>ABS</td>
<td>Almost Blank Subframe</td>
</tr>
<tr>
<td>ACB</td>
<td>Access Class Barring</td>
</tr>
<tr>
<td>ADC</td>
<td>Analog-to-Digital Converters</td>
</tr>
<tr>
<td>ADSL</td>
<td>Asymmetric Digital Subscriber Line</td>
</tr>
<tr>
<td>AF</td>
<td>Amplify-and Forward</td>
</tr>
<tr>
<td>AFE</td>
<td>Analog Front-End</td>
</tr>
<tr>
<td>ANDSF</td>
<td>Access Network Discovery and Selection Function</td>
</tr>
<tr>
<td>ANR</td>
<td>Automatic Neighbor Relation</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>ARQ</td>
<td>Automatic Repeat Request</td>
</tr>
<tr>
<td>ASA</td>
<td>Azimuth Spread at Arrival</td>
</tr>
<tr>
<td>ASA/LSA</td>
<td>Authorized/Licensed Shared Access</td>
</tr>
<tr>
<td>ASD</td>
<td>Azimuth Spread at Departure</td>
</tr>
<tr>
<td>ATCA</td>
<td>Advanced TCA</td>
</tr>
<tr>
<td>AWGN</td>
<td>Additive White Gaussian Noise</td>
</tr>
<tr>
<td>BBU</td>
<td>Baseband Unit</td>
</tr>
<tr>
<td>BC</td>
<td>Broadcast Channel</td>
</tr>
<tr>
<td>BCE</td>
<td>Beam Coverage Enhancement</td>
</tr>
<tr>
<td>BICM</td>
<td>Bit Interleaved Coded Modulation</td>
</tr>
<tr>
<td>BLER</td>
<td>Block Error Rates</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>CA</td>
<td>Carrier Aggregation</td>
</tr>
<tr>
<td>CAPEX</td>
<td>CApital EXpenditure</td>
</tr>
<tr>
<td>CBR</td>
<td>Coded Bidirectional Relaying</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>CC</td>
<td>Component Carrier</td>
</tr>
<tr>
<td>CCO</td>
<td>Coverage Capacity Optimization</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative Distribution Function</td>
</tr>
<tr>
<td>CF</td>
<td>Compress-and Forward</td>
</tr>
<tr>
<td>CGHV</td>
<td>Carrier Grade HyperVisor</td>
</tr>
<tr>
<td>C-GW</td>
<td>Control Gateway</td>
</tr>
<tr>
<td>CHD</td>
<td>Cooperative Hybrid Diversity</td>
</tr>
<tr>
<td>C-ITS</td>
<td>Cooperative Intelligent Transport System</td>
</tr>
<tr>
<td>CM</td>
<td>Coded Modulation</td>
</tr>
<tr>
<td>CMT</td>
<td>Cosine-modulated Multi-tone</td>
</tr>
<tr>
<td>CN</td>
<td>Core Network</td>
</tr>
<tr>
<td>COC</td>
<td>Cell Outage Compensation</td>
</tr>
<tr>
<td>COD</td>
<td>Cell Outage Detection</td>
</tr>
<tr>
<td>CoMP</td>
<td>Coordinated Multi-point transmission / reception</td>
</tr>
<tr>
<td>COMPA</td>
<td>Control, Orchestration, Management, Policies, Analytics</td>
</tr>
<tr>
<td>CoMPT</td>
<td>Coordinated Multiple Beam Transmission</td>
</tr>
<tr>
<td>CoRe</td>
<td>Constellation Rearrangement</td>
</tr>
<tr>
<td>COTS</td>
<td>Commercial off-the-shelf</td>
</tr>
<tr>
<td>CP</td>
<td>Cyclic Prefix</td>
</tr>
<tr>
<td>CPRI</td>
<td>Common Public Radio Interface</td>
</tr>
<tr>
<td>CQI</td>
<td>Channel Quality Indicator</td>
</tr>
<tr>
<td>CR</td>
<td>Cognitive Radio</td>
</tr>
<tr>
<td>CRC</td>
<td>Cyclic Redundancy Check</td>
</tr>
<tr>
<td>C-RAN</td>
<td>Cloud RAN</td>
</tr>
<tr>
<td>C-RAN</td>
<td>Centralized RAN</td>
</tr>
<tr>
<td>CRE</td>
<td>Cell Range Extension</td>
</tr>
<tr>
<td>CRS</td>
<td>Cell Reference Signal</td>
</tr>
<tr>
<td>CS/CB</td>
<td>Coordinated Scheduling/Beamforming</td>
</tr>
<tr>
<td>CSD</td>
<td>Cooperative Source Diversity</td>
</tr>
<tr>
<td>CSMA/CA</td>
<td>Carrier Sensing Multiple Access/Collision Avoidance</td>
</tr>
</tbody>
</table>
CTD  Cooperative Transmit Diversity
CWDM  Coarse Wavelength Division Multiplexing
D2D  Device-to-Device
DAC  Digital-to-Analog Converters
DASH  Dynamic Adaptive Streaming over HTTP
DBS  Dynamic Beam Switching
DBT  Discontinuous Beam Transmission
DeNB  Donor eNB
DF  Decode-and-Forward
DFT  Discrete Fourier Transform
D-GW  Data Gateway
DL  Down Link
DM  Diagnostic Monitor
DMM  Distribute Mobility Management
DPDK  Data Plane Development KIT
DPS  Dynamic Point Selection
D-RAN  Distributed RAN
DSA  Dynamic Spectrum Access
DSL  Digital Subscriber Line
DU/RU  Digital Unit/Radio Unit
DVB-T  Digital Video Broadcasting - Terrestrial
DVB-S2  Digital Video Broadcasting - Satellite - Second Generation
DWDM  Dense Wavelength Division Multiplexing
EAB  Extended Access Barring
ECMA  European Computer Manufacturers Association
EHF  Extremely High Frequency
eICIC  enhanced InterCell Interference Coordination
EPON  Ethernet Passive Optical Network
EPC  Evolved Packet Core
ePDG  Evolved Packet Data Gateway
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESA</td>
<td>Elevation Spread at Arrival</td>
</tr>
<tr>
<td>ESD</td>
<td>Elevation Spread at Departure</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunications Standards Institute</td>
</tr>
<tr>
<td>FBMC</td>
<td>Filter Bank Multicarrier</td>
</tr>
<tr>
<td>FCC</td>
<td>Federal Communication Commission</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency Division Duplexing</td>
</tr>
<tr>
<td>FD-MIMO</td>
<td>Full Dimension MIMO</td>
</tr>
<tr>
<td>FDR</td>
<td>Full Duplex Radio</td>
</tr>
<tr>
<td>FER</td>
<td>Frame Error Rates</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>FMC</td>
<td>Fixed mobile Convergence</td>
</tr>
<tr>
<td>FMT</td>
<td>Filtered Multi-tone</td>
</tr>
<tr>
<td>F-OFDM</td>
<td>Filtered OFDM</td>
</tr>
<tr>
<td>FP7</td>
<td>The Seventh Framework Programme</td>
</tr>
<tr>
<td>FPGA</td>
<td>Field Programmable Gate Arrays</td>
</tr>
<tr>
<td>FQAM</td>
<td>Frequency and Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>FRN</td>
<td>Fixed Relay Node</td>
</tr>
<tr>
<td>FSK</td>
<td>Frequency Shift Keying</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full Width at Half Maximum</td>
</tr>
<tr>
<td>GDB</td>
<td>Geo-location DataBase</td>
</tr>
<tr>
<td>GPON</td>
<td>Gigabit Passive Optical Network</td>
</tr>
<tr>
<td>GPP</td>
<td>General Purpose Processor</td>
</tr>
<tr>
<td>GTP</td>
<td>GPRS Tunneling Protocol</td>
</tr>
<tr>
<td>GW</td>
<td>Gateway</td>
</tr>
<tr>
<td>H/W</td>
<td>Hardware</td>
</tr>
<tr>
<td>H2H</td>
<td>Human-to-Human</td>
</tr>
<tr>
<td>HARQ</td>
<td>Hybrid Automatic Repeat reQuest</td>
</tr>
<tr>
<td>HBM</td>
<td>Host based mobility</td>
</tr>
<tr>
<td>HC-RAN</td>
<td>Hotel C-RAN</td>
</tr>
<tr>
<td>HDTV</td>
<td>High Definition Television</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>HetNet</td>
<td>Heterogeneous Network</td>
</tr>
<tr>
<td>HIP</td>
<td>Host Identity Protocol</td>
</tr>
<tr>
<td>HO</td>
<td>Handover</td>
</tr>
<tr>
<td>h-RATs</td>
<td>heterogeneous Radio Access Networks</td>
</tr>
<tr>
<td>HS2.0</td>
<td>Hotspot 2.0</td>
</tr>
<tr>
<td>HSS</td>
<td>Home Subscriber Server</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>ICI</td>
<td>Inter-Cell Interference</td>
</tr>
<tr>
<td>ID</td>
<td>Identifier</td>
</tr>
<tr>
<td>IDFT</td>
<td>Inverse Discrete Fourier Transform</td>
</tr>
<tr>
<td>IETF</td>
<td>Internet Engineering Task Force</td>
</tr>
<tr>
<td>IF</td>
<td>Intermediate Frequency</td>
</tr>
<tr>
<td>IFFT</td>
<td>Inverse Fast Fourier Transform</td>
</tr>
<tr>
<td>IFOM</td>
<td>IP Flow Mobility</td>
</tr>
<tr>
<td>IFW</td>
<td>Integrated Femtocell and Wi-Fi</td>
</tr>
<tr>
<td>IMS</td>
<td>IP Multimedia Subsystem</td>
</tr>
<tr>
<td>IMT</td>
<td>International Mobile Telecommunications</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>IOTA</td>
<td>Isotropic Orthogonal Transform Algorithm</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>ISM</td>
<td>Industry, Science and Medical</td>
</tr>
<tr>
<td>ITS</td>
<td>Intelligent Transport Systems</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>JP</td>
<td>Joint Processing</td>
</tr>
<tr>
<td>KPI</td>
<td>Key Performance Indicators</td>
</tr>
<tr>
<td>LAA</td>
<td>Licensed Assisted Access</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>LC-MTC</td>
<td>Low Cost Machine Type Communication</td>
</tr>
<tr>
<td>LDPC</td>
<td>Low Density Parity Check</td>
</tr>
<tr>
<td>LISP-MN</td>
<td>Locator Identifier Separator – Mobile Node</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>LIPA</td>
<td>Local IP Access</td>
</tr>
<tr>
<td>LOS</td>
<td>Line-of-Sight</td>
</tr>
<tr>
<td>LPWA</td>
<td>Low-Power Wide-Area</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>LTN</td>
<td>Low Throughput Network</td>
</tr>
<tr>
<td>M2M</td>
<td>Machine-to-Machine</td>
</tr>
<tr>
<td>MAA</td>
<td>Modular Antenna Arrays</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MAICS</td>
<td>Network Assisted Interference Cancellation and Suppression</td>
</tr>
<tr>
<td>MANET</td>
<td>Mobile Ad-hoc Network</td>
</tr>
<tr>
<td>MANO</td>
<td>Management and Orchestration</td>
</tr>
<tr>
<td>MAPCOM</td>
<td>Multiple Access PDN Connectivity</td>
</tr>
<tr>
<td>MBB</td>
<td>Mobile BroadBand</td>
</tr>
<tr>
<td>MCL</td>
<td>Maximum Coupling Loss</td>
</tr>
<tr>
<td>MCN</td>
<td>Macro Cell Node</td>
</tr>
<tr>
<td>MCR</td>
<td>Mutually Cooperative Relaying</td>
</tr>
<tr>
<td>MC-RAN</td>
<td>Macro C-RAN</td>
</tr>
<tr>
<td>MCS</td>
<td>Modulation and Coding Scheme</td>
</tr>
<tr>
<td>mDU</td>
<td>MHN Digital Unit</td>
</tr>
<tr>
<td>MEC</td>
<td>Mobile Edge Computing</td>
</tr>
<tr>
<td>MeNB</td>
<td>Master eNB</td>
</tr>
<tr>
<td>METIS</td>
<td>Mobile and wireless communications Enablers for Twenty-twenty (2020) Information Society</td>
</tr>
<tr>
<td>MF</td>
<td>MobileFlow</td>
</tr>
<tr>
<td>MGW</td>
<td>Multimedia Gateway</td>
</tr>
<tr>
<td>MHN</td>
<td>Mobile Hotspot Network</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple Input Multiple Output</td>
</tr>
<tr>
<td>MIP</td>
<td>Mobile Internet Protocol</td>
</tr>
<tr>
<td>MISO</td>
<td>Multiple Input Single Output</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MLB</td>
<td>Mobility Load Balance</td>
</tr>
<tr>
<td>MME</td>
<td>Mobility Management Entity</td>
</tr>
<tr>
<td>MMSE</td>
<td>Minimum Mean Squared Error</td>
</tr>
<tr>
<td>MPLS-TP</td>
<td>Multi-Protocol Label Switching – Transport Profile</td>
</tr>
<tr>
<td>MRN</td>
<td>Mobile Relay Node</td>
</tr>
<tr>
<td>MRO</td>
<td>Mobility Robustness Optimization</td>
</tr>
<tr>
<td>mRU</td>
<td>MHN Radio Unit</td>
</tr>
<tr>
<td>MS</td>
<td>Mobile Station</td>
</tr>
<tr>
<td>MSS</td>
<td>Mobile Soft Switch</td>
</tr>
<tr>
<td>mTE</td>
<td>MHN Terminal Equipment</td>
</tr>
<tr>
<td>MTC</td>
<td>Machine Type Communication</td>
</tr>
<tr>
<td>MU-MIMO</td>
<td>Multi-User MIMO</td>
</tr>
<tr>
<td>MUST</td>
<td>Multiuser Superposition Transmission</td>
</tr>
<tr>
<td>NaF</td>
<td>Network coding after Fragmentation</td>
</tr>
<tr>
<td>NaP</td>
<td>Network coding after Padding</td>
</tr>
<tr>
<td>NBM</td>
<td>Network Based Mobility</td>
</tr>
<tr>
<td>NCT</td>
<td>New Carrier Type</td>
</tr>
<tr>
<td>NFV</td>
<td>Network Function Virtualization</td>
</tr>
<tr>
<td>NG-POP</td>
<td>Next Generation Point of Presence</td>
</tr>
<tr>
<td>NLOS</td>
<td>Non-Line-of-Sight</td>
</tr>
<tr>
<td>NOMA</td>
<td>Non-Orthogonal Multiple Access</td>
</tr>
<tr>
<td>OAM</td>
<td>Operations, Administration and Maintenance</td>
</tr>
<tr>
<td>OBSAI</td>
<td>Open Base Station Architecture Initiative</td>
</tr>
<tr>
<td>OF</td>
<td>OpenFlow</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency Division Multiplexing</td>
</tr>
<tr>
<td>OFDMA</td>
<td>Orthogonal Frequency Division Multiple Access</td>
</tr>
<tr>
<td>OLT</td>
<td>Optical Line Terminal</td>
</tr>
<tr>
<td>OMA</td>
<td>Orthogonal Multiple Access</td>
</tr>
<tr>
<td>ONE</td>
<td>Open Network Environment</td>
</tr>
<tr>
<td>ONU</td>
<td>Optical Network Unit</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>OP</td>
<td>Outage Probability</td>
</tr>
<tr>
<td>OPEX</td>
<td>OPEration Expenditure</td>
</tr>
<tr>
<td>ORI</td>
<td>Open Radio Interface</td>
</tr>
<tr>
<td>OTN</td>
<td>Optical Transport Network</td>
</tr>
<tr>
<td>OTT</td>
<td>Over-The-Top</td>
</tr>
<tr>
<td>OVS</td>
<td>Open vSwitch</td>
</tr>
<tr>
<td>PAM</td>
<td>Pulse Amplitude Modulation</td>
</tr>
<tr>
<td>PAN</td>
<td>Personal Area Network</td>
</tr>
<tr>
<td>PAPR</td>
<td>Peak to Average Power Ratio</td>
</tr>
<tr>
<td>PCFICH</td>
<td>Physical Control Format Indicator Channel</td>
</tr>
<tr>
<td>PCI</td>
<td>Physical Cell Identifier</td>
</tr>
<tr>
<td>pCNF</td>
<td>Physical Core Network Function</td>
</tr>
<tr>
<td>PDCCH</td>
<td>Physical Downlink Control Channel</td>
</tr>
<tr>
<td>PDCP</td>
<td>Packet Data Convergence Protocol</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PDN</td>
<td>Packet Data Network</td>
</tr>
<tr>
<td>PDU</td>
<td>Protocol Data Unit</td>
</tr>
<tr>
<td>P-GW</td>
<td>Packet Data Network Gateway</td>
</tr>
<tr>
<td>PHICH</td>
<td>Physical H-ARQ Indicator Channel</td>
</tr>
<tr>
<td>PMIP</td>
<td>Proxy Mobile IP</td>
</tr>
<tr>
<td>PMSE</td>
<td>Program Making and Special Events</td>
</tr>
<tr>
<td>PNC</td>
<td>Physical-layer Network Coding</td>
</tr>
<tr>
<td>PNF</td>
<td>Physical Network Function</td>
</tr>
<tr>
<td>PON</td>
<td>Passive Optical Network</td>
</tr>
<tr>
<td>POTN</td>
<td>Packet Optical Transport Network</td>
</tr>
<tr>
<td>PPN</td>
<td>Poly-Phase Network</td>
</tr>
<tr>
<td>pRNF</td>
<td>Physical Radio Network Function</td>
</tr>
<tr>
<td>ProSe</td>
<td>Proximity-based Service</td>
</tr>
<tr>
<td>PU</td>
<td>Primary Users</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>QoE</td>
<td>Quality of Experience</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RAN</td>
<td>Radio Access Network</td>
</tr>
<tr>
<td>RAT</td>
<td>Radio Access Technology</td>
</tr>
<tr>
<td>RLC</td>
<td>Radio Link Control</td>
</tr>
<tr>
<td>RLOC</td>
<td>Routing Locator</td>
</tr>
<tr>
<td>RN</td>
<td>Relay Node</td>
</tr>
<tr>
<td>ROADM</td>
<td>Reconfigurable Optical Add-Drop Multiplexer</td>
</tr>
<tr>
<td>RoE</td>
<td>Radio over Ethernet</td>
</tr>
<tr>
<td>RoF</td>
<td>Radio over Fiber</td>
</tr>
<tr>
<td>ROI</td>
<td>Return on Investment</td>
</tr>
<tr>
<td>RPS</td>
<td>Reduced Power Subframe</td>
</tr>
<tr>
<td>RRC</td>
<td>Radio Resource Control</td>
</tr>
<tr>
<td>RRH</td>
<td>Remote Radio Head</td>
</tr>
<tr>
<td>RRM</td>
<td>Radio Resource Management</td>
</tr>
<tr>
<td>RS</td>
<td>Relay Station</td>
</tr>
<tr>
<td>S/W</td>
<td>Software</td>
</tr>
<tr>
<td>SA</td>
<td>System Architecture</td>
</tr>
<tr>
<td>SAS</td>
<td>Spectrum Access System</td>
</tr>
<tr>
<td>SBT</td>
<td>Single Beam Transmission</td>
</tr>
<tr>
<td>SC-FDMA</td>
<td>Single Carrier Frequency Division Multiple Access</td>
</tr>
<tr>
<td>SCI</td>
<td>Spatial Coordinate Interleaving</td>
</tr>
<tr>
<td>SCM</td>
<td>Spatial Channel Model</td>
</tr>
<tr>
<td>SCMA</td>
<td>Sparse Code Multiple Access</td>
</tr>
<tr>
<td>SCN</td>
<td>Small Cell Node</td>
</tr>
<tr>
<td>SC-PTM</td>
<td>Single Cell Point-to-Multipoint</td>
</tr>
<tr>
<td>SCR</td>
<td>Simple Cooperative Relaying</td>
</tr>
<tr>
<td>SCTP</td>
<td>Stream Control Transmission Protocol</td>
</tr>
<tr>
<td>SDMN</td>
<td>Software Defined Mobile Network</td>
</tr>
<tr>
<td>SDN</td>
<td>Software Defined Networking</td>
</tr>
</tbody>
</table>
SDO Standard Developing Organization
SDU Service Data Unit
SeNB Secondary eNB
S-GW Serving Gateway
SHF Super High Frequency
SIC Serial Interference Cancellation
SINR Signal to Interference and Noise Ratio
SIPTO Selected IP Traffic Offloading
SLA Service Level Agreement
SLS System Level Simulation
SMT Staggered Modulation Multi-tone
SNR Signal-to-Noise Ratio
SOLDER Spectrum Overlay through aggregation of heterogeneous DispERsed bands
SON Self Organizing Network
SONET/SDH Synchronous Optical Network/Synchronous Digital Hierarchy
STC Space Time Coding
STR Simultaneous Transmission and Reception
SU Secondary Users
SWSC Sliding-Window Superposition Coding
TCA Telecommunications Computing Architecture
TCM Trellis Coded Modulation
TCO Total Cost of Ownership
TCP Transmission Control Protocol
TDD Time Division Duplexing
TDM Time Division Multiplexing
TM Transmission Mode
TP Transmission Point
T-SDN Transport SDN
TTI Transmission Time Interval
TVHT TV High Throughput
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVWS</td>
<td>TV White Space</td>
</tr>
<tr>
<td>UDN</td>
<td>Ultra Dense Network</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>UFA</td>
<td>Ultra Flat Architecture</td>
</tr>
<tr>
<td>UHD</td>
<td>Ultra High Definition</td>
</tr>
<tr>
<td>UL</td>
<td>Up Link</td>
</tr>
<tr>
<td>UMi</td>
<td>Urban Micro-cell model</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunication System</td>
</tr>
<tr>
<td>UPA</td>
<td>Uniform Planar Array</td>
</tr>
<tr>
<td>URI</td>
<td>Unified Radio Interface</td>
</tr>
<tr>
<td>V2I</td>
<td>Vehicle to Infrastructure</td>
</tr>
<tr>
<td>V2P</td>
<td>Vehicle to Pedestrian</td>
</tr>
<tr>
<td>V2V</td>
<td>Vehicle to Vehicle</td>
</tr>
<tr>
<td>V2X</td>
<td>Vehicle to Everything</td>
</tr>
<tr>
<td>vCNF</td>
<td>Virtual Core Network Function</td>
</tr>
<tr>
<td>vEPC</td>
<td>Virtualized EPC</td>
</tr>
<tr>
<td>VGA</td>
<td>Variable Gain Amplifier</td>
</tr>
<tr>
<td>VM</td>
<td>Virtual Machine</td>
</tr>
<tr>
<td>VNF</td>
<td>Virtual Network Function</td>
</tr>
<tr>
<td>VPL</td>
<td>Vehicular Penetration Loss</td>
</tr>
<tr>
<td>VR</td>
<td>Virtual Reality</td>
</tr>
<tr>
<td>vRAN</td>
<td>Virtualized Radio Access Network</td>
</tr>
<tr>
<td>vRNF</td>
<td>Virtual Radio Network Function</td>
</tr>
<tr>
<td>WARP</td>
<td>Wireless Open-Access Research Platform</td>
</tr>
<tr>
<td>WCDMA</td>
<td>Wideband Code Division Multiple Access</td>
</tr>
<tr>
<td>WDM</td>
<td>Wavelength Division Multiplexing</td>
</tr>
<tr>
<td>WFA</td>
<td>Wi-Fi Allicance</td>
</tr>
<tr>
<td>WiMAX</td>
<td>Worldwide Interoperability for Microwave Access</td>
</tr>
<tr>
<td>WINNER</td>
<td>Wireless World Initiative New Radio</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
<tr>
<td>Acronyms</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>WPAN</td>
<td>Wireless Personal Area Network</td>
</tr>
<tr>
<td>WRAN</td>
<td>Wireless Regional Area Network</td>
</tr>
<tr>
<td>WRC</td>
<td>World Radiocommunication Conference</td>
</tr>
</tbody>
</table>
## Appendix 2 Contributors

### Wireless Technology Committee

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kang, Chung Ku</td>
<td>Korea Univ.</td>
<td>Chair</td>
</tr>
<tr>
<td>Kang, Jiwon</td>
<td>LG Electronics</td>
<td>Vice Chair</td>
</tr>
<tr>
<td>Kwun, Jonghyung</td>
<td>Samsung</td>
<td>Vice Chair</td>
</tr>
<tr>
<td>Lee, Jun Hwan</td>
<td>ETRI</td>
<td>Vice Chair</td>
</tr>
<tr>
<td>Park, Byungseong</td>
<td>Ericsson-LG</td>
<td>Vice Chair / Editor</td>
</tr>
<tr>
<td>Kwon, KiBurn</td>
<td>ITL</td>
<td>Secretary</td>
</tr>
<tr>
<td>Chen, Yan</td>
<td>Huawei</td>
<td>Member</td>
</tr>
<tr>
<td>Choi, Eunyoung</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Choi, Sooyong</td>
<td>Yonsei Univ.</td>
<td>Member</td>
</tr>
<tr>
<td>Choi, Youngjune</td>
<td>Ajou Univ.</td>
<td>Member</td>
</tr>
<tr>
<td>Chung, Heesang</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Hong, Seung-Eun</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Hur, Sooyoung</td>
<td>Samsung</td>
<td>Member</td>
</tr>
<tr>
<td>Kim, Hasung</td>
<td>KT</td>
<td>Member</td>
</tr>
<tr>
<td>Kim, Jaeyoul</td>
<td>Samsung</td>
<td>Member</td>
</tr>
<tr>
<td>Kim, Joongheon</td>
<td>Intel</td>
<td>Member</td>
</tr>
<tr>
<td>Kim, Kwangtaik</td>
<td>Samsung</td>
<td>Member</td>
</tr>
<tr>
<td>Kim, Young-Han</td>
<td>UC San Diego</td>
<td>Invited</td>
</tr>
<tr>
<td>Kwon, Sang Wook</td>
<td>Samsung</td>
<td>Member</td>
</tr>
<tr>
<td>Kwon, Tae Soo</td>
<td>SEOULTECH Univ.</td>
<td>Member</td>
</tr>
<tr>
<td>Park, Juho</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Park, Michael</td>
<td>Qualcomm</td>
<td>Member</td>
</tr>
<tr>
<td>Park, Mooneo</td>
<td>InterDigital Asia</td>
<td>Member</td>
</tr>
<tr>
<td>Park, Oksun</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Yeh, Choong-il</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Yuk, Young-soo</td>
<td>Nokia</td>
<td>Member</td>
</tr>
</tbody>
</table>
## Network Technology Committee

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jo, Changkil</td>
<td>LG UPlus</td>
<td>Chair</td>
</tr>
<tr>
<td>Song, Pyung Jung</td>
<td>ETRI</td>
<td>Vice Chair</td>
</tr>
<tr>
<td>Park, Jungshin</td>
<td>Samsung</td>
<td>Vice Chair</td>
</tr>
<tr>
<td>Kim, Jungsoep</td>
<td>LG UPlus</td>
<td>Vice Chair</td>
</tr>
<tr>
<td>Koo, Yeonsang</td>
<td>LG UPlus</td>
<td>Secretary</td>
</tr>
<tr>
<td>Chen, Yan</td>
<td>Huawei</td>
<td>Member</td>
</tr>
<tr>
<td>Cho, HeeJeong</td>
<td>LG Electronics</td>
<td>Member</td>
</tr>
<tr>
<td>Chung, Hwan Seok</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Joo, Bheom-Soon</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Kim, Changki</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Kim, Sang Ki</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Park, Dan</td>
<td>Nokia</td>
<td>Member</td>
</tr>
<tr>
<td>Rhy, Hee Yeal</td>
<td>Ericsson-LG</td>
<td>Member</td>
</tr>
<tr>
<td>Ryu, Sunheui</td>
<td>Samsung</td>
<td>Member</td>
</tr>
<tr>
<td>Song, Jongtae</td>
<td>ETRI</td>
<td>Member</td>
</tr>
<tr>
<td>Yuk, Youngsoo</td>
<td>Nokia</td>
<td>Member</td>
</tr>
<tr>
<td>Zhenbin</td>
<td>Huawei</td>
<td>Member</td>
</tr>
</tbody>
</table>
## Revision History

<table>
<thead>
<tr>
<th>Date</th>
<th>Version</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 2015</td>
<td>1.0</td>
<td>First release of 5G Forum Technology White Paper</td>
</tr>
<tr>
<td>March 2016</td>
<td>2.0</td>
<td>Second release of 5G Forum Technology White Paper</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Responsible Committees:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wireless Technology and Network Technology Committees</td>
</tr>
</tbody>
</table>
5G Vision, Requirements, and Enabling Technologies

2016

5G White Paper