Call: H2020-ICT-2014-2
Project reference: 671660

Project Name:
Flexible Air iNTerfAce for Scalable service delivery wiThin wIreless Communication networks of the 5th Generation (FANTASTIC-5G)

Deliverable D3.2
Final report on the holistic link solution adaptation

Date of delivery: 30/04/2017
Start date of project: 01/07/2015
Version: final
Duration: 24 months
Abstract

The FANTASTIC-5G project advocates a single air interface for 5G, which is flexible, versatile, scalable and efficient in order to address the requirements of the IMT 2020. The flexibility featured by the FANTASTIC-5G air interface allows the support of a multitude of services and applications with diverse requirements, and shall be ready to provide solutions for future use cases not yet foreseen. This document reports on the recent outcomes of the research on link design, which are presented in a concise form, including a summary of the research motivation, the final outcomes and topic conclusions. The involved topics include the signal design as well as the framework of frame and physical layer procedure design, which provides a quite complete concept of a holistic physical layer air-interface design for future 5G system.

____________________________________

1CO = Confidential, only members of the consortium (including the Commission Services)

PU = Public
**Keywords**

| 5G, air interface design, physical layer procedure, massive machine communication, mobile broadband, mission critical communication, multicast broadcast service, multiple service support, vehicular to anything, waveforms, channel coding, frame design, control channel, reference signal design, HARQ. |
Executive Summary

This document reports on the final outcome of the research on radio link design in the FANTASTIC-5G project. The research on radio link design focuses on solutions on PHY and MAC layer towards a holistic air interface design, which is intended to enable the system to flexibly adapt its current configurations in response to the diverse requirements of the various service types to be supported by the 5G system. In contrast to 3GPP, where these service types have been defined as eMBB, mMTC and uMTC, the project defined five so-called “core services”, namely mobile broadband (MBB), massive machine communication (MBB), mission critical communication (MCC), vehicular to anything (V2X) and broadcast/multicast service (BMS) -- however, those can be clearly mapped to the former service types defined by 3GPP.

The research has been divided into two main areas: 1) signal design and 2) frame design and link level procedures; each of those areas is devoted its own chapter. In the context of signal design, fundamental brick stones constituting air interface signaling have been investigated in detail, such as waveforms, channel coding, modulation schemes, MIMO and schemes for reducing the peak to average power ratio (PAPR). Waveform research has attracted a lot of attention at the dawn of 5G. A key contribution of this project is thus the provision of an overview on all potential candidates with its advantages and drawbacks (this was presented already in Deliverable D3.1), as well as their common evaluation and comparison based on an evaluation framework reflecting key aspects of future use cases and applications. Corresponding recommendations for the 5G waveform design are given as a conclusion of these evaluations. Channel coding is advanced towards low-latency communications, embracing Turbo, LDPC and Polar codes, and different solutions have been compared in corresponding evaluations. Novel modulation schemes are proposed for improving energy efficiency for cell-edge users and spectral efficiency for high SNR users, respectively. MIMO research covers the compatibility with new waveforms and proposes an adaptive antenna scheme for attaining high spectral efficiency for cars moving at high speed. Finally, novel schemes for PAPR reduction are introduced, and recommendations for their application in a service specific context are given.

The chapter on frame design and link level procedures firstly provides a detailed description of the design principles and recommendations steering this research. It then presents novel solutions meeting those principles and thus facilitating a holistic service-adaptive design. The investigated topics cover the overall frame design, control channel design, design of reference signals, enhanced HARQ schemes and refined PHY procedures, such as random access schemes. All the solutions set a strong focus on the multi-service context of 5G systems and thus allow for manifold service-specific adaptations. A fundamental component for a flexible air interface design has been introduced here as the “tiling concept”, where the radio resources are partitioned into different tiles defined in the time/frequency space, which can then be individually configured according to a particular service’s needs.

This deliverable provides a thoughtful analysis on the main technical components that can be composed towards a holistic solution for the 5G air-interface, enabling readers to get a clear view on the possible solutions, their trade-offs and possibilities for adaptations.
Table of Contents

1 Introduction.................................................................................................................. 17
  1.1 Objective of the document..................................................................................... 17
  1.2 Structure of the document..................................................................................... 18

2 Holistic solution for signal design.............................................................................. 19
  2.1 New waveform: Comparison and recommendation ............................................. 19
     2.1.1 Motivation ...................................................................................................... 19
     2.1.2 Overview of FANTASTIC-5G new waveform candidates.............................. 19
     2.1.3 Waveform Comparisons................................................................................ 21
     2.1.4 Conclusions and recommendation.................................................................. 22
  2.2 Channel coding ...................................................................................................... 24
     2.2.1 Research scope constraints and requirements............................................... 24
     2.2.2 Low latency channel codes............................................................................ 25
     2.2.3 Enhanced Turbo codes (IMTA)........................................................................ 26
     2.2.4 eMBB scenario: standardization discussions and conclusions....................... 28
     2.2.5 Next step: URLLC and mMTC scenarios.................................................... 30
  2.3 Enhanced modulation ............................................................................................ 31
     2.3.1 FQAM for interference limited scenarios with inactive subcarriers utilization... 31
     2.3.2 NUC for high modulation orders.................................................................... 33
     2.3.3 Conclusion and recommendation.................................................................... 33
  2.4 MIMO .................................................................................................................... 34
     2.4.1 Compatibility of MIMO with new waveforms............................................... 34
     2.4.2 Making adaptive MIMO compatible with fast moving connected vehicles thanks to the predictor antenna ........................................................................ 34
     2.4.3 Connecting objects in high data rate with a single carrier modulation: it might work with Massive MIMO ................................................................. 36
     2.4.4 Conclusion and recommendation.................................................................... 37
  2.5 PAPR reduction techniques ................................................................................... 38
     2.5.1 Research scope and candidates’ introduction............................................... 38
     2.5.2 Comparison scenarios and final results.......................................................... 39
     2.5.3 Conclusion and recommendation.................................................................... 39

3 Holistic solution on frame design and link level procedures.................................... 41
  3.1 Motivation............................................................................................................... 41
  3.2 Design principles and recommendations............................................................... 41
  3.3 Topics and detailed proposals................................................................................ 46
     3.3.1 Numerology and frame structure..................................................................... 46
     3.3.2 Physical Downlink control channel................................................................. 52
     3.3.3 Reference signal design.................................................................................. 61
     3.3.4 HARQ.............................................................................................................. 65
     3.3.5 PHY procedures............................................................................................. 70

4 Conclusions.................................................................................................................. 74

5 References.................................................................................................................... 78

6 Appendix....................................................................................................................... 87
  6.1 Link performance comparison of new waveform candidates.................................. 87
     6.1.1 Waveform specifications.................................................................................. 87
     6.1.2 Simulator Calibration...................................................................................... 95
     6.1.3 Waveform comparisons.................................................................................. 99
  6.2 BF-OFDM............................................................................................................... 105
6.3 Low Latency Channel Coding ................................................................. 107
6.4 Enhanced Turbo Codes for 5G ................................................................. 109
6.5 FDPC PAPR reduction technique ............................................................. 118
6.6 Control channel design ........................................................................ 125
6.7 HARQ enhancements ............................................................................ 132
6.8 PHY procedures: details and results ..................................................... 135
  6.8.1 Blind detection of users’ activity in random/grant-free access .......... 135
  6.8.2 New Radio Physical Random Access Channel Sequence Design .... 136
  6.8.3 Massive MIMO assisted random access for 5G IoT ......................... 142
  6.8.4 Embedded air interface (EAI) for multiplexing MBB and MCC traffic: Principle and analysis ................................................................. 149
  6.8.5 Harmonization of proposed PHY procedures ................................... 157
List of Figures

Figure 2-1: Performance comparison of the improved turbo code with the LTE turbo code in AWGN channel in terms of Block Error Rate vs Eb/N0. QPSK modulation, block size $K = 6000$ bits ($K = 6016$ bits for LTE TC), 8 decoding iterations of the Max-Log-MAP algorithm, tail-biting termination. ................................................................. 27

Figure 2-2. Block Error Rate performance of the improved turbo code in AWGN channel for coding rates 1/5 ($K = 100$) and 8/9 ($K = 96$). BPSK modulation, 8 decoding iterations. Comparison with LTE turbo code ($K = 96$). ........................................................................................................... 28

Figure 2-3: Utilizing the inactive in FQAM symbol to transmit to another user .................... 32

Figure 2-4: Throughput comparison of the proposed scheme with conventional approaches ... 33

Figure 2-5: The “wall of speed” measured by simulations for two different scenarios ............ 36

Figure 2-6: Single carrier modulation and multi-carrier modulation partially overlapping in frequency and separated through massive MIMO and Tx/Rx beamforming (the “object” is an MTC device or a high data rate connected object) ................................................................. 37

Figure 3-1: Basic illustration of the resource block grouping (tiling concept) for the uplink...... 47

Figure 3-2: Numerology examples and multiplexing option of CSI-RS (downlink) or sounding RS (uplink) for multiple cells................................................................. 50

Figure 3-3 Mini-slot structure, ......................................................................................... 51

Figure 3-4 Different approaches of tuning the frame structure’s TTI ...................................... 51

Figure 3-5: (Left) Capacity drop as Doppler shift increases for different SNRs. (Right) Capacity increase as SNR increases for different TTI durations ........................................ 52

Figure 3-6: REG with different DMRS patterns, DMRS overhead in the REG: (a) 1/3, (b) 1/3, (c) 1/2, ..................................................................................................................... 53

Figure 3-7: Localized search space of NR-PDCCH with 4 ALs: 8 AL1 candidates, 4 AL2 candidates, 2 AL4 candidates and 1 AL8 candidate. Control resource set: 16 RBs, 2 OFDM symbols. ..................................................................................................... 55

Figure 3-8: Distributed search space of NR-PDCCH with 4 ALs: 8 AL1 candidates, 4 AL2 candidates, 2 AL4 candidates and 1 AL8 candidate. Control resource set: 16 RBs, 2 OFDM symbols. Each smallest block represents a REG .................................................................................. 55

Figure 3-9: per-REG transmit antenna pair cycling, an AL1 NR-PDCCH of 4 REGs. 4 DMRS ports (coloured REs) associated with 4 transmit antenna ports: 0, 1, 2 and 3................................. 56

Figure 3-10: per-RB transmit antenna pair cycling, an AL1 NR-PDCCH of 4 REGs. 4 DMRS ports (colored REs) associated with 4 transmit antenna ports: 0, 1, 2 and 3................................. 57

Figure 3-11: Multi stage PDCCH structure.......................................................................... 58

Figure 3-12: Different allocation strategies of the MS-PDCCH, frequency diverse (left) and frequency selective (right) ......................................................................................... 59

Figure 3-13: Different aggregation levels for localized multi stage PDCCH, UIDS AL=1 (top) and UIDS AL=2 (bottom) .............................................................................................. 60

Figure 3-14: Illustration of the concept of sub tiling; left: conventional, e.g., CRS or potential NR CSI RS with dense grid of reference signals, where each resource element is completely occupied by one single AP1; right: sub-tiled CSI RS with same grid density, but shared usage of resources ........................................ 63
Figure 3-15: Crosstalk for UE moving with 3 km/h and a grid density of 4 CSI RS per TTI of 1ms and a sequence length of 16. Left: virtual beamformer over 32 cm = 3 λ; right: single UE antenna ................................................................. 64

Figure 3-16. Average relative retransmission size using a multi-bit feedback with respect to the single-bit ACK/NACK feedback ........................................................................................................ 68

Figure 3-17: Proposed channel structure: grant-free access area divided into Activity Area and Data Area (example with Activity Area of two OFDM symbols). .................................................. 70

Figure 3-18: Frame structure supporting multiplexing of eMBB and URLLC traffic .............. 72

Figure 3-19 UL and DL resource maps for UEs of different scenarios ................................ 74

Figure 6-1 PSD of UF-OFDM for DL filter applied per PRB (red) and DC filter applied per subband allocation (blue) for 4PRBs ........................................................................................................ 87

Figure 6-2 Time domain symbol of UF-OFDM for DC filter applied per PRB (red) and DL filter applied per subband allocation (blue) for 4PRBs ................................................................. 88

Figure 6-3 Transmit and receive window roll-off. For our WOLA simulations: L_{wt} = L_{wr} = 72.89

Figure 6-4 Transmit windowing for one OFDM symbol. In our simulation settings with N=1024, LCP = 72 samples the overlapped extension is 36 samples on each side. ......................... 90

Figure 6-5 Latency discussion FC-OFDM vs. CP-OFDM ....................................................... 91

Figure 6-6 Transmit and receive filter for P-OFDM ............................................................. 92

Figure 6-7: Block diagram of simulation chain for waveform calibration and comparison. ..... 95

Figure 6-8: BLER performance of CP-OFDM for case 1 ......................................................... 97

Figure 6-9: BLER performance of CP-OFDM for case 2 ......................................................... 97

Figure 6-10: BLER performance of SC-FDMA for case 2 ...................................................... 98

Figure 6-11: BLER performance of CP-OFDM for case 3 ...................................................... 98

Figure 6-12: Illustration of scenario 1 and scenario 3: UL asynchronous transmission ........ 99

Figure 6-13: Illustration of scenario 2: Downlink high speed scenario ......................... 99

Figure 6-14 BLER performance of Scenario 1-1 ................................................................. 101

Figure 6-15 BLER performance of Scenario 1-2 ................................................................. 101

Figure 6-16 BLER performance of Scenario 2 ................................................................. 102

Figure 6-17 BLER performance of Scenario 3-1 ................................................................. 103

Figure 6-18 BLER performance of Scenario 3-2 ................................................................. 103

Figure 6-19 BLER performance of Scenario 3-3 ................................................................. 104

Figure 6-20 BLER performance of Scenario 3-4 ................................................................. 104

Figure 6-21: BF-OFDM transmitter diagram ................................................................. 105

Figure 6-22 BF-OFDM receiver diagram ................................................................. 106

Figure 6-23 - Required E_b/N_0 to achieve a target BLER with channel codes of rate R=1/2 on AWGN channel with BPSK modulation for different message lengths k≤512, together with Shannon’s BPSK constrained capacity, and the Gaussian approximation of Polynaskiy’s finite length bound ................................................................. 108

Figure 6-24 - Decoder Complexity in terms of number of basic operations for different message lengths. ........................................................................................................ 108
Figure 6-25: Possible representation of the ARP interleaver structure, with \( S(0), \ldots, S(Q-1) = 0 \) and \( K = 4Q \), when using CRSC constituent codes. ................................................................. 110

Figure 6-26: (a) Different inter-layer shifts for layer \( l = 0 \), with \( K = 4Q \). (b) Possible intra-layer shifts for layer \( l = 0 \), with \( T_0 = 0 \) and \( K = 4Q \) ................................................................. 110

Figure 6-27: Comparison of EXIT charts computed from the complete data frame and from punctured and unpunctured parity positions at the Signal-to-Noise (SNR) decoding threshold of the TC over the AWGN channel. ........................................................................................................... 112

Figure 6-28: Extrinsic information exchange between constituent codes of the TC at \( E_b/N_0 = 1.6 \) dB with 16 TC iterations, \( K = 1504 \), and \( R = 2/3 \) under uniform or DPC uniform interleaving over the AWGN channel. ...................................................................................................................... 113

Figure 6-29: Frame error rate performance comparison between the different ARP interleaver configurations and the LTE in AWGN channel with a max of 16 decoding iterations of the MAP algorithm for \( R = 2/3, \ R = 4/5, \ K = 1504, \) and CRSC(1,15/13)\(_s\) constituent codes. TUB = Truncated Union Bound. ................................................................. 115

Figure 6-30: Performance comparison of the enhanced turbo code with the Multi-Edge LDPC code in AWGN channel for coding rates ranging from 1/5 to 8/9 in terms of BLock Error Rate vs Es/N0. QPSK modulation, block size \( K \) around 100 bits. .............................................................................................. 116

Figure 6-31: Performance comparison for the first transmission of the enhanced turbo code with the polar code in AWGN channel for coding rates ranging from 1/5 to 8/9 in terms of required Es/N0 to achieve 1% and 0.1% of BLER. QPSK modulation, block size \( K \) ranging from 32 to 1024 bits. .......................................................................................................... 117

Figure 6-32: Performance comparison for up to four retransmissions of the enhanced turbo code with the polar code in AWGN channel for coding rates ranging from 1/12 to 2/3 in terms of required Es/N0 to achieve 1% and 0.1% of BLER. QPSK modulation, block size \( K \) ranging from 32 to 1024 bits. .......................................................................................................... 117

Figure 6-33: Principle of the FDPC algorithm and its implementation in the transmission chain. ........................................................................................................................................................................ 119

Figure 6-34: Basic PCTS structure used by FDPC algorithm ................................................................. 119

Figure 6-35: FDPC algorithm structure. .................................................................................................... 120

Figure 6-36: FDPC and PCTS compared complexity relative to FFT ................................................... 120

Figure 6-37: Constellation modification of type A, B and C ................................................................. 121

Figure 6-38: FDPC CM/A results compared with two advanced clipping algorithms .......... 121

Figure 6-39: FDPC CM/A + CM/B results compared with an advanced clipping process..... 122

Figure 6-40: FDPC CM/C results for a 256 QAM extended constellation ........................................... 122

Figure 6-41: Modulus of the RF signal, before and after FDPC processing ......................................... 123

Figure 6-42: Proposed modified algorithm .............................................................................................. 123

Figure 6-43: Compared simulation results for OFDM and OQAM modulation ................................. 124

Figure 6-44: REG Localized vs. distributed NR-PDCCH transmission, control resource set: 16 RBs and 2 control symbols (i.e., 2 REGs) per RB, 2x2 SFBC ................................................................. 126

Figure 6-45: SFBC vs. per-RE precoder cycling, control resource set: 16 RBs and 2 control symbols (i.e., 2 REGs) per RB, 2 Tx, 2 Rx ......................................................................................................... 127

Figure 6-46: 2 Tx SFBC vs. SFBC with per-RB transmit antenna pair cycling for 4 Tx, control resource set: 16 RBs and 2 control symbols (i.e., 2 REGs) ................................................................. 127
Figure 6-47: UIDS MD and FA rates dependency on ALs and SINR, SINR calculated at the antenna (blue), SINR after Rx processing (green) .......................................................... 128
Figure 6-48: FTP performance results for different feedback/resource unit granularity (in PRBs) and different scheduling modes for EPDCCH .......................................................... 130
Figure 6-49: FTP UE throughput CDF for different feedback unit and resource unit granularities (in PRBs) and different scheduling modes for EPDCCH .......................................................... 130
Figure 6-50: EPDCCH blocking probability for FTP with different file sizes and different EPDCCH scheduling modes .................................................................................. 131
Figure 6-51: EPDCCH search space configurations and performance ................................ 131
Figure 6-52: EPDCCH error CDF at high system load, coarse resource granularity (left), fine resource granularity (right) .............................................................................. 132
Figure 6-53: Throughput vs. SNR using two-bit feedback messages ............................... 133
Figure 6-54: CCDF of the number of transmissions with and without quantizing the feedback messages .............................................................................................................. 133
Figure 6-55: CDF of the first retransmission size ratio with and without quantizing the feedback messages .............................................................................................................. 133
Figure 6-56: HARQ latency for UEs at different coverage conditions for regular and early feedback (FB) .............................................................................................................. 134
Figure 6-57: Absolute PAF value of a ZC-sequence of length 31. Scale of τ- and ν-axis is normalized to the sequence length. The τ-axis is log scaled ........................................ 138
Figure 6-58: Absolute PAF value of a Delay-Doppler shifted m-sequence of length 31. Scale of τ- and ν-axis is normalized to the sequence length. The τ-axis is log scaled .............................................................................. 139
Figure 6-59: Left: False alarm and Right: Missed detection rate as function of SNR and mobile speed: Dashed lines are for ZC-sequences and solid ones for m-sequences. .......... 141
Figure 6-60: Timing Estimation Error CDF rate as function and mobile speed at 0 dB SNR: Dashed lines are for ZC-sequences and solid ones for m-sequences. .......... 141
Figure 6-61: Left: False alarm and Right: Missed detection rate as function of SNR and mobile speed: Dashed lines are for ZC-sequences and solid ones for m-sequences. .......... 142
Figure 6-62: Timing Estimation Error CDF rate as function and mobile speed at 0 dB SNR. ...... 142
Figure 6-63: Proposed random access signal structure from the UE in 5G IoT PRACH. ........ 143
Figure 6-64: Preamble and data are sent together. If successful, eNB sends ACK to the UE. .. 143
Figure 6-66: Preamble and data are sent together. If not successful, eNB estimates the number of users .......................................................... 144
Figure 6-67: Snapshot of an estimation of the number of colliding users (K = 2) .................. 144
Figure 6-68: General UE procedure of enhanced IoT random access .................................... 145
Figure 6-69: General eNB procedure of enhanced IoT random access ............................... 146
Figure 6-70: Number of random access attempts of IoT nodes before resource granted .. 148
Figure 6-71: Percentage of wasted UL resources when multiple UL grants are used .......... 148
Figure 6-72: Timing diagram of scheduling and transmission .............................................. 149
Figure 6-73: Mapping scheme of Turbo encoder in LTE .................................................... 150
Figure 6-74: Multiplexing scheme based on FDMA .......................................................... 150
Figure 6-75: CDF of User Number arrived in one TTI ...................................................... 151
Figure 6-76: MCC user numbers vs. residual BLER (1ms latency limited) ........................................ 152
Figure 6-77 Potential gains with informing the eMBB user about its traffic being punctured. 154
Figure 6-78 Throughput performance with notification of puncturing information (Green – notified, Red- Unnotified) ........................................................................................................ 155
Figure 6-79 CB-Level HARQ performance ............................................................................ 156
Figure 6-80 Performance of eMBB ...................................................................................... 157
Figure 6-81 CDFs of access delays of eMBB and MMC of the harmonized PHY procedure.. 158
Figure 6-82 PMF of access delay of MCC traffic for the harmonized PHY procedure......... 158
Figure 6-83 Downlink throughputs of eMBB and MCC. Left: Performance of baseline solution with no system bandwidth sharing. Middle: Performance of EAI with TB level retransmission. Right: Performance of EAI with CB level retransmission ...................................................... 159
List of Tables

Table 2-1 Summary of complexity comparison of waveform proposals. .................................. 21
Table 2-2 Summary of latency of waveform proposals. .......................................................... 22
Table 2-3: \( C_{wall} [km/h \times GHz] = v_{wallf} [km/h] \times f [GHz] \) ........................................... 35
Table 2-4: \( v_{wallf} \) .................................................................................................................. 35
Table 3-1: Numerology alignment priority table for each multi-cell coordination technique .... 49
Table 3-2 RTT for different DL TTI sizes and feedback duration.......................................... 66
Table 6-1 Complexity comparison of UF-OFDM and CP-OFDM ........................................... 88
Table 6-2 Waveform parameters of FC-OFDM....................................................................... 90
Table 6-3 Waveform parameters of FBMC ............................................................................. 92
Table 6-4 Waveform parameters of BF-OFDM ..................................................................... 93
Table 6-5 Parameters for waveform calibration test cases....................................................... 95
Table 6-6: Parameters for waveform comparison scenarios ................................................ 100
Table 6-7 Complexity evaluation normalized by the complexity of the received FFT............ 106
Table 6-8: Best CRSC distance spectrum for each DPR, corresponding codeword multiplicities
at distance \( d, \alpha(d) \), and puncturing masks (0 = punctured, 1 = unpunctured), \( R =2/3, M =8 \). 113
Table 6-9: CRSC distance spectrum of the DPR-2/8 mask when one additional data bit is
punctured, \( \alpha(d) \) is the multiplicity of codewords at distance \( d \). Considered parity puncturing
mask is 11000001 (0 = punctured, 1 = unpunctured).......................................................... 114
Table 6-10: Best candidates for ARP interleaver with the different configurations, \( S_{min} = 45, g = 8, S(0) = 0, R =2/3 \) and \( K =1504 \). ........................................................................... 114
Table 6-11: Estimated distance spectrum for the best ARP interleavers in AWGN channel with
corresponding multiplicities \( \alpha(d) \) and cumulated input weight at \( d_{min} = d_0, w_0 \). \( R =2/3 \) and
\( K =1504 \). ......................................................................................................................... 115
Table 6-12: Simulation Parameters ...................................................................................... 125
Table 6-13: Simulation Parameters ...................................................................................... 128
Table 6-14: Full buffer performance results for different feedback unit (CSIT) and resource
allocation unit (Alloc) granularities (in PRBs) and different scheduling modes for EPDCCH 129
Table 6-15: Error percentage in estimating the number of users. ........................................ 136
Table 6-16: Link-Level Simulation Parameters for NR-PRACH ........................................... 140
Table 6-17 Table comparison between the current 3GPP random access and this method ..... 144
Table 6-18 Simulation settings.............................................................................................. 147
Table 6-19: Simulation parameters for MCC........................................................................... 151
Table 6-20: Schemes simulated for comparison .................................................................. 152
Table 6-21: Simulation Assumptions.................................................................................... 153
Table 6-22: Schemes simulated for comparison .................................................................. 153
Table 6-23 Simulation Assumptions .................................................................................... 154
Table 6-24: Simulation Assumptions........................................................................................................ 155
Table 6-25 Simulation Assumptions ........................................................................................................ 156
Table 6-26 Schemes simulated for comparison ...................................................................................... 156
Table 6-27: Assumptions for access delay evaluation of the harmonized solution......................... 157
Table 6-28: TTI length assumptions for throughput evaluation of the harmonized solution.... 158
List of Acronyms and Abbreviations

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>Antenna Port</td>
</tr>
<tr>
<td>3D</td>
<td>3 Dimensions----</td>
</tr>
<tr>
<td>3G</td>
<td>3rd Generation-----</td>
</tr>
<tr>
<td>3GPP</td>
<td>3rd Generation Partnership Project</td>
</tr>
<tr>
<td>5G</td>
<td>5th Generation</td>
</tr>
<tr>
<td>ACK</td>
<td>ACKnowledgement</td>
</tr>
<tr>
<td>AMC</td>
<td>Adaptive Modulation and Coding</td>
</tr>
<tr>
<td>ARQ</td>
<td>Automatic Repeat reQuest</td>
</tr>
<tr>
<td>AoA</td>
<td>Angle of Arrival</td>
</tr>
<tr>
<td>BLER</td>
<td>Block Error Rate</td>
</tr>
<tr>
<td>BMS</td>
<td>Broadcast/Multicast Service</td>
</tr>
<tr>
<td>BF-OFDM</td>
<td>Block Filterd OFDM</td>
</tr>
<tr>
<td>CB</td>
<td>Code Block</td>
</tr>
<tr>
<td>CO</td>
<td>COnfidential</td>
</tr>
<tr>
<td>CoMP</td>
<td>Coordinated Multipoint Transmission</td>
</tr>
<tr>
<td>CRS</td>
<td>Common Reference Signal</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel State Information</td>
</tr>
<tr>
<td>CP</td>
<td>Cyclic Prefix</td>
</tr>
<tr>
<td>DL</td>
<td>DownLink</td>
</tr>
<tr>
<td>DS</td>
<td>Delay Spread</td>
</tr>
<tr>
<td>eMBB</td>
<td>enhanced Mobile Broad Band</td>
</tr>
<tr>
<td>eNB</td>
<td>Enhanced Node B</td>
</tr>
<tr>
<td>END</td>
<td>Equivalent Noise Degradation</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunications Standards Institute</td>
</tr>
<tr>
<td>EVA</td>
<td>Extended Vehicular A model</td>
</tr>
<tr>
<td>EVM</td>
<td>Error Vector Magnitude</td>
</tr>
<tr>
<td>FANTASTIC-5G</td>
<td>Flexible Air iNTerfAce for Scalable service delivery wiThinwIrelessCommunication networks of the 5th Generation</td>
</tr>
<tr>
<td>FC-OFDM</td>
<td>Flexibly Configured OFDM</td>
</tr>
<tr>
<td>FDPC</td>
<td>Frequency Domain Partial Construction</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>HARQ</td>
<td>Hybrid ARQ</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>IR</td>
<td>Internal Report</td>
</tr>
<tr>
<td>KPI</td>
<td>Key Performance Indicator</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>M2M</td>
<td>Machine-to-Machine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MBB</td>
<td>Mobile BroadBand</td>
</tr>
<tr>
<td>MC</td>
<td>MultiCarrier</td>
</tr>
<tr>
<td>MCC</td>
<td>Mission Critical Communications</td>
</tr>
<tr>
<td>MCS</td>
<td>Modulation coding scheme</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple-Input Multiple-Output</td>
</tr>
<tr>
<td>MMC</td>
<td>Massive Machine Communications</td>
</tr>
<tr>
<td>MMSE</td>
<td>Minimum Mean Square Error</td>
</tr>
<tr>
<td>MTC</td>
<td>Machine Type Communications</td>
</tr>
<tr>
<td>MUSIC</td>
<td>Multiple Signal Classification</td>
</tr>
<tr>
<td>NACK</td>
<td>Negative ACKnowledgement</td>
</tr>
<tr>
<td>NR</td>
<td>New Radio</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency Division Multiplexing</td>
</tr>
<tr>
<td>PA</td>
<td>Power Amplifier</td>
</tr>
<tr>
<td>PAPR</td>
<td>Peak to Average Power Ratio</td>
</tr>
<tr>
<td>PHY</td>
<td>Physysical layer</td>
</tr>
<tr>
<td>P-OFDM</td>
<td>Pulse-shaped OFDM</td>
</tr>
<tr>
<td>PPP</td>
<td>Public Private Partnership</td>
</tr>
<tr>
<td>PRACH</td>
<td>Physical Random Access Channel</td>
</tr>
<tr>
<td>PU</td>
<td>Public</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RA</td>
<td>Random Access</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>Research and Development</td>
</tr>
<tr>
<td>RACH</td>
<td>Random Access Channel</td>
</tr>
<tr>
<td>RAN</td>
<td>Radio Access Network</td>
</tr>
<tr>
<td>RAR</td>
<td>Random Access Response</td>
</tr>
<tr>
<td>RAT</td>
<td>Radio Access Technology</td>
</tr>
<tr>
<td>RE</td>
<td>Resource Element</td>
</tr>
<tr>
<td>Rel</td>
<td>Release</td>
</tr>
<tr>
<td>RS</td>
<td>Reference Signal</td>
</tr>
<tr>
<td>RX</td>
<td>Reception</td>
</tr>
<tr>
<td>SC</td>
<td>Single Carrier</td>
</tr>
<tr>
<td>SI</td>
<td>Study Item</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal-to-Interference plus Noise Ratio</td>
</tr>
<tr>
<td>SISO</td>
<td>Single Input Single Output</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>sTTI</td>
<td>Shortened Transmission Time Interval</td>
</tr>
<tr>
<td>TA</td>
<td>Timing Advance</td>
</tr>
<tr>
<td>TB</td>
<td>Transport Block</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>TC</td>
<td>Turbo Codes</td>
</tr>
<tr>
<td>TDL</td>
<td>Tapped-Delay-Line model</td>
</tr>
<tr>
<td>TR</td>
<td>Technical Report</td>
</tr>
<tr>
<td>TS</td>
<td>Technical Specification</td>
</tr>
<tr>
<td>TTI</td>
<td>Transmission Time Interval</td>
</tr>
<tr>
<td>TX</td>
<td>Transmission</td>
</tr>
<tr>
<td>UC</td>
<td>Use Case</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>UF-OFDM</td>
<td>Universal Filtered Orthogonal Frequency Division Multiplexing</td>
</tr>
<tr>
<td>UL</td>
<td>UpLink</td>
</tr>
<tr>
<td>uMTC</td>
<td>Ultra-reliable and low-latency MTC</td>
</tr>
<tr>
<td>V2V</td>
<td>Vehicle-to-Vehicle</td>
</tr>
<tr>
<td>V2X</td>
<td>Vehicle-to-anything</td>
</tr>
<tr>
<td>W-OFDM</td>
<td>Windowed-OFDM</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package or Working Party</td>
</tr>
<tr>
<td>ZF</td>
<td>Zero-Forcing</td>
</tr>
</tbody>
</table>
1 Introduction

The FANTASTIC-5G project advocates a single air interface for 5G, which is flexible, versatile, scalable and efficient in order to address the requirements of the IMT 2020. The flexibility featured by the FANTASTIC-5G air interface allows the support of a multitude of services and applications with diverse requirements, and shall be ready to provide solutions for future use cases not yet foreseen. As detailed in [FAN16-D21], the 5 core services constituting the 5G system are specified as Mobile Broadband (MBB), Massive Machine Communications (MMC), Mission Critical Communications (MCC), Broadcast/Multicast Communications (BMS) and Vehicular to anything communications (V2X).

The research on link design in the project focuses on the PHY and MAC layer research. The objective is to provide a holistic link solution that can be flexibly adapted to five core services MBB, MMC, MCC, BMS and V2X. This flexible adaptation to different services works under the assumption that the air interface has knowledge on the service type of each data flow and its requirements, facilitating an efficient cross layer optimization. The main focus of the radio link research is on the technical aspects that are tightly related to future 5G standardization.

The work on link design is divided into two research tasks:

- T3.1 - signal design, and
- T3.2 - frame design and PHY layer procedure.

T3.1 contains the following research topics: new waveform design; enhanced channel coding, enhanced modulation schemes; MIMO design; PAPR reduction techniques. T3.2 focuses on the flexible frame design, including the support for different numerology, user-centric control channel and reference signal design, Hybrid Automatic Repeat Request (HARQ) and service-adapted PHY layer procedure.

1.1 Objective of the document

This document reports on the recent research outcomes of link design. In the first year of the project, significant progress has been attained within T3.1, notably for the waveform design, which is close to the finalization state. For the second year, extensive simulation comparisons are conducted for all these waveform candidate schemes. In this report, we present our final summary and conclusions. Another important topic of this task is the channel coding. After a thorough investigation for an enhanced Turbo code in this first year, we provide more detailed design specifications in this report. Moreover, we analyze the latest 3GPP NR agreements and provide our comments. Besides, more updates on the MIMO and PAPR reduction techniques are reported. On the other hand, the work in T3.2 has been conducted under a common framework, resulting in a design principle, aiming for a holistic service-adaptive solution. Moreover, different topics, e.g. frame structure, control channel, HARQ, etc. are interconnected and jointly designed. In this report, we give all the important updates and the recommendations from the project viewpoint.

The collected results in this report, together with the previous deliverable, i.e. D3.1, will give a full picture of the research work in this work-package including the motivations triggering our research and the findings of 2 years’ work. Finally, we draw the project conclusions or the recommendations based on these research outcomes. Structure of the document...
1.2 Structure of the document

This document contains four main chapters. Chapter 2 reports on the research findings of T3.1. In section 2.1, the main focus here is to report on the waveform comparison framework, allowing the candidate waveform schemes to be compared among each other under agreed simulation assumptions. The presented framework contains two steps: comparison and recommendations. In sections 2.1.1 and 2.1.2, we summarize the main motivations and the waveform candidates considered during the project. In section 2.1.3, we report on the detailed comparison framework, including the link level Block Error Rate (BLER) performance, modem complexity and latency, based on which, we draw our project conclusions and recommendations in section 2.1.4. In section 2.2 we summarize the outcomes of channel coding investigations. The low latency oriented coding schemes are compared in section 2.2.2. Moreover, our findings on the enhanced Turbo code is reported in section 2.2.3, together with the comparison with the competing solutions LDPC and Polar code. In section 2.3, enhanced modulation schemes are presented. The outcomes of MIMO has two directions, on one side we conclude the MIMO-compatibility with respects to variant new waveform solutions (section 2.4.1). On the other side the efficiency of adopting adaptive antenna for moving cars is reported in section 2.4.2. Finally, in section 2.5, the PAPR related topics are summarized. Chapter 3 reports on the outcomes of the work in T3.2. Different from chapter 2, chapter 3 is organized under a common framework, which is due to the fact that T3.2 related topics, i.e. frame structure and numerology, physical downlink control channel, reference signal, HARQ and PHY layer procedures, are highly interconnected. It means that there is a common motivation and in particular a holistic design principle for all topics within this task. Section 3.1 and 3.2 summarize the motivation and the design principles, together with the project recommendations. While starting from section 3.3, detailed proposed new solutions or enhancements are reported. Finally, some conclusions are highlighted in chapter 4. Note that in order to ensure a concise main text, resulting into a good readability; some of the lengthy simulation analyses are given in the annex chapter, for those who are interested in more performance details.
2 Holistic solution for signal design

2.1 New waveform: Comparison and recommendation

2.1.1 Motivation

In order to support heterogeneous services with diverse requirements in 5G new radio (NR) systems, several new waveform solutions, as one of the key components of air interface design, have been proposed in [FAN16-D31]. Generally speaking, these waveform candidates are OFDM-based waveforms with either subband-wise or subcarrier-wise filtering functionalities, addressing the main requirements of supporting flexible numerologies, asynchronous transmission, and high mobility, etc.

Currently in 3GPP discussion, it has been agreed that for enhanced mobile broadband (eMBB) and ultra-reliable and low-latency communication (URLLC) services operated in the bands up to 40 GHz, new radio (NR) supports CP-OFDM based waveform and DFT-S-OFDM based waveform (complementary to CP-OFDM waveform, at least for eMBB uplink for up to 40GHz) [R1-167963, R1-1610485]. In addition, it states that in-band frequency multiplexing of different numerologies is supported in NR for both DL and UL, at least from the network perspective. Spectral confinement techniques (e.g. filtering, windowing, etc.) at the transmitter need to be transparent to the receiver from RAN1 perspective. 3GPP has also agreed that additional waveforms may be supported by NR for other services (e.g. mMTC).

The new waveform proposals in FANTASTIC-5G go beyond the scope of the 3GPP agreements, targeting scenarios that are not yet focused within the first phase of 3GPP NR. These new waveform candidates are designed aiming at the support of services beyond eMBB, such as MMC and V2X, which are promising application services identified by 3GPP. In this deliverable, we will provide the detailed comparison of waveform proposals from various perspectives, such as link performance, complexity, and latency. Conclusions and recommendations for the new waveform application will be drawn at the end of this section based on the comparisons made.

2.1.2 Overview of FANTASTIC-5G new waveform candidates

New waveform candidates proposed in this project are basically OFDM-based waveforms with filtering functionalities, enabling a flexible PHY design of a multi-service air interface. Specifically, they allow partitioning the system bandwidth into separate sub-bands, whose PHY parameters (e.g., subcarrier spacing, cyclic prefix, or specific filter coefficients) can be individually configured according to the requirements of a service. The waveforms thus enable the in-band coexistence of different services, each of those being assigned to its own subband with tailored characteristics.

Waveform candidates proposed and analysed can be grouped into two categories, namely:

- Subband-wise filtered waveform:
  - **Universal Filtered OFDM (UF-OFDM)**: [R1-165014, R1-165013] Inspired by the fact that practical allocations in wireless cellular multi-user systems happen in groups of adjacent subcarriers, UF-OFDM has been introduced to make use of filters being applied to groups of subcarriers (e.g. successive resource blocks spanning 12, 24, 48 subcarriers). UF-OFDM can be applied to a ZP-OFDM or a CP-OFDM signal, as indicated in [R1-165014, R1-1609564]
  - **Filtered OFDM (F-OFDM)**: While in UF-OFDM the filter length is constrained to the length of the CP overhead, an alternative version of UF-OFDM is called F-OFDM, which uses a filter length going beyond the CP duration [AJM15, ZJC+15].
• **Block-filtered OFDM (BF-OFDM):** Combines most of the advantages of the OFDM and FBMC waveforms at the price of a slight complexity increase at the transmitter side while keeping a simple OFDM receiver [Ger17]. Spectral localization and performance in multi-user scenario will be enhanced compared to conventional OFDM, and conventional equalization and MIMO techniques can be considered as the typical OFDM receiver can be kept [Dem17].

• Subcarrier-wise filtered waveform (aka. windowing):
  
  o **Flexibly Configured OFDM (FC-OFDM):** It makes the waveform flexibly configurable, so that configurations can be tailored to the service requirements. FC-OFDM applies a multiple mode processing, e.g. multi-carrier mode, DFT-spreading mode and zero-tail DFT-spreading mode, etc., before the IFFT transform, these different modes can be multiplexed in frequency or in time domain. The resulting time domain signal yields a set of multiplexed waveforms [R1-164619]. Moreover, the post-IFFT employs a windowing process, further enhancing the spectral confinement [R1-166594]. On the other hand, the receiver processing can select either time-domain windowing or frequency-domain windowing implementation structures, depending on the delay spread statistics [R1-166595].

  o **Pulse shaped OFDM (P-OFDM):** Following the idea to fully maintain the signal structure of CP-OFDM, P-OFDM allows for the use of pulse shapes other than the rectangular pulse to balance the localization of the signal power in time and frequency domain [YZB16, ZSG+17]. It entails windowed OFDM as a special case, but offers more degrees of freedom by allowing succeeding symbols to overlap by a factor that may be chosen freely.

  o **Frequency spreading Filter-Bank Multi-Carrier/Filter-Bank Multi-Carrier (FS-FBMC/FBMC):** The prototype filters in an FBMC-OQAM system are generated by applying frequency sampling technique instead of a polyphase network based implementation. This feature provides a significant advantage when the channel is exhibiting large delay spread or for the case of synchronization mismatch [DBC+14].

  o **FBMC with QAM signaling (QAM-FBMC):** FBMC system that supports complex-domain QAM symbol mapping while keeping the symbol transmission rate equal to the maximum time-frequency product (TF=1) at the expense of a relaxed orthogonality constraint. At least two filter-bank bases are used, which are applied to successive subcarriers in an alternating manner [KKY+15].

  o **Zero-Tail-spreading OFDM (ZT-DFT-s-OFDM):** A modified version of the single-carrier frequency division multiple access (SC-FDMA) waveform. There are zeroes placed at the edges of the input to the Discrete Fourier Transform (DFT) block [BTS+13].

The principle and detailed descriptions of above-mentioned waveform candidates (excluding BF-OFDM) are given in [FAN16-D31], while those regarding BF-OFDM are addressed in 6.1.1. Selected waveforms are involved in the performance comparison, namely UF-OFDM, FC-OFDM, P-OFDM, FS-FBMC/FBMC and BF-OFDM. In accordance with 3GPP waveform discussions, we also evaluated the performance of windowed Overlap Add (WOLA) [Qua15], which is considered as one of the windowing schemes in 3GPP. Note that part of the waveform candidates investigated in this project were also the waveform candidates considered in 3GPP.
2.1.3 **Waveform Comparisons**

### 2.1.3.1 Link performance

For a fair comparison between different waveform proposals stemming from different partners in the project, we conducted the following three steps for link performance comparison:

- Agree on the test cases for calibration of different simulators with an implementation of a common reference waveform (e.g. CP-OFDM, SC-FDMA). The test cases are aligned with the 3GPP waveform evaluation discussions that were ongoing in parallel.
- Each waveform partner calibrates the simulator according to the agreed test case.
- Define simulation scenarios for comparison, and each partner provides results for corresponding new waveform proposal and reference waveform (i.e. CP-OFDM).

Regarding the selection of scenarios, we target non-eMBB scenarios primarily, with focusing on requirements of the new services and their coexistence in the same band. Hence, flexible numerology coexistence and support for asynchronicity are of our particular interest. Besides, we adapt these scenarios in alignment with the parallel discussions in 3GPP [R1-163558, R1-166031] and considered more challenging settings if appropriate. These three scenarios are:

- Scenario 1: Uplink asynchronous transmission (Single numerology case).
- Scenario 2: Downlink high speed transmission (Single numerology case).
- Scenario 3: Uplink synchronous transmission (mixed numerology case).

Some general observations from our link performance results are:

- Uplink asynchronous transmission (Single numerology case): for small guard bands, sub-band-wise filtering schemes outperform windowing schemes, while FBMC with OQAM signaling provides best spectral confinement. However, with increasing the guard band, the performance curves of all schemes converge.
- For downlink high speed scenario, windowing schemes and BF-OFDM provide slight gain compared to CP-OFDM in the high-SNR region, while FBMC-OQAM is subject to a performance loss in high-SNR region. This is because all these schemes have better frequency localization which leads to reduced inter-carrier interference (ICI). However, for FBMC-OQAM the interference due to channel delay spread becomes more predominant compared with Doppler effect. This is due to the fact FBMC-OQAM uses longer symbols in time domain, which causes a performance loss when the channel exhibits a short coherence time.
- Uplink synchronous transmission with mixed numerology: when using low to mid modulation order, sub-band based schemes outperform windowing schemes for small guard band, while the gap gradually diminishes for larger guard band. When using high modulation order, sub-band-wise filtering scheme outperforms windowing scheme for small guard band only in high SNR region. The performance of both schemes converge for larger guard band, while FBMC with OQAM signaling performs best in this case.

Detailed simulation setup, waveform specifications, and performance results are detailed in 6.1.3.

### 2.1.3.2 Complexity comparison

The complexity is evaluated in terms of the number of real multiplications for CP-OFDM and new waveform. Complexity analysis for different waveform proposals are given below:

<table>
<thead>
<tr>
<th>Waveform</th>
<th>Complexity compared to CP-OFDM (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UF-OFDM</td>
<td>1.2 (#Suband =12, B=1)</td>
</tr>
</tbody>
</table>

Table 2-1 Summary of complexity comparison of waveform proposals.
2.1.3.3 Latency comparison

We evaluate latency performance of different waveform proposals, while the latency is directly related to the symbol length going beyond the symbol period.

Table 2-2 Summary of latency of waveform proposals.

<table>
<thead>
<tr>
<th>Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOLA</td>
</tr>
<tr>
<td>FC-OFDM</td>
</tr>
<tr>
<td>P-OFDM</td>
</tr>
<tr>
<td>FBMC</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>BF-OFDM</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

2.1.4 Conclusions and recommendation

General note: 3GPP NR requires the spectral confinement techniques, e.g. windowing/filtering to be transparent to the receiver. We call this specification transparent requirement (STR), which naturally compromises the spectral confinement of the radio signal.

Advantages and drawbacks of windowing and sub-band filtering:

- Subcarrier-wise filtering (aka. windowing)
  - Advantages:
    - Windowing at TX and RX can be implemented with low complexity, e.g. complexity increase is less than 2% for FC-OFDM and P-OFDM windowing operation.
• Windowing reflects a per-subcarrier filtering, hence window design is independent of allocated sub-band location and bandwidth
• Windowing operation only needs to be run once, both for contiguous as well as for distributed sub-band allocation
• Windowing does not cause sub-band edge EVM regrowth, rendering it suitable especially for narrow-band transmission
• Transparent windowing (verifying STR) is doable for window lengths not substantially exceeding the symbol interval, e.g. [R1-166594]

- Drawbacks:
  - Transparent windowing constrains the window length and results in compromised spectral confinement.
  - OQAM-based scheme does not allow a straightforward reuse of certain MIMO-OFDM schemes, e.g. SFBC and non-linear MIMO.

- Sub-band-wise filtering
  - Advantages:
    - Transparent filtering (verifying STR) is always doable without restrictions
    - For constant filter tail length, sub-band filtering can achieve a steeper slope of the power at the sub-band edges compared to windowing
    - Filter length is less constrained by STR than windowing, since interference imposed by the filter tails on surrounding symbols is minor or negligible (filter matched at the receiver is not mandatory to suppress interference). This leads to enhanced spectral confinement.
  - Drawbacks
    - Sub-band filter design depends on the allocated sub-band location and bandwidth, i.e. specific set of filter coefficients are needed for each sub-band location and bandwidth
    - Sub-band filtering schemes have relatively higher implementation complexity compared to windowing; however, the gap varies depending on specific filtering schemes (e.g. UF-OFDM, F-OFDM and BF-OFDM)
    - Sub-band filtering causes EVM re-growth at the edges of the sub-band; however precoding can be used to alleviate this issue.

Conclusions:
Note that the conclusion is drawn from our waveform comparison results, which are based on the project simulation assumptions, although, we cannot guarantee that the simulation cases cover all aspects relevant for 5G. We aim at giving some important insights and helping to select the suitable waveform schemes.

• Under the STR condition, sub-band filtering-based schemes allow for steeper slopes at the sub-band edges in f-domain, yielding better spectral confinement than windowing-based schemes
• OQAM signaling based FBMC with overlapping symbols provides best confinement and highest spectral efficiency, but it cannot meet the STR and suffers a compatibility loss with OFDM (e.g. not MIMO-friendly)
• Sub-band filtering-based schemes require relatively high Tx-Rx modem complexity compared to windowing-based schemes (except for FBMC-OQAM, where the symbol rate is doubled and thus two FFTs are required).
• From our comparisons: for small guard bands, sub-band based schemes clearly outperform windowing schemes. However, for larger guard bands, the performances of both schemes converge. Hence, eventually the size of guard bands needed to attain the desired interference isolation is decisive for selecting the appropriate waveform.
Recommendations

- OQAM signaling, according to the latest SoTA, still requires re-design of MIMO, synchronization sequence and some control signals. Thus, we think that it is not ready for 5G yet.
- We recommend to consider both sub-band filtering-based and windowing-based schemes for 5G in various situations
  - Sub-band filtering should be considered in the following cases,
    - enhanced spectral confinement level is required; (especially if small guard bands are desired to be used at high spectral efficiency)
    - allocated sub-band is contiguous; (sub-band size should be larger than a minimum size)
    - sub-band allocation is almost static; (allowing to change / update filter coefficients rarely)
    - reasonable amount of TX antennas (since the signal at each antenna needs to be filtered individually)
  - Windowing should be considered in the following cases,
    - Moderate spectral confinement level is needed
    - Larger guard bands are a reasonable choice, either if spectral efficiency is not the main target or if high interference isolation is desired
    - Allocated sub-bands are distributed in the system band; (support of narrow-band transmissions)
    - Dynamic sub-band allocation is needed (note that here we refer to sub-bands assigned to different services and thus need isolation, aiming to avoid inter-service interference due to e.g. time mis-alignment or different numerology, and not classical multi-user isolation. For example, eMBB certainly needs only one filter for service isolation, which may be adapted, though, should the band allocation between eMBB and other services change).
    - large amount of antennas

2.2 Channel coding

2.2.1 Research scope constraints and requirements

A Study on New Radio Access Technology was approved in RAN#71 meeting [RP-160671]. The channel coding scheme is a fundamental component for fulfilling the different requirements of next generation radio access technologies [3GPP-38.913]. The current LTE coding schemes have not been designed for the new requirements of the identified usage scenarios such as, in 3GPP terminology, eMBB (enhanced mobile broadband), which is equivalent to MBB in our project; mMTC (massive machine type communication), which is equivalent to MMC in our project; and URLLC (ultra-reliable low latency communication), which is equivalent to MCC in our project.

Each identified usage scenario has specific requirements related to the channel coding scheme [3GPP-38.913]:

- **eMBB**: for this usage scenario, the channel coding scheme should support a large range of data rates, from very low data rates for instant messages or control signaling to very high data rates (20 Gbps for downlink(DL) and 10 Gbps for uplink (UL)) with improved error correction capabilities and at the price of a reasonable implementation cost. The coding scheme should be flexible in terms of data block sizes and coding rates. The different latency requirements for data and control planes should also be
satisfied, that is 4 ms (UL and DL) for user plane latency and 10 ms for control plane latency.  
- **URLLC/MCC**: the channel coding scheme must fulfill high reliability on small packets with very low latency. The latency target was set at 0.5 ms for UL and DL whereas the URLLC reliability requirement for one transmission of a packet is $10^{-5}$ for 32 bytes with a user plane latency of 1ms.  
- **mMTC/MMC**: the use cases involving a massive distribution of sensors and actuators require the channel coding scheme to support small packets (a few dozen to hundred bits), with energy-efficient encoding and decoding which is necessary for mMTC long-life devices.

Two contributions related to channel coding were carried out in FANTASTIC-5G. The first one specifically targets low latency communications: several families of codes, including convolutional, turbo, LDPC and polar codes were considered and compared. On the opposite, the second contribution focuses on the turbo code family while targeting all NR scenarios: the goal is to upgrade the existing LTE turbo code in order to cope with the NR requirements.

The contribution descriptions are followed by a synthesis of the ongoing discussions related to channel coding in 3GPP and by some thoughts and recommendations for the scenarios still being dealt with.

### 2.2.2 Low latency channel codes

One way to achieve low latency is to use channel codes with short lengths. However, the choice of the channel codes for short lengths becomes critical as the error correction performance degrade with the decreasing block length. Numerical evaluations indicate that Turbo codes used in LTE are not performing well in the short message length regime, and the error correction performance can be improved by utilizing other modern channel codes. The following channel coding schemes have been considered in 3GPP for NR:

- **Tail-Biting Convolutional Codes**:
  - Advantages: Already included in LTE. Simple encoder and decoder structure.
  - Disadvantages: Performance decreases with increasing message length. High memory required for good performance.

- **LTE Turbo Codes**:
  - Advantages: Mature technology, already included in LTE. Allows simple rate adaptation and HARQ.
  - Disadvantages: Relatively poor error correction performance in short length regime compared to other candidates.

- **Binary LDPC Codes**:
  - Advantages: Mature technology, already included in different standards, good performance.
  - Disadvantages: Rate matching and HARQ support are not as simple as Turbo codes.

- **Non-Binary LDPC Codes**:
  - Advantages: Excellent performance with no error floor.
  - Disadvantages: Very high decoder complexity.

- **Polar Codes**:
  - Advantages: Excellent performance for short lengths with no error floor, scalable decoder complexity depending on the decoder list size.
  - Disadvantages: Relatively new technology and hence no mature solutions for HARQ exist.

- **Outer Erasure Codes**
o Advantages: Improvement in error floor performance
o Disadvantages: Additional complexity

The evaluation of the presented schemes (except for the outer erasure codes) can be found in the appendix.

**Standardization Status:**

In RAN#87, 3GPP decided to use LDPC codes for eMBB data channels (mainly due to their suitability for high throughput applications) and Polar codes for control channels (due to their excellent performance for short message lengths). There have been no agreements for scenarios with low latency requirements (such as URLLC), however it is naturally that the eMBB codes will be strong candidates also for URLLC.

**Conclusion and Recommendations:**

We consider Polar codes as a strong candidate for low latency applications. Although it is a relatively new technology, Polar code show excellent performance for short message lengths with no error floors, and their scalable decoder complexity (depending on the decoder parameters) allows them to be used in simple and complex receivers. This also allows the decoder performance to improve with technological improvement without changing the codes defined in the standard.

### 2.2.3 Enhanced Turbo codes (IMTA)

**Motivation**

With the increasing number of communications scenarios that have to be supported by the upcoming 5G standard, new requirements related to error correcting codes have emerged. As already stated in [FAN16-D31], the conventional FEC coding component of LTE/LTE-A is not designed to answer favorably to service requirements with stringent reliability and latency constraints. A known issue for the LTE Turbo Code (TC) resides in its poor performance at low error rates when transmitting data with coding rates higher than 1/3. This is due to the well-known “error floor” effect, which can be observed when the TC is punctured with the rate matching mechanism [CNB+08]. This detrimental effect results in the frequent resort to retransmissions through the HARQ mechanism. Moreover, these codes were not originally designed targeting best short packet performance.

Therefore, we have made some changes to the LTE codes to make it able to cope with the new requirements introduced by 5G scenarios. Actually, provided that they would be able to guarantee lower error rates when punctured, TCs could remain promising channel coding candidates for 5G due to the maturity of the technology and since they can offer, at least, a partial backward compatibility with LTE. Indeed, the impact of introducing additional coding solutions on the chip area would be greater than reusing the same family of codes than LTE, since NR and LTE are foreseen to co-exist for a long period of time. In addition, this family of codes offers advantages regarding support for HARQ, frame size and code rate flexibility and regarding the decoding complexity for low rate scenarios that represent most of the used cases from network statistics.

**Performed technical work**

The performed work consists of several technical contributions addressing first the improvement in performance at low error rates, on the one hand, and for short frame sizes, on the other hand [GBA+16]. As introduced by Berrou et al. [BGT93], the minimum Hamming distance \((d_{\text{min}})\) of a TC is not only defined by its constituent encoders but also fixed through the TC interleaver. Therefore, we have investigated the joint design of interleavers and puncturing patterns for TCs in order to guarantee low error floors and good convergence thresholds. As a result, a new puncturing constraint related to parity puncturing is proposed for the design of TC interleavers. The work focused on the Almost Regular Permutation (ARP) interleaver model.
[BSD+04], adopted in standards like IEEE 802.16 WiMAX. Trellis termination was also modified to avoid the spectral efficiency loss of tail bits.

At a second phase, aspects related to flexibility were studied to propose efficient rate-compatible tail-biting turbo codes with 1-bit codeword flexibility implementing the previously proposed interleaver and puncturing schemes. Efficiency resides in the absence of any unexpected behavior such as the performance of a shorter frame size better than a larger one.

The technical details regarding all the contributions are provided in the Appendix section.

To show the validity of the proposed enhanced turbo code solution, performance comparisons were made. The scenario of the improvement in the error floor is first chosen. Large frame sizes of 6,000 information bit-frames were used since it was commonly admitted that error floors are more difficult to avoid in this case. LTE turbo code is used as a reference.

As observed in Figure 2-1, large coding gains exceeding two decades in most cases were observed for the enhanced turbo codes. The largest gains can be observed for high coding rates. In addition, in some cases, a small gain is also observed in the upper part of the curves. Block Error Rates (BLER) of $10^{-5}$ are achieved without change in the slope of the curves.

Simulations were also carried out for short frame sizes. Figure 2-2 shows that large gains of several dBs can then be observed for the proposed enhanced TC with respect to LTE turbo code for short frame information sizes of around 100 bits for low and high coding rates. $10^{-8}$ of BLER is achieved without any change in the slope of the curves.

More than eight TDOC contributions to 3GPP RAN1 NR group were submitted for the proposed TCs. Simulation results for eMBB data channel and URLLC scenarios are provided in the Appendix.
2.2.4 **eMBB scenario: standardization discussions and conclusions**

Binary turbo (LTE and the proposed enhanced version), Low Density Parity Check (LDPC) and concatenated polar codes were the main considered candidates proposed for adoption in 3GPP RAN1 for NR eMBB data channel. Agreed technical observations regarding the 3 families were drawn and captured in [R1-1610878].

Performance results were collected in [R1-1610600]. However, it has not been possible to draw conclusions directly from these results, owing to different views on the implementation complexities and possible enhancements. Technical concerns were raised regarding the efficient and performant support of IR-HARQ for the LDPC and polar codes whereas the ability of turbo codes to support IR-HARQ is considered as well established.

The following is a summary of the agreed observations about complexity, performance, maturity, flexibility, strengths and weaknesses for each family of codes captured in the NR chairman’s notes [R1-1610878].

Regarding complexity, it was agreed that computational complexity is far from being an accurate representative of final hardware complexity and power efficiency. Indeed, it was acknowledged that memory requirements and accesses have a large impact on chip area and power consumption respectively.

LDPC codes are considered as widely implemented in commercial hardware supporting several Gbps throughput with attractive area and energy efficiency while supporting flexibility and features that are largely limited when compared to the corresponding requirements for NR. Actually, the area efficiency reduces for lower coding rates and the complexity increases with increasing flexibility for this family of codes. Moreover, despite the ability of achieving large parallel decoding degrees, some of this parallelism may not be exploited for all code block lengths and code rates for NR, resulting into a penalizing impact on energy and area efficiency. To conclude, hardware implementation with attractive area and energy efficiency is considered challenging when simultaneously targeting the peak throughput and flexibility requirements of NR.

Polar codes are considered implementable, although there are currently no commercial implementations, and in relation to NR, there are some concerns linked to the maturity and the...
availability of decoding hardware. In addition, most existing work in literature is related to successive interference cancellation decoders and not list-based decoders that enable the excellent performance of this family of codes. For list decoders, the implementation complexity increases with increasing list size, especially with larger block sizes. Moreover, the area efficiency reduces for shorter block lengths and lower coding rates. List-4 decoder was agreed as implementable for all codeword sizes. However, in practice, most simulations considered list-8 decoders that could be argued implementable for short frame sizes. To conclude, for decoding hardware that can achieve acceptable latency, performance and flexibility, there are some concerns about the area efficiency and energy efficiency that are achievable with polar codes and about the maturity of the technology.

Turbo codes are widely implemented in commercial hardware, supporting HARQ and flexibility similar to what is required for NR, but not at the high data rates or low latencies required for NR. In fact, turbo codes meet the flexibility requirements of NR with the most attractive area and energy efficiency except at higher throughputs, and particularly at lower code rates and lower block lengths. Generally, the area and energy efficiency is constant when varying the puncturing and repetition rate. Another advantage resides in the fact that the decoding complexity increases linearly with the information block size for a given mother code rate. However, there are concerns that implementation with attractive area and energy efficiency is challenging when targeting the highest throughput requirements of NR. Therefore, two proposals were put forward by the proponents of turbo codes: the first one considered using turbo codes as a single solution in NR. The second considered designing a turbo decoder capable of decoding both LTE and, at least, lower information block sizes ($K \leq 6144$ bits) of NR. For the high throughput case of NR, a LDPC code can be designed with a limited flexibility or equivalently for few combinations of code rates ($R > 1/2$) and frame sizes ($K > 6144$ bits). This proposal would have the benefit of combining the advantages of each family (turbo and LDPC) of codes without bearing the burden of their drawbacks. Indeed, it was shown in [R1-1608584] that this combination of codes could answer favourably especially in terms of complexity all the requirements of NR.

As a conclusion to the NR chairman’s notes in [R1-1610878], we can state that each family of codes presents its challenges when trying to satisfy simultaneously all the requirements of NR. Therefore, it is quite difficult to clearly identify an all-around favorite without performing a joint thorough analysis of performance and complexity/latency taking into account real implementations and not estimates of it. Due to timing constraints, the framework for such a comparison was not agreed and individual technical contributions were used as a basis for the selection process.

At a later stage, a proposal of combining polar codes for short frame sizes and LDPC codes for longer frame sizes was also put forward. It was motivated by the fact that the decoding complexity of list decoding increases with larger frame sizes. The willingness to accept the use of LDPC codes for larger frame sizes by the proponents of both polar and turbo codes, was used to adopt LDPC codes for large frame sizes spanning an interval of values and supporting a flexibility level well beyond what was foreseen at the start. Afterwards, it was proven in [R1-1611260] that there is no LDPC code with the minimum cycle length or girth greater than or equal to 6 for the smallest info block lengths and high rates of NR. This results into a large performance penalty when using iterative decoding based on belief propagation. Consequently, the promising candidates for short frame sizes were limited to turbo and polar codes.

Performance comparisons between the proposed enhanced turbo code and the polar code proposals taking into account IR-HARQ support were provided in [R1-1613347]. The corresponding observations showed that enhanced turbo code with Max-Log-MAP decoding and 8 iterations offers comparable BLER performance to the polar code with list-8 decoding for rates $R > 2/3$ and slightly better performance for lower rates at the first transmission. In subsequent transmissions for HARQ support, enhanced turbo code offers better performance than rate-compatible polar code. Gains exceeding 0.5 dB can be observed in this case. In
addition, steeper BLER curves were observed for the enhanced turbo code predicting larger gains if lower error rates were to be targeted.

Complexity assessment on the hardware implementation of channel decoders for short block lengths was presented in [R1-1612306] for the three families of codes. It was shown that despite having lower computational complexity at many coding rates \( R \), flexible list-4 polar decoder ASICs have inferior area-efficiency compared to state-of-the-art turbo decoder ASICs. On the other hand, flexible LDPC decoder ASICs suffer from degraded area- and energy-efficiency at low coding rates and short block lengths. Turbo decoder ASICs have superior area- and energy efficiency.

Finally, regardless of potential technical drawbacks, a compromise was found that led to the adoption of LDPC codes for eMBB data channel and polar codes for the control channel.

2.2.5 **Next step: URLLC and mMTC scenarios**

The choice of the coding solution for the URLLC and mMTC scenarios remains an open issue. While agreed simulation conditions for these two scenarios in 3GPP are quite different from those of eMBB data channel, they are partly related to the comparisons performed for short frame sizes. In fact, frame sizes lower than 1024 bits were considered for coding rates from \( R = 2/3 \) down to \( R = 1/12 \). Error rates of \( 10^{-4} \) to \( 10^{-5} \) of BLER are targeted.

Taking into account the performance results provided in [R1-1613347] comparing the polar and the enhanced turbo code, we can clearly identify these two families of codes as strong candidates for URLLC and mMTC from the performance point of view with a slight edge for turbo codes showing improved performance for these targeted low rates as shown in [R1-1702856]. LDPC codes cumulate two main drawbacks: the first lies in the large performance penalty (more than 1.0 dB for short frame sizes and low rates as shown in [R1-1702856] and [R1-1610314]) in some cases, the second in the fact that decoding complexity increases by orders of magnitude when decreasing the coding rate when compared to the two other families of codes.

Obtained results are conditioned by the simulation parameters, especially by the corresponding decoders. It could be argued that with improving technology, additional complexity could be tolerated and longer list sizes can be used for polar codes that can lead to improving their performance. However, the same argument applies for the proposed turbo codes. In fact, list decoding with or without combination with outer CRC code for turbo codes is also possible which would lead to large improvements in performance especially for short frame sizes as was shown in [R1-1610931]. Indeed, [HS16] shows that the BLER performance of list-32 decoding for polar codes coincides with the Maximum Likelihood (ML) performance of the code, while the classical Max-Log-MAP decoding for LTE turbo codes is around 0.6 to 0.8 dB away from the corresponding ML performance. In addition, when combining with an outer CRC code, a gain exceeding 2.0 dB can be achieved in some cases of short frame sizes for the LTE turbo code as shown in [WJX+13].

From the complexity point of view, turbo codes present an undeniable advantage for such low rates as \( R = 1/12 \) since the decoding complexity of this family of codes scales linearly with the information block size \( K \) and not with the codeword size \( N \). This is not the case for both polar and LDPC decoders. Actually, it was shown in [R1-1702856] that the proposed turbo code has lower computational complexity than both LDPC and polar codes (for lower rates) while showing better performance for the considered simulation parameters for URLLC.

**Conclusion for URLLC and mMTC**

When targeting low rates and short frame sizes for URLLC, polar and turbo codes emerge as the most promising candidates from the error correcting performance point of view. In fact, LDPC codes cumulate the drawbacks of short cycles (decoding problem with BP) and of complexity (increasing with lower rates). To identify the best technical choice, in-depth complexity
comparisons going beyond simple computational complexity, would have to be carried out at comparable performance. The framework for such comparisons should be clearly set and agreed between the parties proposing these coding solutions.

Finally, the selection process for 5G coding solution has launched a wave of new proposals for the 3 families of codes. Current studies are focusing on improving the decoding efficiency of turbo decoders when targeting high throughput scenarios. A large number of studies are also focusing on the design and implementation efficiency of polar codes/decoders. Therefore, important improvements are being made and a thorough investigation taking into consideration performance and hardware complexity between these two strong candidates represented by turbo and polar codes should be performed when the selection process should take place.

2.3 Enhanced modulation

Modulation schemes play an important role in enhancing the system spectral efficiency and coverage. In this chapter, two modulation schemes will be proposed to address the 5G requirements.

2.3.1 FQAM for interference limited scenarios with inactive subcarriers utilization

In the conventional QAM modulation that is adopted in LTE, data symbols are transmitted on all the subcarriers that are assigned for data transmission in the resource block. On the other hand, frequency quadrature amplitude modulation (FQAM) represents a combination of frequency shift keying (FSK) and QAM, where information bits conveyed by selecting one active subcarrier among a subset of candidates and modulating the selected subcarrier with a QAM symbol [F5G-D31]. $M$-ary FQAM carries $Q = \log_2(M)$ information bits by selecting one subcarrier among $M_F$ subcarriers and modulating the selected subcarrier with an $M_Q$-ary QAM constellation. Hence, the modulation order of the resulting FQAM symbol is given by $M = M_F M_Q$. The motivation for FQAM is coming from the fact that the worst-case distribution of interference-plus-noise in terms of channel capacity is the Gaussian distribution [SA13]. Having some of the subcarriers inactive in FQAM results in inter-cell interference with non-Gaussian distribution, which will improve the channel capacity. This advantage is most prominent in interference limited scenarios, such as at cell-edge where the inter-cell interference is the dominant factor. Detailed interference analysis for FQAM with comparison to QAM modulation is presented in [F5G-D41]. The analysis has shown that even with dense small-cells deployment, FQAM keeps the non-Gaussian distribution feature of inter-cell interference. The cell-edge users’ throughput can be significantly improved using FQAM, as shown in [F5G-D31, AMN16]. From the performance evaluation presented in [F5G-D31, AMN16], one can conclude that FQAM represents a promising modulation technique to enhance the system performance in heavy interference environment.

However, although it provides good performance in interference scenarios, FQAM modulation suffers from low bandwidth efficiency compared to QAM scheme. The shortcoming of FQAM is that it underutilizes the available resources by keeping part of the available subcarriers empty/inactive. We propose a method to utilize the inactive subcarriers in the FQAM symbol to serve another user with good channel conditions [AB16]. In this method, the cell-edge user will be served with FQAM scheme, and another cell-centre user will be served with QAM modulation on the inactive subcarriers of the FQAM symbols, as shown in Figure 2-3. The cell-centre user will be served with lower transmission power compared to the power on the active subcarrier to enable the signals’ detection at both users, and to not affect the interference distribution at the cell-edge users. At the receiver of the cell-centre user, the active subcarrier is detected first (e.g. by using energy detection methods) then the data on the active subcarriers is retrieved. On the other hand, as the power on the inactive subcarriers is relatively low, conventional detection methods can be used at the cell-edge user. The power ratio between the
two powers can be adjusted to achieve the system’s requirements, such as ICI (i.e. maintain the non-Gaussian distribution), throughput and QoS.

![Diagram showing power levels to cell-edge user and cell-centre user.](image)

**Figure 2-3: Utilizing the inactive in FQAM symbol to transmit to another user.**

Using numerical simulations, we show that by properly selecting the power levels on the active and inactive subcarriers, the non-Gaussian distribution of the ICI at the cell-edge can be maintained, which is essential to achieve a good performance with FQAM scheme [AB16]. In addition, we observed that the distribution of the ICI for the cell-centre user in the proposed transmission scheme deviates from the Gaussian one as well. Thus, detection methods that are not based on Gaussian assumption can be utilized to enhance the users’ performance. The results show that the system total throughput can be significantly enhanced using the proposed method compared to LTE-QAM and conventional FQAM schemes. With the proposed scheme, the system throughput is increased by 130%-140% compared to conventional FQAM.

Another advantage of FQAM is that it can enable blind detection of users’ activity in the scenarios when the base-station is not aware of how many users are transmitting on the same resources (details of the scheme provided in Section 3.3.5.1).

To compare the achieved spectral efficiency by the proposed method to utilize the inactive subcarriers in FQAM, Figure 2-4 shows the achieved throughput for the considered schemes for different cell-edge user distances, where the throughput is determined by the number of successfully received frames by the users. For comparison, we consider two benchmark schemes: 1) LTE system, where the resources are exclusively allocated to the cell-edge user with QAM used as the modulation scheme, and will be referred to as LTE-QAM in the figures, 2) Conventional FQAM scheme, where no data is transmitted on the unselected subcarriers. The x-axis represent the relative distance of the cell-edge user from the maximum distance (the cell-centre user location is fixed). A system consisting of 19 cells is considered with 500m inter-site distance to capture the impact of the inter-cell interference, where FQAM is most effective. We adopt the 5 MHz bandwidth LTE system's parameters, which consist of 25 resource blocks each with 12 subcarriers and maximum transmission power of 46 dBm at the base-station. ITU pedestrian B channel model is adopted for generating the fast fading. We implement turbo code of coding rate 1/3 and QPSK modulation, and $M_f=4$ for FQAM. For the proposed scheme, the results will be presented for one cell-edge user and one cell-centre user (at distance half of the cell-edge distance). For the benchmark schemes, as only one user is served on a set of subcarriers in each cell, only the results for the cell-edge user will be presented.

Firstly, it can be noticed that LTE-QAM modulation has a very bad performance due to the inter-cell interference at the cell-edge. FQAM shows very desirable performance in the high inter-cell interference region. In the proposed scheme, the cell-edge user performance will be slightly degraded due to the increased interference from serving the cell-centre user on the inactive subcarriers. However, the total system throughput is increased by being able to serve another user on the resources. With the proposed scheme, the system throughput is increased by 130%-140% compared to conventional FQAM. In addition to the gain in the system’s
throughput offered by the proposed scheme, the capability of supporting more users in the system is highly advantageous for 5G communications systems.

Figure 2-4: Throughput comparison of the proposed scheme with conventional approaches.

2.3.2 NUC for high modulation orders

In LTE, the bit to symbol mapper is a simple uniform QAM mapper. Uniform QAM constellations are easy to map and de-map. However, there is no information theory basis for this choice and these constellations can be shown to be far from the Shannon limit. Non-Uniform Constellation (NUC) has received attention as a tool to improve the performance of uniform constellations with minor decoder complexity increase. In the NUC the constellation points are no further required to be in a uniform or rectangular shape. NUC can be obtained by optimizing the alphabet in order to maximize the BICM capacity subject to the power constraint. The work carried out up to now on NUC has always concentrated on the single channel BICM capacity optimization. Most of the time the BICM capacity is optimized to perform well for the AWGN channel. The design can be performed as well using the Rayleigh channel as a basis by using the probability density function of the Rayleigh distribution. However, in mobile communications, users experience different channel conditions. Thus, it is important that the selected NUC has a good performance across different channels and different SINR waterfalls. We have proposed a novel algorithm to design multichannel NUC and analyze its performance compared to the single channel NUC and uniform constellations [MAA16]. We have shown that, on average, this method outperforms the single channel single SNR design method used up to now. Simulation results show that, compared to the uniform constellation (used in LTE), a potential improvement of up to 1.1 dB for 256QAM without additional complexity is expected. Higher gains are expected for higher modulation orders. The advantage of NUC is more prominent in high modulation orders. Thus, NUC can be utilized for services with high data rate such as MBB and BMS services.

2.3.3 Conclusion and recommendation

As it has been shown in the previous sections, FQAM and NUC outperform current modulations schemes used in LTE, and bring advantages for the 5G system. It is recommended to use FQAM to serve users that suffer from high interference levels (such as cell-edge users) and QAM for users with good channel conditions (cell-centre users). The inactive subcarriers in each FQAM symbol should be utilized to serve another user with relatively lower transmission power as
detailed in Section 2.3.1. A user feedback (similar to CQI in LTE) can be used to determine the interference level at the user and decide which modulation to be used (FQAM or QAM). Also, it is recommended to use FQAM in uplink for grant-free and random access schemes to enable blind detection at the base-station as explained in Section 3.3.5.1.

For BMS and MBB services, it is recommended to utilize NUC for at least high modulation orders (64QAM, 256QAM, etc.) to enhance the achieved throughput. As it is not expected to have high modulation orders for MMC, MCC and V2X, NUC may not be very beneficial for these services.

2.4 MIMO

In D4.2, system level aspects of multiple input multiple output (MIMO) are addressed by the project. The role of the current sub-section is to complete this work with the physical layer aspects of MIMO:

- The feasibility of MIMO with new waveforms is addressed in section 2.4.1;
- Providing a robust and high data rate link to fast moving connected vehicles thanks to adaptive MIMO and the predictor antenna is addressed in section 2.4.2;
- Providing high data rate to connected objects thanks to Massive MIMO and single carrier modulation is addressed in section 2.4.3.

2.4.1 Compatibility of MIMO with new waveforms

The system integration of enhanced multiple input and multiple output schemes is addressed in section 4.1 of D4.2. This work shows that these enhanced MIMO schemes, with and without cooperation, boost the network spectral efficiency and users’ data rates. However, for the sake of simplicity in the system level performance evaluation, all these MIMO schemes are proposed on top of an OFDM layer. In the current sub-section, we check that the new waveforms presented in section 2.3 and are compatible with these MIMO schemes.

The following waveforms proposed by the project are compatible with any MIMO scheme originally design for OFDM, as they all provide orthogonality between sub-carriers in the complex domain: UF-OFDM and F-OFDM; ZT-s-OFDM; BF-OFDM; P-OFDM and FC-OFDM (in a QAM version only).

Only FS-FBMC OQAM implies modifications on the MIMO schemes. For instance, during the project, we have shown that a block-wise time reversal Alamouti scheme for FS-FBMC (see pages 52-54 of D3.1) and applying soft-input decoding to FS-FBMC receiver [DDR+16] manage to reach similar performance in terms of BER vs SNR than OFDM based solutions and better performance than PPN-based FBMC receiver. The modified Alamouti scheme works with blocks of symbols and requires guard time between blocks to guarantee the “Alamouti” orthogonality. In these studies, we have demonstrated the benefits of the proposed scheme combined with FS receivers, especially when the inter-carrier spacing is large (small size of FFT). Increasing the inter-carrier spacing allows lowering the size of the FFT that also reduces the complexity and the number of simultaneous active carriers. This point has a direct impact on the PAPR that is also decreased. This makes the proposed scheme particularly relevant for MMC scenarios, in which, power consumption, complexity and robustness are key parameters.

2.4.2 Making adaptive MIMO compatible with fast moving connected vehicles thanks to the predictor antenna

In the section 2.3 of D4.2, system level aspects of mobility are addressed. New handover solutions are provided. In the system level simulations used to assess the network performance at high velocity, the link is assumed to be robust, at the physical layer. In the current sub-section, we present a physical layer scheme proposed by the project to guarantee this robustness.
Adaptive MIMO schemes (such as adaptive beamforming and spatial multiplexing) are not naturally robust to speed. At low speed, they exploit the channel state information at the transmitter (CSIT), to reduce the “cost” (in terms of spectrum usage and radiated power) of the delivered data rate. Unfortunately, beyond a limiting velocity (which depends on the carrier frequency), the network can no longer exploit such information (which becomes outdated). Typically, beyond a speed limit that depends on the carrier frequency the network falls back to open loop MIMO schemes (such as diversity) which are less cost efficient (in terms of spectrum and power usage). This situation is described metaphorically as follows: the network hits “the wall of speed” [PHT+15]. A simple channel prediction technique has been proposed in the section 3.4.4 of D3.1 to solve this problem. It is based on the predictor antenna concept [PHS+15]. It only requires one additional antenna on the car roof.

For the first time, in this project, we determine by simulation the location of the “wall of speed” in the speed and frequency domains. We also measure to which extent the “wall of speed” can be “pushed farther” in the speed and frequency domains, thanks to channel prediction. Our study is detailed in [PHS+16] and summarized in the current deliverable. In this study, we assume Time Division Duplex mode, channel reciprocity based downlink 256x2 Zero Forcing (ZF)-MIMO, downlink 256x1 maximum ratio transmission (MRT) beamforming and a 1 ms delay between channel measurement in the uplink and downlink data transmission based on channel state information. With these assumptions, we show that the afore mentioned speed limit \( v_{wall} \) is the following function of the carrier frequency \( f \):

\[
v_{wall}(f) = \frac{C_{wall}}{f},
\]

where \( C_{wall} \) is a constant. A system with a large \( C_{wall} \) value is more robust to velocity. Table 2-3 shows the values of \( C_{wall} \) which have been determined by simulation for various adaptive MIMO schemes and channel prediction assumptions. Note that as expected, ZF-MIMO is less robust to velocity than MRT MISO.

<table>
<thead>
<tr>
<th>MIMO</th>
<th>Without Prediction</th>
<th>With Predictor Antenna</th>
<th>Gain from Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>256x2 ZF-MIMO</td>
<td>60 km/h x GHz</td>
<td>180 km/h x GHz</td>
<td>( \times 3 ) higher velocity or higher carrier frequency can be supported</td>
</tr>
<tr>
<td>256x1 MRT-MIMO</td>
<td>180 km/h x GHz</td>
<td>735 km/h x GHz</td>
<td>( \times 4 ) higher velocity or higher carrier frequency can be supported</td>
</tr>
</tbody>
</table>

Speed limits derived for 1.2 GHz, 3.5 GHz and 6 GHz carrier frequencies are provided as examples in Table 2-4.

<table>
<thead>
<tr>
<th>MIMO</th>
<th>Without Prediction</th>
<th>With Predictor Antenna</th>
</tr>
</thead>
<tbody>
<tr>
<td>256x2 ZF-MIMO</td>
<td>50 km/h, for ( f = 1.2 ) GHz</td>
<td>150 km/h, for ( f = 1.2 ) GHz</td>
</tr>
<tr>
<td></td>
<td>17 km/h, for ( f = 3.5 ) GHz</td>
<td>51 km/h, for ( f = 3.5 ) GHz</td>
</tr>
<tr>
<td></td>
<td>10 km/h, for ( f = 6 ) GHz</td>
<td>30 km/h, for ( f = 6 ) GHz</td>
</tr>
<tr>
<td>256x1 MRT-MISO</td>
<td>150 km/h, for ( f = 1.2 ) GHz</td>
<td>612 km/h, for ( f = 1.2 ) GHz</td>
</tr>
<tr>
<td></td>
<td>51 km/h, for ( f = 3.5 ) GHz</td>
<td>210 km/h, for ( f = 3.5 ) GHz</td>
</tr>
</tbody>
</table>
Figure 2-5 illustrates the “wall of speed”. As illustrated in Figure 2-5-a), without channel prediction, beyond 50 km/h, the network stops using ZF-MIMO 256x2, falls back to diversity and consumes twice more spectrum; whereas with channel prediction this happens only at 150 km/h. As illustrated in Figure 2-5-b), without channel prediction, beyond 30 km/h, the network stops using MRT-MISO 256x1, falls back to diversity and consumes 20dB more power; whereas with channel prediction this happens only at 122.5 km/h.

![Figure 2-5: The “wall of speed” measured by simulations for two different scenarios](image)

Table 1: The “wall of speed” measured by simulations for two different scenarios

<table>
<thead>
<tr>
<th>Speed (km/h)</th>
<th>Frequency (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>f=6 GHz</td>
</tr>
<tr>
<td>122.5</td>
<td>f=6 GHz</td>
</tr>
</tbody>
</table>

In [PHS+16], we also show that at 3.5GHz, both 256x2 ZF-MIMO or 256x1 MRT MISO can be used without ever hitting the wall of speed in dense urban areas, where the speed is limited at 50 km/h.

This first study shows that the wall of speed (its location in the speed domain and its height in the cost domain) depends on many factors. Early studies on the predictor antenna [PHH13] have analysed the impact of the number of transmit antennas and the impact of the propagation (line-of-sight versus non line-of-sight propagation) on MRT beamforming. These first studies indicate that these parameters may have no impact on the location of the wall in the speed domain, but rather a strong impact on the height of the wall in the cost domain (i.e. power consumption). Such studies should be updated and extended to all MIMO schemes being currently standardized at the 3GPP, such as the grid of beam (GoB).

### 2.4.3 Connecting objects in high data rate with a single carrier modulation: it might work with Massive MIMO

The project has made separate in-depth analysis of new MIMO schemes (in D4.2) and new multi-carrier waveforms (in current deliverable). As explained in section 2.4.1, they are compatible with each other and match the current expectations on 5G.

However, we foresee that in following releases of 5G, the need for high data rate in low cost connected objects will be stronger.

We believe that the use of simple devices does not necessarily prevent from reaching high data rates, if most of the signal processing is performed by the base station. Indeed, in [PHB+13][PHT+13] hundreds of Mb/s are delivered to a device using a single carrier modulation signal, a single receive antenna and a single-tap receiver (even less complex than a GSM receiver). The base station pre-processes the signal (using the transmit matched filter) and exploits a large number of transmit antennas to delivered to the device an almost “echo free” signal.
We also believe that the use of a (true) single carrier modulation (i.e. without FFT and IFFT) does not prevent from using multi-carrier modulations as well. Indeed in [FCH13], several links using single carrier modulations are managed even though they partially overlap in the frequency domain. The interference between links is mitigated thanks to beamforming. Therefore, as illustrated in Figure 2-6, single carrier modulation (or other non-orthogonal modulations) could be re-introduced in a 5G multi-carrier system. Partial overlapping in the frequency domain could be dealt by the use of massive MIMO and beamforming (in both the uplink and downlink directions).

Figure 2-6: Single carrier modulation and multi-carrier modulation partially overlapping in frequency and separated through massive MIMO and Tx/Rx beamforming (the “object” is an MTC device or a high data rate connected object)

2.4.4 Conclusion and recommendation

The following conclusions have been derived:

- All the new waveforms proposed by the project (except FS-FBMC OQAM) in the current deliverable are compatible with the enhanced MIMO schemes proposed in D4.2.
- Regarding high mobility, in general, thanks to the predictor antenna, the networks hits the “wall of speed” (i.e. stops using adaptive MIMO) at higher velocities. Simulation results show that, for 256x2 ZF-MIMO and 256x1 MRT-MISO, respectively, speeds or carrier frequencies which are 3 and 4 times higher, respectively, can be supported thanks to the predictor antenna.

The following recommendations have been derived:

- Regarding high mobility, the use of more advanced prediction will push even further the “wall of speed” to higher carrier frequencies and higher speeds.
• First studies indicate that the wall of speed and its sensibility to parameters such as the number of antennas or the propagation environment need to be characterized for all MIMO techniques being studied at the 3GPP.
• Regarding the longer term evolution of 5G, massive MIMO is identified as a potential enabler to mix service-specific waveforms. For instance, single carrier modulation links with high data rates (Mbits/s) could be introduced for high data rate connected objects. The interference due to partial overlapping in frequency between single carrier and multi-carrier modulations would be reduced by transmit/receive beamforming.

2.5 PAPR reduction techniques

PAPR is an important factor in the design of an air interface for future communication systems since it has a high impact on the transceiver complexity. Generally, a high PAPR requires highly linear power amplifiers (expensive) or a large input back-off (energy inefficient) in order to protect the signal from severe degradation due to non-linearity effects introduced by the power amplifier. In the current LTE uplink transmission, single carrier based DFTs-OFDM is used for UL in order to increase power efficiency. There are several schemes discussed for PAPR reduction with different degree of complexity. Due to different core services within 5G, it is required to differentiate between device types (which correspond to core services) and requirements on PAPR reduction in terms of complexity, energy and/or spectral efficiency. However, single-carrier vs. multi-carrier based transmission should not be made solely on the PAPR performance, but rather be settled after comprehensive evaluation for different channel characteristics and scenarios, in terms of PAPR, power amplify back-off, and link performance, etc. Within this report, guidelines / conditions are provided from Fantastic 5G perspective to select specific PAPR reduction techniques for specific service types.

2.5.1 Research scope and candidates’ introduction

Within F5G, several PAPR reduction schemes have been investigated and proposed applicable for different 5G scenarios. In the following, a summary of the contributions is provided.

Two-stage PAPR reduction: It is a newly proposed set of PAPR reduction schemes, whereas, evident from the naming, the reduction is done in two separate stages. The first stage consists in applying one of the iterative clipping based schemes (i.e., TR, ACE, TRACE); a further clipping is the second (and final) stage. This means that it is a version of the iterative schemes, where after the predefined maximum number of iterations is reached a final clipping is applied giving the final output [ED+16].

Modified selected mapping: It is quite appealing to take the overlapping of the FBMC/OQAM time symbols, into consideration for improved PAPR reduction. Thus, we have devised a new scheme that we called “modified SLM (mSLM). This scheme considers $U^K$ hypotheses instead of just $U$ as for the regular SLM, and hence a great increase in computational complexity is expected, as well as a larger gain in PAPR reduction [ED+16].

Soft clipping for single-/multi carrier: Concerning pulse shaped OFDM technique as an example of multi-carrier scheme, the PAPR level of P-OFDM is comparable to the one of CP-OFDM and higher than DFTs-OFDM. There are three standard approaches for Low PAPR design of P-OFDM systems. Firstly, we can design the multi-carrier systems by adapting the subcarrier spacing according to channel frequency selectivity and PAPR target. Secondly, the idea of transforming into a single-carrier transmission can be applied to pulse shaped OFDM, similar to DFT-spread OFDM systems. Thirdly, for constrained modulation order, signal clipping approach is considered as a simple and efficient method to reduce the PAPR level. Although this measure induces a signal error vector magnitude (EVM) loss, it only introduces marginal
link performance loss; as long as the EVM is kept in limits to still facilitate a reliable transmission [FAN16-D31].

**FDPC:** The FDPC algorithm uses a basic structure, so-called PCTS (Pre-Constructed Temporal Signal) which appears in the form of a Feedback loop within mapping symbols Feed forward correction structure. The PAPR reduction on an MC symbol is calculated during an MC symbol period with a latency of one symbol duration. For each subcarrier, the algorithm computes the correction of the corresponding mapping symbol, which will be applied to this symbol before IFFT and will also be reintroduced in the loop for the calculation of the mapping symbol correction of the next sub-carrier. The proposed algorithm can be either transparent to receiver, i.e. the receiver does not need to know the specific PAPR reduction algorithm; or it should specify the constellation mapping fashion in the standard for the receiver being able to demodulate, e.g. CM/C. More details can be found in Appendix 6.5.

### 2.5.2 Comparison scenarios and final results

In the following table, 5G-relevant core service types and corresponding requirements on PAPR reduction are summarized.

<table>
<thead>
<tr>
<th>Core service type</th>
<th>Requirements for technical candidate to be selected</th>
</tr>
</thead>
</table>
| **Mobile Broadband (MBB)**                    | • High modulation order which requires low distortion  
• Energy efficiency is not critical, a higher backoff can be allowed                                               |
| **Massive machine communication (MMC)**       | • Low modulation order may tolerate a higher degree of distortion  
• Energy efficiency is crucial  
• No or minimal backoff plus robust transmission for reliable single shot transmission                           |
| **Mission critical communication (MCC)**      | • Low delay mechanisms  
• Medium complexity  
• Spectral and energy efficiency may be compromised  
• Enable diversity gains to attain high reliability                                                                |
| **Broadcast multicast service (BMS)**         | • Medium up to high modulation orders allow for certain distortion  
• Energy efficiency rather of secondary importance  
• Diversity gains should be enabled to allow reliable reception                                                        |
| **Vehicular-to-anything (V2X)**               | • Low/medium modulation order  
• Low/medium delay tolerance (use case specific)  
• Medium complexity  
• Spectral and energy efficiency may be compromised  
• Diversity gains: large in cellular links, rather small in side-link. Nevertheless, diversity should be exploited for improving reliability. |

### 2.5.3 Conclusion and recommendation

Based on the analysis above, the following mapping from specific requirements of the core service to the prioritized PAPR reduction technique is proposed:

<table>
<thead>
<tr>
<th>PAPR-requirement classes</th>
<th>Service type</th>
<th>Prioritized technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>High modulation order</td>
<td>MBB, MBS</td>
<td>Multicarrier modulation with PAPR reduction</td>
</tr>
<tr>
<td>High complexity</td>
<td></td>
<td>• Allow for complex algorithms</td>
</tr>
<tr>
<td>Category</td>
<td>Example:</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td>Low/medium modulation order + Medium Complexity</td>
<td>FDPC, mSLM / Two Stage</td>
<td></td>
</tr>
<tr>
<td>MCC, V2X</td>
<td>Multicarrier modulation (for sub-6GHz), Single carrier modulation (for high-frequency comm.). Depending on the application, single carrier may be favorable even in sub 6 GHz band. Example: MC with Soft Clipping.</td>
<td></td>
</tr>
<tr>
<td>MMC</td>
<td>Low complex algorithms + use case specific: For narrowband transmission: DFT-spreading based modulation. For MC: low complex algorithms (e.g. OFDM with clipping) for PAPR reduction.</td>
<td></td>
</tr>
<tr>
<td>high reliability + multicarrier</td>
<td>FDPC: allows for degradation-free on reference signals. controlled processing latency.</td>
<td></td>
</tr>
</tbody>
</table>
3 Holistic solution on frame design and link level procedures

In D3.1 [FAN16-D31] we have provided our views related to frame design based on a set of design principles. In the following we will revisit this list and make connections to the work in FANTASTIC-5G related to these items.

3.1 Motivation

The frame design for 5G has to serve the following 5G design target:

A. Enable forward compatibility (i.e. simplifying the inclusion of new use cases and features)
B. Enable multi-service support (i.e. meeting the high heterogeneity of requirements)
C. Enable high spectral, energy, and cost efficiency
D. Enable tight interworking with other RATs (e.g. LTE and mmW)

This leads to a set of rules (in the following called design principles) the frame design should follow. Where applicable we refer to these 4 items in the next sections respectively.

3.2 Design principles and recommendations

Sampling rate, bandwidth support, numerology options, relative timing:

1. Sampling rate, number of subcarriers covering the bandwidth and corresponding subcarrier spacing should be integer multiples of a given basis, to keep system complexity and testing efforts at a reasonable level.

In D3.1 (section 6.5 [FAN16-D31]) we have presented reasonable options for selecting the triple of (1) covered bandwidth, (2) sampling rate and (3) the respective FFT length for various sub carrier spacings. We have kept the list of options rather wide, ranging from 0.2 MHz up to 320 MHz. Which ones are to be applied will in the end depend on the availability of bands and roll-out scenarios. The lower end of these options (e.g. 0.2 MHz) are for example candidates for spectral refarming of 2G bands, while the higher end of this list is more relevant for the capacity layer of the network and potentially need to be applied at carrier frequencies above 6 GHz. The coverage layer of 5G will most likely make use of a similar set of options as 4G did (10 MHz, 20 MHz, 40 MHz) with the potential of going up to 80 MHz or even 160 MHz. The made selections are addressing design target D (e.g. interworking with LTE).

2. The base sampling rate should be aligned with the base sampling rate of LTE, to ease interworking and multi-link functionalities and enable hardware reusability and sharing.

The options we have proposed in D3.1 (section 6.5 [FAN16-D31]) follow this rule for improved coexistence (even in-band), interworking and multi-link processing. Additionally, this helps the reuse of design activities happened in the era of 4G. Again the made selections are to improve D.

3. Reasonable amount of supported bandwidth: 5 MHz, 10 MHz, 20 MHz, 40 MHz and up to 160 MHz (for small cell capacity hot spots), to keep system complexity and testing efforts at a reasonable level, while supporting all reasonable scenarios. Smaller bandwidth options might be envisaged in later standard releases.
As given in point 1 above (and detailed in D3.1, section 6.5 [FAN16-D31]) we have decided to keep the set of options rather wide. The actual down-selection in NR depends on various aspects being outside of the project such as available bands, roll-out strategies and selected carrier frequency.

4. Means to support narrow-band devices by dedicated design of in-band structures are required (e.g. introduction of complementary narrow-band synchronization signals), to successfully integrate MMC services while still enabling MMC devices to be low-cost and long-lasting with a single set of energy sources.

While allowing low-end devices (e.g. sensors) to skip parts of the time- and energy-consuming procedures related to accessing the system (e.g. message exchange during network entry), these devices still need to be able to detect close-by cells and synchronize on frame level. For this it needs to detect the respective synchronization signals regularly being broadcasted by the basestations (in LTE: primary and secondary synchronization signals, PSS and SSS) and detect the general cell configurations (in LTE: MIB and SIB). While for conventional broadband devices in 4G (and potentially in 5G NR as well) PSS and SSS have been transmitted in the center of the band covering about 1 MHz, the wide bandwidth is less suitable for low-end devices. Instead it is beneficial to broadcast another set of synch signals covering a smaller bandwidth as this allows these devices to be more cost- and energy efficient. In section 6.6.3 of D3.1 [FAN16-D31] we have investigated this feature. Naturally, with relying on a smaller bandwidth only, the synchronization accuracy degrades. We have provided means to improve the accuracy without extending the bandwidth. Additionally, the work on filtered waveforms has shown that a reduced accuracy with respect to symbol timing and carrier frequency can be tolerated (e.g. sections 6.1.1.3, 6.1.4.3, 6.1.5.4 in D3.1). These activities are addressing B and C (at the device).

5. The basic 5G subframe length should follow LTE (i.e. have a duration of 1 ms), to ease interworking and multi-link functionalities and enable hardware reusability and sharing.

To enable low latency transmissions, special subframes with lengths being fractions x of 1 ms should be supported (e.g. x=2^N with Integer N, e.g. N = [0, 1, 2, …]).

In section 6.5.2.2 and 6.5.3 of D3.1 [FAN16-D31] we have proposed a set of resource block definitions following a more general rule as given above (i.e. the TTI length options are N multiples/divisions – N being integer – of the baseline of 1 ms). Additionally, the tiling concept being presented in 6.5.1 of D3.1 follows this criterion as well. Applying this rule simplifies the definition of the set of available tile/sub-tile configurations. Being aligned to LTE improves coexistence and simplifies cooperation between 4G and 5G. The design targets being approached here are B and D.

The actual gain the system can make use of when allowing the use of various TTI lengths has been investigated and presented in section 2.4.1 and 2.4.2 of D4.2 [FAN17-D42].

6. Subframe bundling needs to be supported (for longer transmission opportunities), to increase the signalling efficiency (less control signalling required) and to support coverage extension.

The tiling concept as proposed in 6.5.1 of D3.1 [FAN16-D31] has a granularity in time of 1 ms as baseline. Shorter TTI lengths are supported via the selection of sub-tiles and by allowing preemptive scheduling. Longer TTI lengths are supported by bundling consecutive tiles. This means the scheduler of sub-frame n+1 needs to take into account, if in sub-frame n it has been decided to bundle. This imposes some restrictions to the tile scheduler. Additionally, this
requires to use asynchronous HARQ mechanisms to avoid SAW channels (Stop-And-Wait) to be blocked (or vice versa tile-bundling being blocked by active SAW channels). This item is allowing a more efficient inclusion of low-end devices and to make signaling more effective. Thus, the design targets B and C are strengthened.

The actual gain the system can make use of when allowing the use of various TTI lengths has been investigated and presented in sections 2.2.3.2 in D4.1 [FAN16-D41] (related to overhead reduction with making use of subframe bundling), 2.4.1 and 2.4.2 of D4.2 [FAN17-D42].

7. The support of different subcarrier spacings (following rule 1) within a single carrier is one option to be envisaged, to improve respective use cases (e.g. low latency transmissions) and transmissions with extreme channel characteristics (e.g. very high Doppler values), while maximizing multiplexing gains and reducing the number of constraints for e.g. scheduling. For the former an alternative approach is the introduction of minislots containing a lower number of symbols.

The key questions to ask for these items are twofold: (1) How can the system support different subcarrier spacings without imposing severe inter-subband interference and (2) what are the gains the system can make use of with this degree of freedom. The former question is strongly related to our work on waveforms. In [SW15] we have proposed means to efficiently implement this into a system applying UF-OFDM (though, other waveforms being investigated in FANTASTIC-5G are able to do so in a similar manner). The latter question is taken in [SWA16] and section 6.1.5.5 in D3.1 [FAN16-D31]. We have shown, that for extreme Doppler scenarios we can gain up to a factor of two/three/four (low SNR/medium SNR/high SNR scenario) with respect to spectral efficiency. By applying this mechanism design target B is improved.

8. Transmission points are aligned to a common time base, to ease collaborative schemes such as interference coordination.

As soon as different cells are to coordinate with each other on a spectral resource basis - e.g. to deal with intercell-interference – it is beneficial, if these cells have a common time base. To name an example, if a given cell is to mute specific resources to protect transmissions in another cells, it only has to mute a single sub-frame, if the sub-frames of these two cells are time-aligned. If the cells are not time-aligned two sub-frames would have to be muted as it is very likely, that the sub-frame to be protected intersects with two sub-frames of the other cell. WP4 of FANTASTIC-5G has worked on mechanisms related to this in the area of intercell-interference handling. The outcomes are given in section 4.2 of D4.2 [FAN17-D42]. Applying this rule supports the design target C.

Another aspect benefitting of having a common time base is the improved in-band coexistence of 4G and 5G as e.g. outlined in section 6.1.1.5 of D3.1 [FAN16-D31]. This way the design target D is improved.

Lean channel/signal design, time/frequency confined structures, access procedures:

1. Use of multi-carrier signalling, to allow for simple transceiver mechanisms (e.g. one-tap equalization, frequency selective precoding) and multiple transmissions to share the band.

All waveform candidates being proposed in FANTASTIC-5G are of multi-carrier type (or modifications, e.g. by applying DFT precoding). With applying multi-carrier signaling all design targets are fostered as multi-carrier signalling, which allows the dedication of the available spectral resources for various means and with differing characteristics concurrently.
(supporting design targets A and B). This approach has been shown in earlier generations to be highly efficient (supporting design target C) and simplifies coexistence and interworking with e.g. 4G (design target D).

2. The amount of always-on components should be minimized and the actual repetition rate (e.g. for synchronization signals in DL) should be configurable. Apply the on-demand principle as far as possible (e.g. MIB – master information block – is always-on while SIB – system information block - is on-demand), to increase energy efficiency especially in low-load scenarios.

While cell-specific control-plane messaging per se is not part of the project, the contribution in section 4.2.3 of D4.1 [FAN16-D41] (icell coordinated small-cell on/off) is related to this item. With applying this and similar means the design target C is addressed. Additionally, being able to deactivate certain signaling components as required improves forward compatibility (the design target D) and makes the system more energy efficient (the design target C). DL reference symbols in 4G are an example for always-on signals constraining the system in that respect.

3. Highly flexible TDD configurations required for efficiently following the actual traffic/service needs.
4. Subframe options supporting both UL and DL components (control, data, reference signals) should be envisaged (for latency optimized scenarios) in addition to subframe configurations purely containing either DL or UL components.

We have addressed items 3 and 4 in section 3.5.3 in D3.1 [FAN16-D31]. Here, we have proposed four different subframe types (pure UL, pure DL, DL control + data and UL control, DL control and UL control and data). All types include respective reference signals. By dynamically selecting the respective subframe the system is able to follow efficiently the actual traffic needs. This degree of freedom allows the system to be highly efficient (design target C).

WP4 has taken up this in D4.2 [FAN17-D42] section 2.4.1 for the design of efficient scheduling mechanisms.

5. The radio frame (with radio frames consisting of an integer number of subframes) should support different kinds of access mechanisms (scheduled access, contention based access, beam guided access), to support various use cases related to e.g. eMBB and MMC concurrently.

We have investigated different kinds of access mechanisms to support the different kinds of services in D4.1 [FAN16-D41] and D4.2 [FAN17-D42]. These mechanisms range from scheduled access e.g. making use of adaptive TTI lengths and in-band control channels, efficient massive access mechanisms (protocols and detection mechanisms), D2D and beamforming aided procedures. Besides unicast mechanisms we also have investigated broad- and multicast solutions. For allowing the frame to efficiently support different kinds of access mechanisms concurrently in parallel (e.g. scheduled access, massive access and broad/multicast) the tiling concept can be used by defining respective tile types. Overall, this flexibility allows the system to support different kinds of use cases efficiently and thus fosters the design targets B and C.

6. Reference signal design should be configurable to meet various design targets (e.g. to be optimized towards the respective transmission mode; to enable frequency resource blanking – i.e. avoid ‘almost blanked subframes’ as e.g. in LTE).
7. Reference signal design should natively support a wide range of number of antenna ports, to enable later extensions without requiring fundamental redesigns.

The work being reported in section 3.3.3 of this deliverable follows items 6 and 7, describing our approach for the design of DL reference signals. The sub-tiling concept in combination with the use of a respective code strategy (e.g. Walsh Hadamard) allows to easily scale the system according to the number of antenna ports/beams available without introducing excessive overheads. Depending on the state of the device (e.g. related to time variance and frequency selectivity of its channel towards the basestation) it can use the relevant reference symbols to conduct its measurements. The design target C is addressed with applying these techniques.

8. Time frequency confined user-specific control channel design to enable multi-service support and frequency resource blanking and to make 5G forward compatible.

9. Control channel structure, allowing for devices to go immediately to the sleep mode (micro-sleep), to maximize energy efficiency at the device side.

10. Control channel design enabling energy efficient reception without putting severe restrictions to the search space, to minimize energy consumption without putting heavy restrictions to e.g. the scheduling mechanism.

Those three design principles are connected. As presented in sections 3.3.2 of this deliverable and 2.4.1 in D4.2 [FAN17-D42] FANTASTIC-5G promotes to apply in-band user specific control channels for user specific control messages (e.g. DCIs carrying the selected resource configuration, scheduling grants ...). The aspects making this a valuable choice (as e.g. in contrast to the PDCCH design in 4G) is the fact, that the actual control messages can use different formats (as e.g. a eMBB services requires different control means than a MCC service does). Additionally, by not multiplexing control messages of multiple users into a single structure and instead appending it to the respective data transmission allows to use the same link enhancing methods (e.g. interference coordination, rank 1 precoding, etc. as e.g. described in section 4.2 of D4.2 [FAN17-D42]), does not require to configure the transmission mode according to the weakest link and allows to share reference symbols between control and data. So, with applying this means the system is more efficient in general (design target C) and the respective control elements can be designed according to the respective use case (design target B). Finally, the inclusion of new use cases and features is improved as the number of constraints is minimized (design target A). The work related to the user-ID signals (UIDS) as presented in section 6.6.1 of D3.1 [FAN16-D31] is targeting a control channel structure with 3 parts: 1) UIDS (specific preamble), 2) fixed-sized FEC block (basic control) carrying the beginning of the DCI and 3) variable-sized FEC block (main control) carrying the remainder of the DCI (and the start of user data if any). The UIDS is to allow the device to quickly and efficiently detect its messages. As the device only requires to scan for its preamble instead of having to fully decode all control messages down to the CRC check (as e.g. in LTE) the search space is not required to be restricted (i.e. the device is able to check all possible positions) and thus does not restrict the scheduler improving design target C.

11. Avoid non-elastic transmission mechanisms (e.g. configurable asynchronous HARQ instead of fixed synchronous HARQ as e.g. used in LTE UL), for the system to be more flexible e.g. related to dynamic TDD and deployments with centralized structures.

Non-elastic transmission mechanisms have the advantage of lower overheads (as the non-elasticity provides implicit knowledge about relevant characteristics of the transmission which not have to be communicated). At the same time, they add constraints to the system (e.g. to the scheduler) preventing it to fully make use of the flexibility options described earlier (e.g. related to variable TTI definitions and dynamic TDD configurations). Asynchronous HARQ has been
tackled in FANTASTIC-5G in section 3.3.4 of this deliverable. The application of this is for example described in section 2.4.1 of D4.2 [FAN17-D42]. Design targets A, B and C are improved with applying these principles.

12. Allow for specific traffic types with highest priority (e.g. related to MCC) to ‘hijack’ allocations originally being dedicated to other traffic types (e.g. eMBB), to avoid wasting resources, while still supporting MCC services with highest efficiency.

Having to occasionally transmit messages requiring ultra-high reliability with very low response times (use cases in the area of MCC; respective 3GPP terminology: URLLC, ultra-reliable low latency communications) while having at the same time a constant flow of ‘conventional’ MBB traffic to support, opens up the potential issue of inefficient use of the available resources. A straight-forward but inefficient solution is to persistently keep parts of the transmission band reserved for potential MCC messages. As they may or may not occur a reasonable wastage of resources is unavoidable. A more advanced way of treating this is discussed in section 2.4.3 of D4.2 [FAN17-D42]. Here, all spectral resources are available for non-latency critical transmissions. When a request to transmit a latency critical message occurs, the system is allowed to ‘hijack’ resources already being scheduled to carry ‘conventional’ traffic. This can be done by either puncture the less critical message (i.e. in DL not modulate the respective resource elements with the data related to the less critical message) or to allow the latency critical message to be superimposed using a much higher transmission power to ensure successful transmission (e.g. in UL by allowing both devices to transmit their data). As both mechanisms increase the chance for failed transmissions of the ‘conventional’ message, means to recuperate this have to be applied (see section 2.4.3 in D4.2 [FAN17-D42]). With allowing the system to make use of this mechanisms both the multi-service support is improved (design target B) and the overall system is more efficient (design target C).

13. Enable robust and scalable contention based access, to integrate MMC services with highest efficiency.

Robust and scalable contention based access requires both efficient protocol design, advanced code design and sophisticated PHY layer processing mechanisms to detect and separate the transmitted messages. FANTASTIC-5G has worked on all these items. Details are included in section 3.2 of D4.1 [FAN16-D41] and of D4.2 [FAN17-D42] and in related appendices. As the overall system is more efficient with serving the different services according to their respective needs and characteristics, these mechanisms address both the design targets B and C.

3.3 Topics and detailed proposals

3.3.1 Numerology and frame structure

In D3.1 [FAN16-D31] we have provided our proposals related to reasonable frame design choices for 5G taking the anticipated characteristics into account. We have elaborated how the frame design requires to be designed to allow for efficient multi-service support both for FDD and TDD. These designs have been holistic in the sense of including all relevant items such as control fields, data and reference symbols. We have introduced the so-called tiling concept including reasonable configurations for the basic setting such as resource block definitions, bandwidth and related TTI tuning. In this deliverable we will provide further details related to more specific aspects such as resource block grouping and multi-cell alignment for mixed numerologies, mini-slot structures and service specific TTI tuning.
3.3.1.1 Resource block grouping and multi-cell alignment for mixed numerologies

In principle there are several options to deal with several different numerologies inside the same carrier. The two extreme cases are:

- Each physical resource block (PRB) may have a different numerology. This provides maximum flexibility. However, this comes with high signalling overhead and non-negligible inter-carrier interference or potential guard band overhead. Furthermore, it has a higher implementation complexity. E.g., in the case of subband-filtered waveforms, the receive filter has to be adjusted to all possible combinations.
- The carrier is split into a contiguous portions per appearing numerology. This is very simple and has minimal guard band overhead. A disadvantage of this option is that neither using a particular numerology at different parts of the band for diversity nor fully exploiting frequency selective scheduling gains are supported with this solution.

In this section we discuss an adjustable alternative between the two extreme cases: Resource block grouping, which we also have called the tiling concept. The basic idea of the tiling concept has been addressed in previous FANTASTIC-5G deliverables: A physical resource block group (RBG) (“tile”) can be seen as a configurable larger resource group chunk with homogeneous numerology or other PHY/MAC parameters (such as TTI length, waveform parameters, such as active DFT-spreading, or tight/relaxed time-frequency alignment). Different RBGs may have different numerologies and parameters. This enables NR to provide a configurable air interface.

Since the RBGs are larger groups of resource blocks with identical numerology, they also help to further reduce residual inter-carrier interference as compared to having different numerology for each resource block. Further discussion of resource block properties including guard bands is done in [R1-167260].

Example RBG sizes are 720 kHz or 1440 kHz in frequency and 1 ms in time (this corresponds to 12 subcarriers and 14 symbols). Figure 3-1 shows an example portion of the band using the tiling concept.

![Figure 3-1: Basic illustration of the resource block grouping (tiling concept) for the uplink.](image)

The following set of rules apply:

- An RBG is defined as a time-frequency region of constant size
• The radio frame consists of consecutive RBGs in time- and frequency direction
• An RBG is characterized by the numerology parameters used, TTI duration, symbol duration and subcarrier spacing, number of subcarriers per resource block
• Derived parameters are: Number of symbols per TTI, number of TTIs per tile, number of subcarriers per tile and number of resource blocks per tile
• An RBG consists of consecutive PRB in time- and frequency direction
• A PRB is the smallest possible allocation unit
• RBG configuration could make use of profiles, sets of configuration parameters and features of the radio system. In future NR releases, new profiles can be added. Each 5G profile (containing a set of supported tile types etc.) can use its own types of tiles, i.e. a new profile can provide a new tile type appropriate for a certain purpose not known today.
• Two different RBG types; one with and one without CSI-RS
• RBGs with or without guard bands at its upper and/or lower end, as discussed in [R1-167260].

Resource blocks for different numerologies have the same overall number of resource elements, to simplify PRB usage w.r.t. coding performance and for LTE similarity. As depicted by Figure 3-1 and discussed in [R1-167260] this leads to different absolute PRB sizes in time and frequency (i.e. in ms and kHz).

The basic usage of RBG tiling simplifies also a 2-D scheduling problem with different TTI lengths: The scheduler pre-allocates RBGs according to service and user demands in a time structure given by the tile sizes, e.g. 1ms. Each RBG uses its own TTI length and numerology from a predefined set. The actual "1-D" scheduling is then carried out within the tiles. In case the pre-allocated URLLC tiles are not sufficient, a pre-emption is possible as discussed in [FAN16-D31].

Discussion of multi-cell coordination
There are several options for the level of multi-cell coordination of tiles which are available and its potential handling. In principle, the following coordination options (discussed in detail further below) are possible:

• No coordination
• Implicit coordination
• Partial coordination
• Full coordination

Different service requirements (e.g. low latency, high speed) lead to different parameterization of the time-frequency allocation of the user (e.g. different subcarrier spacings and symbol lengths, respectively). During the connection setup, a user informs the network about its service requirements. This set of requirements and the derived physical parameters are the basis of an implicit (without signaling exchange between transmission points) or explicit (with signaling exchange between TPs) coordination between neighboring TPs with the aim to minimize mutual interference between cells. The assignment of radio resources to the UEs is based on the coordination. As an example: UEs with certain numerology are preferably scheduled in a certain frequency sub-band. The principle applies for both uplink and downlink. Further details can be found in [R1-167261].

We suggest using at least partial numerology alignment with alignment priorities based on the required Tx/Rx processing within the tiles, as illustrated in the following table:
Table 3-1: Numerology alignment priority table for each multi-cell coordination technique

<table>
<thead>
<tr>
<th>Technique</th>
<th>Numerology Alignment priority</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blanking of resources</td>
<td>0 (none)</td>
<td></td>
</tr>
<tr>
<td>Single cell transmission</td>
<td>0 (none)</td>
<td></td>
</tr>
<tr>
<td>Interference rejection combining</td>
<td>1 (low)</td>
<td>Spatial suppression of interference also brings benefits in non-aligned numerology</td>
</tr>
<tr>
<td>(Multi-cell linear MMSE receive combining)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inter-cell interference cancellation</td>
<td>2 (high – should be aligned)</td>
<td>Frequency-domain based SIC techniques are only reasonable when the same numerology is used in the cells of interest. Time-domain IC techniques are possible without numerology alignment.</td>
</tr>
<tr>
<td>Coherent joint transmission</td>
<td>3 (highest - must be aligned)</td>
<td>Coherent combining of symbols across different cells demands same numerology</td>
</tr>
</tbody>
</table>

Multi-cell reference signal- and sounding design and alignment based on RBG

In the mixed numerology case the properties of pilot and sounding sequences across multiple cells when different numerologies are used in different cells is an issue which needs to be addressed. When certain well defined cross-correlation properties are anticipated, e.g. supported by Zadoff-Chu sequences, there needs to be a proper way of alignment between different cells.

An appealing solution is when CSI-RS are time multiplexed with control/data transmissions, cf. Figure 3-2. The numerology of CSI-RS is typically aligned among TPs, while the control/data/DMRS numerology is regularly chosen independently and is not aligned among TPs except when e.g. UE signals are transmitted by multiple TPs (DL CoMP) or to simplify SIC processing (as discussed above in the table above). Even within a single TP different control/data numerologies may be applied on the same time frequency resources, e.g. when different UEs are served simultaneously on the same resources using spatial multiplexing (MU-MIMO). The time multiplexing is motivated from the target of avoiding inter-(sub-)carrier interference, as CSI-RS and data numerologies may differ. Note: Especially when multiple CSI-RS numerologies are applied within a multi-TP system, there may exist boundaries between regions applying a specific CSI-RS numerology. Therefore, having TPs with different CSI-RS numerology on the very same time-frequency resource, although not preferred, may still occur, but with the drawback of hampering inter-TP channel measurements.
3.3.1.2 Mini-slot structure for 5G

In mini-slot structure, there are less OFDM symbols than in ordinary slot/sub-frame. There are two main motivations for the use of mini-slots: latency reduction and unlicensed band operation.

The length of an ordinary slot is 7 or 14 OFDM symbols. Therefore, useful lengths of mini-slot could be between 1 and 6 symbols. It is required that there are appropriate control symbols in the mini-slot. The use of mini-slots for latency reduction is especially important for 15 kHz subcarrier spacing, due to longer OFDM symbols.

In unlicensed system, it is beneficial for the user to occupy the channel as soon as possible, since there may be other communication systems (such as Wi-Fi) competing for the same resources. The use of mini-slots helps in this, since it provides more frequent transmission opportunities.

One scalable mini-slot design is preferred. The same structure should be scalable for the following situations: different slot lengths, duplexing schemes (FDD/TDD), different GP lengths in the case of TDD, different CP lengths, and for both licensed and unlicensed operation.

Furthermore, the mini-slot should not exceed slot/subframe boundaries (since it would create scheduling dependencies between slots/subframes), and FDM can be used for data/control multiplexing (unlike in the case of ordinary slot).

Half-duplex constraint will set boundary conditions for mini-slot allocations in TDD. This means that when a regular slot is in DL phase, also mini-slot needs be in the DL phase, and when regular slot is in UL phase, mini-slot needs to be in UL phase, respectively. Supporting DL-UL switching within a min-slot would result in additional non-negligible overhead.

The length of the mini-slot should be variable and the starting position flexible. This approach supports all foreseeable scenarios, and supports forward compatibility.

Figure 3-3 shows three options how mini-slot can be arranged. We give examples to two slots: DL only slot (left) and DL slot with bi-directional control (right). In the first option, the mini-slot has fixed length and fixed starting position. Here the length of mini-slot is one symbol, which results in high control and DMRS overhead. In option 2, the mini-slot length is two symbols. Option 3 is most flexible, since both mini-slot length and starting position are variable, but control signalling is needed to inform about the mini-slot lengths.

Figure 3-2: Numerology examples and multiplexing option of CSI-RS (downlink) or sounding RS (uplink) for multiple cells.
3.3.1.3 TTI tuning for multi-service

The diverse requirements of 5G use cases impose a need for a flexible overall communications stack in terms of latency and reliability. On the PHY layer, the latency of delivering a message is determined by a combination of factors such as available bandwidth, channel acquisition delay, processing delay etc. A key parameter of the PHY layer affecting the overall latency is the Transmission Time Interval (TTI). In LTE systems, the typical TTI is fixed to 1 ms regardless of the latency requirement of the service. For 5G, obviously there is a need to tune the TTI to fit into the total latency budget. There are two main approaches for tuning the TTI [IX16]:

a) Self-contained frame: Keeping the number of symbols per TTI fixed and modifying the symbol duration.

b) Symbol-wise frame: Keeping the symbol duration fixed and tuning the number of symbols per TTI.

\[
C = \log(1 + \frac{(1 - \epsilon)\gamma_0}{1 + \epsilon\gamma_0})
\]
Where $\varepsilon$ is the averaged MSE over all data symbols and $\gamma_0$ is the nominal SNR. Figure 3-5 shows the resulting theoretical capacity under different channel conditions. For the simulations, we assume a negative exponential power delay profile channel with RMS delay of 0.1 $\mu$s and nominal SNRs of 5, 10, and 15 dB. We consider a TTI duration of 250 $\mu$s for the self-contained frame structure and 213 $\mu$s TTI (3 OFDM symbols) for the symbol-wise frame structure. As shown on the left plot, the self-contained frame structure has a higher capacity up to a certain Doppler shift after which the symbol wise frame starts offering a higher capacity. On the right plots, the capacities of the frame structures are plotted versus SNR under low and high mobility. As shown, the symbol-wise frame structure is more robust to mobility; however, in low mobility situations, the self-contained frame structure seems to offer a near linear increase in capacity as the SNR increases. One can conclude from the results that for scenarios with high Doppler shifts (V2X, mmW) using the subcarrier spacing tuning approach with preamble reference symbol will not be suitable. Hence, some sort of modification of the pilot pattern should be introduced. If not possible, then the LTE-based approach of symbol-wise tuning should be adopted. On the other hand, using a preamble pilot reduces the complexity of the receiver which shortens the processing delay and the overall latency.

Figure 3-5: (Left) Capacity drop as Doppler shift increases for different SNRs. (Right) Capacity increase as SNR increases for different TTI durations.

### 3.3.2 Physical Downlink control channel

Due to diverse performance requirements of the 5G services, optimal radio resource allocation and scheduling is one of the most challenging tasks in the 5G radio system design. Similar to LTE physical downlink control channel (PDCCH), the NR-PDCCH has been agreed in [RAN1-86b] to perform downlink data scheduling and uplink data assignment. This means all downlink control information (DCI) relevant to data scheduling, such as parameters used for radio resource allocation, link adaptation, hybrid ARQ and advanced MIMO operation etc, are conveyed by NR-PDCCH.

After establishing a radio resource control (RRC) connection with serving base station, a UE will be configured with some search space (SS) of PDCCHs, from which UE monitors possibly scheduled PDCCH in every transmission time interval (TTI). Typically, PDCCH SS includes a number of PDCCH candidates to be potentially transmitted in the current TTI, UE performs blind decoding in each TTI for all PDCCH candidates in the SS. If a PDCCH candidate is correctly decoded by e.g., a successful CRC decoding, the UE shall continue to receive the data channel scheduled by the PDCCH, otherwise, UE can potentially enter a “micro-sleep” mode to save power consumption. SS design is clearly an important topic for PDCCH design, it has direct impact on the control channel capacity, spectrum and overall system operation efficiency.
To achieve stable system operation, it is very important to ensure the reliable reception of control channel. The failure of control channel reception leads UE to lose the scheduled data packet in the current TTI, furthermore, consecutive failures of certain number of PDCCH receptions will trigger the radio link failure procedure for the UE to perform cell-reselection which is quite costly procedure and considerably deteriorates the quality of user experience. As such, transmit schemes, e.g., transmit diversity techniques, to ensure reliable reception is of very importance for PDCCH design as well. As a result, we focus on the efficient NR-PDCCH SS design as well as transmit diversity schemes in this section.

The section is organized as follows. In Section 3.3.2.1 basic control channel structure and search space design are presented. In Section 3.3.2.2 the transmit diversity schemes as well as enhanced high order transmit diversity technique are detailed. In Section 3.3.2.3 simulation results of proposed transmit schemes are given. Finally, section 3.3.2.4 concludes the section.

3.3.2.1 Control channel structure and search space

In this section, basic control channel structure and search space design are addressed.

**Basic control channel structure**

In 3GPP NR [RAN1-86b], resource element group (REG), which is comprised of 12 consecutive resource elements (RE) located in an OFDM symbol, has been adopted as the basic control channel building block. Moreover, like in LTE, control channel entity (CCE), consisting of a number of REGs, is defined as smallest unit of a scheduled PDCCH transmission, i.e., the granularity of the PDCCH link adaptation. As such, each NR-PDCCH is transmitted by using one or several CCEs depending on channel quality estimation of the respective UE at the base station. The number of CCEs employed by a PDCCH is called aggregation level (AL) of the PDCCH. Currently, similar to LTE, several ALs, namely 1, 2, 4, 8 or even 16 and 32 are considered by 3GPP NR study. A particular AL is designed to meet certain coverage requirement. It is obvious that PDCCH with low AL will be used for UEs with high SNR, and high AL for UEs with low SNR.

![Figure 3-6: REG with different DMRS patterns, DMRS overhead in the REG: (a) 1/3, (b) 1/3, (c) 1/2.](image)

To enable channel estimation for coherent demodulation, some REs in the REG are allocated for the demodulation reference signal (DMRS). The possible DMRS placements are depicted in Figure 3-6. Each column in Figure 3-6 represents a REG of 12 REs. The REs in green stands for the DMRS associated with antenna port 0, and the REs in yellow for DMRS with antenna port 1. As illustrated in Figure 3-6 (a)-(c), three DMRS patterns with different RS overhead can be considered. DMRS pattern with large density, e.g., as shown in Figure 3-6-(c), can be employed for channels with large frequency selectivity or in low SNR situation. On the contrary, DMRS pattern with less density, e.g., shown in Figure 3-6-(a) and (b), is more suitable for high SNR and/or less frequency selective channels. It is envisioned that REG with configurable DMRS pattern can provide additional trade-off between channel estimation performance and effective coding gain obtained from the usable REs for control symbol allocation.
Control channel radio resources

In LTE, radio resources allocated for control channels are explicitly signalled to UEs in the cell. For ordinary eMBB services, two types of control channels, namely PDCCH and enhanced PDCCH (EPDCCH) [3GPP-36.211], are supported in LTE. The control resource region used for PDCCH spans the first several OFDM symbols in each TTI of 1ms, the number of which is signalled by a dedicated cell-specific physical channel on a TTI basis. As a result, the PDCCH region in a TTI can be dynamically changed according to the actual control resource demand in the TTI. The radio resources for EPDCCH, which corresponds to a set of resource blocks (RB), defined as a resource grid of 12 subcarriers in frequency and 14 OFDM symbols in time [3GPP-36.211] corresponding to one TTI or subframe of 1ms, is configured by a UE-specific high layer signaling.

In 3GPP NR, control resources shall be configured in a more flexible manner, and aim to achieve the advantages of both PDCCH and EPDCCH while mitigating the limitations of them. As a consequence, it is decided that control resources of NR-PDCCH shall include a set of RBs in frequency, and the time duration in terms of number of control symbols and occurring periodicity of which can be configurable. Such configurable TTI of control channel enables various scheduling time granularities which depend on the respective service requirement. For instance, since the TTI of URLLC service should be much smaller than that of eMBB service, it is plausible that control resource TTI of two services shall be different.

Search space design

As mentioned earlier, each UE shall be configured with a control channel SS defined over the control resources described above, in which all the potential PDCCH candidates are defined. To support link adaptation, the SS include PDCCH candidates with different ALs. As a result, the SS can be viewed as a union of several sub-search spaces (sub-SS), each of which is comprised of a number of respective PDCCH candidates at a certain AL. For example, the SS can contain four sub-SSes, corresponding to ALs 1, 2, 4 and 8, respectively. Such search space principle has already been specified in LTE, and it continues to be employed in 3GPP NR. The dimension of the SS, in terms of number of PDCCH candidates, should be designed properly by taking into account several interacted factors, namely, targeted control channel blocking probability, control resource utilization and UE blind decoding complexity. It is clear that the larger the dimension of SS is, the less control channel blocking probability and control resource utilization, and the larger UE blind decoding complexity. As such, SS design needs to achieve a good trade-off among these different aspects. Ideally, this could be formulated as an optimization problem, the solution to which can be obtained to meet the desired optimization criteria. However, this is beyond the scope of this section.

In addition to different coding gains obtained from various ALs, receive performance of NR-PDCCH can be also improved by other advanced techniques depending on the availability of channel state information (CSI) of UE at the base station. Provided that the base station has the knowledge about UE channel condition over the system bandwidth, it is advantageous that base station can transmit the NR-PDCCH in the frequency part where UE has good channel quality, to achieve frequency selective scheduling gain. Moreover, if base station has the UE CSI feedback about the preferred beamforming/precoding vector in certain frequency part, beamforming gain can be further achieved when the preferred beamforming operation is applied to the scheduled NR-PDCCH.

Due to the fact that the accurate CSI may not be available at base station for some UEs in the cell, NR-PDCCH SS design needs to take these situations into account as well. It is plausible that NR-PDCCH should support both localized and distributed transmissions. By localized transmission, the NR-PDCCH is transmitted in a frequency localized manner, this would enable NR-PDCCH to benefit from frequency selective scheduling gain and better beamforming/precoding gain if possible. On the other hand, distributed NR-PDCCH transmission in frequency is able to provide more frequency diversity for the scheduled NR-PDCCH. It is natural that localized transmission is more suitable for the situation where base...
station has more accurate CSI of the UE, while in the absence of sufficient CSI, it is more advantageous to apply distributed transmission for NR-PDCCH.

Figure 3-7: Localized search space of NR-PDCCH with 4 ALs: 8 AL1 candidates, 4 AL2 candidates, 2 AL4 candidates and 1 AL8 candidate. Control resource set: 16 RBs, 2 OFDM symbols.

Figure 3-8: Distributed search space of NR-PDCCH with 4 ALs: 8 AL1 candidates, 4 AL2 candidates, 2 AL4 candidates and 1 AL8 candidate. Control resource set: 16 RBs, 2 OFDM symbols. Each smallest block represents a REG.

Two examples of localized and distributed NR-PDCCH SSes are illustrated in Figure 3-7 and Figure 3-8, respectively. It is shown from the figures that control resource set is comprised of 16 RBs, each of which includes 12 subcarriers and 2 control OFDM symbols, i.e., 2 REGs. Four sub-SSes, corresponding to four ALs, namely 1, 2, 4 and 8, are defined over the control resource set. With the definition of CCE of 4 REGs in these two examples, there exist 8 AL1 candidates shown in the 1st column of the respective figure, 4 AL2 candidates in the 2nd column, 2 AL4 candidates in the 3rd column, and 1 AL8 candidate in the 4th column defined in the SS. As shown in Figure 3-7, each NR-PDCCH candidate in same color is allocated in frequency localized manner so as to enable frequency selective scheduling and precoding bundling over CCEs allocated in consecutive RBs. In contrast to the localized NR-PDCCH, each NR-PDCCH candidate in same color illustrated in Figure 3-8 is transmitted in several RBs evenly distributed over the control resource set to achieve maximum available frequency diversity. To better understand the performance differences, these two types of transmission schemes are evaluated and compared in Section 3.3.2.2.

3.3.2.2 Transmit diversity schemes

In addition to the adjustable coding gain by virtue of different ALs and trade-off between frequency diversity and selective scheduling gain which can be realized by two transmission
schemes detailed above, transmit diversity techniques are also very important for improving the NR-PDCCH reception performance so as to increase the overall control channel capacity and spectrum efficiency.

Currently two transmit diversity schemes, namely spatial-frequency block codes (SFBC) and per-RE precoder cycling have been adopted in LTE for PDCCH and EPDCCH [3GPP-36.211], respectively. From achievable transmit diversity point of view, based on the well-known principle of space-time block codes (STBC), SFBC can achieve the maximum transmit diversity. Since STBC/SFBC always requires a pair of input transmit symbols to create the output coded block, this constraint causes some difficulty when applying STBC/SFBC to the source data stream with odd/uncertain number of input symbols. Due to the rate matching around those legacy LTE signals, EPDCCH at a given aggregation level, has an uncertain number of available REs in a RB. This makes it hard to apply SFBC to EPDCCH. As such, per-RE precoder cycling, whereby REs in a REG are alternatively associated with one of two DMRS antenna ports, has been adopted as the transmit diversity scheme for EPDCCH. It is clear that STBC/SFBC can achieve better transmit diversity than per-RE precoder cycling. For example, in case of 2 transmit antennas, STBC/SFBC can achieve transmit diversity of order 2 for each allocated RE, on the contrary each RE is transmitted from only one antenna port by per-RE precoder cycling scheme. As a result, from each RE standpoint, per-RE precoder cycling does not provide additional transmit diversity although the overall transmit diversity of scheduled EPDCCH is 2.

In LTE, PDCCH supports transmit diversity with four transmit antenna ports by virtue of combination of SFBC and frequency switched transmit diversity [3GPP-36.211], i.e., two pairs of REs in a LTE REG are transmitted from 2 different transmit antenna pairs. To achieve same maximum order of transmit diversity, it is envisioned that NR-PDCCH should also support transmit diversity schemes with 4 tx antenna ports. In light of the principle of LTE high order transmit diversity scheme, transmit antenna pair cycling can be also applied to increase the maximum transmit diversity order to 4. Specifically, two types of transmit antenna pair cycling, based on which level the antenna pair cycling is performed, namely per-REG, and per-RB, can be considered. They are detailed as follows.

![Figure 3-9: per-REG transmit antenna pair cycling, an AL1 NR-PDCCH of 4 REGs. 4 DMRS ports (coloured REs) associated with 4 transmit antenna ports: 0, 1, 2 and 3.](image)

The per-REG transmit antenna pair cycling is shown in Figure 3-9. As illustrated in Figure 3-9, a NR-PDCCH is comprised of 4 REGs, e.g., AL1 candidate, and allocated in two RBs. It is shown that each REG of the NR-PDCCH contains two DMRS ports associated with antenna port pairs (0, 1) and (2, 3) alternatively. As a consequence, the NR-PDCCH are transmitted from 4 antenna ports.
The per-RB transmit antenna pair cycling is shown in Figure 3-10. As illustrated in Figure 3-10, an AL1 NR-PDCCH is comprised of 4 REGs and allocated in two RBs. Specifically, two REGs of the NR-PDCCH are allocated in each RB. It is shown from the figure that only the first REG of the NR-PDCCH in the allocated RB contains DMRS. REGs in the first RB, i.e., RB #1 in Figure 3-10, are transmitted from antenna port pair (0, 1) and REGs in the second RB, i.e., RB #2 in Figure 3-10, from antenna port pair (2, 3).

It is clear that per-REG transmit antenna pair cycling provides more diversity than per-RB transmit antenna pair cycling. However, per-REG transmit antenna pair cycling has more DMRS overhead. Therefore, these two schemes furnish the trade-off between required diversity gain and effective coding gain depending on the number of REs available for transmitting NR-PDCCH symbols. Simulation results can be found in Appendix 0.

3.3.2.3 Multi-stage physical downlink control channel

In the previous section the important role of the NR-PDCCH has been outlined as well as one option of implementation regarding preconfigured search spaces for PDCCH reception by the UE. However, such predefined search spaces impose certain constraints on the network, the BS scheduler and also on the UE. The configuration of user specific search spaces introduces further signaling overhead in the network, why EPDCCH configurations are typically kept semi-static. Accordingly, the resource allocation in the BS scheduler can be restricted in its flexibility to allocate DCIs properly according to the frequency selective criterion, to support multiple numerologies or fragmented spectrum. Last, the configured SS for a UE can be subject to control channel blocking when SS resources are shared among UEs which can be very sensitive for low latency traffic. Significant control channel blocking probability is known to be present with LTE PDCCH and to a minor degree also with LTE EPDCCH. This section focuses on an alternative implementation of a user-specific control channel design which addresses the aforementioned disadvantages of the solution with search spaces.

This section introduces the design of the multi-stage control. Then in Appendix section 6.6 we present some system level simulations results to show the performance of the new scheme vs LTE EPDCCH. In section 3.3.2.4 we conclude with some final remarks.

The goal of the multi-stage PDCCH (MS-PDCCH) for NR is two-fold: alleviate the need for control search spaces and the need to trade-off between allocation flexibility and control overhead. With the MS-PDCCH, a BS scheduler can just allocate resources for control in the same way as it allocates data resources and it can allocate both arbitrarily and not restricted by a control search space or resource allocation type as in LTE. Nonetheless, UEs may be configured with specific (per RNTI) search spaces in frequency and time. The most likely use is of such search spaces is two-fold: 1) for low end devices that cannot receive and search the full carrier bandwidth in every TTI for DCIs, and 2) to exclude spectrum and/or (sub)frames that are used...
by other services, slices, numerologies and/or RATs and thereby avoid UEs pointless search for DCIs therein. Furthermore, MS-PDDCH employs a user-specific design: control and user data are multiplexed individually per UE – which is also baseline for 3GPP NR and in contrast to LTE EPDCCH that multiplexes control of multiple UEs. As a result, a single allocation per UE carries both control and data. Advantages of the overall design are that it avoids the blocking problem of search-space restricted control allocations and it allows full flexibility to the BS scheduler. Furthermore, it does not create interdependencies among the PDCCH designs of different UEs, which benefits the future evolution of the air interface as it alleviates the backward compatibility issues incurred by such interdependencies. The drawbacks are that at least one full PRB is allocated per UE, also for control only transmissions, that some additional information need to be included into a DCI to realize the allocation flexibility and last not least that the pilot overhead needs to increase to limit a UE’s blind decoding effort even for an unlimited respectively non-existing search space. However, the simulation results presented in Appendix section 6.6 show that for a typical setting these drawbacks are overcompensated by its benefits.

Figure 3-11 shows the structure of the MS-PDCCH and its three stages. The consequence of a SS-less allocation, without any countermeasures, would be that a UE would have to do blind decoding attempts on the whole supported bandwidth in order to find a potentially sent DCI in every TTI. This would impose very high processing effort on the UE. To overcome such high computational effort at the UE side, the first stage of the MS-PDCCH, the so called user identification signal (UIDS), is introduced to down-select the blind decoding candidate resources that potentially carry a DCI for the UE. The UIDS is a unique pilot pattern of the UE that is used to do simple correlation based matching of potential MS-PDCCH resources with the signals received by the UE right at the antenna without any further receiver processing of the signal [FAN16-D31]. A search for the UIDS is far less computational intensive than a blind decoding attempt of the full DCI. The false alarm probability of the UIDS is expected to be adapted such that the overall effort to find a potential DCI will not exceed that of conventional SS-limited DCI allocation. Therefore, the UIDS supports different aggregation levels to adapt to the UE’s link quality. For the smallest aggregation level, i.e. for UEs with good link quality, the UIDS can be allocated to e.g. one OFDM symbol within one PRB. Higher aggregation levels are supported by repeated allocation of the UIDS either in frequency domain, time domain or both.

Figure 3-11: Multi stage PDCCH structure

The second stage is constituted by the basic control forward error correction (FEC) block. Basic control is collocated with UIDS, reusing it for channel estimation. After the UE identifies a PRB as a potential MS-PDCCH resource candidate (following correlation of the received UIDS) it
performs blind decoding on the associated resource elements (RE) carrying basic control. Basic control supports a limited number of modulations and coding schemes (MCS), corresponding to aggregation levels, to adapt to UE’s link quality. From knowledge of the link quality on the respective resources, the UE can choose a subset of reasonable basic control MCSs for blind decoding.

Depending on the number of bits required to transmit a DCI, basic control may contain information on the location and MCS of further PDCCH resources belonging to the same DCI, represented with the last stage, the main control FEC block. When the presence of main control is indicated in basic control, the UE also decodes main control and proves correct DCI reception by calculating jointly over both decoded basic and main control the CRC and adding modulo-2 the UE’s radio network temporary identifier (RNTI). In case of a blind decoding attempt due to a false alarm (FA) detection on the UIDS or in case either one of the basic or main control FEC block been decoded erroneously, the CRC check for the DCI will fail. Missed detection (MD) of the UIDS as well as erroneously decoded basic or main control will cause a control channel error and thereby loss of a complete downlink transmission, whereas a FA only causes processing overhead at the UE. The design principle is therefore to keep UIDS MD rates and thereby PDSCH resource wastage respectively spectral efficiency loss caused by non-detected DCIs below an acceptable level, typically below 1%. At the same time, false alarms and thereby blind decoding attempts should be at a reasonable FA rate of about 10%, to downselect the PRBs in which to attempt blind decoding by one order of magnitude compared to without UIDS. A UE should attempt blind decoding in order of decreasing UIDS correlation. The UE stops after it has found (i.e. successfully decoded) the (configurable) maximum number of DCIs per each of its configured RNTIs (LTE: max. 1 DCI per RNTI per TTI) or when the (device class specific) maximum number of supported blind decoding attempts has been reached. MD and FA are linked together by the choice of the decision threshold for UIDS detection. Selecting the threshold value for UIDS detection in favour of low MD rates will inevitably increase FA rates [FAN16-D31].

There are two types of allocations of the MS-PDCCH depending on the type of channel quality information (CQI) that is available to the BS scheduler shown in Figure 3-12. The scheduler can allocate the MS-PDCCH in a frequency diverse (distributed) manner when it has only wideband CQI available or it can allocate the MS-PDCCH in a frequency selective (localized) manner, when it has more accurate CQI available.

Figure 3-12: Different allocation strategies of the MS-PDCCH, frequency diverse (left) and frequency selective (right)
Link adaptation of the MS-PDCCH is done with different aggregation levels (AL) for every stage separately as illustrated in Figure 3-13. Increasing UIDS AL is done by repetition of the UIDS in frequency or time dimension or both. ALs of basic and main control elements is done via different MCS. Hence, it is very important to select the MCSs and the correct block sizes for the basic control element such that the block error rate (BLER) of the basic control element is aligned with MD rate for the UIDS, i.e. the UIDS MD rate for the SINR at 1 % target BLER of the chosen MCS for the basic control element should be also at least 1 % or less. For detection efficiency reasons, as the UIDS has a minimum size and it determines the lower performance limit, several basic control MCSs can be associated with one UIDS AL to benefit in the higher SINR region. The simulation performance is reported in Appendix section 6.6.

Figure 3-13: Different aggregation levels for localized multi stage PDCCH, UIDS AL=1 (top) and UIDS AL=2 (bottom)

3.3.2.4 Conclusions

This section presents detailed solutions to NR-PDCCH design in regard to search spaces with different transmission schemes and transmit diversity approaches. In particular, localized and distribution transmission based search spaces are provided, and performance gain of distributed transmission compared to localized transmission is demonstrated by simulation results. Moreover, SFBC and per-RE precoder cycling based transmit diversity schemes are compared. Based on performance comparison, SFBC seems to be a preferred scheme. Finally, high order transmit diversity methods based on antenna pair cycling are also proposed. The considerable performance gain exhibited from simulation results clearly motivates the support of high order transmit diversity.
On the other hand, we have presented the concept of a user-specific multi-stage PDCCH. The proposed MS-PDCCH consists of three different stages: The UIDS pilot sequence and the fixed-sized basic control and the variable-sized main control FEC block, respectively. The UIDS drastically reduces UEs’ blind decoding effort, alleviating the need to introduce PDCCH search spaces. Two FEC blocks instead of one allow a variable-sized DCI that individually adapts to the signalled user data allocation size, providing full scheduling flexibility (per single PRB) especially for large frequency-selective allocations with spatial multiplexing as well as low control overhead in case of small or simple wideband single layer allocations. Performance comparisons with LTE EPDCCH demonstrate the advantages of the MS-PDCCH design: It avoids the EPDCCH blocking problem, showing higher system throughput due to lower blocking probability and control overhead as well as due to improved frequency selectivity of both MS-PDCCH itself and the user data allocation. Last not least it avoids scheduler complexity normally incurred by search spaces and, in conjunction with its user-specific design which multiplexes control information and user data UE-individually, provides greater overall system flexibility and thereby improved future proofness (e.g. no need to maintain PDCCH backwards compatibility between releases, integration of new PHYs within the radio carrier bandwidth not constrained by PDCCH search spaces).

3.3.3 Reference signal design

Reference signals have a variety of uses in wireless communication systems. In D3.1 [FAN16-D31] we have proposed a preamble based design, which can be used for user identification and symbol demodulation. Those reference symbols are user specific and confined to the respective allocation. In this deliverable, we present our work related to reference signal design for massive MIMO. Those common (in contrast to user specific) reference symbols are used for channel estimation and beam selection.

3.3.3.1 Reference signal design for massive MIMO

Massive antenna arrays are one of the key 5G components. In D4.2 in a typical example there are 32 antenna ports per cell, e.g., based on a grid of beams concept to support multi-user MIMO with ten or more simultaneously active UEs per cell. In case of coordinated multipoint transmission (CoMP) or joint transmission CoMP (JT CoMP) there will be multiples of 32 antenna ports from the multiple cells. With the increase in the number of antenna ports, the pilot overhead becomes an important issue.

Typically, the pilot overhead is also proportional to the maximum UE speed as well as the rms delay spread of the radio channel defining the coherence time and frequency and accordingly the required grid density of CSI RSs. For high speed UEs up to 250 km/h LTE provides a dense scattered grid of 8 out of 168 resource elements per physical resource block (PRB) for the common reference signals (CRS). The resulting pilot overhead of about 4.7% per antenna port (AP) scales under the assumption of orthogonal pilots from one AP to, e.g., 32 APs to infeasible 150%. Note, in new radio there will be no CRS anymore, but from a physical layer point of view the CSI RS will have a similar dense grid with an accordingly similar overhead.

For that reason, for new radio the strategy is to avoid any always on CRS and instead to provide different CSI RSs densities in different frequency subbands with suitably adapted RS grid densities.

Still there are some obvious drawbacks as the concept will limit MU scheduling gains, needs a high reconfiguration effort and requires UE specific information about channel characteristics. In addition, for high speed UEs the number of supported APs has to be low.

In addition, there are some not so obvious drawbacks as subband specific CSI RSs will make channel measurements over the full channel component bandwidth more difficult. Fisher information increases with larger bandwidth, providing accordingly higher channel estimation and prediction accuracy, and would be especially useful for massive MIMO and JT CoMP.
3.3.3.1.1 Sub-tiling and coded CSI-RS

The novel concept of sub tiling is proposed here as it allows for a dense grid of pilots for a high number of APs, while keeping the overhead close to that of a single antenna port.

The key idea of sub tiling is illustrated in Figure 3-14. On the left, there is the typical frequency and time grid of CRS signals density for one single antenna port AP1, leading to an overhead of about 4.7% per PRB. The given density covers the ‘worst case’ time and frequency variation of all possible radio channels.

As we want to serve fast moving and frequency selective UEs from the same pilot grid, the sub tiled grid has to be kept dense in time and frequency. But, as illustrated in Figure 3-14right - for sub tiling each AP uses only a fraction of a resource element (RE) and only a fraction of the power of that RE. As an example, in case of 16 APs only 1/16th of the RE might be utilized by each sub tiled AP specific CSI RS and thereby the total overhead for 16 APs remains the same as in the case of a single conventional AP.

This comes not for free and reduction in the resource utilization means degradation in the channel estimation quality. However, the UEs are typically either time variant (high speed UEs in a motor high way) or frequency selective (nomadic or slow moving UEs within urban NLOS areas), that is, the high speed UEs have typically flat frequency response and slow moving UEs have low time variance. Therefore, the fast or slow moving UEs can combine for example 16 sub tiled pilots per AP over frequency or time, direction respectively. That way UEs can get a dense CSI estimation in the converse domain, e.g., time or frequency.

It is important to note that the sub tiled CSI RSs are cell specific and constantly transmitted - typically - over the full frequency bandwidth or a suitably defined large subband. UEs adapt to different channel conditions by choosing the optimum combining strategy, i.e., either in time or in frequency domain (or in a certain combination of time and frequency).

The sub tiling concept has the following advantages: The overhead for 16, 32 or even more APs remains low or moderate in the range of some 5%, similar as that for a CRS for a single AP. There is a common pilot pattern for the entire cell and the UEs can individually decide, based on their channel conditions, to combine the pilot across time or frequency or both. Furthermore, this allows the UEs to utilize the full bandwidth for channel estimation and it avoids the sub-band specific reconfiguration of pilot pattern.

For implementation of the sub tiling concept different options can be considered, but most simple might be the suitable allocation of Walsh Hadamard (WH) codes to pilot resource elements. Each code relates to one AP and the WH codes are allocated in matrix form so that full inter code (inter AP) orthogonality can be achieved either in time or in frequency domain. Note other codes with similar properties fulfilling other side constraints like constant amplitude etc. can be considered as well.
Sub tiling provides a certain elasticity with respect to UE mobility and frequency selectivity, but in case of a grid of beam (GoB) concept with 32 beams per cell and cooperation areas of, e.g., nine cells one would need ideally a number of $N = 288$ orthogonal CSI RSs. The according length of the sub tiling codes would violate the typical coherence time and/or frequency.

Coded CSI RS - as explained in more detail in [ZSB16] - exploit the spatial structure generated by the beamformers of the GoB, which results in sparse sets of relevant channel components per UE, where each UE receives a different set of sparse channel components depending on its location. Properly designed non-orthogonal Vandermonde like sequences of length $K$ are mapped to the $N = 288$ APs. The rank of the UE individual channel matrices is in this case equal to $K$ so that in principle each UE can estimate its $K << N$ sparse subset of relevant channel components by a Moore Penrose inversion. So, the non-orthogonal coded CSI RS sequences reduce the CSI RS overhead by $K/N$, which can be small in case of high sparsity levels and large $N$.

3.3.3.1.2 Performance analysis

One of the main limitations of sub-tiling as well as coded CSI RS is the inter-code interference generated in case of time or frequency variations within the sequence length. For sub-tiling the sequence length is that of the WH code, while for coded CSI RSs it is that of the non-orthogonal Vandermonde like sequences. Sub-tiling requires a scattered grid similar to that of the CRSs, while coded CSI RSs are assumed to be sparse in time as they should serve mainly nomadic users with low mobility. For that reason, a clustered allocation over certain adjacent time and frequency bins is preferable.

To get a better understanding of the inter code interference a measured radio channel in a NLOS urban macro scenario has been evaluated firstly assuming a single UE antenna and secondly for a virtually beamformed radio channel, generating an antenna aperture of about 3 degree for the given carrier frequency of 2.6 GHz. In Figure 3-15 the typical crosstalk $XT$ for ten different subcarriers is plotted over 50 UE movements. For a single UE antenna, there are significant $XT$ variations due to the strong channel fluctuations over time. With the virtual UE beamformer these fluctuations are reduced and accordingly the mean crosstalk $XT$ is lowered to $< -30$dB, being small compared to the single UE omni antenna case with $XT$ close to $-20$dB.

As already mentioned for coded CSI RSs a clustered allocation of the non-orthogonal sequences is beneficial. In that case one could place length $K=14$ sequences within one TTI of 1ms. That
way a crosstalk below -20 to -30dB could be achieved even for the single antenna case. For larger sequence length one could concatenate 2 or 4 length 14 sequences on adjacent subcarriers so that an overall length of \( K = \text{degree} 4 \times 14 = 56 \) can be achieved with low crosstalk. In case of more narrow UE beamformers even higher sequence length or higher mobility can be supported.

The coded CSI RSs have been applied to the system level analysis of the interference mitigation framework as explained in D4.2 with a cooperation area of nine cells and 32 beams per cell or accordingly 288 beams overall. The cooperation area scheduled simultaneously 90 out of 172 UEs per PRB which resulted for ideal CSI in a gross spectral efficiency of 47 bit/s/Hz/cell. With channel estimation for the 288 beams based on coded CSI RSs with a sequence length of \( K = 64 \) this spectral efficiency reduced to 40.6 bit/s/Hz/cell or by about 8.5%. This is quite a good result considering that 64 CSI RSs every 5ms results in an overhead of only 7.5%, which might have to be doubled for channel re-estimation.

Note, these simulation results are preliminary providing a first high level understanding of the potential of these concepts, but more analysis is ongoing and needed.

Figure 3-15: Crosstalk for UE moving with 3 kmh and a grid density of 4 CSI RS per TTI of 1ms and a sequence length of 16. Left: virtual beamformer over 32cm = 3 \( \lambda \); right: single UE antenna

### 3.3.3.2 Conclusions

With sub tiling and coded CSI RS two novel concepts for channel estimation have been elaborated leading to a significantly different solution compared to concepts known from LTE. There are some real benefits as the high elasticity with respect to UE mobility for sub-tiling. More importantly coded CSI RS allow estimation of high number of, e.g., \( N = 288 \) of relevant channel components with low to moderate overhead of only 10% together with high estimation accuracy (proven for example by the high spectral efficiency for the nine cells scenario).

For implementation, the allocation of CSI RSs to resource elements is very important and leads to novel design targets as one has to ensure sufficiently low inter-code interference. Accordingly, for coded CSI reference signals a clustered allocation of length \( K \) sequences is beneficial or even mandatory.

Note, the concept of relevant channel components assumes a two-step approach for the channel estimation. Semi-statically, e.g., every 100 to 500ms a fully orthogonal set of CSI RSs for all \( N=288 \) APs is being provided so that each UE can identify its individual set of relevant channel components plus report it to the gNB. Periodically, e.g., every 5 to 10ms non-orthogonal CSI RS of shortened length \( K/N \) are being transmitted so that UEs can follow the ‘small scale’ channel variations for their individual sparse sets of relevant channel components.

In addition to the common reference symbols being presented here, the wireless system requires further user specific reference symbols for user identification and demodulation. For this we have presented in D3.1 [FAN16-D31] a preamble based concept combining both functionalities. While this work has targeted the implementation of user specific control channels, the reference symbols can be used for the demodulation of the collocated data channel as well.
3.3.4 HARQ

In this subsection, the main design principle for an HARQ concept able to cope with the identified flexibility requirements is presented. Further enhancement for improving latency or resource utilization is also presented.

3.3.4.1 General design principles

HARQ is adopted in cellular standards such as HSPA and LTE, and both chase combining (CC) and incremental redundancy (IR) options are considered. Today’s LTE standard relies on HARQ with a given number of parallel stop-and-wait (SAW) channels, assuming synchronous HARQ operation for the uplink, and asynchronous operation for the downlink. In the synchronous option, the HARQ processes are always transmitted at fixed timings, while in the asynchronous option they can be transmitted without any predefined order, thus enabling a larger flexibility. Obviously, asynchronous HARQ comes at the expense of an increased overhead since the HARQ process information should be signalled, while in the synchronous case this is implicitly known. A simple Boolean HARQ feedback from the receiver to the transmitter, carrying the positive acknowledgement (ACK), or negative acknowledgement (NACK), is assumed. The HARQ scheme adopted in LTE was mainly designed with objective of adding robustness towards fast link adaptation imperfections (e.g. against erroneous channel quality information), as well as to improve the basic spectral efficiency of the system by allowing transmissions at a higher block error rate (BLER) target.

As mentioned in section 3.2, in 5G the necessity of supporting a plethora of new services and applications with challenging and conflicting targets in terms of throughput, latency and reliability, calls for a flexible HARQ design where the timing of the control loop can be optimized on a link basis. The Round trip time (RTT) of a HARQ process can be defined as the time it takes from beginning of a transmission until the retransmission of the same data or transmission of a new data can be performed over the same SAW channel. For instance, latency limited users exploiting e.g. ultra-reliable low-latency communications (URLLC) services require extremely short RTTs, which can be instead relaxed for enhanced mobile broadband users. In general, the trends in current 5G design calling for a flexible HARQ timing are the following:

- **Different Transmission Time Interval (TTI) sizes:** The transmission to/from different users may likely include variable TTI sizes to best match the individual user’s service and radio conditions.

- **Asymmetric link operation:** The TTI size may in many cases be different for a user’s uplink and downlink transmission. From a coverage perspective, especially a macro cell-edge user may be required to transmit with longer TTI sizes to maintain its uplink coverage, while still being schedulable with short TTI size in the downlink due to the higher available transmit power.

- **Latency in the network implementation:** For more advanced network implementations, the lower physical layer processing may be physically separated from the processing of the rest of the transmitter/receiver chain. Examples include cases with Remote Radio Heads (RRHs), where the fronthaul connection between the RRH and centralized processing unit may be subject to certain latencies (e.g., if implemented as fronthaul over Ethernet).

- **Decoupled uplink and downlink:** While traditional cellular designs assume the same cell association for the uplink and downlink, recent studies have indicated promising benefits from allowing a user to be associated with different cells in the two link directions [BHL+16]; e.g. being connected a small cell in the uplink, while receiving data in the downlink from a macro-cell. Support for decoupled
uplink/downlink cell associations challenges the HARQ control loop timing, due to the potential backhaul latency of the communication between the involved cells.

- **Flexible Time Division Duplex (TDD):** The use of highly flexible TDD operation with arbitrary per-cell switching between downlink and uplink transmission on a fast time resolution also challenges the HARQ timing, as the relative timing of the HARQ feedback as compared to payload transmissions become a variable.

Given this, asynchronous HARQ in both link direction is assumed as a starting point. As mentioned above, this comes at the expense of additional signaling. The required timing flexibility inevitably leads to a large number of combinations in terms of number of SAW channels. Table 3-2 shows the RTT value for different downlink data transmission TTI sizes versus different uplink feedback durations. The numbers are provided for a Frequency Division Duplex (FDD) scenario where uplink resources are available at every subframe for feedback transmission. No fronthaul latency is assumed. For the sake of simplicity, the processing time at both the eNB and the UE is assumed to equal 0.1 ms. A minimum TTI duration of 0.125 ms is set. Such a short TTI size may consist of two OFDM symbols of 0.0625 ms. The feedback transmission duration can then vary from 0.0625 ms (i.e., one symbol) for UE’s with very good coverage up to 1 ms for extreme coverage-challenged UE’s. From the LTE link budget studies, it was found that the uplink control channel (PUCCH) transmitted during a 1 ms time-interval has a coverage range of 1.4 km for suburban non line of sight conditions (NLOS) [3GPP-36.912]. Therefore, we assume maximum uplink transmission duration of 1 ms and will further adopt the propagation delay based on the coverage range of <1.4 km. It is shown in Table 3-3 that the HARQ RTT can be reduced to 0.5 ms by transmitting over short TTI of 0.125 ms. The former is significantly shorter than the 8 ms HARQ RTT for the LTE FDD scenario.

![Table 3-2 RTT for different DL TTI sizes and feedback duration](image)

Furthermore, Table 3-2 shows the minimum number of parallel SAW channels corresponding to each RTT to avoid HARQ stall. The number of SAW channels must follow the number of consecutive downlink transmissions that can be performed during a RTT duration; e.g., with RTT of 1.5 ms while transmitting over TTI’s of 0.25 ms it is necessary to have at least 6 parallel SAW channels active in the MAC layer. However, a long ACK/NACK duration in the uplink can reduce the number of active SAW channels. For instance in the example in Table 3-2 when ACK/NACK duration is 1 ms, the number of active SAW channels is reduced to only 2. This can significantly reduce channel utilization efficiency. Bundling together the feedback of three HARQ packets increases the number of SAW channels to 6 and allows a continuous DL transmission of data.

### 3.3.4.2 Feedback enhancements

Different feedback enhancement can be considered for the holistic link solution, with the aim of improving resource utilization or reducing the latency. In this section, we describe two approaches based on an enriched feedback message to convey extra information for optimizing...
the retransmission, as well as the option of an early feedback for the sake of shortening the round trip time of an HARQ process.

3.3.4.3 Enriched feedback with multi-codeblock indexing

The carrier bandwidth for 5G below 6 GHz is expected to be up to 100 MHz, while it may reach values of GHz for millimeter wave bands. Hence, scheduling a user over the full transmission bandwidth results in very larger transport block size (TBS), demanding a huge memory to store samples from past HARQ retransmissions. As in LTE a TB, if larger than a certain limiting number of information bits (e.g., 6144 bits), gets segmented into several code blocks (CB’s) so that each code block size (CBS) is smaller or equal to that limit. The segmentation process is considered as an effective approach to limit the complexity of encoding/decoding which is known to increase with the block length.

The feedback message for a HARQ process is traditionally a rigid Boolean type value that conveys ACK or NACK. On the receiver side, each CB will be decoded separately and a positive acknowledgement (ACK) will be transmitted in case all the CBs are correctly decoded; otherwise the feedback will be a negative acknowledgement (NACK). A single failed CB inevitably leads to the retransmission of the whole TB, while it could suffice to perform the partial retransmission of the failed CBs.

In case a larger number of feedback bits can be transmitted, it becomes possible to index the failed CBs such that those becomes to only ones being retransmitted, avoiding unnecessary usage of radio resources.

It is worth to mention that not all the CBs need to be indexed in the practice assuming realistic BLER targets. Let us assume for instance that the decoding of each CB in the TB is successful with probability $1 - \rho$ or fails with probability $\rho$, i.e., modelled with independent and identically distributed Bernoulli random variables. For a TB segmented into 50 CB’s, and target BLER of 10% at first transmission, approximately 95% of NACK’s will be as a result of only one failed CB’s while 4% of NACK’s will be result of two failed CB’s. In order to report indexes of one or two failed CB’s in this example, 6 or 11 bits are required, respectively. An attractive feedback design will further consist of a e.g. 2 bit header message, identifying the type of information that the feedback conveys. The header points to four different feedback messages, e.g. as follows:

11. ACK – in case of decoding success for the TB
01. NACK (followed by one CB index) – in case only one CB fails in decoding
10. NACK (followed by two CB indexes) – in case only two CB’s fail in decoding
00. NACK (followed by DSI) – in case of more than two failed CB’s

The resultant multi-bit feedback has a total size of 8 or 13 bits, and the resources used for retransmissions can be reduced by ~93% and ~97% respectively. Figure 3-16 shows the average relative resource savings by using a similar setup for different TBS’s with respect to total feedback size. Note that a higher resource utilization efficiency gain is expected e.g. by sending DSI over the feedback in the case feedback header indicates 00 (i.e., all of the CB’s must be retransmitted). Transmitter then uses the DSI-rich feedback for an adaptive-length redundancy matching as explained in the previous section which will result in utilizing less resources on average for retransmissions.
3.3.4.4 Enriched feedback with variable block length IR

In fixed IR schemes, the size of the retransmission keeps fixed, which may lead to a waste of time and frequency resources if the receiver does not need so much redundancy to correctly decode a failed transmission. With the aim of improving resource utilization we focus on a variable block length IR scheme, where an enriched feedback is used for signalling the size of the retransmission, which depends on the amount of estimated mutual information already available at the receiver. The size of the \((i+1)\)-th transmission when binary Turbo codes are employed, can be inferred from the following inequality [WWD+14]:

\[
N_{i+1} I_{M_s}(SNR_{i+1}) > N_i \log_2(M_s)R - \sum_{j=1}^{i} N_j I_{M_s}(SNR_j),
\]

where symbols are mapped into a \(M_s\)-ary modulation alphabet, \(R\) is the code rate and the duplet \((N_j,SNR_j)\) respectively denotes the effective SNR and size in terms of symbols associated with the \(j\)th transmission. The function \(I_{M_s}(SNR_j)\) is the bit-interleaved coded modulation (BICM) capacity at \(SNR_j\) assuming that symbols are drawn from the \(M_s\)-ary modulation alphabet. Since the modulation keeps unchanged, the number of code bits that should be delivered in the \((i+1)\)th transmission is \(N_i = \log_2(M_s)N_{i+1}\). Conversely, the condition that governs the design with non-binary low density parity check (NB-LDPC) codes becomes [CNM+17]:

\[
N_{i+1} \Bar{I}_{i+1} > N_i \Bar{I}_{th} - \sum_{j=1}^{i} N_j \Bar{I}_j.
\]

In notation terms, \(N_i\) accounts for the \(i\)th transmission size in terms of code symbols and \(\Bar{I}_i\) denotes the average mutual information estimated in the \(i\)th transmission. The average mutual information required to achieve the target BLER with the selected code rate is given by \(\Bar{I}_{th}\), which can be easily obtained from the reference BLER curves in AWGN. The details to compute the average mutual information are provided in [CNM+17]. The most general case where the modulation is selected on a per RB basis has been covered.

Unfortunately, when the \(i\)th transmission fails, the metrics \(\Bar{I}_{i+1}\) and \(SNR_{i+1}\) cannot be measured yet. In this case, the size of the next transmission will be computed using the available information from the last transmission. Hence, previous inequalities can only be used for low and moderate mobility scenarios.

To make an efficient use of the resources, \(N_i\) should be set as the minimum value that satisfies the inequalities. Then, we can easily compute the minimum number of RBs that are necessary to transmit the additional redundancy. In practice, full resolution is not available, and the number
of RBs that can be allocated is predetermined. For instance, if \( l \) bits are used to convey the ACK/NACK information, then \( 2^{l-1} \) NACK messages can be defined, each associated with a predefined number of RBs. Therefore, the most spectral efficient strategy consists in selecting the NACK level that allows transmitting the required redundancy with the minimum number of RBs. Numerical results are provided in the appendix.

### 3.3.4.5 Early feedback

The processing time at the UE could become a significant bottleneck for achieving the challenging latency targets set for certain 5G applications. In current LTE receivers, it is estimated that around 60% of the processing time is spent by turbo decoding, while the remaining time is spent by OFDM processing, equalization, soft-demodulation, etc. Predicting whether the decoding will be successful prior to the decoding itself, would enable the UE to anticipate its feedback transmission while running the decoding in its pipeline. The HARQ RTT can be reduced, with significant benefits in terms of latency. The prediction can be performed upon reception of the TTI, or even a portion of a TTI, e.g. only a limited number of OFDM symbols. It can be easily shown that the predicting the outcome of the decoder before decoding occurs can lead to a latency reduction in case the following condition is satisfied:

\[
\text{UE}\_\text{processing}\_\text{time} > \text{TTI}\_\text{duration} - 2 \times \text{Propagation}\_\text{Delay}.
\]

In case the condition above is not satisfied, there would be sufficient time for performing the decoding before the next opportunity for feedback transmission occurs.

The decoder outcome can be predicted from the log-likelihood ratios (LLR) of its input bits, which are obtained from the soft-demodulation of the equalized data symbols. The proposal in [BKP+16] is based on estimating the uncoded bit error rate (BER) from such likelihood ratios, and mapping such estimate to reference coded BER curves of a specific code. Though the performance is evaluated for turbo codes, the principle can be generalized to other coding solutions. Simulation results show the possibility of obtaining a correct prediction in more than 90% of the cases. However, the average rate of erroneous estimates is not negligible. False positives occur when an early ACK is generated for a codeblock which is not going to be correctly decoded, while false negatives occur when an early NACK is generated for a successfully decoded codeblock. False positives could result in higher layer retransmissions. This leads to a latency increase. Conversely, false negatives may cause unnecessary retransmissions, which affect the throughput without harming the latency. False positives are then considered more critical for LLC cases. The solution in [BKP+16] assumes that their occurrence should be minimized, at the expense of an eventual higher rates of FNs.

Some numerical results are provided in the Appendix 6.7.

### 3.3.4.6 Conclusions

Given the necessity of supporting diverse services and network implementations, HARQ can significantly benefit from a flexible timing of the control loop. Envisioned 5G ultra-low latency applications may require the usage of an early feedback to be transmitted prior to finalizing the decoding. Further, the concept of the feedback is expected to take a leap forward from its Boolean nature and provide further information for optimizing the retransmissions. In particular, the enriched feedback can be used for indexing specific code block retransmissions, avoiding retransmission of the entire transport block and therefore improving resource utilization. Such resource saving can be also achieved by exploiting the variable IR concept, i.e. using the enriched feedback to signal the redundancy version of a transport block which is likely to ensure correct decoding given the already accumulated mutual information at the receiver.
3.3.5 PHY procedures

3.3.5.1 Proposed PHY procedures

Blind detection of users’ activity in random/grant-free access

It has been agreed in 3GPP that grant-free multiple access should be studied for mMTC and URLLC use cases in the 5G study item [R1-162189, R1-163983]. In grant-free access, the users can transmit to the base-station without prior scheduling grant. One of the challenges in grant-free access is that the base-station will not be aware of how many users are transmitting to the base-station, at a given time, in grant-free manner. Without the knowledge of number of active users (users simultaneously transmitting to the base-station on the same resources), the base-station will not be able to retrieve/decode the users’ data. Thus, it is essential to develop a scheme that enables blind detection of user activity at the base-station.

We propose a novel method that enables the base-station to accurately detect the number of users currently accessing the system in a grant-free manner, by capitalizing on a unique feature in FQAM [Ali16a]. As the number of active subcarriers in each FQAM symbol depends on the number of transmitting users, it will give information to the base-station on how many users are transmitting. Thus, in the proposed method, the base-station will measure the number of active subcarriers in the grant-free access area and infer/deduce the number of active users accordingly. Once the base-station detected the number of transmitting users, it will retrieve the data by applying the suitable detection and decoding procedures. Using numerical simulations, we show that by using the proposed method the number of active users can be blindly detected at the base-station.

In the proposed method, the grant-free access area will be divided into two regions, as illustrated in Figure 3-17. On the first region, which will be referred to as Activity Area, FQAM modulation is used to enable detecting the users’ activity. The modulation on the second region of the grant-free access area, which will be referred to as Data Area, can be selected by the network based on the system requirements, hence, any modulation scheme can be used (FQAM, QAM, etc.). Please note that both regions will carry users’ data, however, the modulation used on the Activity Area will facilitate blindly detecting the number of users transmitting on the grant-free access area. In addition, the grant-free access area can consist of the Activity Area only. The available resource elements in the Activity Area will be divided into \( U = N_a/M_f \) subsets/groups, and each subset of subcarriers will represent one FQAM symbol, where \( N_a \) denotes the number of resource elements in the Activity Area. The modulation and coding parameters (i.e. \( M_f, M_Q, \) code rate, etc.) the users should use on the Activity Area in each grant-free access area is pre-defined by the base-station and known for both; users and base-station.

![Figure 3-17: Proposed channel structure: grant-free access area divided into Activity Area and Data Area (example with Activity Area of two OFDM symbols).](image-url)
On the arrival of new data, each user will select one of the grant-free access areas to transmit on. The user will use the modulation and coding parameters of the Activity Area in the selected grant-free access area. In the data transmitted on the Activity Area, the user will include; i) modulation and coding information of the Data Area, ii) its user ID (and other identification information if needed), iii) small amount of data (depending of the Activity Area size and the modulation and coding scheme). In the data transmitted on the Data Area, the user can include its user ID (if not included in the Activity Area) and any other payload data. Details and simulation results of the blind detection of users’ activity in random/grant-free access can be found in Appendix 6.8.1.

New Radio Physical Random Access Channel Sequence Design

Transmission of a random access (RA) preamble in uplink is one of the first steps of a terminal to get access to the radio network. Therefore, sets of signature sequences are defined that show good auto-correlation and low cross correlation characteristics. Another important requirement is a low susceptibility against impairments that originate in time and frequency domain in order to guarantee low false-alarm and missed-detection probabilities. Zadoff-Chu (ZC) sequences that are used in 4G for RA, are primarily designed to show high robustness in time-dispersive channels, but not in channels that are simultaneously frequency-dispersive.

For 5G New Radio we propose a new class of sequences for the random access channel (RACH) that show excellent auto- and cross-correlation properties in a time-frequency dispersive channel. The new sequences are derived by circular Delay-Doppler shifts (time-frequency) of m-sequences. These sequences show a higher robustness against frequency impairments and a significantly reduced false-alarm probability compared to legacy ZC-sequences. In addition, much more sequences can be uniquely detected at the base station receiver. Important key-performance metrics such as protocol latencies, upload probabilities or cell-coverage are improved [R1-1612296]. The implementation effort of the proposed new class of sequences is rather small: at the transmitter side the device must be enabled to construct the new sequences in the digital domain, while the receiver side must be enabled to perform correlations with the new set of sequences. Details and simulation results of circular Delay-Doppler shifts (time-frequency) of m-sequences can be found in Appendix 6.8.2.

Massive MIMO assisted random access for 5G IoT

The massive MIMO assisted random access for 5G IoT aims at solving the massive access problem in M2M/IoT. It includes the enhanced design of the signal structure of the UE on IoT PRACH and an improved procedure which enables potential one-stage transmission to reduce latency. Additionally, with a massive number of antennas, the eNB is able to separate users with the same preamble in the spatial domain. As a result, user number estimation based on AoA estimation is performed at the eNB.

The inclusion of user data in the random access signal enables possibilities of one-stage transmission, which can significantly reduce latency. If one-stage transmission is not successful, multiple RARs will be transmitted by the BS in the two-stage transmission. The number of RARs is determined by an AoA estimation of users. It has been demonstrated via simulations that the proposed method can reduce access delay while minimizing waste resources. Design details and simulation results of the massive MIMO assisted random access for 5G IoT can be found in Appendix 6.8.3.

Embedded air interface (EAI) for multiplexing eMBB and MCC traffic

In this research work, we consider the simultaneous support of two typical services in the same band, namely MCC and eMBB. Specifically, eMBB sets focus on attaining high spectral efficiency while MCC requires strict latency and reliability constraints to be fulfilled (e.g. a packet needs to be transmitted successfully with 99.999% reliability within 1ms). Hence, special design principles should be followed when designing frame structure and scheduling. In general,
eMBB builds on longer TTI (e.g. 1ms in LTE) due to long data packets and moderate delay tolerance, while MCC relies rather on short TTIs (e.g. 0.25ms or even short TTI of 0.125ms) in order to achieve low latency. In addition, short TTI is is well supported by MCC since its traffic is typically constituted of small packets.

From the viewpoint of a smooth evolution, it is reasonable to operate MCC service in a coexistent way with eMBB service, especially in the initial phase of 5G. We therefore focus on the investigation of how to design an air interface simultaneously supporting eMBB and MCC service in an efficient manner. For simplification, we consider the same numerology, i.e. CP length and subcarrier spacing, for both services, while different TTI length is used. Just as an example, we assume LTE frame structure for eMBB, while a shorter TTI length is chosen for MCC, depending on the packet size (e.g. one symbol TTI).

![Frame structure supporting multiplexing of eMBB and URLLC traffic](image)

Figure 3-18: Frame structure supporting multiplexing of eMBB and URLLC traffic

There are some key aspects investigated in the field of scheduling and signaling (in case of DL):

1) Resource sharing among eMBB and MCC. Since MCC traffic is unpredictable and has higher priority than eMBB traffic, MCC scheduling may override at any time resource allocations that have already been scheduled for eMBB traffic. This way, resources in the resource grid may be “punctured” by the MCC scheduler.

2) Design of the scheduling algorithm for both eMBB and MCC.

3) Informing the UE operating eMBB service on the resource puncturing:
   a) Without being informed on the punctured resources, eMBB UE will attempt to decode the punctured resource, yielding significant performance degradation.
   b) If being informed, eMBB UE may avoid using the punctured resource.

Conservative allocation may not satisfy both the low-latency and high reliability requirements of MCC. Therefore, MCC is permitted to override the eMBB resources through puncturing, i.e., MCC could “steal” some OFDM symbols of the resources allocated to eMBB service. However, eMBB would suffer from a performance loss in that case. To guarantee the performance of eMBB, the following schemes should be applied:

   a. With being informed on the punctured resource, eMBB UE may avoid using the punctured resource;
   b. The data punctured by MCC would be excluded after reception at eMBB UE, given that punctured symbols have been reported to that UE;
   c. Since MCC in many cases will only puncture a few code blocks (CBs) of eMBB service, eMBB UE only requires retransmission of the wrong CB instead of the whole transmit block (TB), thus saving downlink resources;
   d. Mapping MCC service data to those resources occupied only by parity bits of eMBB would greatly improve eMBB performance.

Details and simulation results of EAI for multi-service can be found in Appendix 6.8.5.
3.3.5.2 Harmonization of proposed PHY procedures

3GPP has been considering three key use cases for 5G new radio (NR). These use cases include enhanced mobile broadband (eMBB), massive machine type communications (mMTC) which is also known as massive machine communications (MMC), and ultra-reliable and low latency communications (URLCC) which is also known as mission critical communications (MCC).

The target for peak data rate for eMBB should be 20Gbps for downlink and 10Gbps for uplink. Besides peak data rates, 5G NR is expected to require lower latencies and higher reliability. The target for control plane latency should be 10ms. For URLLC the target for user plane latency should be 0.5ms for UL, and 0.5ms for DL. For eMBB, the target for user plane latency should be 4ms for UL, and 4ms for DL. A general URLLC reliability requirement for one transmission of a packet is $10^{-5}$ for X bytes (e.g., 20 bytes) with a user plane latency of 1ms [3GPP-38.913]. It should be noticed that the traffic from a node can lie into multiple use cases. An electrocardiogram (ECG) sensor, for instance, can be categorized as MMC and MCC at the same time when the patient is in emergency and the ECG sensor alarm is triggered. In this case, requirements from both use cases need to be satisfied.

To satisfy different requirements on throughput, delay, and reliability of these 5G use cases, a harmonized enhanced physical layer (PHY) procedure maximizing the full use of resources in spatial/frequency/time domains, was proposed in this section. This harmonized enhanced PHY procedure includes proposed methods described in Section 3.3.5.1.

Harmonized solution

In order to enable this harmonized PHY procedure, the base station (BS) is required to support:

- massive MIMO: resolving various user equipments (UEs) who multiplex their random access (RA) signals in the spatial domain during the RA procedure in the uplink
- shortened TTI (sTTI): various services multiplexed together in the time and frequency domain using different TTI lengths in the downlink
- embedded air interface (EAI): certain sTTIs punctured for eMBB in order to carry MCC payloads and ensure the efficient and reliable transmissions of diverse services. The BS should be able to inform resource puncturing pattern to eMBB and MCC UEs.
- Zadoff-Chu (ZC) sequences or Cyclic Delay-Doppler shifted M (CDDSM) sequences: sequences to be used as preambles in the RA procedure.
- blind detection of users’ activities: detection performed in either the spatial domain with massive MIMO or the frequency domain with FQAM to estimate the number of colliding users

Requirements for UEs:

- eMBB UEs: should be able to identify puncture patterns and needs to support retransmitting the wrong code blocks (CB) (instead of transport blocks) if 'puncture and inform' mode is applied; should be able to support data transmission with RA preamble as shown in Section 6.8.3, i.e. the random access signal will consist of the preamble and payload data;
- MCC UEs: should be able to receive data from sTTIs; should be able to support data transmission with RA preamble, i.e., 1-stage random access protocol; should be aware of reserved preambles for MCC traffic;
- MMC UEs: should be able to support data transmission with RA preamble

An example of the harmonized solution is depicted in Figure 3-19. UEs with different services initiate RA in the UL using the signal structure in Section 6.8.3. MCC UEs will transmit their RA signals in the MMC PRACH resources, which was known as the narrowband physical random access channel (NPRACH) in earlier LTE releases [3GPP-36.211]. The preamble in a RA signal is a ZC sequence or a CDDSM sequence. In 5G NR, the MMC PRACH is not necessarily identical to NPRACH. MCC and eMBB UEs will transmit their RA signals in
PRACH. As the RA signals consist of payload data, which are independent to each other, the BS can attempt to resolve the number of colliding users with massive MIMO in the spatial domain. Additionally, if the data payload is modulated using FQAM, blind detection in the frequency domain can be applied as well. The benefit of using the signal structure in Section 6.8.3 is that it allows potential one-stage transmission and can reduce the access delay even when preamble collisions occur. Then, the BS will allocate resources to the UEs in the UL for radio resource control (RRC) connection establishment. Once the RRC connections are set up, resources in the DL can be assigned to UEs for data transmission. For the scenarios where the base-station is not aware of the number of users simultaneously transmitting on the same resources (e.g. grant-free access and users’ collision in random access), it is recommended to use tone-switching modulation (such as FQAM) to enable blind detection of users activity at the base-station using the method explained in Section 3.3.5.1 (see Figure 3-17).

Figure 3-19 UL and DL resource maps for UEs of different scenarios

In the downlink, the BS responds to MMC UEs like legacy LTE. However, in terms of eMBB and MCC traffic, the BS may enable EAI as presented in Section 6.8, data for MCC services can be multiplexed in both the time and frequency domain with eMBB traffic. This multiplexing can be achieved by puncturing certain sTTIs, which is illustrated in Figure 3-19 (left). To ameliorate the performance loss of eMBB caused by puncturing, MCC users need to inform puncturing resources to eMBB service to avoid using the punctured resource. Consequently, the data punctured by MCC need to be excluded at the receiver side of eMBB UE with the notified information regarding the puncturing positions. Noticing that MCC may only puncture a few code blocks of eMBB, code-block-Level HARQ is performed for eMBB users where they only to feedback index of the wrong CBs, and then to retransmit the wrong CB instead of whole transport blocks. In addition, mapping MCC service data to parity bits of eMBB is proposed to greatly improve eMBB reliability.

To further ensure that MCC nodes can have access to resources instantly, a small subset of preambles can be reserved for MCC purposes. When the traffic is with normal priority, UEs use normal preambles for RA. On the other hand, when the traffic is critical or with high priority, UEs can use the reserved preambles for RA and the BS will assign dedicated resources for this traffic.

4 Conclusions

In this report, we have summarized the findings of the research on link design. Although the conclusions are given in each topic-dedicated section, we summarize the main take-away of this
report, enabling the readers to get a quick overview on outcomes of the respective topic. The following topics are grouped according to the task groups.

**Task 3.1: signal design**

**Waveform:** various waveform candidates are investigated and compared in this project, including the main streams being considered in 3GPP. Several conclusions are drawn after a careful analysis of the pros and cons of each candidates and comparison results. The complex orthogonality is deemed as the most important factor to make the waveform be friendly with all other components, e.g. MIMO, frame structure, reference signal, etc. Therefore, OQAM signaling based scheme such as FBMC-OQAM is regarded as an immature solution for 5G. While, among QAM based waveform candidates, neither subband-wise filtering nor subcarrier-wise filtering (aka. windowing), can be a one-size-fits-all solution. Thus, we recommend the favorable cases for selecting these solutions (section 2.1.4).

**Channel coding:** The enhanced Turbo code design has been simulated compared to LTE Turbo, showing that the enhanced Turbo has remarkable gains. It is shown that the enhanced Turbo code can meet all 5G requirements except for extreme high throughput. We have prepared a detailed description of the gains as well as a narrative of 3GPP channel coding progress, including the comparison observations among candidates: how were the decisions made and the future potential use cases for the enhanced Turbo codes.

**Enhanced modulation:** FQAM and NUC are presented, which bring advantages for the 5G system. It is recommended to use FQAM to serve users that suffer from high interference levels (such as cell-edge users) and QAM for users with good channel conditions (cell-centre users). The inactive subcarriers in each FQAM symbol should be utilized to serve another user with relatively lower transmission power as detailed in section 2.3.1. For BMS and MBB services, it is recommended to utilize NUC for at least high modulation orders (64QAM, 256QAM, etc.) to enhance the achieved throughput. As it is not expected to have high modulation orders for MMC, MCC and V2X, NUC may not be very beneficial for these services.

**MIMO:** This topic contains several subtopics. One of those is towards the MIMO compatibility with new waveforms, we have proved that the QAM based waveforms have all complex orthogonality hypotheses which make them fully compatible with MIMO; while OQAM based waveforms are not compatible with MIMO. But nevertheless, the solutions to make MIMO transmission work on the top of OQAM-based waveforms are developed. Another topic aims at enhancing high mobility transmission quality using the predictor antenna. The networks hits the “wall of speed” (i.e. stops using adaptive MIMO) at higher velocities. Simulation results show that, for 256x2 ZF-MIMO and 256x1 MRT-MISO, respectively, speeds or carrier frequencies which are 3 and 4 times higher, respectively, can be supported thanks to the predictor antenna.

**PAPR reduction techniques:** In the current LTE uplink transmission, single carrier based DFTs-OFDM is used for UL in order to increase power efficiency. There are several schemes discussed for PAPR reduction with different degree of complexity. Due to different core services within 5G, it is required to differentiate between device types (which correspond to core services) and requirements on PAPR reduction in terms of complexity, energy and/or spectral efficiency. However, single-carrier vs. multi-carrier based transmission should not be made solely on the PAPR performance, but rather be settled after comprehensive evaluation for different channel characteristics and scenarios, in terms of PAPR, power amplify back-off, and link performance, etc. Within this report, guidelines / conditions are provided from Fantastic 5G perspective to select specific PAPR reduction techniques for specific service types.

**Task 3.2: frame design and PHY layer procedure**

**Frame design:** A tiling concept is introduced as a holistic design principle, constituting the key component for a flexible air interface design. Therein, the radio resources are partitioned into different tiles defined in the time / frequency space, each of which can be configured individually by a set of important physical layer parameters such as subcarrier spacing, TTI, waveform etc. Moreover, multi-cell numerology alignment is also discussed for different
advanced UE operations. Variable TTI is another important feature to be supported by the 5G system. And the concept of mini-slots has been proposed, which is also topic being discussed in 3GPP NR, for the support of low latency communication. The key aspects of mini-slots, such as its realization and constraints for half-duplex operation, are also detailed. Moreover, two approaches of implementing tunable TTIs, namely self-contained frame and symbol-wise frame, which are based on adjustable subcarrier spacing (i.e. subcarrier spacing scaling) and variable number of OFDM symbols (i.e. OFDM symbol number scaling) per TTI, respectively, are compared. It is shown from the simulation results that the symbol-wise frame structure is more robust to mobility; however in low mobility situations, the self-contained frame structure seems to offer a near linear increase in capacity as the SNR increases.

**Control channel design**: We have presented the NR-PDCCH design in regard to search spaces with different transmission schemes and transmit diversity approaches. In particular, localized and distribution transmission based search spaces are provided, and performance gain of distributed transmission compared to localized transmission is demonstrated by simulation results. Moreover, SFBC and per-RE precoder cycling based transmit diversity schemes are compared. Based on performance comparison, SFBC seems to be a preferred scheme. Finally, high order transmit diversity methods based on antenna pair cycling are also proposed. The considerable performance gain exhibited from simulation results clearly motivates the support of high order transmit diversity. On the other hand, we have presented the concept of a user-specific multi-stage PDCCH. The proposed MS-PDCCH consists of three different stages: The UIDS pilot sequence and the fixed-sized basic control and the variable-sized main control FEC block, respectively. The UIDS drastically reduces UEs’ blind decoding effort, alleviating the need to introduce PDCCH search spaces. Moreover, it avoids scheduler complexity normally incurred by search spaces and, in conjunction with its user-specific design which multiplexes control information and user data UE-individually, provides greater overall system flexibility and thereby improved future proofness.

**Reference signal design**: In order to reduce the overhead of CRS and instead use CSI RSs more flexibly for all purposes, e.g. high mobile UEs, we have proposed the concept of sub-tiling by which a dense grid of pilots in time and frequency is generated and in order to reduce the overhead each pilot uses only a fraction of the resource element for each access point. The sub-tiled CSI RSs are cell specific and constantly transmitted - typically - over the full frequency bandwidth. We have also investigated how to enhance the detection of control channel and data for low-power communications, which enables the UE to achieve a target false alarm and missed-detection probability even under relatively low SNR [FAN-D31].

**HARQ**: In this project we have a common view that the HARQ procedure for 5G shall follow an asynchronous manner for both downlink and uplink. Moreover, the flexibility is the key feature being pursued while designing HARQ procedure, for which the design shall target different Transmission Time Interval (TTI) sizes, asymmetric link operation, latency and decoupled uplink and downlink. Moreover, FANTASTIC-5G partners also looked into different ACK/NACK feedback algorithms. Among them, some aim to save resource efficiency, e.g. adaptive-length redundancy matching with enriched feedback and multiple bits ACK/NACK feedback, indicating the retransmission code block index only instead of the whole transport block. Other algorithm, such as early feedback, aims to reduce the processing latency.

**PHY procedure**: We have investigated massive MIMO assisted random access for 5G IoT, aiming at solving the massive access problem in M2M/IoT. With a massive number of antennas, the eNB is able to spatially separate users using the same preamble on the PRACH. As a result, the number of colliding users can be estimated based on AoA estimation at the eNB. Using this property, firstly, we propose an enhanced signal structure design for IoT random access, which helps effective low-latency transmission. Secondly, it provides an improved random access procedure enabling potential one-stage transmission to reduce latency. General physical layer procedures for UE and eNB are also provided. To enable this technique, we have also detailed the technology how to estimate the number of users in spatial domain using massive
antenna at eNB. For the scenarios where the base-station is not aware of the number of users simultaneously transmitting on the same resources (e.g. grant-free access and users’ collision in random access), it is recommended to use tone-switching modulation (such as FQAM) to enable blind detection of users activity at the base-station using the method explained in Section 3.3.5.1 (see Figure 3-17). In relation to the PRACH, we provided an enhancement on the PRACH sequence design which takes into account the dually dispersive channel. Moreover, we studied how to operate MCC service in a coexistent way with MBB, with a focus on how to design efficient coexistent air interface. It includes frame structure and resource partition design, scheduling and signalling procedure, code-block-level HARQ, data mapping. Through comprehensive system-level simulations we have observed that with informed resource puncturing, MBB UE may avoid using the punctured resource and these resources may also be excluded at the MBB receiver. MBB only needs to feedback and retransmit a few punctured code blocks instead of the whole transport block data in order to save the HARQ effort. Additionally, mapping MCC service data to parity bits of MBB can greatly improve MBB link performance.
5 References


[3GPP-36.141] LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) conformance testing (3GPP TS 36.141) 01/2010.


Imperfect CSIR,” in European Conference on Networks and Communications, 12-15 June 2017.


R. Garzón-Bohórquez, C. Abdel Nour and C. Douillard, "Improving Turbo Codes for 5G with parity puncture-constrained interleavers," 9th Int. Symp. on Turbo Codes and Iterative Information Processing (ISTC), Brest, 2016, pp. 151-155.


Helmling, Michael and Scholl, Stefan: Database of Channel Codes and ML Simulation Results. University of Kaiserslautern, 2016. URL: www.uni-kl.de/channel-codes.


Yong Jiang, Yi Wang, “Adaptive weight peak-cancelling scheme for OFDM systems”; the 18th annual IEEE Symposium PIMRC’07.


[R1-162890] 3GPP RAN1 meeting #84b, April, 2016, “5G Waveforms for the Multi-Service Air Interface below 6 GHz”, Nokia, Alcatel-Lucent Shanghai Bell

[R1-163558] 3GPP RAN1 meeting #84bis, April, 2016, “Way forward on assumptions for waveform evaluation”, Huawei

[R1-164635] 3GPP RAN1 meeting #85, May, 2016, “Improved LTE turbo codes for NR”, IMT, Orange.

[R1-164698] 3GPP RAN1 meeting #85, May, 2016, “LDPC - Performance evaluation”, Qualcomm.

[R1-165012] 3GPP RAN1 meeting #85, May, 2016, “New radio waveforms for the Multi-Service Air Interface below 6 GHz”, Nokia, Alcatel-Lucent Shanghai Bell

[R1-165013] 3GPP RAN1 meeting #85, May, 2016, “Initial uplink performance results for the New Radio Waveforms below 6GHz”, Nokia, Alcatel-Lucent Shanghai Bell

[R1-165014] 3GPP RAN1 meeting #85, May, 2016, “Subband-wise filtered OFDM for New Radio below 6GHz”, Nokia, Alcatel-Lucent Shanghai Bell

[R1-165859] 3GPP RAN1 meeting #85, May, 2016, “Way forward on further evaluation assumptions for NR waveform”, Huawei, HiSilicon, InterDigital, NTT DOCOMO, LGE

[R1-165989] 3GPP RAN1 meeting #85, May, 2016, “Way forward on calibration assumptions for NR waveform”, Huawei, HiSilicon, NTT DOCOMO, Nokia

[R1-166004] 3GPP RAN1 meeting #85, May, 2016, “Response LS on realistic power amplifier model for NR waveform evaluation”

[R1-166594] 3GPP RAN1 meeting #86, August, 2016, “Discussion on the window processing of FC-OFDM and WOLA”, Orange.

[R1-166595] 3GPP RAN1 meeting #86, August, 2016, “FC-OFDM enhanced receiver design for very long delay spread”, Orange.

[R1-167244] 3GPP RAN1 meeting #86, August, 2016, “Uplink performance results with extended settings for the New Radio waveforms below 6 GHz”, Nokia, Alcatel-Lucent Shanghai Bell
[R1-167260] 3GPP RAN1 meeting #86, August, 2016, “Resource block and guard band arrangement supporting mixed numerology”, Nokia, Alcatel-Lucent Shanghai Bell

[R1-167261] 3GPP RAN1 meeting #86, August, 2016, “On resource block grouping and multi-cell coordination aspects for mixed numerology support”, Nokia, Alcatel-Lucent Shanghai Bell

[R1-167413] 3GPP RAN1 meeting #86, August, 2016, “Enhanced Turbo Codes for NR: Implementation Details”, IMT, Orange.


[R1-163558] 3GPP RAN1 meeting #84bis, April, 2016, “Way forward on assumptions for waveform evaluation”, Huawei


[R1-166031] 3GPP RAN1 meeting #85, May, 2016, “Way forward on High Speed Link Level evaluation”, Cohere, AT&T, CMCC, Deutsche Telekom, InterDigital, National Taiwan University, Orange, Spreadtrum, Telefonica, Telstra.


[R1-1610314] 3GPP RAN1 meeting #86bis, October, 2016, “FEC performance comparison for short frame sizes for NR”, IMT, Orange.

[R1-1610404] 3GPP RAN1 meeting #86bis, October, 2016, “FC-OFDM discussion based on RAN1#86 waveform agreements”, ORANGE

[R1-1610600] 3GPP RAN1 meeting #86bis, October, 2016, “Updated Summary of Channel Coding Simulation Data Sharing for eMBB Data Channel”, Interdigital.[R1-1610878] 3GPP RAN1 meeting #86bis, October, 2016, “Chairman's notes of AI 8.1.3 on channel coding and modulation for NR”, Nokia.


[R1-1608584] 3GPP RAN1 meeting #86bis, October, 2016, “Complementary turbo and LDPC codes for NR, motivated by a survey of over 100 ASICs”, AccelerComm.

[R1-1610931] 3GPP RAN1 meeting #86bis, October, 2016, “Way Forward on channel codes for NR eMBB data”, IMT, Orange, AccelerComm, Ericsson, LG, NEC.


[R1-1613347] 3GPP RAN1 meeting #87, November, 2016, “WF on channel codes for NR short block length eMBB data”, AccelerComm, Ericsson, Orange, IMT, LG Electronics, NEC.


6 Appendix

6.1 Link performance comparison of new waveform candidates

To evaluate the waveform proposals, FANTASTIC-5G has agreed on simulation scenarios for comparison. For a fair comparison, different simulators with different waveform candidates are first calibrated with an implementation of a common reference waveform (e.g. CP-OFDM, SC-FDMA). The test cases for calibration are aligned with parallel on-going 3GPP waveform evaluation discussions.

In the following, we will first provide the calibration test cases and results are presented to align individual simulator with different waveform proposals. Secondarily, we address the detailed waveform specifications for selected waveform proposals involved in comparison acti firstly provide the details of three waveform comparison scenarios in this project, which targets at providing a quantitative overview of new waveform properties for non-eMBB scenarios. Secondly, Last but not least, waveform comparison results are described for each scenarios.

6.1.1 Waveform specifications

We provide in this section the specifications for each waveform candidates used in our comparisons.

6.1.1.1 UF-OFDM

Waveform parameters

UF-OFDM exploits the different Dolph-Chebyshev (DC) filters for different numerologies: For 15 kHz subcarrier spacing and 4 PRBs case, filter order is \( L=72 \) and \( \text{SLA}=75\,\text{dB} \); for 30 kHz subcarrier spacing and 2 PRBs, filter order is \( L=37 \) and \( \text{SLA}=37\,\text{dB} \). The pre-equalization technique is applied according to [5GNOW15], and matched filter is also applied at the receiver.

In UF-OFDM, filter is applied per subband allocation. As shown in Figure 6-1 and Figure 6-2, this leads to a higher choice of the SLA and therefore a lower side lobe level in the frequency domain and a shorter filter response in the time domain. The effect is more visible in larger allocations.

![Figure 6-1 PSD of UF-OFDM for DL filter applied per PRB (red) and DC filter applied per subband allocation (blue) for 4PRBs](image)

Figure 6-1 PSD of UF-OFDM for DL filter applied per PRB (red) and DC filter applied per subband allocation (blue) for 4PRBs
Figure 6-2 Time domain symbol of UF-OFDM for DC filter applied per PRB (red) and DL filter applied per subband allocation (blue) for 4PRBs

For additional illustrations regarding transmit symbol sequence and receive window placement see [R1-165013].

Latency

With the choice of the DC filter equal to the ZP length, UF-OFDM symbol has the same length as that of a CP-OFDM symbol.

Complexity

Complexity analysis for UF-OFDM:

- Transmitter Complexity: The transmitter complexity is discussed in detail in [FAN16-D31] and [R1-165014]. Several ways to implement the transmitter exist. For small allocation sizes, look-up table supported solutions can achieve lesser complexity than basic CP-OFDM. Table 6-1 in [FAN16-D31] gives a basic overview.

Table 6-1 Complexity comparison of UF-OFDM and CP-OFDM

<table>
<thead>
<tr>
<th>Waveform and Implementation Variant</th>
<th>Number of subbands B=1</th>
<th>Number of subbands B=50</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP-OFDM, FFT-based</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CP-OFDM, LUT-based, M=2, P=2 (Appendix A in [FAN16-D31])</td>
<td>0.294</td>
<td>-</td>
</tr>
<tr>
<td>UF-OFDM, LUT-based, M=2, P=2 (Appendix A in [FAN16-D31])</td>
<td>0.315</td>
<td>-</td>
</tr>
<tr>
<td>UF-OFDM, time domain, G=1</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>UF-OFDM, time domain, G=3</td>
<td>3.7</td>
<td>3.7</td>
</tr>
<tr>
<td>UF-OFDM, frequency domain</td>
<td>2.4</td>
<td>9.5</td>
</tr>
<tr>
<td>UF-OFDM, segmentation technique (subband size = 12)</td>
<td>1.2</td>
<td>2.35</td>
</tr>
<tr>
<td>UF-OFDM, segmentation technique (subband size = 16)</td>
<td>1.2</td>
<td>1.2</td>
</tr>
</tbody>
</table>
A novel UF-OFDM technique has been investigated by Telecom Bretagne [FAN16-D52] in order to further reduce the computational complexity while preserving the signal quality of the UF-OFDM baseline solution. The proposed technique exploits two main ideas in order to reduce the computational complexity of the UF-OFDM implementation. The first idea consists of separating the processing of each subband and each subcarrier, avoiding redundant operations. In fact, the subband processing requires an IFFT of size $K$ (the total number of subband) for each subcarrier which allows to reduce the computational complexity.

Then, the complexity is further reduced by exploiting a second idea which consists of decomposing the UF-OFDM symbol into 3 distinct parts: a prefix part, a core part and a suffix part. By using the proposed segmentation into subbands and subcarriers, it can be shown that the core part of the UF-OFDM symbol can be computed using a windowing operation followed by an IFFT of size $Q$ (the subband size) for each subband. Furthermore, the suffix part is deduced by subtracting the prefix part from the core part. Since this technique mainly relies on the segmentation of the subband processing and the subcarrier processing, it is referred to as “UF-OFDM, segmentation technique” in the table above.

Such segmentation assumes that the total number of subcarriers $N$ must be a multiple of the subband size $Q$ ($\text{mod}_Q(N) = 0$) if there is more than one subband allocated (for instance, $N = 1024, Q = 16$). In the other cases, the UF-OFDM signal must be up-sampled which increases the computational complexity. However, this complexity remains lower than most of the state-of-the-art techniques, while preserving the signal quality of the original UF-OFDM technique. More details can be found in [FAN16-D52].

- **Receiver Complexity:** In contrast to past publications, stating a 2-$N$ FFT as necessary for UF-OFDM, we have shown in [FAN16-D31] and [R1-165014] that an $N$-point FFT, plus a tail-copy operation which involves a negligible amount of additions, is equivalent. UF-OFDM can additionally use windowing (low-complex) or subband-filtering at the receiver. The subband-filtering can be implemented efficiently in frequency domain using fast convolution techniques, such as overlap and save, as discussed in [R1-165014], close to a dual structure of the low-complex frequency domain transmitter [WS15], applied to the receiver side.

### 6.1.1.2 WOLA

The detailed description of WOLA is given in [Qua15]. In this implementation, no Tail was included for WOLA in TTI transmission. However, a different implementation can assume the tails to be included (or partially included) and hence creates an additional latency, with the advantage of better ICI suppression. WOLA applies half Overlap at transmitter and receiver, and raised cosine window weighting function with window length $L_{\text{wslTX}} = L_{\text{wslRX}} = 72$ samples. Window functions and transmit windowing are illustrated in Figure 6-3 and Figure 6-4, respectively.

![Figure 6-3 Transmit and receive window roll-off. For our WOLA simulations: $L_{\text{wslTX}} = L_{\text{wslRX}} = 72$.](image-url)
Figure 6-4 Transmit windowing for one OFDM symbol. In our simulation settings with $N=1024$, $LCP = 72$ samples the overlapped extension is 36 samples on each side.

From the Latency perspective, as mentioned earlier, since the tails of each TTI are discarded before transmission, WOLA has same latency compared to CP-OFDM.

6.1.1.3 FC-OFDM

Waveform parameters

Waveform parameters of FC-OFDM used in the comparison are given in Table 6-2.

**Table 6-2 Waveform parameters of FC-OFDM**

<table>
<thead>
<tr>
<th>Numerology</th>
<th>15 KHz subcarrier spacing</th>
<th>CP length 72 samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FFT size 1024</td>
<td></td>
</tr>
<tr>
<td>CP-OFDM</td>
<td>no windowing/filtering</td>
<td></td>
</tr>
<tr>
<td>FC-OFDM</td>
<td>window coefficients</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f[k] = \begin{cases} a[k] &amp; k \in [0, L - 1] \ 1 &amp; n \in [L, M - 1] \ b[k] &amp; k \in [M, M + L - 1] \end{cases}$,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>where $\alpha = -\alpha \left(\frac{k-M}{L}\right)^3 \quad \text{and} \quad \beta = \sqrt{1 - \beta^2 [k + M]^3}$, $(\alpha, \beta) = (5, 2)$,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L = 144, M = 1024$</td>
<td></td>
</tr>
</tbody>
</table>

Latency

The FC-OFDM symbol is longer than CP-OFDM symbol by CP samples due to the windowing (illustrated in Figure 6-5). The consecutive FC-OFDM symbols are overlapped in time domain by CP samples, which lead to the same overhead as CP-OFDM case. The latency increase for FC-OFDM compared to CP-OFDM is thus the duration of 1 CP.
**Observation:** The latency increase for FC-OFDM is equivalent to CP duration.

**Complexity**

The complexity increase for FC-OFDM compared to CP-OFDM is due to the windowing. Based on the waveform parameters in Table 6-2, the complexity in terms of the number of real multiplications for CP-OFDM and FC-OFDM is given below.

<table>
<thead>
<tr>
<th></th>
<th>FFT process</th>
<th>window process</th>
<th>complexity based on parameters in Table 6-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP-OFDM</td>
<td>$4M \log_2 M$</td>
<td>0</td>
<td>40960</td>
</tr>
<tr>
<td>FC-OFDM</td>
<td>$4M \log_2 M$</td>
<td>$4T_c$</td>
<td>41536</td>
</tr>
<tr>
<td>Complexity ratio</td>
<td></td>
<td></td>
<td>0.986</td>
</tr>
</tbody>
</table>

**Observation:** The complexity increase for FC-OFDM compared to CP-OFDM is less than 2%.

6.1.1.4 P-OFDM

**Waveform parameters**

To restrict the latency introduced by P-OFDM, we use the short pulse design and apply matched filtering at the transceiver. The transmit and receive filters are illustrated in Figure 6-6, with the filter length 1168 and FFT size 1024. The filter is optimized according to the algorithm in [YZB16] with $TF = 15/14$ and $K = 16/15$. Note that P-OFDM here yields the similar operation as windowing function due to the use of short pulse. Theoretically P-OFDM can also exploit longer pulse shaped design providing better time-frequency localization property.
Latency
Due to the application of short pulse shape, P-OFDM can be considered as the windowed OFDM here. Analogous to the analysis of FC-OFDM, the latency increase for P-OFDM is equivalent to CP duration due to the overlap.

Complexity
Analogous to the analysis of FC-OFDM, complexity introduced by P-OFDM here is due to windowing. Consequently, the complexity increase for P-OFDM compared to CP-OFDM with short pulse shape given in Figure 6-6 is less than 2%.

6.1.1.5 FBMC
Waveform parameters of FBMC used in the comparison are given in Table 6-3. Here FBMC is implemented in FS-FBMC framework.

<table>
<thead>
<tr>
<th>OFDM parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FFT size</td>
<td>1024</td>
</tr>
<tr>
<td>CP length</td>
<td>72</td>
</tr>
<tr>
<td>Sampling Frequency</td>
<td>15.36MHz</td>
</tr>
<tr>
<td>RB bandwidth</td>
<td>180kHz</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FBMC parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter parameters</td>
<td>PHYDYAS, K=4 for scenario 1 and 3</td>
</tr>
<tr>
<td></td>
<td>PHYDYAS K=2 for scenario 2</td>
</tr>
</tbody>
</table>

Latency
For FBMC, the total burst length can be expressed as \( (N_S + K - \frac{1}{2})T \) where K is the overlapping factor, T is the symbol duration and N_S are the number of symbols in the burst. For
OFDM, the total burst length can be expressed as $N_S(T + T_g)$. Compared to CP-OFDM the penalty brought by FBMC is thus roughly $(K - \frac{1}{2}) T$.

**Complexity**

The complexity of FBMC can also be evaluated and compared with CP-OFDM. A simple way to estimate complexity is to evaluate the total number of real multiplications for both options. Assuming split-radix implementation of the IFFT and FFT, the number of real multiplications per multicarrier symbol is given by [Nog11]:

$$C_{FBMC} = 4(2N_{FFT} + (N_{FFT}(\log_2(N_{FFT}) - 3) + 4) + 2KN_{FFT})$$

Normalized with respect to the complexity of the CP-OFDM transceiver, the complexity for FBMC transceiver is then:

- Scenario 1 and scenario 3: 4.86
- Scenario 2: 3.71

### 6.1.1.6 BF-OFDM

**Waveform parameters**

BF-OFDM is parametrized by the filter bank FFT size (M), the carrier bandwidth, the OFDM precoding FFT size (N) and the CP size. The proposed waveform is scalable, and different configurations can be used to optimize the performance w.r.t a given indicator. For a given carrier number M, better frequency resolution can be obtained by increasing the number of subcarrier N per carrier. On the contrary, one can imagine decreasing the number of subcarrier per carrier or increasing the carrier bandwidth to reduce the Rx FFT size in order to allow better resistance in high mobility scenarios. Prototype filter can also be optimized to fulfill a target requirement.

Different parametrizations have been proposed for BF-OFDM, in terms of subcarrier resolution (linked to the precoding stage) and prototype filter length. For scenario 1 and 3, we have set two configurations, where we have used two different subcarrier resolutions (I and II, where II has a slightly higher complexity due to the resolution increase). For scenario 2 where we wanted to keep the symbol duration short, we have used a lower subcarrier resolution and a shorter prototype filter impulse response.

Waveform parameters of BF-OFDM used in the comparison are given in Table 6-4. Waveform parameters of BF-OFDM used in the comparison are given in Table 6-4. In BF-OFDM the sampling frequency is directly derived from the carrier spacing. To maintain compatibility with CP-OFDM, we have set a carrier bandwidth equal to an OFDM PRB. As a consequence, the sampling frequency (180kHz x 128 = 23.04MHz) is different from the one of OFDM (15kHz x 1024=15.36MHz).

<table>
<thead>
<tr>
<th>OFDM parameters</th>
<th>BF-OFDM parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFT size</td>
<td>1024</td>
</tr>
<tr>
<td>CP length</td>
<td>72</td>
</tr>
<tr>
<td>Sampling Frequency</td>
<td>15.36MHz</td>
</tr>
<tr>
<td>RB bandwidth</td>
<td>180kHz</td>
</tr>
</tbody>
</table>

Dissemination level: Public Page 93 / 159
### Scenario 1 and scenario 3

<table>
<thead>
<tr>
<th>[(M, N, CP size), Rx FFT size]</th>
<th>Config I: [(128,48,4), 3072]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier bandwidth</td>
<td>180kHz</td>
</tr>
</tbody>
</table>

### Scenario 2

<table>
<thead>
<tr>
<th>[(M, N, CP size), Rx FFT size]</th>
<th>[(64,32,2), 1024]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier bandwidth</td>
<td>360kHz</td>
</tr>
</tbody>
</table>

### Common parameters

| Sampling Frequency            | 23.04MHz         |
| Prototype filter              | Gaussian filter, with length K equal to CP size |

### Latency

For BF-OFDM, as demonstrated in [Ger17], the total number of samples in a TTI, with $N_S$ symbols in the TTI is: 

$$N_S(N + N_{CP}) - \frac{1}{2}MN + KM.$$  

whereas for CP-OFDM we would have

$$N_S(N_{FFT} + N_{CP–OFDM}).$$  

This means that no penalty can be achieved if BF-OFDM CP overhead $\frac{M N_{CP}}{M N}$ is similar to the one of OFDM $N_{CP–OFDM} \frac{N_{CP–OFDM}}{N_{FFT}}$.

### Complexity

Complexity evaluation is done by estimating the amount of complex multiply necessary to perform the transmitter function. Let $N_{FFT}$ denotes the FFT size. Concerning OFDM, assuming a Cooley Tukey FFT algorithm, the number of complex multiplication is equal to:

$$C_{OFDM} = \frac{N_{FFT}}{2} \log_2(N_{FFT})$$

The complexity of a BF-OFDM transmitter is the sum of the complexity of the pre-compensation stage, the $M$ IFFT stages and the filter bank stage applied to $N + N_{CP}$ set of samples. By considering that half of the inputs of the IFFT modulators are zeros and the sample repetition provided by the CP, we can express the complexity by [Ger17]:

$$C_{BF–OFDM} = \frac{MN}{2} + M \left( \frac{N}{2} + \frac{N}{2} \log_2 \left( \frac{N}{2} \right) \right) + N \frac{M}{2} \log_2(M) + KMN$$

The double-stage structure of the transmitter accounts for the complexity difference with CP-OFDM. However, the BFOFDM transmitter provides an embedded sub-band filtering that is not the case of for CP-OFDM. Besides, additional complexity (with the pre-distortion) has been added in order to reduce the receiver to a simple FFT. The increase in complexity is thus justified and reasonable considering those two aspects.

Normalized with respect to the complexity of the received FFT, we have then the normalized following complexity for BF-OFDM transmitter:

- Scenario 1 and scenario 3, case 1: 3.73
- Scenario 1 and scenario 3, case 2: 3.59
- Scenario 2: 3.2
6.1.2 Simulator Calibration

The challenge to evaluate different waveform candidates simulated by different simulators is how to align and achieve a fair comparison. To deal with it, FANTASTIC-5G partners who are interested in waveform calibration have agreed on a set of calibration cases, implemented CP-OFDM or SC-FDMA for each case in individual simulator and provided calibration results. Note that the partners who provide the calibration results are not limited to the new waveform proposers.

The block diagram of simulation chain is given in Figure 6-7. After the Turbo encoding, coded bits are first mapped to symbols and resources according to subcarrier spacing, MCS, FFT and CP length, etc. Then waveform modulator is applied, including OFDM modulator or new waveform modulator. The signal is transmitted through multipath channels characterized by a parameterizable channel delay profile [38,900]. At the receiver side, the signal is first demodulated by the corresponding waveform demodulator and resource demapping. Using the ideal channel knowledge, the bits are recovered by MMSE equalizer, max-log demapper, and Turbo decoder. BLER is deduced from the decoder outputs. Note that for simplicity of simulations and avoiding multiple effects on waveform performance, we assume there are no reference signals, pilots, and the control overhead. HARQ is not modeled and SISO channel is assumed. At the receiver side, linear MMSE equalizer with perfect noise variance is applied and Max-Log-MAP with maximum 8 iterations is employed for Turbo decoding.

![Figure 6-7: Block diagram of simulation chain for waveform calibration and comparison.](image)

We have identified three cases for simulator calibration with simulation parameters specified in Table 6-5. The calibration cases are aligned with 3GPP Tdoc [R1-165989, R1-165859] for case 1a, 1b, and 2.

Case 1: DL single numerology case (CP-OFDM).
Case 2: UL single numerology case (CP-OFDM and SC-FDMA)
Case 3: DL mix numerology case (CP-OFDM).

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>Case 1 and Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier Frequency</td>
<td>4 GHz</td>
<td>4 GHz</td>
</tr>
<tr>
<td>System Bandwidth</td>
<td>10 MHz</td>
<td>10 MHz</td>
</tr>
<tr>
<td>TTI length</td>
<td>1 ms</td>
<td>1 ms</td>
</tr>
<tr>
<td>Frame size</td>
<td>10 ms</td>
<td>10 ms</td>
</tr>
<tr>
<td>Frame structure</td>
<td>72 samples for CP, 14 symbols per subframe for 15 KHz case: 72 samples for CP, 14 symbols per subframe for 20 KHz case: 36 samples for CP, 28 symbols per subframe</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Subcarrier spacing</td>
<td>Single numerology case: 15 KHz 1 target UE: 15 KHz 1 interfering UE: 30 KHz only target UE is evaluated. 60 KHz Guard band between target UE and interfering UE</td>
<td></td>
</tr>
<tr>
<td>FFT size</td>
<td>1024 for 15 KHz subcarrier spacing 1024 for 15 KHz subcarrier spacing 512 for 30 KHz subcarrier spacing</td>
<td></td>
</tr>
<tr>
<td>Number of resource blocks</td>
<td>50 (case 1 DL)/4(case 2 UL) 4 target UE/2 interfering UE</td>
<td></td>
</tr>
<tr>
<td>Antenna configuration</td>
<td>1x1 1x1</td>
<td></td>
</tr>
<tr>
<td>MCS</td>
<td>64QAM 1/2 64QAM 1/2</td>
<td></td>
</tr>
<tr>
<td>Channel estimation</td>
<td>Ideal (no pilot) Ideal (no pilot)</td>
<td></td>
</tr>
<tr>
<td>Channel model</td>
<td>TDLc-300 (3kmh) TDLc-300 (3kmh)</td>
<td></td>
</tr>
</tbody>
</table>

The calibration results for three cases are listed in Figure 6-8, Figure 6-9, Figure 6-10 and Figure 6-11, respectively. Note that for each calibration case, not all the partners participated in this exercise but only those who are interested and willing to contribute in the next-round comparisons. As observed from these figures, we reach the convergence in the calibration phase for all calibration cases.
Figure 6-8: BLER performance of CP-OFDM for case 1.

Figure 6-9: BLER performance of CP-OFDM for case 2.
Figure 6-10: BLER performance of SC-FDMA for case 2.

Figure 6-11: BLER performance of CP-OFDM for case 3.
6.1.3 Waveform comparisons

The principles of selecting the waveform evaluation scenarios are two-fold. First, we target at non-eMBB scenarios primarily, which support flexible numerology coexistence and asynchronicity. Second, we adapted these scenarios aligning with the parallel discussions [R1-163558, R1-166031] in 3GPP and considered more challenging settings if necessary. Three scenarios are considered for waveform comparison:

- Scenario 1: UL asynchronous transmission (Single numerology case). This scenario refers to the potential use case in machine-type communications (e.g., MMC or MCC services) where uplink transmission can rely on relaxed time synchronization. Such scheme leads to reduced latency, signaling and power consumption for uplink transmission.

Note that this scenario includes two setups depending on the guard tone bandwidth and timing offset parameters: scenario 1-1 (30 kHz guard tone bandwidth and timing offset 0/128/128 samples) and scenario 1-2 (120 KHz guard tone bandwidth and timing offset 0/128/512 samples).

![Figure 6-12: Illustration of scenario 1 and scenario 3: UL asynchronous transmission](image)

- Scenario 2: Downlink high speed transmission (Single numerology case) which reflects the requirement of V2X services with high mobility terminals. This scenario is illustrated in Figure 6-13, with detailed parameters given in Table 6-6.

![Figure 6-13: Illustration of scenario 2: Downlink high speed scenario](image)

- Scenario 3: uplink synchronous transmission (mixed numerology case), which is a general use case for a multi-service support of different PHY configurations in the same band. This scenario is illustrated similarly with Figure 6-12, while UE2 and UE3 apply different numerologies as UE1. Detailed parameters are given in Table 6-6.

Note that this scenario includes four setups depending on the number of users, the guard tone bandwidth and modulations: scenario 3-1 (0 kHz guard tone bandwidth, 3UEs, and MCS-16QAM), scenario 3-2 (120 kHz guard tone bandwidth, 3UEs, and MCS-16QAM), scenario 3-3 (0 kHz guard tone bandwidth, 2UEs, and MCS-16QAM and 64QAM), and scenario 3-4 (120 kHz guard tone bandwidth, 2UEs, and MCS-16QAM and 64QAM).
The link performances for above scenarios are illustrated as follows.

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>Scenario 1-1 and Scenario 1-2</th>
<th>Scenario 2</th>
<th>Scenario 3-1 and Scenario 3-2</th>
<th>Scenario 3-3 and Scenario 3-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier Frequency</td>
<td>4 GHz</td>
<td>4 GHz</td>
<td>4 GHz</td>
<td>4 GHz</td>
</tr>
<tr>
<td>System Bandwidth</td>
<td>10 MHz</td>
<td>10 MHz</td>
<td>10 MHz</td>
<td>10 MHz</td>
</tr>
<tr>
<td>TTI length</td>
<td>1 ms</td>
<td>1 ms</td>
<td>1 ms</td>
<td>1 ms</td>
</tr>
<tr>
<td>Frame size</td>
<td>10 ms</td>
<td>10 ms</td>
<td>10 ms</td>
<td>10 ms</td>
</tr>
<tr>
<td>Frame structure</td>
<td>LTE-A (6.7% CP)</td>
<td>LTE-A</td>
<td>LTE-A</td>
<td>LTE-A</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
<td>15 KHz</td>
<td>15 KHz</td>
<td>Mixed numerology case: 15 KHz/30 KHz</td>
<td>Mixed numerology case: 15 KHz/30 KHz</td>
</tr>
<tr>
<td>FFT size</td>
<td>1024</td>
<td>1024</td>
<td>1024 for 15 subcarrier spacing 512 for 30 subcarrier spacing</td>
<td>1024 for 15 subcarrier spacing 512 for 30 subcarrier spacing</td>
</tr>
<tr>
<td>Number of resource blocks</td>
<td>FDMA: 4/4/4 RBs for UE1/UE2/UE3</td>
<td>Single numerology case: 50 PRBs</td>
<td>Mixed numerology case: 2/4/2 RBs for UE1/UE2/UE3 (30/15/30 KHz)</td>
<td>Mixed numerology case: 2/4 RBs for UE1/UE2 (30/15 KHz)</td>
</tr>
<tr>
<td>Number of UEs</td>
<td>3: FDMA (only UE1 is evaluated while UE2 and UE3 are interferers.)</td>
<td>1</td>
<td>3 (only UE2 with 4RB is 3 (only UE2 with 4RB evaluated while is evaluated while UE1 UE1/UE3 are interferers is interferer located at located at the left/right the left side of UE2) side of UE2)</td>
<td></td>
</tr>
<tr>
<td>Guard tone bandwidth</td>
<td>30/120 KHz</td>
<td>Zero</td>
<td>0/120 KHz</td>
<td>0/120 KHz</td>
</tr>
<tr>
<td>Antenna configuration</td>
<td>1x1</td>
<td>1x1</td>
<td>1x1</td>
<td>1x1</td>
</tr>
<tr>
<td>Modulation</td>
<td>16QAM 1/2</td>
<td>16QAM: 1/2</td>
<td>16QAM 1/2</td>
<td>16QAM 1/2 (UE1) and 64QAM 1/2 (UE2)</td>
</tr>
<tr>
<td>Timing offset of interfering user</td>
<td>0/128/128 Samples (UE1/UE2/UE3), 0/128/512 Samples (UE1/UE2/UE3)</td>
<td>zero</td>
<td>zero</td>
<td>zero</td>
</tr>
<tr>
<td>Channel estimation</td>
<td>Ideal</td>
<td>Ideal</td>
<td>Ideal</td>
<td>Ideal</td>
</tr>
<tr>
<td>Channel model</td>
<td>TDL-A DS30ns/TDL-C DS1000ns</td>
<td>EVA 400 kmh</td>
<td>TDL-C DS300ns</td>
<td>TDL-C DS300ns</td>
</tr>
<tr>
<td>Power Offset</td>
<td>0/0/0 dB</td>
<td>0 dB</td>
<td>0/0/0 dB</td>
<td>0/0/0 dB</td>
</tr>
</tbody>
</table>

Table 6-6: Parameters for waveform comparison scenarios
UL Asynchronous Transmissions (16QAM, 30 KHz GCs, 0-128-128 samples)

Figure 6-14 BLER performance of Scenario 1-1.

UL Asynchronous Transmissions (16QAM, 120 KHz GCs, 0-128-512 samples)

Figure 6-15 BLER performance of Scenario 1-2.
Observations from scenario 1-1 and 1-2: for small guard bands, sub-band-wise filtering schemes outperform windowing schemes, while FBMC with OQAM signaling provides best spectral confinement. However, with increasing the guard band, the performance curves of all schemes converge.

![Figure 6-16 BLER performance of Scenario 2.](image)

Observations from scenario 2: for high speed scenario, windowing schemes and BF-OFDM provide slight gain compared to CP-OFDM in the high-SNR region, while FBMC-OQAM is subject to a performance loss in high-SNR region. This is because all these schemes have better frequency localization which leads to reduced inter-carrier interference (ICI). However, for FBMC-OQAM, due to the increased symbol length, the interference due to channel delay spread becomes more predominant compared with Doppler effect.
Figure 6-17 BLER performance of Scenario 3-1.

Figure 6-18 BLER performance of Scenario 3-2.
Figure 6-19 BLER performance of Scenario 3-3.

Figure 6-20 BLER performance of Scenario 3-4.
Observations from scenario 3: when using low to mid modulation order, sub-band based schemes outperform windowing schemes for small guard band, while the gap gradually diminishes for larger guard band. When using high modulation order, sub-band-wise filtering scheme outperforms windowing scheme for small guard band only in high SNR region. The performance of both schemes converge for larger guard band, while FBMC with OQAM signaling performs best in this case.

6.2 BF-OFDM

Block Filtered OFDM waveform will combine most of the advantages of the OFDM and FBMC waveforms at the price of a slightly complexity increase at the transmitter side while keeping a simple OFDM receiver. Spectral localization and performance in multi-user scenario will be enhanced with respect to OFDM and same equalization and MIMO techniques can be considered while keeping typical OFDM receiver. Furthermore, CP insertion ensures circularity of the signal and thus offers the same performance than legacy CP-OFDM in presence of multipath channel. The proposed solution is also scalable, which paves the way for future multi-service scenarios.

**Transmitter**

The transmitter scheme is depicted in Figure 6-21. We denote M the number of carriers, and N the number of subcarriers. There are N subcarriers per carrier. To maintain orthogonality, only N/2 subcarriers are bearing data per carrier (Note that there is no loss in spectral efficiency, since we can play on the value of CP and N to achieve the same spectral efficiency as in LTE, e.g. see subsequent subsections). The subcarrier framing depends on the carrier index parity. For each BF-OFDM symbol, N/2 data are mapped in frequency domain, an IFFT of size N is applied to each carrier, and a CP is appended to ensure the circularity of the received signal. The output of the M stages is then fed to a filter bank parametrized by a prototype filter with an overlapping factor K. As in typical FBMC transmitter, an overlap and sum operation is done in the PolyPhase Network (PPN). It ensures that symbols are transmitted each M/2 samples. A predistortion stage is applied to each subcarrier, and has two objectives: i) ensures that the receiver stage can be based on a single FFT ii) compensates the effect of the distortion (phase and amplitude) of the filter bank. The insertion of a cyclic prefix aims to avoid inter-symbol interference (ISI) and makes the equalization in the frequency domain simple. ISI is the result of the convolution of the multipath channel and the prototype filter. Therefore, a direct link between the minimum CP size and the overlapping factor exists and is expressed as N_{CP} ≥ 2K-1. It should be noted that in practice the size of the CP can be reduced at the price of a negligible increase of the interference level.

![Figure 6-21: BF-OFDM transmitter diagram](image-url)
**Receiver**

The receiver scheme is depicted in Figure 6-22. It consists of selecting a window of size MN/2 each N\(_{CP}\) M/2+MN/2 samples. It is followed by a MN/2 FFT stage. The receiver is thus with low complexity, similar to the one used for a CP-OFDM receiver.

![Figure 6-22 BF-OFDM receiver diagram](image)

**Complexity**

Complexity evaluation is done by estimating the amount of complex multiplications necessary to perform the transmitter function. Concerning OFDM, assuming a Cooley-Tukey FFT algorithm, the number of complex multiplications is equal to:

\[
C_{\text{OFDM}}^{\text{OFDM}} = \frac{N_{\text{FFT}}}{2}\log_2(N_{\text{FFT}}) + C_f^{\text{OFDM}}
\]

where \(C_f^{\text{OFDM}}\) is the complexity associated with the filtering function applied to an OFDM signal.

The complexity of a BF-OFDM transmitter is the sum of the complexity of the pre-compensation stage, the \(M\) OFDM stages and the filter bank stage applied to \(N + N_{CP}\) set of samples where \(N_{CP}\) is the CP length. Using the properties that \(N/2\) samples are zeros at the input of the OFDM stage, the complexity is reduced to:

\[
C_{\text{BF-OFDM}}^{\text{BF-OFDM}} = \frac{MN}{2} + M\left(\frac{N}{2} + \frac{N}{2}\log_2\left(\frac{N}{2}\right)\right) + \frac{(N + N_{CP})M}{2}\log_2(M) + KM(N + N_{CP})
\]

An evaluation of the complexity is done using typical parameter of \(M\) and \(N\). Results are provided in Table 6-7. The complexity is normalized with respect to the complexity of the received FFT. We assume \(K=4\) and \(N_{CP} = 4\). For a typical BF-OFDM scenario where OFDM FFT size is set to \(N = 64\) and FBMC FFT size is set to \(M = 128\), the complexity of the transmitter is 4 times more complex than the received FFT. This complexity increase is reasonable and makes sense if the frequency localization of the waveform is of interest. Indeed, with our proposed waveform, the filtering operation is embedded; contrary to classic OFDM where additional filter stages can dramatically increase the overall complexity.

**Table 6-7 Complexity evaluation normalized by the complexity of the received FFT**

<table>
<thead>
<tr>
<th>M/N</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>5.71</td>
<td>5.31</td>
<td>5.00</td>
<td>4.75</td>
<td>4.55</td>
</tr>
<tr>
<td>32</td>
<td>4.88</td>
<td>4.58</td>
<td>4.35</td>
<td>4.16</td>
<td>4.00</td>
</tr>
<tr>
<td>64</td>
<td>4.39</td>
<td>4.16</td>
<td>3.98</td>
<td>3.82</td>
<td>3.69</td>
</tr>
<tr>
<td>128</td>
<td>4.08</td>
<td>3.89</td>
<td>3.74</td>
<td>3.61</td>
<td>3.50</td>
</tr>
<tr>
<td>256</td>
<td>3.85</td>
<td>3.70</td>
<td>3.57</td>
<td>3.46</td>
<td>3.37</td>
</tr>
</tbody>
</table>
6.3 Low Latency Channel Coding

One way to reduce the channel coding related latency in the receiver chain is to use codes with short lengths. However, the choice of the channel codes for short lengths becomes critical as the error correction performance degrades with the decreasing block length. In this section, we evaluate the performance of some important classes of channel codes for message lengths $k \leq 512$ bits, together with their decoder complexity. For the sake of simplicity, we only consider a single rate ($R=1/2$, block length $n=2k$) on AWGN channel with BPSK modulation, and compare the results with the finite length bound of Polanskiy et al. [PPV], as Shannon capacity is not an accurate metric for the short message length regime. We consider the following classes of channel codes:

a) Convolutional Codes: We consider Tail-biting Convolutional Codes (TBCC) with $m=6$ and $m=10$ memory elements ($S=2^6$ states) with generators $(133,171)$ and $(4672,7542)$, respectively. We use a Viterbi decoder on the trellis extended by a decoding depth $t = 6(m+1)$ on both sides.

b) Turbo Codes: We consider LTE Turbo codes (with $S=2^3$ state constituent convolutional codes) with LTE rate matching, and decode with $I=10$ turbo iterations using scaled-max-log-map decoding (scaling factor 0.75).

c) Binary LDPC Codes: We use protograph based AR3A codes [DDJ05] with average variable node degree $d_v=2.8$ and average check node degree $d_c=4.66$. Protographs are lifted by progressive edge growth and decoded using scaled min-sum message passing decoder (scaling factor 0.8) with $I=100$ iterations.

d) Non-Binary LDPC Codes: We use regular NB-LDPC codes in GF($2^p$) with variable node degree $d_v=2$ and check node degree $d_c=4$ with a parity check matrix of size $M_q \times N_q$ where the coefficients of the parity check matrix is selected randomly from the set of optimized coefficients in [PFD08]. We decode with a message passing decoder in probability domain [CML12] with $I=100$ iterations.

e) Polar Codes: We used Polar+CRC codes where the set of frozen bits are obtained numerically using density evolution [MT09]. We use $z=8$ bit CRC for $k<128$, and $z=16$ bit CRC otherwise. We decode with a successive cancellation list (SCL) decoder with list size $L=32$ and $L=1024$.

The following table shows the decoder complexity calculation for the considered codes in terms of basic operations. For more details, the readers are referred to [ILX16].

<table>
<thead>
<tr>
<th>Channel Code</th>
<th>Additions and Subtractions</th>
<th>Multiplications</th>
<th>Divisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBCC</td>
<td>$6S(k + 2t)$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Turbo Code</td>
<td>$(2l - 1) [(k + m)(2S)(2 + N) + k(4S + 1)] + 4(l - 1)k$</td>
<td>$2(l - 1)k$</td>
<td>-</td>
</tr>
<tr>
<td>Binary LDPC</td>
<td>$ln_{2} d_v$</td>
<td>$I(n - k)(2d_v + 3)$</td>
<td>-</td>
</tr>
<tr>
<td>Non-Binary LDPC GF($2^p$)</td>
<td>$I[N_q d_v(q - 1) + M_q d_c q p ]$</td>
<td>${N_q d_v(d_v - 1)q + 3M_q (d_c - 2)qN_q d_v q }$</td>
<td>$ln_q d_v q$</td>
</tr>
<tr>
<td>Polar + $z$ bit CRC</td>
<td>$\frac{ln}{2} \log n + L(k + n) + 3Ln \log n + L(k + n) + 2Lk \log 2L + (k + z) L$</td>
<td>$Ln \log n$</td>
<td>-</td>
</tr>
</tbody>
</table>

Dissemination level: Public
The following figures show the error correction performance and the complexity analysis of the presented codes. One observes that LTE Turbo codes and binary LDPC codes show similar performance and have a gap to the finite length bound in this regime. More advanced channel coding schemes like Polar+CRC codes with list decoding or non-binary LDPC codes with non-binary message passing algorithm can outperform Turbo and LDPC codes, especially for very short lengths. TBCC with high memory also shows very good performance, however only for very short lengths (k=64) and its performance degrade with increasing message length. One also observes that the excellent performance of Polar+CRC (L=1024) and NB-LDPC codes come at a cost of increased decoder complexity.

Note that the decoding performance of the discussed codes mainly depends on two aspects: code properties and decoder properties. By using more complex decoders, the performance gets close to maximum-likelihood (ML) performance and the code properties (such as minimum distance) become more important. On the other hand, the relatively poor performance of Turbo and LDPC codes in the short message length regime is partly due to the suboptimal iterative decoding method, and may be improved by utilizing more advanced decoders.

Figure 6-23 - Required Eb/N0 to achieve a target BLER with channel codes of rate R=1/2 on AWGN channel with BPSK modulation for different message lengths k≤512, together with Shannon’s BPSK constrained capacity, and the Gaussian approximation of Polynaskiy’s finite length bound.

Figure 6-24 - Decoder Complexity in terms of number of basic operations for different message lengths.
6.4 Enhanced Turbo Codes for 5G

**Topic motivation**

As already stated in [FAN16-D31], the conventional FEC coding component of LTE/LTE-A is not designed to answer favorably to service requirements with stringent reliability and latency constraints. A known issue for the LTE Turbo Code (TC) resides in its poor performance at low error rates when transmitting data with coding rates higher than 1/3. This is due to the well-known “error floor” effect, which can be observed when the TC is punctured with the rate matching mechanism [CNB+08]. This detrimental effect results in the frequent resort to retransmissions through the HARQ mechanism. Moreover, these codes were not originally designed targeting best short packet performance. Therefore, we have made some changes to the LTE codes to make it able to cope with the new requirements introduced by 5G scenarios.

Nevertheless, provided that they would be able to guarantee lower error rates when punctured, TCs could remain promising channel coding candidates for 5G. As introduced by Berrou et al. [BGT93], the minimum Hamming distance ($d_{\text{min}}$) of a TC is not only defined by its constituent encoders but also fixed through the TC interleaver. Therefore, we have investigated the joint design of interleavers and puncturing patterns for TCs in order to guarantee low error floors and good convergence thresholds. As a result, a new puncturing constraint related to parity puncturing is proposed for the design of TC interleavers. The work focused on the Almost Regular Permutation (ARP) interleaver model [BSD+04], adopted in standards like IEEE 802.16 WiMAX.

**Interleaver model and design criteria**

In this study, TC interleaving is defined as follows: the interleaver reads the symbols from an input vector $d$ and writes them to an output vector $d'$ of size $K$ corresponding to the information frame length. A symbol read out from address $\Pi(i)$ in $d$ is written to address $i$ in $d'$, $\Pi(i)$ denoting the interleaver function.

When using Circular Recursive Systematic Convolutional (CRSC) codes [WBR01] as constituent codes of the TC, $d$ and $d'$ can be represented by circles.

The ARP interleaver structure is derived from the Regular Interleaver (RI):

$$\Pi(i) = (P \times i) \mod K$$

Where $P$ is the RI period that must be relatively prime to $K$.

A degree of disorder is introduced into the permutation through a vector of shifts $S$, leading to the ARP:

$$\Pi(i) = (P \times i + S(i \mod Q)) \mod K$$

$S$ has length $Q$ and to guarantee the bijectivity of the ARP interleaver function, $Q$ must be a divisor of $K$. Two main criteria have to be considered for the design of a TC interleaver: the Hamming distance spectrum of the TC and the correlation between the channel information and extrinsic data at the input of each component decoder. Two measurable parameters related to these criteria have been considered: minimum span $S_{\text{min}}$ [BSD+04] and correlation girth $g$ [Gar15].

In order to simplify the interleaver parameters selection, the interleaver structure is divided into different groups of addresses or layers that are progressively placed to complete the interleaver structure. It can be shown that:

$$\Pi(i + Q) \mod Q = \Pi(i) \mod Q$$

Therefore, $Q$ groups of permutation addresses are identified in the whole interleaver size $K$, each one corresponding to a given modulo $Q$ value. Thus, the interleaver parameters can be
selected by dividing the permutation addresses into \( Q \) groups or layers with index \( l = 0, \ldots, Q - 1 \).

In Figure 6-25, the interleaver structure is divided into its \( Q \) corresponding layers and represented in a circle with the addresses of \( d \) and \( d' \) in its inner and outer parts, respectively. The final permutation addresses of each layer \( l \) are defined by choosing its corresponding shift value \( S(l) \). The layer number \( l \) to which address \( i \) in \( d' \) belongs is determined as \( l = i \mod Q \). For each layer \( l \), we propose to express the shift value as:

\[
S(l) = T_l + A_l Q
\]

where \( T_l = 0, \ldots, Q - 1 \) and \( A_l = 0, \ldots, \frac{K}{Q} - 1 \) correspond to the inter and intra-layer shifts for layer \( l \), respectively.

Figure 6-25: Possible representation of the ARP interleaver structure, with \( S(0), \ldots, S(Q - 1) = 0 \) and \( K = 4Q \), when using CRSC constituent codes.

The inter-layer shift \( T_l \) defines the position that a given layer takes among the \( Q \) different layers as shown in Figure 6-26(a), whilst the intra-layer shift \( A_l \) defines the rotation that a given layer performs inside the considered layer as shown in Figure 6-26(b). Thus, the value of \( S(l) \) can be obtained first by selecting the value of inter-layer shift \( T_l \) and then selecting the value of intra-layer shift \( A_l \). This layered construction of ARP interleavers simplifies the validation of minimum span and correlation girth criteria in the interleaver design, since these criteria are verified each time a new layer is placed.

Figure 6-26: (a) Different inter-layer shifts for layer \( l = 0 \), with \( K = 4Q \). (b) Possible intra-layer shifts for layer \( l = 0 \), with \( T_0 = 0 \) and \( K = 4Q \).

Considering the above introduced shift selection method, for a given set of design parameters \( (S_{\text{min}} \) and \( g \) targets – chosen a few points below the upper bounds previously given, \( K, R, \)
polynomial generators, and puncturing mask), the proposed interleaver design strategy involves the following steps:

1) Select the candidate values for $P$: The set of admissible values for $P$ is the group of integers from 1 to $K$ relatively prime to $K$. Only the $C$ candidate values for $P$ ensuring a minimum span value greater than or equal to $S_{\text{min}}$ target, considering a RI structure, are selected.

2) Select the $Q$ shift values for each candidate for $P$: For each candidate for $P$, layer $l$ is placed by computing a value for $T_l$ and $A_l$, fulfilling puncturing constraints if any (cf. next section). For this value, the $S_{\text{min}}$ and $g$ are evaluated. If they are equal to or higher than $S_{\text{min}}$ and $g$ targets, one can move on to layer $l + 1$. Otherwise, another value for $S(l)$ has to be evaluated. This process is performed until the whole group of $Q$ shift values are determined.

3) Select the best ARP interleaver candidate: The best candidate for TC interleaver is selected from the group of candidates previously generated by comparing their Hamming distance spectra. The ARP interleaver candidate with the best TC Hamming distance spectrum is chosen.

**Interleaving with puncturing constraints**

Usually, high coding rate TCs are obtained by puncturing parity bits. Nevertheless, it has already been observed that puncturing well-chosen systematic bits can increase the minimum Hamming distance and reduce the convergence threshold of TCs [AR99], [MKK99]. The selection of the systematic bits to be punctured must be performed carefully, in order to prevent the punctured constituent codes of the TC from having a minimum distance equal to one or even zero.

**Puncturing Mask Selection**

In this study, a periodic puncturing pattern with period $M$ is considered. The same CRSC code with coding rate 1/2 is used for the two constituent codes of the TC. Thus, the puncturing mask is composed of two vectors of length $M$, corresponding to the puncturing positions in the data and parity vectors. The puncturing mask is defined according to the target code rate $R$ of the TC and to the puncturing period $M$. For given $R$ and $M$, the data puncturing rate $DPR$ can take $M + 1$ different values: $DPR = m/M$, $m = 0, \ldots, M$. However, the values of $DPR$ are restricted to those ensuring a CRSC code rate $R_c$ smaller than 1, to be able to reconstruct $d$ from the encoded sequence. The puncturing mask design procedure proposed in this work involves the following steps:

1) Find the best puncturing pattern for each $DPR$ value: The best puncturing mask for each $DPR$ value is the one generating the best CRSC Hamming distance spectrum (i.e., highest distance values in the first spectrum terms and minimal number of codewords at these distances).

2) Carry out a mutual information exchange analysis to select a restricted set of puncturing masks: the best puncturing masks in terms of convergence performance are selected. Then, using the chosen puncturing masks, the error rate performance of the TC under uniform interleaving is analyzed. The mask providing the best tradeoff between convergence threshold and error floor performance is finally selected.

**Determination of Data and Parity Puncturing Constraints**

To avoid poor puncturing patterns in the second (interleaved) constituent CRSC code of the TC, a Data Puncture-Constrained (DPC) interleaver must guarantee the same data puncturing pattern in both constituent CRSC codes [CNB+08]. During the turbo decoding process, extrinsic information from a given constituent decoder is generated based on its received parity sequence and is sent to the other constituent decoder via the interleaver/deinterleaver as a priori information on data. The extrinsic information computed from unpunctured parity positions is expected to be more reliable than the one generated from punctured parity positions. In order to illustrate this conjecture, we have plotted a conventional EXIT chart for a parity-only punctured version of a TC. As shown in Figure 6-27, the EXIT curve obtained from data at unpunctured
parity positions has a more opened tunnel than the one obtained from data at punctured parity positions. Since extrinsic information is used as a priori information on data, a possible strategy for the interleaver construction involves connecting the positions with highly reliable extrinsic information to the positions with less reliable extrinsic information, which are more prone to errors. This connection strategy aims to spread the correction capability of the TC over the whole data block.

Figure 6-27: Comparison of EXIT charts computed from the complete data frame and from punctured and unpunctured parity positions at the Signal-to-Noise (SNR) decoding threshold of the TC over the AWGN channel.

For a given puncturing mask, the proposed connection strategy is defined by the following steps:

1) **Identify free parities in a puncturing period M**: Free parities are defined as non-punctured parities corresponding to non-punctured data symbols.

2) **Identify error-prone data positions in a puncturing period M**: To this end, additional data puncturing is introduced at the remaining non-punctured data positions, and the resulting CRSC distance spectrum is evaluated. The number of additional punctured data positions corresponds to the number of free parities. The additional punctured data positions leading to the poorest CRSC distance spectrum (i.e., the lower $d_{\text{min}}$ and the higher number of codewords at that distance or multiplicity) are then labeled as the most error-prone. Note that additional data puncturing is only introduced to identify error-prone data positions and is then removed from the puncturing mask.

3) **Apply the connection strategy**: The proposed strategy involves connecting free parities of a CRSC code to the most error-prone data positions of the other one.

In the ARP model, puncturing constraints are included via the $T_l$ values, to simplify their inclusion, $Q$ is set as a multiple of $M$ (in this study, $Q = M$). Since the layer order of the ARP interleaver is $Q$-periodic, the validation of puncturing constraints in a puncturing period $M$ is a sufficient condition for their validation in the whole data sequence. In the following sections, an interleaver including the proposed parity puncturing constraint on top of the abovementioned data puncturing constraint is named Data and Parity Puncture-Constrained (DPPC) interleaver.

**Application example**

The previous design guidelines were applied to coding rates 2/3 and 4/5, for $K = 1504$ and constituent codes CRSC(1,15/13)$_s$. Error rate performance was simulated in the AWGN channel.
with a BPSK modulation and a maximum of 16 decoding iterations with the maximum a posteriori probability (MAP) algorithm. Only the design for $R = 2/3$ is detailed.

**Puncturing Mask Selection**

Puncturing mask design was carried out for puncturing period $M$ equal to 8. Table 6-8 lists the distance spectrum of the CRSC constituent code for the best mask at each DPR value. Only DPR values with even values of $m$ were considered, leading to symmetric puncturing masks. The analysis of the distance spectra shows that DPR values higher or equal to 6/8 should be avoided since the rate of the constituent CSRC code is greater than 1 for these values of DPR. Then, the convergence behavior with the different puncturing masks is analyzed in the TC structure. Figure 6-28 shows the modified EXIT chart of the TC evaluated at the Signal to Noise Ratio (SNR) decoding threshold of the code with puncturing mask $DPR = 0$. The puncturing mask providing better convergence than the $DPR$-0 mask, due to its (IA, IE) crossing point closer to (1,1), corresponds to $DPR$ value 2/8. In this example, since only one puncturing mask candidate remains, there is no need to plot the TC error rate performance under uniform interleaving to decide between them.

Table 6-8: Best CRSC distance spectrum for each DPR, corresponding codeword multiplicities at distance $d$, $\alpha(d)$, and puncturing masks (0 = punctured, 1 = unpunctured), $R = 2/3$, $M = 8$.

<table>
<thead>
<tr>
<th>DPR</th>
<th>$R_c$</th>
<th>$\alpha(0)$</th>
<th>$\alpha(1)$</th>
<th>$\alpha(2)$</th>
<th>$\alpha(3)$</th>
<th>$\alpha(4)$</th>
<th>Data</th>
<th>Parity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.80</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>89</td>
<td>11111111</td>
<td>10100000</td>
</tr>
<tr>
<td>2/8</td>
<td>0.88</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>62</td>
<td>566</td>
<td>01111110</td>
<td>11000001</td>
</tr>
<tr>
<td>4/8</td>
<td>1.00</td>
<td>0</td>
<td>8</td>
<td>64</td>
<td>482</td>
<td>3616</td>
<td>11110000</td>
<td>11100001</td>
</tr>
<tr>
<td>6/8</td>
<td>1.14</td>
<td>1</td>
<td>36</td>
<td>670</td>
<td>12122</td>
<td>219196</td>
<td>01000100</td>
<td>11011100</td>
</tr>
</tbody>
</table>

Figure 6-28: Extrinsic information exchange between constituent codes of the TC at $E_b/N_0 = 1.6$ dB with 16 TC iterations, $K = 1504$, and $R = 2/3$ under uniform or DPC uniform interleaving over the AWGN channel.
Identification of Parity Puncturing Constraints

In the selected puncturing mask, only the parity at position 1 is a free parity for \( M = 8 \) (see Table 6-8, data and parity positions range from 0 to 7). Table 6-9 lists the different distance spectra obtained including an additional punctured data symbol from positions 1 to 6 in the data puncturing mask 0111110. According to these results, 4 is the most error-prone data position in the puncturing period. Therefore, the interleaver must ensure that data symbols at positions 1 and 4 in \( d \) are interleaved to positions 4 and 1 in \( d' \), respectively.

Table 6-9: CRSC distance spectrum of the \( DPR-2/8 \) mask when one additional data bit is punctured, \( \alpha(d) \) is the multiplicity of codewords at distance \( d \). Considered parity puncturing mask is 11000001 (0 = punctured, 1 = unpunctured).

<table>
<thead>
<tr>
<th>( \alpha(0) )</th>
<th>( \alpha(1) )</th>
<th>( \alpha(2) )</th>
<th>( \alpha(3) )</th>
<th>( \alpha(4) )</th>
<th>Data punct. mask</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1880</td>
<td>1060320</td>
<td>465121494</td>
<td>001111110</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>4000</td>
<td>2003510</td>
<td>671273377</td>
<td>010111110</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>3015</td>
<td>1151175</td>
<td>294778989</td>
<td>011011110</td>
</tr>
<tr>
<td>0</td>
<td>8</td>
<td>140</td>
<td>2229</td>
<td>35176</td>
<td>011101110</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2256</td>
<td>1275364</td>
<td>320347391</td>
<td>011110110</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>3019</td>
<td>1152688</td>
<td>148963135</td>
<td>011111110</td>
</tr>
</tbody>
</table>

Puncture-Constrained Interleaver Design

For \( K = 1504 \), \( S_{\text{min}} \) has a theoretical upper bound of 54. A \( S_{\text{min}} \) target of 80% of the \( S_{\text{min}} \) theoretical upper bound and a \( g \) target of 8 were selected. Afterwards, the interleaver candidates are generated:

1) Selection of the candidate values for \( P \): The maximum achievable value of \( S_{\text{min}} \) for \( K = 1504 \) considering a RI structure is 52. Thus, the best candidates for \( P \) are those allowing a \( S_{\text{min}} \) value between 45 and 52.

2) Selection of the \( Q \) shift values for each candidate for \( P \): The parameters for the ARP interleaver candidates are determined via the construction method previously presented. Three different design configurations have been studied. In the first one, the puncturing mask with Non Data-Punctured (NDP) \( DPR = 0 \) is considered. In the other two, DPC, and DPPC ARP interleavers are considered for the \( DPR-2/8 \) mask. In order to compare the efficiency of the different configurations in finding large \( d_{\text{min}} \) values, 64000 ARP candidates are generated by each configuration. The most efficient configuration is the one with the highest number of generated candidates meeting all the design constraints (girth, span and \( d_{\text{min}} \)).

3) Selection of the best ARP interleaver candidate: Table 6-10 lists the best ARP interleavers generated for each design configuration. For all candidates \( S_{\text{min}} = 45 \) and \( g = 8 \). Their respective distance spectrum is estimated and given in Table 6-11. It is observed that the use of data puncturing allows a larger \( d_{\text{min}} \) value to be reached. Furthermore, the proposed DPPC ARP interleaver allows a reduced \( d_{\text{min}} \) multiplicity compared to the DPC ARP interleaver.

Table 6-10: Best candidates for ARP interleaver with the different configurations, \( S_{\text{min}} = 45 \), \( g = 8 \), \( S(0) = 0 \), \( R = 2/3 \) and \( K = 1504 \).

<table>
<thead>
<tr>
<th>ARP</th>
<th>( P )</th>
<th>( S(1) )</th>
<th>( S(2) )</th>
<th>( S(3) )</th>
<th>( S(4) )</th>
<th>( S(5) )</th>
<th>( S(6) )</th>
<th>( S(7) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDP</td>
<td>399</td>
<td>792</td>
<td>630</td>
<td>829</td>
<td>1010</td>
<td>90</td>
<td>1471</td>
<td>658</td>
</tr>
<tr>
<td>DPC</td>
<td>227</td>
<td>495</td>
<td>998</td>
<td>280</td>
<td>1090</td>
<td>734</td>
<td>361</td>
<td>362</td>
</tr>
<tr>
<td>DPPC</td>
<td>699</td>
<td>289</td>
<td>1452</td>
<td>1292</td>
<td>1349</td>
<td>391</td>
<td>417</td>
<td>874</td>
</tr>
</tbody>
</table>
Table 6-11: Estimated distance spectrum for the best ARP interleavers in AWGN channel with corresponding multiplicities $\alpha(d)$ and cumulated input weight at $d_{min} = d_0 w_{d_0}$.

<table>
<thead>
<tr>
<th>ARP</th>
<th>$w_{d_0}$</th>
<th>$d_0$</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$\alpha(d_0)$</th>
<th>$\alpha(d_1)$</th>
<th>$\alpha(d_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDP</td>
<td>5640</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>1128</td>
<td>4512</td>
<td>7708</td>
</tr>
<tr>
<td>DPC</td>
<td>4324</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>752</td>
<td>1880</td>
<td>5264</td>
</tr>
<tr>
<td>DPPC</td>
<td>2444</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>376</td>
<td>2444</td>
<td>3572</td>
</tr>
</tbody>
</table>

Simulated performance results

Figure 6-29 shows the FER performance in AWGN channel of the 8-state CRSC(1,15/13)$_8$ TC for $R = 2/3$, $R = 4/5$, and $K = 1504$, with the three sets of parameters listed in Table 6-10. In addition, error rate simulation results for the original LTE TC are included for comparison. We observe that DPC and DPPC interleavers achieve better asymptotic performance than the conventional interleaver without data puncturing (NDP), while also displaying a convergence improvement. Compared to the puncturing pattern and the interleaver adopted in LTE, the proposed DPPC interleaver provides a gain of about 0.5 and 0.7 dB in convergence threshold for $R = 2/3$ and $R = 4/5$, respectively and almost 4 decades in error floor.

![Figure 6-29: Frame error rate performance comparison between the different ARP interleaver configurations and the LTE in AWGN channel with a max of 16 decoding iterations of the MAP algorithm for $R = 2/3$, $R = 4/5$, $K = 1504$, and CRSC(1,15/13)$_8$ constituent codes. TUB = Truncated Union Bound.](image)

Additional technical work carried out and contributions to 3GPP

Several technical contributions were submitted to the coding group of the 3GPP RAN1. In [R1-164635], the proposal of an enhanced turbo code based on the technical improvements detailed in the previous section was submitted. In [R1-167413], the details of the design (interleaver model, parameters, puncturing mechanism and patterns) and implementation of the enhanced...
turbo code for the agreed simulation conditions were provided. In [R1 - 167414], the corresponding simulation results and comparisons with respect to the LTE turbo code were presented. These contributions have shown large performance improvements with respect to the LTE turbo code, especially for short frames and for high coding rates, while requiring mild modifications with limited additional complexity to support both LTE and NR. In [R1-1610314], it was shown that the proposed enhanced turbo codes matched the performance of the best LDPC codes of [R1-164698] for large frame sizes while largely outperforming these LDPC codes for the short frame sizes. Indeed, a gain exceeding 1.5 dB was observed for the 100-bit frame size and a coding rate of $R = 8/9$ as shown in Figure 6-30. In addition, the improvements in the high SNR region were confirmed for the enhanced turbo code since error rates as low as $10^{-6}$ of Block Error Rate (BLER) were attained without any change in the slope of the error rate curves.

![Figure 6-30](image)

**Figure 6-30:** Performance comparison of the enhanced turbo code with the Multi-Edge LDPC code in AWGN channel for coding rates ranging from 1/5 to 8/9 in terms of BLER vs Es/N0. QPSK modulation, block size $K$ around 100 bits.

In [R1-1613029], a rate-compatible version of the enhanced turbo code was proposed for both eMBB and URLLC scenarios and their corresponding simulation conditions. It was also shown that additional gains (0.6 to 0.8 dB) can be achieved when performing list-like decoding for turbo codes. Further improvements (up to 2.0 dB) are also possible, especially for short frame sizes when combining the list-like decoding of turbo codes with an outer CRC code. In [R1-1613347], performance comparisons for short frame sizes ($K < 1024$ bits) with the proposed polar codes were performed for the first and successive transmissions. It was shown that enhanced turbo code with Max-Log-MAP decoding and 8 iterations offers BLER performance comparable to the polar code with list-8 decoding for rates $R > 2/3$ and slightly better performance for lower rates at the first transmission as shown in Figure 6-31. In subsequent transmissions for HARQ support, the enhanced turbo code offers better performance than the proposed rate-compatible polar code. Gains exceeding 0.5 dB can be observed in this case, as shown in Figure 6-32. In addition, steeper BLER curves were seen for the enhanced turbo code predicting larger gains if lower error rates would be targeted. In [R1·1703331], the excellent performance of the proposed turbo codes was confirmed for the all simulation conditions of
URLLC. Target BLER of $10^{-5}$ was achieved without any change in the slope of the curves for the simulation conditions agreed for URLLC, including the association with QAM modulations.

Figure 6-31: Performance comparison for the first transmission of the enhanced turbo code with the polar code in AWGN channel for coding rates ranging from 1/5 to 8/9 in terms of required $E_s/N_0$ to achieve 1% and 0.1% of BLER. QPSK modulation, block size $K$ ranging from 32 to 1024 bits.

Figure 6-32: Performance comparison for up to four retransmissions of the enhanced turbo code with the polar code in AWGN channel for coding rates ranging from 1/12 to 2/3 in terms of required $E_s/N_0$ to achieve 1% and 0.1% of BLER. QPSK modulation, block size $K$ ranging from 32 to 1024 bits.

**Conclusion and ongoing work**

A new method to design turbo code interleavers was proposed, which calls for a joint optimization of puncturing patterns and interleaver function. When combined with a layered construction of ARP interleavers, it was shown that significant improvements at low SNRs (especially for short frames) and high SNRs can be achieved by including data and parity puncturing constraints into the interleaver design. Furthermore, it was observed that adding stronger constraints made the interleaver parameter research process more efficient: the
percentage of candidate interleavers with large minimum Hamming distances is increased and the average time to find such interleavers is reduced.

With the proposed method, turbo codes complying with the 5G requirements can now be easily designed. Moreover, the proposed code does not entail any additional complexity compared to the existing LTE turbo code and only very small modifications to the LTE encoder/decoder are required. The extension to design rate-compatible turbo codes based on this technique was performed, thus facilitating the implementation of retransmission mechanisms. The proposed enhanced turbo code was subject to multiple technical submissions to 3GPP. For short frame sizes and targeting URLLC scenario, it compares favourably in terms of performance and complexity to the other coding solutions composed of LDPC and polar codes proposed to 3GPP, making it an excellent candidate for adoption for such a scenario.

6.5 FDPC PAPR reduction technique

The multi carrier (MC) modulations are today commonly used in the radio communication domain. But the PAPR issue has been identified as an important drawback of these systems, which leads to a low energy efficiency in transmission. To overcome it, many PAPR reduction techniques have been studied [XYZ+12], [JW08], [PG09], [KJ03], [LCC+10]; but we propose here a new algorithm and methods to reduce the PAPR regardless of the constellations used and the FFT sizes; that could provide already excellent results being conformed to the actual standards with low complexity and latency and it could also be applied, without any signal distortions, for future standards when high constellation orders transmission or larger FFT sizes are targeted. This algorithm, so-called FDPC (Frequency Domain Partial Construction), can reduce the PAPR of a multi-carrier (MC) symbol by modifying the mapping of the subcarriers in a generalized manner; just depending on the constellation size the method to modify the mapping has to change but is derived by the same FDPC algorithm.

Regards to these systems, the FDPC algorithm reduce the PAPR of a MC signal by acting on symbols that map the sub-carriers, as represented in figure below. The principle of modifying the mapping symbols is an already known technique, so the innovation is the algorithm that can control the changes of these symbols or constellation points, in fact all the type of mapping modifying can be supported by FDPC algorithm, e.g. CD, TR, ACE, CES, etc.. The FDPC works therefore at the transmit side only, and is located between mapping and IFFT processing.
Figure 6-33: Principle of the FDPC algorithm and its implementation in the transmission chain.

The FDPC algorithm uses a basic structure, so-called PCTS (Pre-constructed Temporal Signal) which appears in the form of a feedback loop within a mapping symbols feed forward correction structure. Note that the PCTS structure is not processed in an iterative manner. The PAPR reduction on an MC symbol is calculated during an MC symbol period with a latency of one symbol duration. For each subcarrier, the algorithm computes the correction of the corresponding mapping symbol, which will be applied to this symbol before IFFT and will also be reintroduced in the loop for the calculation of the mapping symbol correction of the next subcarrier, as we can see on this figure:

Figure 6-34: Basic PCTS structure used by FDPC algorithm.

So the FDPC algorithm which uses the PCTS structure can be represented as follow:
Although FDPC structure looks less simple than that of PCTS, but it yields much less implementation complexity (concrete complexity reduction depends on the K/M ratio, K and M reported in the figure). And the following curves give the compared complexity to implement between FDPC and PCTS relative to complexity function of the size of the FFT.

So, compared to other SoTA methods, which deal with the mapping operation, one first advantage of the FDPC algorithm is its relatively low complexity (equivalent to 3~4 FFT operations), which makes it fully implementable in real time in a FPGA of today, or other DSP ICs, without any difficulty for FFT sizes ranging from the small ones up to even 64K. The only drawback compared to PCTS is the higher latency of 3 MC symbols that regard to SoTA is always low or very low.

Another feature, compared to advanced clipping techniques [XYZ+12] in particular, is that this algorithm does not produce any nonlinear signal degradation whatever the level of the researched effectiveness is. Therefore it does not lead to any deterioration on the signal spectral response. Moreover, the correction is conducted in a mapping symbol-wise procedure. This permits in particular to not disturb the reference signals such as pilots, or to freely adjust the correction level when different modulation constellation and level configurations are used in the same MC symbol (e.g. LTE).
As shown in figure hereafter, the FDPC algorithm also has the property to be able to control the constellations in a more generalized manner, covering all the constellation-control techniques reported in the literature (as well as the PRT of the Tone Reservation System [JW08]), which can be categorized into three types below:

- Constellation Modification type A (CM/A): move of all the constellation points in all directions inside their decision sector (noisy like constellation). Systems which lead to this modification are clipping and CD [XYZ+12], [JW08].
- Constellation Modification type B (CM/B): shift of the mapping symbols at the border of the constellation only, towards the outside. A system which uses this modification is ACE [JW08], [KJ03].
- Constellation Modification type C (CM/C): add of extra mapping symbols on the constellation which are dedicated to PAPR reduction. A system which uses this modification is CES [JW08], [LCC+10].

![figure 6-37: Constellation modification of type A, B and C.](image)

The interest of types CM/A and CM/B constellation modification is that they are compatible with existing systems such as DVB-T or LTE or also WiFi. However, type C requires additional specification to the standard.

With a hardware test bench (validation of the PCTS structure) and with the Simulink Matlab platform, multiple PAPR reduction configurations have been evaluated, considering the EVM and END criteria. For the constellation modification of type A, the CCDF curve reported below shows that, in terms of PAPR reduction for a same EVM level, the algorithm gives results as good as the best current techniques, such as advanced clipping, with equivalent or less complexity. Moreover, the FDPC process is a fully linear one, there is no degradation of the spectrum shape, and reference carrier can be saved by the algorithm, producing lower END and better QoS at the receiver side.

![figure 6-38: FDPC CM/A results compared with two advanced clipping algorithms](image)
Type CM/B has also been studied while some standards such as LTE limit the EVM [3GPP 36.141] whereas there is no degradation of the QoS in reception with this constellation modification. In fact it was highlighted that PAPR reduction is quickly limited by the increase of the constellation average power in the CM/B only mode.

Also it appeared that a CM/A+CM/B combination allows to obtain the best results in terms of efficiency and minimization of the END for the same PAPR reduction by comparison with an only CM/A modification; or the PAPR reduction can be significantly augmented for an equivalent END at the receiver; nevertheless this gain can’t be taken into account for standards which use only EVM measurement at the transmitter side (contrary, for example, to DVBT measurement guidelines which, in a better way for QoS, employ END). We can also notice that this type of mapping modification can’t be supported by techniques like clipping.

Figure 6-39: FDPC CM/A + CM/B results compared with an advanced clipping process

And for the CM/C modification, which is not supported by actual standards, this one has however many interests which are to not create EVM nor END and also to be very efficient to reduce the PAPR for high order constellations starting from 16 or 64 QAM to 1024, 2048,…; this, of course, that can’t be supported by clipping.

Thus is shown in the following figure a conventional 256 QAM constellation with extensions for different numbers of point. These extensions allow to obtain a PAPR reduction up to 3.5 dB at CCDF= 10^{-4}, whereas the average power has risen by only 0.24 dB with a high number of extended points. In fact the CM/C modification is a very efficient one, because the number of used extended points remains very low regards to the standard points.

Figure 6-40: FDPC CM/C results for a 256 QAM extended constellation
So, with these capabilities, optimized constellation modifications may be proposed and might be processed by the same FDPC PAPR reduction algorithm with high efficiency; these proposed constellation modification are summarized on the table as follows:

- **QPSK**
  - CM/A + CM/B

- **16 QAM**
  - CM/A+CM/B

- **64QAM**
  - CM/C

- **256QAM**
  - CM/C

**Figure 6-41: Modulus of the RF signal, before and after FDPC processing**

Moreover this algorithm is totally suitable for other MC modulations others than OFDM, and produces entirely similar gains in PAPR reduction; so, we show below Simulink simulation results for two different MC modulations, OFDM and OQAM:

- **Without any non linear degradation** (Spectrum response)
- **No disturbance of the reference carriers**
- **For the same END : up to 2 dB more of PAPR reduction (4.0 / 4.5 dB)** than clipping

- **Without any non linear degradation** (Spectrum response)
- **No EVM nor END**
- **Up to 1.5 dB more of PAPR reduction (3.5 / 4.0 dB)** than clipping in 64 QAM for example
- **Clipping not suitable** for higher constellation orders

**Figure 6-42: Proposed modified algorithm**
Figure 6-43: Compared simulation results for OFDM and OQAM modulation

The following compared results are for the same configuration, but with noise to highlight the low proportion of extended points, resulting in a low increase in average power.

Finally, this algorithm is also fully able to correct multiplexed signals in radio frequency although as compared to a system that operates only on the time signal, the algorithm FDPC works only in the digital processing part before IFFT. But in this case joint correction signals can be carried out only on the condition that they can be processed in a same digital baseband part.

For conclusion we give the following two tables, the first for the actual standards and the second for new standards, which summarize all the advantages, with very low drawback, of this new FDPC PAPR reduction algorithm.
6.6 Control channel design

Simulation results:

In this section, computer simulations have been conducted to study the performances of the presented schemes in previous sections. The detailed simulation parameters are listed in Table 6-12.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel model</td>
<td>TDL-A [6], delay scaling factor: 200ns</td>
</tr>
<tr>
<td>Channel coding</td>
<td>Tail-biting convolutional codes [7]</td>
</tr>
<tr>
<td>Number of base station transmit antennas</td>
<td>2 and 4</td>
</tr>
<tr>
<td>Number of UE receive antennas</td>
<td>2</td>
</tr>
<tr>
<td>Subcarrier spacing [kHz]</td>
<td>15</td>
</tr>
<tr>
<td>System bandwidth [MHz]</td>
<td>10</td>
</tr>
<tr>
<td>DCI payload size [bits]</td>
<td>40</td>
</tr>
<tr>
<td>Control resource RB set for localized transmission</td>
<td>[1:16]</td>
</tr>
<tr>
<td>Control resource RB set for distributed transmission</td>
<td>[1:2:32]</td>
</tr>
<tr>
<td>Number of control symbols</td>
<td>2</td>
</tr>
<tr>
<td>Number of REGs/CCE</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 6-12: Simulation Parameters
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>REG to CCE mapping</td>
<td>Time first</td>
</tr>
<tr>
<td>Aggregation levels</td>
<td>1, 2, 4 and 8</td>
</tr>
<tr>
<td>Transmit diversity scheme</td>
<td>2 Tx: SFBC and per-RE precoder cycling;</td>
</tr>
<tr>
<td></td>
<td>4 Tx: SFBC with RB-level antenna pair cycling.</td>
</tr>
<tr>
<td>Channel estimation</td>
<td>MMSE, no averaging over RBs</td>
</tr>
<tr>
<td>UE receiver</td>
<td>MMSE</td>
</tr>
</tbody>
</table>

To study the performance benefits from frequency diversity provided by distributed NR-PDCCH transmission, we compare the block error rate (BLER) of localized and distributed NR-PDCCH transmissions. For each transmission scheme, 4 ALs, i.e., AL1, AL2, AL4 and AL8 are evaluated. In localized transmission, the control resource set is comprised of the first 16 RBs with 2 REGs in each RB in the system bandwidth. In distributed transmission, the control resource set is comprised of the first 16 RBs with odd RB index. For each AL, the first NR-PDCCH candidates, i.e., the NR-PDCCH with index 1 in each sub-SS corresponding to a certain AL illustrated in Figs. 2 and 3 for localized and distributed transmission, respectively, are transmitted and the BLER thereof are calculated. It is shown from Figure 6-44 that thanks to larger frequency diversity, the distributed transmission outperforms localized transmission by 0.5 to ~1 dB gain at the 1% BLER target. These results confirm the advantage of distributed transmission in the situation where smart localized transmission, e.g., frequency selective NR-PDCCH scheduling and/or preferred beamformed transmission, is not possible due to lack of UE CSI at the base station.

Figure 6-44: REG Localized vs. distributed NR-PDCCH transmission, control resource set: 16 RBs and 2 control symbols (i.e., 2 REGs) per RB, 2x2 SFBC.
As mentioned in Section 3.3.2.2, SFBC and per-RE precoder cycling are two transmit diversity approaches studied during 3GPP NR development. To focus on the effect of transmit diversity, we simulate only localized transmissions with SFBC and per-RE precoder cycling. It is illustrated from Figure 6-45 that due to achieved maximum transmit diversity, SFBC always outperforms per-RE precoder cycling. It is also observed that except AL1, the performance differences between two approaches are quite small. Given this observation, SFBC in comparison with per-RE precoder cycling, should be a recommended transmit diversity approach.

Figure 6-46: 2 Tx SFBC vs. SFBC with per-RB transmit antenna pair cycling for 4 Tx, control resource set: 16 RBs and 2 control symbols (i.e., 2 REGs).
To further study the performance benefits from high order transmit diversity, e.g., diversity order of 4 like in LTE, the performance of SFBC with 2 transmit antennas is compared to that of the SFBC in combination of per-RB transmit antenna pair cycling with 4 transmit antennas depicted in Figure 3-10. For similar reason above, we only simulate the localized transmissions. It is illustrated from Figure 6-46 that compared to 2 transmit antennas, larger transmit diversity from 4 transmit antennas offers considerable performance gain, i.e., more than 1dB, at 1% BLER target. This clearly motivates the support of high order transmit diversity transmission in NR.

Simulation results:

In this section the performance results obtained by system level simulations are presented where we compare the MS-PDCCH introduced above with LTE EPDCCH. Main simulation parameters are listed in Table 6-13.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario</td>
<td>Homogeneous macro (19 3-sector sites, 500 m inter-site distance)</td>
</tr>
<tr>
<td>Channel model</td>
<td>UMa 3D with 3 km/h (instead 30 km/h) (3GPP TS 36.814)</td>
</tr>
<tr>
<td>Traffic model</td>
<td>Full Buffer (20 users per cell in average), FTP (100 kByte file size)</td>
</tr>
<tr>
<td>System Bandwidth</td>
<td>10 MHz at 2 GHz</td>
</tr>
<tr>
<td>Antenna configuration</td>
<td>2 TX x-pol sector-3D, 2 RX x-pol omni-2D</td>
</tr>
<tr>
<td>Scheduler</td>
<td>Proportional Fair (PF), allocation type 0 (3GPP TS 36.213 §7.1.6)</td>
</tr>
<tr>
<td>Precoder</td>
<td>EBF (with sum power constraint and equal power per layer)</td>
</tr>
<tr>
<td>Receiver</td>
<td>IRC</td>
</tr>
<tr>
<td>MIMO</td>
<td>single user MIMO, closed loop</td>
</tr>
<tr>
<td>Overhead</td>
<td>1 OFDM symbol per TTI retained for signaling other control information (CSI-RS, sync)</td>
</tr>
<tr>
<td>CSIT</td>
<td>per sub-band CQI every 5 ms with 6 ms delay</td>
</tr>
<tr>
<td>LTE EPDCCH</td>
<td>DCI allocated within pre-configured search space resources according to TS 36.213 (10 MHz, 6 PRB feedback block and 3 PRB RBG), common SS for all UEs, 12 PRBs large, centered in the BW to match RBG boundaries</td>
</tr>
<tr>
<td>Multi stage PDCCH</td>
<td>DCI allocated to the resources with best reported CQI (out of the resources allocated to the UE, multiplexed with data)</td>
</tr>
<tr>
<td>UIDS</td>
<td>12 QPSK modulated symbols per PRB</td>
</tr>
<tr>
<td>Data target BLER</td>
<td>10% for initial transmission</td>
</tr>
<tr>
<td>Control target BLER</td>
<td>1%</td>
</tr>
</tbody>
</table>

![Figure 6-47: UIDS MD and FA rates dependency on ALs and SINR, SINR calculated at the antenna (blue), SINR after Rx processing (green)](image)

Dissemination level: Public
As mentioned earlier, ALs of UIDS and basic control must be aligned in order to obtain similar performance for the first two stages. In case of a misalignment UIDS will dominate the PDCCH error rate. Such accurate alignment is especially important for higher ALs in the lower SINR region. UIDS MD and FA rates dependency of AL and SINR is illustrated in Figure 6-47. Furthermore, the MD and FA rates are shown for two different locations of the SINR calculation. The reason for this is that UIDS detection is done right at the antenna before receiver processing but the decoding of the associated basic control element includes the receiver gain which is in the order of 4 dB at 1 % MD rate. This must be taken into account when designing MCSs (including correct block sizes) for different ALs of the basic control element which are associated to a certain UIDS AL. The threshold for UIDS detection has been chosen for a FA rate of 10 %, such the FA rate is constant over the full SINR range.

Table 6-14: Full buffer performance results for different feedback unit (CSIT) and resource allocation unit (Alloc) granularities (in PRBs) and different scheduling modes for EPDCCH

<table>
<thead>
<tr>
<th>MS-PDCCH vs. EPDCCH</th>
<th>EPDCCH Mode 1</th>
<th>EPDCCH Mode 2</th>
<th>EPDCCH Mode 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSIT/Alloc</td>
<td>6/3</td>
<td>3/3</td>
<td>1/1</td>
</tr>
<tr>
<td>Spectral efficiency</td>
<td>6/3</td>
<td>4.5%</td>
<td>3/3</td>
</tr>
<tr>
<td></td>
<td>3/3</td>
<td>5%</td>
<td>6/3</td>
</tr>
<tr>
<td>Cell edge throughput</td>
<td>6/3</td>
<td>3.5%</td>
<td>3/3</td>
</tr>
<tr>
<td></td>
<td>3/3</td>
<td>4%</td>
<td>6/3</td>
</tr>
<tr>
<td></td>
<td>3/3</td>
<td>9.5%</td>
<td>6/3</td>
</tr>
<tr>
<td></td>
<td>3/3</td>
<td>15%</td>
<td>6/3</td>
</tr>
<tr>
<td></td>
<td>3/3</td>
<td>0.5%</td>
<td>6/3</td>
</tr>
</tbody>
</table>

Simulations with full buffer traffic have been conducted to assess system performance regarding spectral efficiency and cell edge throughput performance with the two PDCCH variants. The achieved performance gains with the MS-PDCCH design can be seen in Table 6-14. Different feedback and resource granularities, in multiples of PRBs, have been compared. For the investigated system bandwidth of 10 MHz, 3GPP standard specifies a feedback unit granularity of 6 PRBs and a resource unit granularity respectively RBG size of 3 PRBs, that specifies the smallest amount of resources the BS scheduler can assign to a UE (in resource allocation type 0). Also the used DCI formats in LTE are designed to contain such scheduler resource assignments in form of a bitmap that has a fixed size and can only support a fixed resource unit granularity. In contrast to the LTE DCI, for the MS-PDCCH we assume also full system flexibility as the new DCI is scalable in size and is able to support very fine granular resource unit allocations by the BS scheduler.

For EPDCCH two different scheduling modes have been compared. In Mode 1, the left table of Table 6-14, the scheduler always gives priority to LTE EPDCCH allocation over data allocation, whereas for Mode 2 in the right table, the scheduler blocks resources for EPDCCH allocation which have been once assigned as data resource. For scheduling Mode 1 operation, the MS-PDCCH shows higher gains due to a higher control channel overhead of EPDCCH. On the other hand, the advantage of scheduling Mode 1 is a low EPDCCH blocking probability. Contrary to Mode 1, the MS-PDCCH shows less gain versus EPDCCH scheduling Mode 2. In scheduling Mode 2 EPDCCH has higher system performance due to lower EPDCCH overhead but also higher probability of EPDCCH blocking.
To better understand the effect of the PDCCH on the system performance under various load conditions, also a more dynamic FTP type traffic model with 100 kByte file size was used in simulations. The performance metric in the FTP case is the average number of concurrent UEs per cell, i.e. a higher efficiency manifests itself in a lower number of average UEs per cell at a certain system load. As it is shown on the left side of Figure 6-48, the MS-PDCCH shows superior performance under all load conditions and system configurations regarding feedback/resource granularity. For a fully flexible system, up to 40 % less users are observed at high system load. Identical behaviour for the different scheduling modes of EPDCCH can be observed as under full buffer traffic assumption. With Mode 1 system throughput performance is worse but at the same time offers low EPDCCH blocking probability, whereas Mode 2 offers higher system throughput in exchange for higher EPDCCH blocking rates.

On the right of Figure 6-48 the average PDCCH overhead confirms the observations made for the system performance. EPDCCH shows an overhead floor around 6.5 % which corresponds to one scheduling resource (RBG) at 10 MHz system bandwidth. The MS-PDCCH overhead scales more smoothly with the offered load and is in the same order for different feedback unit and resource unit configurations. Increasing the resource granularity will typically increase the MS-PDCCH overhead. The DCI will grow due to a larger number of bits required to describe the user data allocation and a larger user diversity arises per TTI – but there are other effects which offset this overhead. First, a fine granular feedback allows more efficient LA of the MS-PDCCH and second, with increased offered load, the average grant per UE is smaller and requires less signalling bits.
The CDF of the UE throughput in Figure 6-49 is another meaningful representation to show the performance gain of the MS-PDCCH. It shows how the lower average number of UEs per cell translates into higher throughput per UE. There is ~38\% increase in median UE throughput for same system parameters and up to ~130\% increase for a fully flexible system.

Control channel blocking can cause a serious threat to delay sensitive or delay intolerant services. As stated above, for some system configurations and under high system load the blocking probability can reach a high value with EPDCCH. Figure 6-50 shows the blocking probability for EPDCCH for various loads and different file sizes. Under high load the system faces an increased user diversity and also higher blocking probabilities, especially for a small file size. With scheduling Mode 1, a low blocking probability can still be maintained but at the cost of lower throughput performance due to a higher EPDCCH overhead. Nevertheless, a lower system throughput indirectly will cause higher packet delays.

To assess the impact of the search space configuration for EPDCCH, we have compared two different SS configurations disregarding the mandatory system configuration by the 3GPP standard for 10 MHz bandwidth with respect to resource granularity. The first configuration is composed of two SSs per UE, one 8 PRBs wide and one 4 PRBs wide. It is UE individually configured based on CQI feedback and updated every 100 ms. The second configuration is likely composed of two SSs which is common to all UEs and placed equidistantly in the bandwidth as shown in Figure 6-51. Both SS configurations can achieve a low EPDCCH.
blocking probability, however the SS configuration which is common to all UEs shows better system performance when looking at the leftmost graph, i.e. there are less users in the system at same load. This fact is confirmed by the rightmost graph where the common SS shows lower overhead due to better utilization and a higher share level of EPDCCH resources.

Figure 6-52: PDCCH error CDF at high system load, coarse resource granularity (left), fine resource granularity (right)

From a system performance perspective, it is desirable to have a low PDCCH error rate. For this reason, PDCCH LA is adjusted to a target error rate of 1 %. In Figure 6-52 the error rates of the aggregate and the different stages of the MS-PDCCH is compared to EPDCCH. It should be noted that no extensive LA analysis and optimization was performed for the compared PDCCHs. From the left graph of Figure 6-52 it can be observed that UIDS missed detection dominates the control errors of the MS-PDCCH. This is a hint that there is a possible misalignment between UIDS AL and the associated basic control ALs in the lower SINR region. Even so, the aggregate control error probability is still lower for most UEs than in case of EPDCCH. Due to the narrowband PDCCH allocation, for a coarse feedback granularity, the PDCCH LA is more challenging due to possible fading within the large feedback block and due to CQI averaging. In general, to reach a lower control error rate, there must be a more robust LA in the lower SINR region that would create larger control overhead and would degrade EPDCCH performance even more. With fine granular feedback, as can be seen in the left graph, the target control error probability can be reached for more than 99 % of the users in case of the MS-PDCCH.

When the exact PDCCH coding rules are defined for NR and their performance assessment is available, a precise association of UIDS ALs to the corresponding basic control MCSs will be needed.

### 6.7 HARQ enhancements

In the following some numerical results for the variable block length IR concept presented in Section 3.3.4.2.2 are provided. As waveform design, the FBMC/OQAM scheme has been considered, where the prototype pulse is designed according to [Bel01] with an overlapping factor of 4. The link adaptation algorithm and the NB-LDPC code, whose field order is 256, are described in [CNM+16] and [PFD08], respectively. FBMC/OQAM and NB-LDPC have been selected to evaluate the HARQ procedure, because their combination is not straightforward and represent the major challenge. The symbols are drawn from 256-QAM and 16-QAM constellation diagrams. The system parameters have been set as follows: the sampling frequency is $f_s = 15.36$ MHz, each message is transmitted in a 1 ms TTI and 600 out of $M = 1024$ subcarriers are active. The number of RBs is equal to 50 and each RB consists of 12 subcarriers and 7 time slots. Concerning the propagation conditions, the extended vehicular A channel model is simulated to carry out the experimental validation. To estimate the channel, the auxiliary pilot method has been employed [JLR03]. The effects that distort the transmitted
signal in presence of imperfect CSIR, which are characterized in [CNM+17], have been taken into account. To assess the HARQ procedure, a single user has been simulated with two different setups, one with a target BLER of 10% for all (re)transmissions where up to 3 retransmissions are allowed, and another one with a target BLER of 30% for all (re)transmissions where up to 6 retransmissions are allowed. As a result, the final BLER is at least $10^{-4}$ and $2.2 \times 10^{-4}$, depending on the strategy. Concerning the feedback, we have employed a two bit ACK/NACK message, which is sent to the transmitter through an error-free channel. Therefore, 3 NACK levels are used to specify the number of RBs to be used in the next retransmission. In particular, we choose NACK1 = NRB, NACK2 = \lfloor NRB/2 \rfloor and NACK3 = \lfloor NRB/4 \rfloor, where NRB denotes the RBs assigned for the first transmission. It is noteworthy that this mapping allows to relinquish up to 75% of the resources to other users.

From Figure 6-53 it can be inferred that both strategies achieve a similar throughput. As the complementary cumulative distribution function (CCDF) of Figure 6-54 highlights, the more aggressive the strategy is, the higher the number of retransmissions. Figure 6-54 also reveals that the target BLER is guaranteed. In other words, the BLER in the $i$th transmission is equal or lower than $0.1^i$ or $0.3^i$, depending on the strategy. This observation suggests that the sources of distortion are accurately characterized. It must be noticed that the number of transmissions with quantized feedback is lower than that without quantization. By resorting to quantization, the overhead is reduced but the amount of redundancy transmitted is higher than the minimum required. This observation highlights that a conservative approach has been adopted.

![Figure 6-53 Throughput vs. SNR using two-bit feedback messages](image)

![Figure 6-54 CCDF of the number of transmissions with and without quantizing the feedback messages](image)

Finally, to determine the what extent variable block size IR schemes can provide resource savings, Figure 6-55 compares the resources allocated in the first retransmission by different strategies. In the scenario simulated, the 2-bits feedback reduces the amount of resources spent...
for retransmissions to ~25% with respect to one bit feedback, while full resolution feedback would reduce it to less than 10%.

![Figure 6-55 CDF of the first retransmission size ratio with and without quantizing the feedback messages](image)

In the following, we report a numerical evaluation for the early HARQ feedback concept presented in Section 3.3.4.2.3.

The feedback duration is set to 0.0625 ms (short coverage UEs). A downlink TTI duration of 0.125 ms is considered, as well as a (BLER1=10%, Kmax=4) configuration, with 0.7 combining gain. The processing time at BS/UE is set to 0.125 ms. A false negative (FN) rate of 2% and 10% is assumed. A FN rate of 10% is a rather pessimistic case, while the 2% case is compliant to the initial findings in [BKP4+16].

Figure 6-56 displays the probability of having a larger HARQ latency than the value on the x-axis, assuming different feedback types. The presence of FNs increases the probability of a first retransmission, but still achieves latency benefits at low probability. For instance, at 10^{-5} probability, latency of the early feedback with 2% FN rate is still lower than the one of regular feedback. Improving the reliability of the decoder prediction by further minimizing such wrong estimates (including false positives) is a valid research direction for an enhanced HARQ concept.

![Figure 6-56 HARQ latency for UEs at different coverage conditions for regular and early feedback (FB).](image)
6.8 PHY procedures: details and results

6.8.1 Blind detection of users’ activity in random/grant-free access

In order to blindly detect the number of active users in a grant-free access area, the base-station will perform the following steps. Firstly, the base-station will calculate the expected average number of active subcarriers \( \alpha(K, N_a, M_F) \) on the Activity Area for a given number of users (\( K \)) [Ali16a]. As it will be shown later, in addition to the number of active users, the average number of active subcarriers is function of the number of resource elements in the Activity Area (\( N_a \)) and \( M_F \). The expected average number of active subcarriers, \( \alpha(K, N_a, M_F) \), for a given number of users (\( K \)) can be done one time and there is no need to perform it with each transmission time. Then, the base-station will estimate the number of active subcarriers \( \beta \) based on the received signal in the Activity Area. As in FQAM each user will activate one subcarrier in each subset of subcarriers, the number of active resource elements will depend of the number of transmitting users on the same grant-free access area. If there is no signal transmitted on a given subcarrier, then the base-station should receive Gaussian noise only (and inter-cell interference in some scenarios) on that subcarrier. On the other hand, if there is a transmission on a subcarrier (from one or more users), the base-station will receive users’ signal corrupted by noise. Conventional energy detection methods can be used to detect the active subcarriers. Once the number of active subcarriers \( \beta \) is estimated, the base-station will compare it with the expected number of active subcarriers \( \alpha(K, N_a, M_F) \), and estimates the number of active users accordingly.

The estimated number of users (\( K \)) will be passed to the detection and decoding blocks to retrieve the users’ data on the Activity Area. For example, if the base-station estimated that only one user is active, it will perform single-user detection and channel decoding for one user. However, if the base-station detected two users active, it will perform multiuser detection then channel decoding for two users.

The method can be utilized as well to detect the number of colliding users that select the same preamble in random access procedure [Ali16b]. In the 4-steps random access procedure, the allocated resources in response to a random access attempt (used by the users to transmit Msg3) will consist of Activity and Data areas (in a similar way as shown in Figure 3-17). Thus, by monitoring the number of active subcarriers on the Activity area, the base-station will be able to detect the number of colliding users, and retrieve the users’ data accordingly. In addition, in the 2-steps random access procedure [R1-1612468], the payload area will be divided into Activity and Data areas. If multiple users selected the same preamble, the base-station will be able to detect the number of colliding users using the number of active subcarriers on the Activity area. If the base-station was able to detect the data of colliding users, it will send acknowledgment to the users. Otherwise, the base-station can send back several resource assignments (equal to the number of colliding users) where the colliding users can retransmit their data (each user randomly selects on resource assignment).

Here, we provide performance results of the proposed method using grant-free access scheme. We consider a single-cell scenario with users 100 m from the base-station each with 23 dBm transmission power and using FQAM modulation with \( M_F = 4 \). The grant-free access area (pre-defined by the base-station) consists of 2, 4 or 6 resource blocks shared among the users (each resource block consists of 12 subcarriers and 7 OFDM symbols). Half of the resource elements in the grant-free access area will be dedicated to the Activity Area and the other half for the Data Area. Consequently, the Activity Area size will be 1, 2 or 3 resource blocks (84, 168 and 252 respectively). We implement turbo code of 1/3 coding rate and QPSK modulation. We consider the number of users that are simultaneously active (accessing the system in a grant-free manner) as: 1, 2 or 3 users. The base-station is not aware of how many users are active, and it will estimate it using the number of active resource elements in the Activity Area. A performance metric for the blind detection as the percentage of the wrong estimations for the
active number of users will be adopted. For example, if there are two active users and the base-station estimated that more/less number of users, then it is counted as wrong estimation.

The results for the estimation error in the blind detection of users’ activity are presented in Table 6-15 for different Activity Area sizes. These results show the efficiency of the proposed method in detecting the number of active users. As the results show, when only one user is active, the base-station can accurately (without errors) detect the existence of the user. With the increase of the number of active users, the estimation error increases. However, the estimation accuracy is improved by increasing the size of the Activity Area from 1 resource block up to 3 resource blocks. Consequently, the size of the Activity Area can be adjusted by the network based on the predicted number of the users in the system. It can be concluded that the proposed method for blind detection of users’ activity can efficiently enable the base-station to detect the number of active users.

<table>
<thead>
<tr>
<th>Activity Area size</th>
<th>1 User</th>
<th>2 Users</th>
<th>3 Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 resource block</td>
<td>0%</td>
<td>0.063%</td>
<td>7.28%</td>
</tr>
<tr>
<td>2 resource blocks</td>
<td>0%</td>
<td>0.0005%</td>
<td>1.1%</td>
</tr>
<tr>
<td>3 resource blocks</td>
<td>0%</td>
<td>0%</td>
<td>0.14%</td>
</tr>
</tbody>
</table>

6.8.2 New Radio Physical Random Access Channel Sequence Design

Design options for the preamble sequences

In the following we consider two sequence types: a) ZC-sequences as state-of-the art reference and b) cyclic Delay-Doppler shifted M-sequences as the proposed sequence design for 5G New Radio.

ZC-Sequences

In LTE uplink random access preamble sequences are generated from root Zadoff-Chu sequences, which are defined by

\[ x_u[n] = e^{-j \frac{\mu(n+1)}{N}} \], \quad 0 \leq n \leq N - 1

Herein N is a prime sequence length and u is the physical root number. For LTE the sequence length N is 839 and the physical root index ranges from 1 to 838. Its value depends on the broadcasted logical root sequence index and the mapping can be found in [3GPP-36.211]. Usually different physical root indices are assigned to neighbour cells in order to guarantee low cross-correlations between preambles of adjacent cells.

Different preamble sequences from one root sequence are generated by applying cyclic shifts

\[ x_{uv}[n] = x_u[(n + C_v) \mod N] \]

Where the cyclic shift \( C_v \) is given by multiples of the distance \( N_{CS} \) between two preambles

\[ C_v = \begin{cases} vN_{CS} & v = 0, 1, \ldots \lfloor N/N_{CS} \rfloor - 1, N_{CS} \neq 0 \\ N_{CS} = 0 & \end{cases} \]

In our performance analysis (see below) we consider ZC-sequences of length 839 and 1021.

Cyclic Delay-Doppler shifted M-Sequences

The m-sequences are generated via linear-feedback shift registers. In this contribution we consider sequences that originate from a 10th order pseudo noise generator. The generator polynomial shall be given by

\[ g(D) = D^{10} + D^9 + D^8 + D^5 + D^1 + 1 \]
The output of the generator is a binary sequence $b(n)$ of length 1023 that is transformed into a BPSK ($\pm 1$) modulated base sequence $x[n]$. Different bases sequences can be generated by initializing the pn-generator with different values e.g. depending on respective cell IDs. Alternatively, one can introduce a cell-specific base offset. From the base sequence different preamble sequences can be derived by applying circular Delay-Doppler shifts [GUE08] as follows

$$x_{u,v}[n] = x[(n - C_v)mod N] \cdot e^{j2\pi fu N}, \quad 0 \leq n \leq N - 1$$

Where $C_v$ is the cyclic shift defined as an integer multiple of $N_{cs}$. Here, the $N_{cs}$ value adjusts the separability in time domain and should be therefore larger than the maximum expected delay spread. The phase signature parameter $f$ should be selected larger than the maximum expected Doppler spread in the system. The cell Id can be used as root index parameter $u$ in order to guarantee that neighbour cells apply different frequency shifts.

Sequence Correlation Properties

Ambiguity functions (AF) [HE] are widely employed in radar technology to analyse the auto- and cross correlation of a reference signal with Delay-Doppler $(\tau, \nu)$ shifted versions of the same or a different signal. The periodic auto-ambiguity function (PAF) is defined as

$$\chi(\tau, \nu) = \frac{1}{T} \int_0^T u(t) \cdot u^*(t - \tau) e^{-j2\pi\nu(t-\tau)} dt$$

And the periodic cross-ambiguity function as

$$\chi(\tau, \nu) = \frac{1}{T} \int_0^T u(t) \cdot w^*(t - \tau) e^{-j2\pi\nu(t-\tau)} dt$$

In a real system the absolute value of the ambiguity function can be considered as the output of a preamble correlator i.e. the correlator calculates the AF between an unmodified expected signal and a signal that experienced delays due to multipath propagation and Doppler-effect induced frequency shifts due to the terminal’s mobility or offsets in the transmitter’s or receiver’s oscillator. Therefore, the AF is known to be a good measure to characterize a sequence’s ability to be uniquely identified in a time-frequency dispersive channel.

Sequence Correlation Properties of ZC-Sequences

In the following we briefly discuss the PAF for a ZC sequence of length 31. Figure 6-57 shows the absolute value of the PAF as a function of Delay-Doppler shifted versions of a ZC sequence. For discussion the resolution in time and frequency domain has been normalized to the sequence length. In absence of a frequency shift $\nu = 0$ the correlation is maximum at $\tau = 0$ and zero for $\tau \neq 0$. This property had been one motivation to use Zadoff-Chu sequences as preamble sequence for the RACH. However, in the presence of frequency shift this is no longer the case. As can be observed form the Figure 6-57 additional correlation peaks occur for specific frequency and time shifts. These self-images of the original transmitted sequence can result in detection errors at the receiver. For example: Assume that the sequence, transmitted with zero time-frequency shift, experiences a Doppler shift by $\nu = +1$ (corresponding to one subcarrier spacing), then the correlation based receiver will detect a sequence at around $\tau = 10$. The receiver will therefore wrongly interpret the detected sequence as one that has been transmitted with a cyclic time shift equivalent to $\tau = 10$. Actually, we have in this situation two errors. Firstly, the receiver was not able to detect the transmitted sequence i.e. we have a miss-detection event. Secondly, the receiver observed a sequence that very likely had not been transmitted by any terminal within the cell i.e. we have a false-alarm event. A collision may occur is a second
terminal within the same cell transmits a preamble with a cyclic shift equal $\tau = 10$ without frequency uncertainty. A standardized method to resolve ambiguity is to introduce restricted sets [3GPP-36.211]. The drawback of this approach is that a substantial number of preambles are no more available. These observations are the drivers for the search and proposal of enhanced sequences.

Figure 6-57: Absolute PAF value of a ZC-sequence of length 31. Scale of $\tau$- and $\nu$-axis is normalized to the sequence length. The $\tau$-axis is log scaled.

Sequence Correlation Properties of Cyclic Delay-Doppler shifted M-Sequences

Figure 6-58 shows the absolute value of the PAF as function of time-frequency shifted versions of a Delay-Doppler shifted m-sequence. The two-dimensional PAF shows almost ideal correlation behaviour, which manifests in a thumbtack like peak in the centre and small correlation values elsewhere. This feature makes the new sequence design principle very attractive, because it can be expected that false alarm errors can be reduced substantially. In addition, we expect a significantly increased robustness against frequency impairments and that we can allow more preambles to be used within a cell.
Figure 6.8 Absolute PAF value of a Delay-Doppler shifted m-sequence of length 31. Scale of τ- and ν-axis is normalized to the sequence length. The τ-axis is log scaled.

Link-Level Performance Evaluation

In this section we briefly compare the link-level performance between the legacy ZC and the novel m-sequences. The key simulation parameters are listed in Table 6-16. The key performance metrics are a) the false alarm probability, b) the missed detection probability and c) the timing estimation error.
### Table 6-16: Link-Level Simulation Parameters for NR-PRACH

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier Frequency</td>
<td>4 GHz</td>
</tr>
<tr>
<td>Channel Model</td>
<td>CDL-C</td>
</tr>
<tr>
<td>Delay Scaling</td>
<td>100 ns</td>
</tr>
<tr>
<td>Circular Angle Spread at BS after angle scaling (including</td>
<td>ASD: 25°</td>
</tr>
<tr>
<td>subrays)</td>
<td>ZSD: 1°</td>
</tr>
<tr>
<td>Circular Angle Spread at UE after angle scaling (including</td>
<td>ASA: 60°</td>
</tr>
<tr>
<td>subrays)</td>
<td>ZSA: 5°</td>
</tr>
<tr>
<td>Circular Mean Angle at BS after angle scaling (including subrays)</td>
<td>Uniformly distributed</td>
</tr>
<tr>
<td>Circular Mean Angle at UE after angle scaling (including subrays)</td>
<td>AoD: [-30°,30°]</td>
</tr>
<tr>
<td>Antenna Configuration at BS</td>
<td>(M,N,P) = (1,1,2) with omni-directional antenna element</td>
</tr>
<tr>
<td>Mechanical downtilt at BS</td>
<td>0°</td>
</tr>
<tr>
<td>Antenna Configuration at the UE</td>
<td>(1,1,2) with omni-directional antenna element</td>
</tr>
<tr>
<td>Antenna port virtualization</td>
<td>No beamforming and no beam selection</td>
</tr>
<tr>
<td>Frequency Offset</td>
<td>0.05 ppm at TRP, 0.1 ppm at UE</td>
</tr>
<tr>
<td>UE speed</td>
<td>3 km/h, 120 km/h</td>
</tr>
<tr>
<td>Initial timing Offset</td>
<td>Uniformly distributed [0,100 ms] i.e. assuming a maximum cell radius of 14.4 km.</td>
</tr>
<tr>
<td>Max. tolerated Timing Estimation Error</td>
<td>+/-1.2us for 30 kHz data SCS and 2.3us data CP</td>
</tr>
<tr>
<td>Preamble Detector</td>
<td>FCME with parameterization set such that false alarm is 0.1% if input at receiver is noise only</td>
</tr>
<tr>
<td>FFT size</td>
<td>2048</td>
</tr>
<tr>
<td>Sampling Frequency</td>
<td>30.72 MHz</td>
</tr>
<tr>
<td>T_SEQ</td>
<td>24576</td>
</tr>
<tr>
<td>T_CP</td>
<td>3158</td>
</tr>
<tr>
<td>Sequences Length</td>
<td>Case 1: ZC-Sequences 1021 and M-Sequences 1023</td>
</tr>
<tr>
<td></td>
<td>Case 2: ZC-Sequences 839 and M-Sequences 511</td>
</tr>
</tbody>
</table>

Case 1: In this case we consider sequences of similar length i.e. length 1021 for ZC- and 1023 for m-sequences. From the Figure 6-59 and Figure 6-60 we see that the differences between both sequence types in the missed detection rates and timing estimation errors are small. But as can be observed from the Figure 6-59 the false alarm rates for m-sequences are about factor 100 smaller than for the ZC- sequences for SNRs greater than -22 dB.
Case 2: In this case we consider sequences of smaller length i.e. length 839 for ZC- and 511 for m-sequences (9th order polynomial). From the false alarm rates in Figure 6-59 (left) we see that the m-sequences perform much better than the ZC-sequences in the high SNR regime. The difference here is about a factor 10. A very large difference is observed in the missed detection rates in Figure 6-59 (right): In a low speed (3 km/h) scenario, the minimum required SNR for <1% missed detection probability is about -23 dB for the m-sequences and about -10 dB for the ZC-sequences. The difference is even larger in a high speed (120 km/h) scenario: For m-sequences the minimum required SNR is now about -21 dB, but for the ZC-sequences the target missed detection rate of <1% cannot be achieved. The reason for this failure is a bad timing estimation error for the ZC-sequences in high speed scenarios. Very often the timing estimation error missed the <1.2µs target (see Figure 6-61) and this increases the missed detection probability. It is very important to note that even though the m-sequences used here are shorter and occupy a smaller transmission bandwidth they clearly outperform the ZC-sequences.
6.8.3 Massive MIMO assisted random access for 5G IoT

**UE signal structure for enhanced IoT random access**

In this method, the signal structure from the UE is redesigned. The user data will be divided into $N$ parts and the signal structure consists of the preamble and the user data together. The preamble and user data parts are transmitted in turn on the PRACH. This is illustrated in Figure 6-63. It should be noticed that data signals from different users are assumed independent, which forms the basis for AoA estimation. If there is no preamble collision, the user data will be transmitted successfully via this signal structure. As a result, latency of UE access is then reduced. On contrary, if preamble collision occurs, eNB can estimate how many users are colliding in the spatial domain using the user data.
An additional requirement in IoT is longer battery life in M2M communication and IoT battery lives are expected to be ten years. This requires highly power efficient transmission schemes. In practice, the user data can be modulated with constant envelope modulation schemes such as Quadrature Phase Shift Keying (QPSK), such that power efficient amplifiers can be applied in the M2M/IoT nodes as the Peak to Average Power Ratio (PAPR) is low.

**General enhanced IoT random access procedure**

The general enhanced IoT random access procedure in this method has two possible situations. One is that the UE data sent along with the preamble is successfully decoded by the eNB. This happens most likely when preamble collision did not happen. In this case, a one-stage transmission is finished. The procedure of this situation is depicted in Figure 6-64. The benefit of this is that the latency of transmission is largely reduced.

On the other hand, when there are multiple UEs using the same preamble, the one-stage method will fallback to two-stage transmission, i.e., the UE requests resources in the first stage and transmits data in the second stage as shown in Figure 6-65. In current LTE systems, the eNB will only transmit one RAR back to UEs even if there are multiple UEs using the same preamble. In this method, the eNB will try to estimate the number of UEs using the same preamble by distinguishing AoAs from different UEs. Then the eNB will send the corresponding number of RARs to the UEs. As a result, more uplink resources will be assigned.
to multiple UEs to avoid RACH overload. Each UE randomly selects one RAR and the remaining steps are the same as in current 3GPP. Although there are still chances that multiple UEs select the same RAR, the collision probability is further reduced by extra randomness.

![Diagram](image)

**Figure 6-66** Preamble and data are sent together. If not successful, eNB estimates the number of users.

![Graph](image)

**Figure 6-67** Snapshot of an estimation of the number of colliding users (K = 2).

A comparison between the current 3GPP random access procedure and the proposed method is given in **Table 6-17**. The benefit of this enhanced procedure is that the UE procedure is similar to the current 3GPP, without significant adjustments.

**Table 6-17** Table comparison between the current 3GPP random access and this method.

<table>
<thead>
<tr>
<th></th>
<th>Current 3GPP</th>
<th>This method</th>
</tr>
</thead>
<tbody>
<tr>
<td>UE initiates random access</td>
<td>Preambles only</td>
<td>Preamble and data</td>
</tr>
<tr>
<td>Number of stages of random</td>
<td>Two stages: UE uses</td>
<td>Possible one stage: UE sends</td>
</tr>
<tr>
<td>access</td>
<td>random access, and then sends data</td>
<td>random access and data at one stage. If not successful, then back on two stages automatically</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>eNB estimates number of users</td>
<td>No</td>
<td>In the spatial domain</td>
</tr>
<tr>
<td>Multiple antenna at eNB</td>
<td>Optional</td>
<td>Required</td>
</tr>
<tr>
<td>eNB sends multiple RARs</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Tradeoff of this method</td>
<td>-</td>
<td>Since some preamble slots are replaced by data, only UEs closer to eNB will have a higher probability of benefiting from this method.</td>
</tr>
</tbody>
</table>

**General UE procedure**

The general UE procedure in the proposed enhanced random access method is shown in Figure 6-68. A UE randomly selects a preamble and then forms the random access signal with its data according to Figure 6-63. After transmitting the random access signal, the UE will wait for eNB’s response. If the eNB responds an acknowledgement (ACK), it means that the data transmission has been successful. Otherwise, the UE will receive multiple RARs from the eNB and then randomly select one among them. Next, the UE will transmit Msg3 to the eNB and then receive Msg4. If Msg4 is successfully acknowledged then the connection is established. Otherwise, the UE needs to back off before another random access attempt.

![Figure 6-68 General UE procedure of enhanced IoT random access.](image)

**General eNB procedure**

The general eNB procedure in the proposed enhanced random access method is shown in Figure 6-69. Because of the novel random access signal structure, the eNB needs to remove preambles and then concatenate data signal. Then, the eNB will try to decode the data signal. If the data signal is successfully decoded, then the eNB will feedback ACK to the UE. Conversely, if the data signal is not successfully decoded, there are two possible situations. One is that there is only one UE using the preamble, but the channel condition is so bad that the eNB cannot decode the data signal correctly. The other is that there are multiple UEs using the same preamble and the interference among them is too large for the eNB to correctly decode the data signal. Neither situation will impact the user number estimation based on AoA estimation. Therefore, the eNB
will transmit $M$ RARs to the UE, where $M$ is the estimated number of users. After receiving Msg3 from UEs, the eNB will acknowledge the request from the UE if it is successful.

![Figure 6-69 General eNB procedure of enhanced IoT random access.](image)

One key distinction to the current 3GPP standard procedure is the insertion of user number estimation based on AoA estimation. How well UEs are distinguished in the spatial domain depends on the number of antennas equipped at the eNB. With a massive number of antennas at the eNB, the eNB can estimate AoAs more accurately. It should be noticed that, even if two UEs are closely located and the eNB fails to distinguish them in the spatial domain, the eNB will send one RAR to both UEs. Therefore, it cannot be worse than the current 3GPP procedure.

**Detailed eNB procedure on estimating user number $M$**

Assume that the concatenated received data symbols at the eNB are $x(t)$ and the steering vectors of the eNB from $K$ ($K \leq Q \leq N$) different AoAs $\theta_1, \theta_2, ..., \theta_K$ are $a(\theta_1), a(\theta_2), ..., a(\theta_K)$. Unlike the regular multiple signal classification (MUSIC) algorithm, which assumes that the number of sources is known to the base station, the eNB here needs to assume the number of colliding users is $Q$. A typical value of $Q$ can be set as $Q=N/2$. It should be noticed that the concatenated received data symbol may contain data symbols from multiple UEs. In addition, data symbols from different UEs are assumed independent. The received signal at the eNB is

$$x(t) = As(t) + n(t)$$

where $A = [a(\theta_1) \ a(\theta_2) \ ... \ a(\theta_K)]$ is the steering matrix at the eNB, $s(t)$ is aggregate data symbols of from UEs, and $n(t)$ is the Gaussian noise vector. The eNB will perform the MUltiple MUSIC algorithm to obtain the power angular spectrum (PAS) $P(\theta)$.

**Step 1:** Compute the $N$-by-$N$ correlation matrix $R$ by

$$R = \text{E}\{ x(t) x^H(t) \}$$

**Step 2:** Eigen value decomposition on $R$

$$R = \boldsymbol{U} \Sigma \boldsymbol{U}^H$$

where $\boldsymbol{U}$ is the $N$-by-$N$ eigen vector matrix and $\Sigma$ is the diagonal eigen value matrix.
Step 3: Partition the eigen vector matrix by

$$U = [U_Q \, \bar{U}_Q]$$

where $\bar{U}_Q$ is the $N$-by-$(N-Q)$ eigen vector matrix whose columns correspond to the noise subspace.

Step 4: Search through all angles $\theta$, the PAS can be obtained by

$$P(\theta) = \frac{1}{a^H(\theta)U_Q}.$$  

Step 5: Find the estimated AoA set $B = \{\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_Q\}$ in the PAS by

$$\frac{dP(\theta)}{d\theta} = 0$$

Step 6: Remove $\tilde{\theta}_l$ in $B = \{\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_Q\}$ if $P(\tilde{\theta}_l)$ is less than a certain threshold $\varepsilon$. Then the estimated number of users $M = |B|$.

**Simulation results**

Simulations are performed in this section to test the performance of the proposed method of enhanced IoT RACH. The general settings are listed in Table 6-18.

<table>
<thead>
<tr>
<th>Simulation settings</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of preambles</td>
<td>64</td>
</tr>
<tr>
<td>Number of initial users</td>
<td>200</td>
</tr>
<tr>
<td>Number of new users in each transmission period</td>
<td>10</td>
</tr>
<tr>
<td>Number of received antennas</td>
<td>16</td>
</tr>
<tr>
<td>Detection probability</td>
<td>0.9</td>
</tr>
<tr>
<td>Signal to noise ratio</td>
<td>5dB</td>
</tr>
</tbody>
</table>

An important metric for RACH, as shown in Figure 6-70, is the number of random access attempts in order to obtain uplink resources from the eNB. The proposed method is compared with other two constant multiple RARs methods, i.e., 1 RAR for each preamble, 3 RARs for each preamble, and RARs for each preamble. It can be observed that one UL grant per preamble results in larger numbers of random access attempts, which is not good for reducing access latencies. On the other hand, 3 or 4 RARs per preamble reduce the number of random access attempts as there are more UL resources assigned to UEs. The proposed method provides a similar performance to the 4 RARs method. Also, as the number of antennas increases, the number of access attempts drops due to higher spatial resolution.
Besides the number of random access attempts, it is also essential to access how many assigned UL grants are effective and how many are wasted, as a measure of efficiency. When 4 UL grants are assigned to each preamble, although the number of random access attempts is the smallest, 50% of the assigned UL resources are actually wasted. This will largely reduce spectral efficiencies in the UL channel. When 3 UL grants are assigned to each preamble, the percentage of wasted UL resources declines to 38%. However, when the proposed method is applied, this number shrinks to 28% because the AoA estimation is able to compute the estimate number of random access UEs.

![Figure 6-70. Number of random access attempts of IoT nodes before resource granted.](image)

![Figure 6-71 Percentage of wasted UL resources when multiple UL grants are used.](image)
6.8.4 Embedded air interface (EAI) for multiplexing MBB and MCC traffic: Principle and analysis

In order to describe the key issues of EAI, we use Figure 6-72 to illustrate the frame structure and schedule timeline for multiple services of eMBB and MCC.

![Figure 6-72: Timing diagram of scheduling and transmission](image)

As shown in Figure 6-72, in case of DL, eMBB has the longer TTI, amounting to 4 times that of MCC, yielding a longer scheduling interval. If we use the same scheduling procedure to MCC as used in eMBB, it will suffer from an extra latency. In more detail, when a MCC packet arrives as shown in the figure, it may be scheduled immediately and transmitted in the next sTTI (short TTI). However, limited by the aligned schedule procedure, if the resource of TTI n has already been allocated to eMBB, the said MCC packet may not be transmitted before TTI n+1, which leads to a latency of about 4x sTTI’s.

To guarantee the performance of eMBB when parts of the transmission have been punctured, potential schemes for improvement are listed below:

1) Notifying puncturing information

Without being informed whether eMBB resources have been punctured by MCC service, eMBB UE would still attempt to decode the data in the corresponding resources. The decoded data would likely be erroneous if been interfered. More seriously, the erroneously decoded data would still be involved in HARQ combining process, resulting in degradation of the retransmission performance. Thus, notifying the eMBB UE on the resources punctured by MCC service is proposed. The data punctured by MCC would be excluded after being received by eMBB UE, and would not be considered for channel decoding.

2) Code Block-Level HARQ

When retransmission is requested, the whole transmit block (TB) is retransmitted in LTE. Transmit block size (TBS) of eMBB is expected to be large in 5G, while TBS of MCC is small and the typical size is 50 bytes. Hence, when MCC user punctures eMBB user, only a few code blocks (CBs) in eMBB will be affected. Retransmitting the whole large TB of eMBB will induce a serious waste of resources. The idea of CB-Level HARQ is only to feedback the index of the erroneous CB, and then to retransmit the erroneous CB instead of the whole TB. eMBB UE needs to feedback the location of erroneous CB in uplink.

3) Mapping MCC data to parity bits of eMBB

The Turbo encoder, adopted by LTE, segments the whole TB into several CBs, each of them consisting of systematic bits and parity bits, see Figure 6-73. It would be better to map MCC service data to parity bits of eMBB.
The Turbo encoder, adopted by LTE, segments the whole TB to several CB, each of them consists of systematic bits and parity bits, see Figure 6-73. It would be better to map MCC service data to parity bits of eMBB.

\[ \text{S: Systematic bits} \]
\[ \text{P: Parity bits} \]

![Figure 6-73: Mapping scheme of Turbo encoder in LTE](image)

**Performance Evaluation for MCC**

In this section, we provide results about the MCC benefitting from EAI. We use a system simulation where multi-users in a single cell are considered, where MCC users are supported with the required reliability of 99.999% under several bandwidths constraints.

Because of the big difference in throughput and allowed transmission delay for the two services, one simple way to solve the problem is using FDM to divide the band into sub-bands dedicated for eMBB and MCC, respectively, enabling each service to use an individual TTI length, as shown in Figure 6-74.

![Figure 6-74: Multiplexing scheme based on FDMA](image)

This opens up the question how to distribute the bandwidth (statically or dynamically). MCC traffic does have non-stationary arrival period and needs to transmit immediately when a packet arrives. Hence, if reserving the bandwidth according to the peak payload of MCC, severe waste of resources occurs during periods of low activity. On the contrary, conservative allocation (e.g. according to average payload) may not satisfy both the low-latency and high reliability requirements at any time.
Given a Poisson distribution with average event number 1 (in one TTI) as example, as shown in Figure 6-67, at least 7 resource units are needed to guarantee the resource allocation probability without collision within 1 TTI to be smaller than 1e-05 (and thus fulfilling 99.999% reliability). Hence, the average resource usage ratio is as low as almost 1/7.

In order to avoid the waste of resources, an efficient joint allocation scheme based on resource pooling, which is denoted as embedded air interface (EAI), is proposed to improve the resource utilization, while satisfying the requirements of low-latency and high reliability for MCC. One of its key point is to transmit the MCC data at any time by puncturing the already scheduled eMBB resources.

The features of EAI can be summarized as follows:

1) Maximum eMBB capacity: The whole bandwidth could be fully utilized by eMBB and there is no waste of bandwidth.
2) Maximum MCC capacity: The whole bandwidth could be utilized by MCC.
3) Satisfying the low latency requirement of MCC: MCC data packet is capable to be transmitted within one OFDM symbol if sufficient bandwidth is offered.

Simulation assumptions: single cell and multiuser scenario; each user has the same average SNR but undergoes independent small-scale channel fading; packet arrival model follows Poisson distribution with the average arrival period of 10ms and packet size of 256 bits. The scheduler runs once per sTTI scheduling the MCC user with the higher priority based on queuing. If the delay due to queuing is bigger than “1ms - transmission delay”, the packet will be discarded and regarded as lost.

Simulation will assess the probability of successful packet transmission within 1ms for different numbers of accessing users. With increasing the number of users, the collision probability increases, which will lead to higher probability of packet failure.

Simulation Parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR</td>
<td>-3dB</td>
</tr>
<tr>
<td>sTTI</td>
<td>2 OFDM symbols</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
<td>15 kHz</td>
</tr>
</tbody>
</table>
Packet Model | Poisson arrival with lambda=10ms and the size 256bits
---|---
Antenna config. | 2x1
Transmit Mode | SFBC
Channel Model | ETU 3km/hr
AMC | ON
CQI Delay | 4 sTTI
MaxTransNum | 2
HARQ period | 4 sTTI

**Table 6-20: Schemes simulated for comparison**

<table>
<thead>
<tr>
<th>Case</th>
<th>MCC Bandwidth</th>
<th>MBB Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case1: FDM</td>
<td>5MHz</td>
<td>15MHz</td>
</tr>
<tr>
<td>Case2: FDM</td>
<td>10MHz</td>
<td>10MHz</td>
</tr>
<tr>
<td>Case3: EAI</td>
<td>20MHz (Sharing)</td>
<td></td>
</tr>
</tbody>
</table>

**Simulation Results**

![Figure 6-76: MCC user numbers vs. residual BLER (1ms latency limited)](image)

Figure 6-76 shows the performance of packet error probability within 1ms latency for different bandwidth allocations.

**Observations:**

1) The bandwidth required by MCC is rather large for meeting the strict reliability and latency requirements. In Case1 (5 MHz bandwidth for MCC), the residual BLER of 1e-05 cannot be attained at all.

2) With more bandwidth being reserved for MCC, more users per bandwidth unit can be supported;

3) The result shows that EAI can obtain a gain of roughly factor 4 compared to FDM schemes.
Performance Evaluation for eMBB

In this section, we provide some results about different schemes aiming at guaranteeing the performance of eMBB.

Investigated mechanism: Informing resource puncturing mode

Simulation assumptions:

1) Simplified SLS (system level simulation) using single cell and multiuser scenario, each user has equal average SNR but independent small scale fading;
2) For the sake of simplification, a fixed number of resources being punctured is assumed and no retransmissions are considered for MCC.

Simulation Parameters

<table>
<thead>
<tr>
<th>Table 6-21: Simulation Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR</td>
</tr>
<tr>
<td>System bandwidth</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
</tr>
<tr>
<td>MIMO</td>
</tr>
<tr>
<td>Transmitter mode</td>
</tr>
<tr>
<td>Frame structure of MBB</td>
</tr>
<tr>
<td>MBB TTI</td>
</tr>
<tr>
<td>MCC sTTI</td>
</tr>
<tr>
<td>Channel Model</td>
</tr>
<tr>
<td>AMC</td>
</tr>
<tr>
<td>MCC traffic and Puncture mode</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6-22: Schemes simulated for comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) no puncturing</td>
</tr>
<tr>
<td>2) Puncture (w/o inform)</td>
</tr>
<tr>
<td>3) Puncture and inform</td>
</tr>
</tbody>
</table>

Simulation Results

Figure 6-77 shows the throughput of eMBB of EAI compared with corresponding FDM schemes. “FDM” curve means that eMBB cannot use reserved bandwidth so that the throughput in this reserved bandwidth may become zero. From the figure, we can observe that EAI with informing the eMBB UE on the puncturing is more efficient than EAI without informing.
Figure 6-77 Potential gains with informing the eMBB user about its traffic being punctured.

Based on the simulation, we also found that the retransmission ratio increases significantly due to the resource puncturing from MCC especially at high SNR. Although only part of the eMBB allocation is punctured, a high waste of resources occurs, when the complete transmission block needs to be retransmitted. Hence, EAI for eMBB traffic can be optimized by applying adaptive HARQ (i.e. retransmitting code-blocks instead of transmission blocks).

Observations:

1) Informing the eMBB user is necessary for EAI, achieving about 90% gain.
2) Retransmission ratio increases significantly due to the resource of eMBB being punctured especially at high SNR.
3) Although only part of the eMBB allocation is punctured, high waste of resources occurs, if complete transmission blocks need to be retransmitted.

Investigated mechanism: Notifying puncturing information

Simulation assumptions:

It is considered that a MCC service data packet is generated every 2ms and one data packet punctures 36 resource blocks (RBs) of eMBB resources for the duration of a sTTI. Notifying puncturing information to eMBB UE could be facilitated by signaling the puncturing position of MCC service data in the resource map. The eMBB data punctured by MCC would be excluded after reception, and would not be used for data decoding.

Simulation Parameters

<table>
<thead>
<tr>
<th>SNR</th>
<th>0 to 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>System bandwidth</td>
<td>20 MHz</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
<td>15 KHz</td>
</tr>
<tr>
<td>MIMO</td>
<td>2x2</td>
</tr>
<tr>
<td>Transmitter mode</td>
<td>LTE TM2, rank=1</td>
</tr>
<tr>
<td>Frame structure of MBB</td>
<td>the same as LTE</td>
</tr>
</tbody>
</table>

Table 6-23 Simulation Assumptions
Simulation Results

Figure 6-78 Throughput performance with notification of puncturing information (Green – notified, Red- Unnotified)

Figure 6-78 shows that for an SNR of 10 dB, the scheme with notification of puncturing position can improve the throughput up to 40% compared to the scheme without notification.

Investigated mechanism: CB-Level HARQ

Simulation assumptions:
Two MCC traffic requests appear within one TTI (1ms) randomly. Every MCC request punctures 36 RBs for the duration of one sTTI (= 2 OFDM symbols). eMBB UE feeds back the location of erroneous CBs and eNB retransmits all those CBs. In enhanced TB-Level HARQ, eNB notifies eMBB UE which CB is punctured and will retransmit the entire TB.

Simulation Parameters

<table>
<thead>
<tr>
<th>Table 6-24: Simulation Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR</td>
</tr>
<tr>
<td>System bandwidth</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
</tr>
<tr>
<td>MIMO</td>
</tr>
<tr>
<td>Transmitter mode</td>
</tr>
<tr>
<td>Frame structure of MBB</td>
</tr>
<tr>
<td>MBB TTI</td>
</tr>
</tbody>
</table>
Simulation Results

As observed in Figure 6-79, simulation results show that compared to the HARQ TB-Level, the HARQ CB-Level can attain a gain of about 60%.

Investigated mechanism: Mapping MCC data to parity bits of eMBB

Simulation assumptions:

We assume 10 MCC UEs are multiplexed with 1 eMBB UE. In the first case, MCC UEs’ data are mapped to resources carrying parity bits of eMBB traffic only. In the second case for comparison, MCC UEs’ data are continuously mapped to a frequency resources allocated to eMBB traffic, which will puncture almost certainly the systematic bits.

Simulation Parameters

<table>
<thead>
<tr>
<th>SNR</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>System bandwidth</td>
<td>20 MHz</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
<td>15 KHz</td>
</tr>
<tr>
<td>MIMO</td>
<td>2x2</td>
</tr>
<tr>
<td>Transmitter mode</td>
<td>LTE TM2, rank=1</td>
</tr>
<tr>
<td>Frame structure of MBB</td>
<td>the same as LTE</td>
</tr>
<tr>
<td>MBB TTI</td>
<td>1ms</td>
</tr>
<tr>
<td>Channel Model</td>
<td>ETU</td>
</tr>
<tr>
<td>AMC</td>
<td>On</td>
</tr>
</tbody>
</table>

Table 6-26 Schemes simulated for comparison

<p>| case 1      | 0 MCC UE multiplex with eMBB UE |</p>
<table>
<thead>
<tr>
<th>case 2</th>
<th>10 MCC UE multiplex with eMBB UE, puncture parity bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>case 3</td>
<td>10 MCC UE multiplex with eMBB UE, continuously puncture</td>
</tr>
</tbody>
</table>

**Simulation Results**

As shown in Figure 6-80, the performance differs substantially if parity bits are punctured instead of systematic bits. Hence, it would be better to map MCC service data to resources occupied by parity bits of eMBB traffic.

6.8.5 **Harmonization of proposed PHY procedures**

Simulation results of the harmonized solution are discussed in this subsection. The results are compared to LTE baseline.

**Access delay:** For the evaluation of access delay of the proposed harmonized PHY procedure, simulation assumptions are listed in Table 6-27. A total number of 56 preambles are used for RA of eMBB and MCC traffic, 2 out of 56 are dedicated to MCC traffic in order to minimize the delay of MCC. Moreover, MMC traffic uses different frequency-time resources to perform RA, and hence all 64 available preambles can be used. It should be noted that the number of MMC users may be much larger than that of the other two traffic types in general.

<table>
<thead>
<tr>
<th></th>
<th>eMBB</th>
<th>MCC (reserved)</th>
<th>MMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of preambles</td>
<td>54</td>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td>Number of initial users</td>
<td>100</td>
<td>2</td>
<td>200</td>
</tr>
<tr>
<td>Number of new users in each transmission period</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Number of receive antennas</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detection probability</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal to noise ratio</td>
<td>5dB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It can be seen in Figure 6-81 that 45% of the MMC users are able to achieve RA in the first attempt, while the percentage for LTE reference is 20%. Similarly, 58% of the eMBB users are able to achieve RA in the first attempt, while the percentage for LTE reference is 35%.
Figure 6-81 CDFs of access delays of eMBB and MMC of the harmonized PHY procedure

Also, Figure 6-82 shows the PMF of access delay of MCC traffic for the harmonized PHY procedure. The proposed harmonized procedure reduces the access delay for MCC users. The success rate for the first attempt increases from 70% for the LTE reference to 90% for the proposed scheme.

Figure 6-82 PMF of access delay of MCC traffic for the harmonized PHY procedure

eMBB throughput: Downlink throughputs for the harmonized solution are depicted in Figure 6-83. To evaluate throughput performance, different TTI lengths were assumed for eMBB and MCC traffic as listed in Table 6-28. In the baseline solution, when there is no bandwidth sharing between eMBB and MCC, the eMBB throughput drops significantly as the number of MCC user increases. However, when bandwidth sharing is enabled as in the harmonized PHY procedure, significant eMBB and MCC throughput gains are both achievable.

Table 6-28: TTI length assumptions for throughput evaluation of the harmonized solution

<table>
<thead>
<tr>
<th></th>
<th>eMBB</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTI length</td>
<td>1ms</td>
<td>0.125ms</td>
</tr>
</tbody>
</table>
Figure 6-83 Downlink throughputs of eMBB and MCC. Left: Performance of baseline solution with no system bandwidth sharing. Middle: Performance of EAI with TB level retransmission. Right: Performance of EAI with CB level retransmission.