



iJOIN

INFSO-ICT-317941



INFSO-ICT-317941 iJOIN

Deliverable 2.2

Definition of PHY layer approaches that are applicable to RANaaS and a holistic design of backhaul and access network

Editor:	Jens Bartelt, TUD
Deliverable nature:	Deliverable
Suggested readers:	iJOIN GA
Due date:	October 31th, 2014
Delivery date:	October 31th, 2014
Version:	1.0
Total number of pages:	133
Reviewed by:	GA members
Keywords:	PHY State-of-the-art, RAN PHY, Backhaul PHY, RANaaS, Functional Split, iJOIN Architecture
Resources consumed	51.96 PM

Abstract

This deliverable defines the set of candidate technologies (CT) that are analysed inside the Work Package 2 (WP2) and that are applicable for a dense deployment of small cells and a joint optimization of access and backhaul networks. The evaluation method and a baseline for performance comparison are defined for each CT. Following this, the obtained evaluation results are described and the compliance to the iJOIN objectives is addressed. Besides, more detailed investigations concerning the functional split options as well as precoder and decoder implementation in a cloud platform based on General Purpose Processors (GPP) are provided. Finally, the virtual eNB implementation aspects concerning in particular interfaces, information exchange and constraints to be respected are analysed in detail.

List of authors

Company	Author
IMC	Umer Salim (umer.salim@intel.com)
TUD	Jens Bartelt (jens.bartelt@ifn.et.tu-dresden.de)
UoB	Dirk Wübben (wuebben@ant.uni-bremen.de) Henning Paul (paul@ant.uni-bremen.de) Guang Xu (xu@ant.uni-bremen.de) Ban-Sok Shin (shin@ant.uni-bremen.de)
UNIS	Lei Zhang (lei.zhang@surrey.ac.uk) Atta Quddus (a.quddus@surrey.ac.uk) Efsthios Katranaras (e.katranaras@surrey.ac.uk)
CEA	Valentin Savin (valentin.savin@cea.fr) Matteo Gorgoglione (matteo.gorgoglione@cea.fr)
SCBB	Massinissa Lalam (massinissa.lalam@sagemcom.com)
TI	Boldi Mauro Renato (mauro.boldi@telecomitalia.it) Bruno Melis (bruno1.melis@telecomitalia.it)

History

Modified by	Date	Version	Comments
Jens Bartelt	October 31 th , 2014	1.0	Final Version

Table of Contents

List of authors.....	2
History.....	3
Table of Contents.....	4
List of Figures.....	6
List of Tables.....	8
Abbreviations.....	9
List of Symbols.....	13
1 Executive Summary.....	17
2 Introduction.....	18
2.1 Motivation and Background.....	18
2.2 Key Contributions.....	18
3 Functional Split and Virtual eNB.....	21
3.1 iJOIN Architecture.....	21
3.2 Functional Split.....	22
3.2.1 Functional Split Options.....	22
3.2.1.1 Downlink.....	24
3.2.1.2 Uplink.....	31
3.2.1.3 Numerical Example and Summary.....	33
3.2.2 Functional Split Investigation for Main Blocks for DL and UL.....	35
3.2.2.1 Precoder @ Cloud.....	36
3.2.2.2 Decoder @ Cloud.....	40
3.2.3 Decoder Implementation on Cloud Platform.....	41
3.3 Virtual eNB Implementation.....	44
3.3.1 veNB and 3GPP compliance.....	44
3.3.2 Further description of veNB.....	46
3.4 Physical layer constraints.....	46
4 iJOIN Physical Layer Candidate Technologies.....	51
4.1 CT2.1: In Network Processing.....	51
4.1.1 Technical description.....	51
4.1.2 Implementation in the iJOIN architecture.....	56
4.1.3 Evaluation of the CT.....	59
4.2 CT2.2: Multi-Point Turbo Detection.....	65
4.2.1 Technical description.....	65
4.2.2 Implementation in the iJOIN architecture.....	67
4.2.3 Evaluation of the CT.....	69
4.3 CT2.3: Joint Network-Channel Coding.....	72
4.3.1 Technical description.....	72
4.3.2 Implementation in the iJOIN architecture.....	74
4.3.3 Evaluation of the CT.....	77
4.4 CT2.4: Sum-Rate and Energy-Efficiency metrics of DL CoMP with backhaul constraints.....	82
4.4.1 Technical description.....	82
4.4.2 Implementation in the iJOIN architecture.....	89
4.4.3 Evaluation of the CT.....	91
4.5 CT2.5: Partially Centralized Inter-Cell Interference Coordination.....	94
4.5.1 Technical description.....	94
4.5.2 Implementation in the iJOIN architecture.....	96
4.5.3 Evaluation of the CT.....	98
4.6 CT2.6: Data Compression over RoF.....	100
4.6.1 Technical description.....	100
4.6.2 Implementation in the iJOIN architecture.....	100
4.6.3 Evaluation of the CT.....	102
4.7 CT2.7: Millimetre Wave Backhauling.....	105
4.7.1 Technical description.....	105
4.7.2 Implementation in the iJOIN Architecture.....	108

4.7.3	Evaluation of the CT.....	110
5	Overall Evaluation	112
6	Summary and Conclusions.....	114
	Acknowledgements and Disclaimer	115
	Appendix I Input/Output Interfaces	116
	Appendix II Evaluation Methodology	122
II.1	Evaluation methodology	122
II.2	Scenario 1: Outdoor deployment	122
II.3	Scenario 2: Indoor deployment	125
	References	127

List of Figures

Figure 3-1: Functional Architecture of PHY processing.....	21
Figure 3-2: Project-wide functional split options.....	23
Figure 3-3: PHY-layer detailed functional split options.....	23
Figure 3-4: Power adjustment operation at the iSC in the LTE case.....	29
Figure 3-5: The total CSI delay of the precoding operation.....	38
Figure 3-6: Exploiting parallelism in multiprocessor/multicore platforms.....	41
Figure 3-7: FER and average number of decoding iterations, WiMAX LDPC with rate 1/2 and 5/6.....	43
Figure 3-8: Multi-codeword decoder with synchronized threads.....	43
Figure 3-9: Average Throughput as function of the number of cores, WiMAX LDPC with rate 1/2 and 5/6.....	43
Figure 3-10: Spectral efficiency and computational complexity results for turbo decoding on the UoB RANaaS testbed.....	44
Figure 3-11 System architecture with virtual eNodeB (veNB) consisting of a central RANaaS platform and several iSCs.....	45
Figure 4-1: Multi-user scenario in a dense deployment of iSCs.....	52
Figure 4-2: Functional split B.1 for CT2.1.....	56
Figure 4-3: Functional split C for CT2.1.....	57
Figure 4-4: BER and FER of an LTE uplink transmission with MCS 7 and a bandwidth of 1.4 MHz.....	61
Figure 4-5: Throughput of an LTE uplink transmission with MCS 7 and a bandwidth of 1.4 MHz.....	61
Figure 4-6: BER of an LTE uplink transmission with MCS 7 and a bandwidth of 1.4 MHz with perfect (dotted) and noisy (solid) backhaul links.....	61
Figure 4-7: Floating point operations per subcarrier and per iSC for a system with $N_{iSC} = 6$, $N_R^{iSC} = 2$, $N_{UE} = 4$, $N_T^{UE} = 2$	63
Figure 4-8: Considered backhaul topologies for an example of $N_{iSC} = 4$ iSCs.....	63
Figure 4-9: FER and throughput over total communication overhead for considered backhaul topologies... ..	64
Figure 4-10: MPTD Simulation Set-up.....	65
Figure 4-11: MPTD System Model.....	66
Figure 4-12: HARQ modelling in the Monte-Carlo framework.....	66
Figure 4-13: Turbo-detection principle.....	67
Figure 4-14: MPTD performed at the RANaaS platform using the J1 interface.....	68
Figure 4-15: SPTD performed in each iSC with possible cooperation using the X2 interface.....	68
Figure 4-16: FER vs SNR (a) and Average number of transmissions vs SNR (b) for the QPSK MCS.....	70
Figure 4-17: FER vs SNR (a) and Average number of transmissions vs SNR (b) for the 16QAM MCS.....	71
Figure 4-18: FER vs SNR (a) and Average number of transmissions vs SNR (b) for the 64QAM MCS.....	71
Figure 4-19: Proposed scenario for joint network-channel coding.....	72
Figure 4-20: Joint network-channel graph.....	73
Figure 4-21: Example of mirroring of a cycle of length 4.....	73
Figure 4-22: Joint network-channel graph with Repeat Accumulate (RA) structure.....	74
Figure 4-23: Functional Split C for CT2.3.....	75
Figure 4-24: Functional Split B.2 for CT2.3.....	76
Figure 4-25: Baseline scenario, no direct UEs-iSC ₂ links.....	77
Figure 4-26: FER for WiMax and (2,4)-LDPC codes with different network graph constructions (symmetric scenario).....	79
Figure 4-27: FER for WiMax and (2,4)-LDPC codes with different network graph constructions (asymmetric scenario).....	79
Figure 4-28: Example of non-binary network check-node.....	80
Figure 4-29: FER for WiMax and (2,4)-LDPC codes with binary and NB network coding and random interleaving.....	80
Figure 4-30: Analysis of the percentage of transmitted bits from the relay.....	81
Figure 4-31: Grouping of iSCs to veNBs operated on a same RANaaS data centre.....	82
Figure 4-32: veNB division for planar deployment and each veNB is shaped as a hexagon.....	85
Figure 4-33: veNB area containing a single UE at \hat{G} and two contribution regions of the desired signal power.....	85

Figure 4-34: functional split A.3 for CT2.4.....	89
Figure 4-35: Output SINR for ZF-based planar deployment versus available computational resource factor q_c	92
Figure 4-36: Sum-rate for ZF-based planar deployment versus available computational resource factor q_c ..	93
Figure 4-37: The total delay and their component versus available computational resource factor q_c	93
Figure 4-38: iJOIN architecture with both J1 links between the RANaaS and the iSCs and J2 links between the iSCs	94
Figure 4-39: Two possible practical scenarios leading to a hierarchical CSIT configuration.....	95
Figure 4-40: Ergodic sum rate achieved in a wireless network with 4 single-antennas iSCs and 4 single-antennas UEs with only two iSCs having access to a perfect CSI.....	99
Figure 4-41: Functional split A.3	100
Figure 4-42: Area throughput versus available BH capacity per area in Gbit/s/km ²	104
Figure 4-43: iSC density versus available BH capacity per area in Gbit/s/km ²	104
Figure 4-44: Wireless BH architecture.....	105
Figure 4-45: CT2.7 system model	105
Figure 4-46: Overview of the approach for joint design of access and backhaul.....	106
Figure 4-47: Joint encoding procedure.....	107
Figure 4-48: Distribution of received symbols after erroneous backhaul	108
Figure 4-49: Functional split A.1:All digital processing in RANaaS.....	108
Figure 4-50: Functional split A.4: Detection and decoding in RANaaS	109
Figure 4-51: BER when using encoded BH (green lines) as compared to an uncoded BH with the ERD (blue lines) and with a SISODQ (red lines).....	111
Figure 4-52: Throughput when using encoded BH (green lines) as compared to an uncoded BH with the ERD (blue lines) and with a SISODQ (red lines)	111

List of Tables

Table 3-1: System parameters for BH load calculations	24
Table 3-2: System parameters used in the numerical example.....	33
Table 3-3: Summary of functional split options.....	35
Table 4-1: List of iJOIN PHY Candidate Technologies (CTs)	51
Table 4-2: Overview of algorithmic steps of presented DiCE variants and PALCE per iteration	55
Table 4-3: PHY Measurement placement and exchange links for CT2.1	58
Table 4-4: Backhaul control information of CT 2.1	58
Table 4-5: Backhaul payload of CT 2.1	59
Table 4-6: Number of floating point operations per subcarrier and per iSC	62
Table 4-7: Communication overhead per iteration and subcarrier for different topologies	64
Table 4-8: PHY Measurement placement and exchange links for CT2.2	68
Table 4-9: Backhaul control information of CT 2.2	69
Table 4-10: Backhaul Load of CT 2.2.....	69
Table 4-11: PHY Measurement placement and exchange links for CT2.3	76
Table 4-12: Backhaul control information of CT 2.3	76
Table 4-13: Backhaul Load of CT 2.3.....	77
Table 4-14: PHY Measurement placement and exchange links for CT2.4	90
Table 4-15: Backhaul control information of CT 2.4.....	90
Table 4-16: Backhaul Load of CT 2.4.....	91
Table 4-17: PHY Measurement placement and exchange links for CT2.5	97
Table 4-18: Backhaul control information of CT 2.5.....	97
Table 4-19: Backhaul Load of CT 2.5.....	98
Table 4-20: PHY Measurement placement and exchange links for CT2.6	101
Table 4-21: Backhaul control information of CT 2.6.....	101
Table 4-22: Backhaul Load of CT 2.6 for payload.....	102
Table 4-23: PHY Measurement placement and exchange links for CT2.7	109
Table 4-24: Backhaul control information of CT 2.7.....	109
Table 4-25: Backhaul Load of CT 2.7.....	110
Table 5-1: Categorization and compatibility of WP2 CTs	112
Table 5-2: Qualitative impact of WP2 CTs on iJOIN objectives	113
Table 6-1: Requested Input of WP2 CTs.....	116
Table 6-2: Provided Output of WP2 CTs	118
Table 6-3: Explanation of Parameters	119
Table 6-4: 3GPP outdoor deployment assumptions	123
Table 6-5: Mapping of the iJOIN WP2 assumptions to the 3GPP outdoor model.....	124
Table 6-6: ITU indoor deployment assumptions.....	125
Table 6-7: Mapping of the iJOIN WP2 assumptions to the 3GPP indoor model.....	126

Abbreviations

3GPP	3 rd Generation Partnership Program
A/D	Analogue to Digital
ADMM	Alternating Direction Method of Multipliers
AF	Amplify-and-Forward
ASIC	Application Specific Integrated Circuit
AWGN	Additive White Gaussian Noise
BBU	Base Band Unit
BER	Bit Error Rate
BH	Backhaul
BS	Base Station
BP	Belief Propagation
CB	Coordinated Beamforming
CDF	Cumulative Distribution Function
CF	Compress-and-Forward
CFO	Carrier Frequency Offset
CoMP	Coordinated Multi-Point
CP	Cyclic Prefix
CPRI	Common Public Radio Interface
QCI	Channel Quality Indicator
C-RAN	Centralized RAN
CRC	Cyclic Redundancy Check
CRS	Cell Reference Signal
CS	iJOIN Common Scenario
CSCG	Circularly Symmetric Complex Gaussian
CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
CT	Candidate Technology
CW	Codeword
D/A	Digital to Analogue
DCI	Downlink Control Information
DCSI	Distributed Channel State Information
DF	Decode-and-Forward
DFT	Discrete Fourier Transform
DiCE	Distributed Consensus-Based Estimation
DL	Downlink
DMRS	Demodulation Reference Signal
DPB	Dynamic Point Blanking
DPC	Dirty Paper Coding
E-LMMSE-IRC	Enhanced Linear Minimum Mean Square Error with Interference Rejection Combining
eNB	Evolved Node B
E/O	Electrical to Optical
EPRE	Energy Per Resource Element
ERD	Error Resilient Decoder
EPC	Evolved Packet Core
FD	Frequency Domain
FDD	Frequency Division Duplex
FD-RoF	Frequency Domain Radio over Fibre
FEC	Forward Error Correction
FER	Frame Error Rate
FFT	Fast Fourier Transform
FLOPS	Floating Point Operations
FPGA	Field Programmable Gate Array
GF	Galois Field

GNSS	Global Navigation Satellite System
GP-HW	General Purpose Hardware
GPS	Global Position System
GPP	General Purpose Processors
HARQ	Hybrid Automatic Repeat Request
HeNB	Home eNB
I	In-phase (sample)
ICI	Inter-Cell Interference
ICIC	Inter-Cell Interference Coordination
IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
i.i.d.	Independently and Identically Distributed
iJOIN	Interworking and JOINT Design of an Open Access and Backhaul Network Architecture for Small Cells based on Cloud Networks
iNC	iJOIN Network Controller
InH	Indoor / Hotspot
INP	In-Network Processing
IR	Incremental Redundancy
IRC	Interference Rejection Combining
iSC	iJOIN Small Cell
ISI	Inter Symbol Interference
iTN	iJOIN Transport Node
ITU	International Telecommunication Union
iveC	iJOIN virtual eNB Controller
JNCC	Joint Network-Channel Coding
JT	Joint Transmission
LLR	Log Likelihood Ratio
LDPC	Low-Density Parity-Check
LOS	Line Of Sight
LTE	Long Term Evolution
LS	Least Square
MAC	Medium Access Control
MARC	Multiple-Access Relay Channel
MCS	Modulation and Coding Scheme
MIMO	Multiple-Input Multiple Output
MMSE	Minimum Mean Square Error
mmWave	Millimetre Wave
MP	Message Passing
MPI	Message Passing Interface
MPTD	Multi-Point Turbo Detection
MRT	Maximum Ratio Transmission
MS	Min-Sum
MUD	Multi-User Detection
NAIC	Network-Assisted Interference Cancellation
NB	Non-Binary
NC	Network Coding
nfX-DiCE	Noise-free exchange Distributed Consensus-Based Estimation
OBSAI	Open Base Station Architecture Initiative
O/E	Optical to Electrical
OFDM	Orthogonal Frequency Division Multiplexing
OpenMP	Open Multi-Processing
OTA	Over The Air
PALCE	Priority-based Augmented Lagrangian Cooperative Estimation
PAPR	Peak to Average Power Ratio
PCI	Physical Cell Identity
PCFICH	Physical Control Format Indicator Channel

PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channel
PDF	Probability Density Function
PEG	Progressive Edge-Growth
PER	Packet Error Rate
PHICH	Physical HARQ Indicator Channel
PHY	Physical Layer
PL	Path Loss
PMI	Precoder Matrix Index
PPB	Parts Per Billion
PRACH	Physical Random Access Channel
PRB	Physical Resource Block
PSS	Primary Synchronization Signal
PTP	Precision Time Protocol
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
Q	Quadrature (sample)
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RA	Repeat-Accumulate
RAN	Radio Access Network
RANaaS	RAN-as-a-Service
RAP	Radio Access Point
RAT	Radio Access Technology
RB	Resource Block
RBG	Resource Block Group
RF	Radio Frequency
RE	Resource Element
RI	Rank Indicator
RLC	Radio Link Control
RN	Relay Node
RO-DiCE	Reduced Overhead Distributed Consensus-Based Estimation
RoF	Radio over Fibre
RRC	Radio Resource Control
RRH	Remote Radio Head
RRM	Radio Resource Management
RS	Reference Signal
RS-EPRE	Reference Signal EPRE
SC	Small Cell
SC-FDMA	Single-Carrier Frequency Division Multiple Access
SINR	Signal to Interference and Noise Ratio
SISODQ	Soft-Input Soft-Output DeQuantizer
SNR	Signal to Noise Ratio
SP	Sum Product
SPTD	Single-Point Turbo Detection
SSS	Secondary Synchronization Signal
SRS	Sounding Reference Signals
TBS	Transport Block Size
TD	Time Domain
TDD	Time Division Duplex
THP	Tomlinson Harashima Precoding
TM	Transmission Mode
UE	User Equipment
UL	Uplink
veNB	virtual eNB

V-BLAST	Vertical Bell Laboratories Layered Space-Time
WiMax	Worldwide Interoperability for Microwave Access
WP	Work Package
ZF	Zero Forcing

List of Symbols

System Parameters

a_{veNB}	area of veNB
A_{net}	area of network
c_0	speed of light
C_{BH}	backhaul capacity per unit area
C_{fb}	CSI feedback capacity from UE to iSC
d_c	check-node degree
d_y	bit-node degree
$d_{ue_to_iSC}$	distance from UE to iSC
$d_{iSC_to_RANaaS}$	distance from iSC to RANaaS
D_P	payload message size
D_S	signalling message size
E_A	energy per data RE in OFDM symbols that do not contain RS
E_B	energy per data RE in OFDM symbols that contain RS
E_{RS}	energy per RS RE
f_s	sampling frequency
f_c	carrier frequency
f_D	Doppler shift
F_{FLOPS}	RANaaS total computational capability (FLOPS)
$J_0(\cdot)$	zero-order Bessel function of first kind
K_{multi}	real-valued operations of multiplication for ZF precoding
K_{add}	real-valued operations of addition for ZF precoding
N_{ant_tot}	total channel coefficient to be estimate at UE
N_{block}	no. of bits in one code block
N_{CB}	no. of codebook elements
$N_{CSI_per_fb}$	no. of channel coefficients feedback from UE to iSC each time
N_D	no. of bits transmitted from destination iSC to RANaaS
N_{fb_time}	total number of times for channel coefficients feedback
N_{info}	no. of information bits in one code block
N_I	no. of inputs, total no. of transmit antennas
N_{iSC}	no. of iSCs
$N_L^{UE,u}$	no. of spatial transmit layers by UE u on UL
N_L	no. of transmitted layers
N_{MCS}	no. of bits used to encode the MCS information
N_{MOD}	no. of bits used to encode the modulation scheme information
N_O	no. of outputs, total no. of receive antennas
N_{PA}	no. of bits used to encode the power adjustment information
N_Q	no. of quantization bits
N_{Q_CSI}	no. of quantization bits for CSI

N_R^{iSC}	no. of receive antennas at iSC
N_R^{UE}	no. of receive antennas at UE
$n_{RB,u}^j$	no. of resource blocks allocated to the u -th user served by the j -th iSC
N_{RA}	no. of bits used to encode the resource allocation information
N_{RB}	no. of available resource blocks
N_s	no. of signals to be transmitted on the backhaul link
N_{sc}	no. of subcarriers available for transmission
$N_{sc,u}^j$	no. of subcarriers allocated to the u -th user served by the j -th iSC
N_R	no. of bits transmitted from relaying iSC to either destination iSC or RANaaS
N_T^{iSC}	no. of transmit antennas at iSC
N_T^{UE}	no. of transmit antennas at UE
N_{TM}	no. of bits used to encode the transmission mode information
N_{veNB}	no. of virtual eNBs in the network
N_{UE}	no. of UEs
\mathcal{N}_j	Set of all neighbouring iSCs of the iSC j
\mathcal{N}_j^+	Set of all neighbouring iSCs of the iSC j and itself
\mathcal{N}_{j+}	Set of neighbouring iSCs i of the iSC j for $i > j$
\mathcal{N}_{j-}	Set of neighbouring iSCs i of the iSC j for $i < j$
OF	oversampling factor
P	parameters used to encode the resource allocation information
q	cardinality of the Galois Field
q_c	RANaaS computational resource division factor
Q_m	modulation order
R_t	radius of network
$S_{sc,u}^j$	set of subcarriers allocated to the u -th user served by the j -th iSC
S_{sc}^j	set of all subcarriers allocated the users served by the j -th iSC
T_s	symbol duration
T_{SUB}	subframe duration
T_f	one transmission duration of CF relay
T_{50}	channel coherence time at 50%
Δt	total CSI delay of the precoding operation
$\Delta t_{chan_est_UE}$	channel estimation delay at the UE
Δt_{prop_total}	total round trip propagation delay
Δt_{fb}	channel feedback delay
$\Delta t_{process_RANaaS}$	processing delay at RANaaS
$\Delta t_{process_iSC}$	processing delay at iSC
Δt_{Tx}	total Tx processing time in LTE system
γ	overhead introduced by BH protection coding
κ	path-loss exponent
λ	temporal correlation factor

ρ_{UE}	density of UE
ρ_{iSC}	density of iSC
$\rho_{T_{50}}$	correlation corresponding to channel coherence time T_{50}
$(\sigma_u^{(j)})^2$	variance of the quantization noise of the CSI relative to user u at the iSC j
ζ_1	number of equivalent addition operation times for each division operation
ζ_2	number of equivalent addition operation times for each multiplication operation

General Parameters

B	bandwidth
C	capacity
D	data rate
N_{It}	no. of iterations
R_c	coding rate
S	spectral efficiency
η	throughput
\mathcal{E}	encoder, code
FEC_{BH}	forward-error correction on backhaul connection
SNR_{BH}	SNR on backhaul connection

Signals

\mathbf{a}	information word
\mathbf{c}	code word
\mathbf{G}	channel path-loss
h	non-binary coefficient of the parity-check matrix
\mathbf{h}	channel vector
\mathbf{H}	channel matrix
\mathbf{n}	noise vector
\mathbf{s}	modulated symbols
\mathbf{U}	detector matrix
\mathbf{v}	backhaul information word
\mathbf{w}	backhaul code word
\mathbf{H}	parity-check matrix
$\mathbf{w}_{i,u}$	precoding vector for the u -th UE in the i -th veNB (cooperative cluster)
\mathbf{W}	precoding matrix
\mathbf{W}_i	precoding vector for the i -th veNB (cooperative cluster)
\mathbf{y}	receiver signal vector
\mathbf{x}, \mathbf{z}	transmitter signal vector
$\delta_u^{(j)}$	quantization noise of the CSI relative to user u at the iSC j

General Notation of Symbols

a	scalar
\mathbf{a}	vector
\mathbf{a}^T	transpose
\mathbf{a}^H	transpose conjugate
\mathbf{A}	matrix
\tilde{a}	soft estimate
\hat{a}	hard estimate
\mathbf{a}'	distorted (received) vector

Indices

i, j, ν	iSC
k	iteration
l	symbol within frame
n	carrier
u	user

1 Executive Summary

Within iJOIN, promising candidate technologies (CT) as well as the initial definition of the iJOIN architecture were previously investigated in D2.1 [36] for physical layer (PHY) aspects, in D3.1 [31] for medium access control (MAC) / radio resource management (RRM) aspects, in D4.1 [32] for network layer aspects, and in D5.1 [17] for the overall system definition. This deliverable now provides a more detailed analysis of the physical layer for the set of CTs investigated inside the Work Package 2 (WP2). It describes their applicability to the iJOIN architecture as well as their corresponding implementation, how iJOIN's objectives can be achieved by these approaches and which practical constraints and requirements are met. Furthermore, preliminary performance results are provided.

Section 3 recapitulates important aspects of the iJOIN architecture and discusses general topics investigated in WP2. The functional split approaches which determine the execution of processing in the iJOIN small cell (iSC) or in the RAN-as-a-Service (RANaaS) platform and directly influence the required backhaul (BH) data rate are reviewed. Besides, more detailed investigations concerning the precoder and decoder implementation in a cloud platform based on general purpose processors (GPP) are also provided. The precoder and decoder functions are fundamental components of the physical layer for the downlink and uplink, respectively [36]. The precoder is a key building block for the implementation of coordinated transmission techniques in the downlink and its placement in the RANaaS makes possible to consider centralized cooperation schemes that address the interference coming not only from neighbouring cells. The decoder is the last functional component of the physical layer at the receiver side in the uplink. The implementation in the RANaaS poses a significant demand in terms of computational resources, even though the availability of multi-core processors and the inherent decoding parallelism of modern forward error correction (FEC) techniques open new perspectives for the cloud implementation. The virtual evolved node B (veNB) concept is also analysed in Section 3.

In Section 4, promising PHY CTs for a dense deployment of small cells and a joint optimization of access and backhaul networks are analysed in detail. The CTs' scenarios, system models, and approaches are described and the impact on the implementation, namely the load induced on the backhaul and the required message exchange, are provided. The methodology and the baseline for evaluation are introduced as well, before some preliminary results are discussed.

Finally, for the purpose of a comparable performance evaluation among the different CTs, common simulation scenarios and a set of harmonized parameters have been defined and can be found in the appendix. In particular two simulation scenarios based on 3GPP reference scenarios are considered: an indoor scenario corresponding to the iJOIN Common Scenario (CS) 4, which could be an airport or a shopping mall, and an outdoor scenario corresponding mainly to the CS2 (e.g. a square) or to the CS3 (wide-area continuous coverage). Using these scenarios, a methodology for comparing the results of different CTs is given in Section 5.

2 Introduction

2.1 Motivation and Background

Future mobile communications systems will have to support an exponential growth of traffic caused by a tremendous increase of mobile devices and the emergence of new applications. The expected increase of traffic can only be supported by a significant change of wireless network architecture. The physical layer has to meet the excessive throughput requirements from upper layers. This can be achieved through dense deployment of small cells. Overlaying the existing cellular networks with small cells has three main advantages. Firstly, the distances between access points (small cell) and user equipments decreases and thus path-loss is reduced significantly. Hence maintaining the same quality of service, modulation order and coding rate can be increased due to the improved signal to interference and noise ratio (SINR) which results in significant gains in data rate. Secondly, through smaller cell sizes the available spectrum is reused more often per area. Thirdly, small cell deployment causes each user to be in the vicinity of multiple small cells. This gives a unique possibility of collaboration and cooperation between the neighbouring cells. Thus, multiple cells can transmit/receive in a cooperative way. Such transmission schemes, known widely as “network MIMO”, lead to huge throughput benefits. However, they require sophisticated cooperative transmission techniques in the downlink to cope with the strong inter-cell interference and advanced receiver approaches in the uplink to detect the user messages by their spatial separation.

In order to meet these tremendous requirements, the iJOIN small cell (iSC) and the RAN-as-a-Service (RANaaS) concept have been introduced in this project as practical solutions to achieve the improvements promised by the theoretical results. The developed approach allows for a flexible cooperative processing among iSCs and a (partially) centralized processing in the RANaaS. The communication between these entities is provided by the backhaul network. In order to maximize the overall system performance, both the radio access and the backhaul network have to be jointly optimized. In the iJOIN project, the J1 logical link is introduced to represent the backhaul links between RANaaS and iSCs while backhaul links between iSCs are realized by logical J2 links.

The main objective of WP2 in iJOIN is to investigate physical layer approaches for a joint design of access and backhaul network with the aim of increasing the overall throughput, energy efficiency, and utilisation efficiency of very dense small cell networks with (partially) centralized processing. In the first deliverable D2.1 [36], we described the corresponding state-of-the-art from a physical layer perspective and presented a set of innovative candidate technologies. In this deliverable, these technologies are further investigated, applied to the developed iJOIN architecture and evaluated. Furthermore, the definition of the virtual evolved node B (veNB) is investigated from PHY layer perspective, principle functional splits are analysed, and practical implications by cloud implementations are discussed. In order to evaluate each CT with respect to the envisioned performance improvements, parameters for two typical scenarios are defined.

2.2 Key Contributions

The RANaaS concept proposed by the iJOIN project allows classical functionalities, usually performed within an evolved node B (eNB), to be partially or fully executed in a cloud platform. This paves the way for new physical layer approaches, where the coordinated processing over several small cells allows to overcome the severe inter-cell interference that arises in dense deployment scenarios. The key innovation in WP2 is the flexible cooperation of iSCs in PHY-layer processing, utilizing not only centralized, but also distributed approaches like in-network processing or network coding, which stand in contrast to the more rigid concepts of centralized RAN (C-RAN). Also, iJOIN envisions the implementation of PHY-layer processing (e.g. turbo decoding) on commodity hardware, as compared to the dedicated hardware used in current eNBs or CRAN baseband-pools. Here we briefly summarize the main novelty of the CTs previously defined in D2.1 [36] and concisely show the update in the current phase of the project:

CT2.1 In-Network Processing (INP) investigates distributed implementation of multi-user detection among densely deployed iSCs. Distributed multi-user detection can be achieved with the help of the distributed consensus-based estimation (DiCE) [2], algorithm based on a least square (LS) optimization problem with a consensus constraint which has been presented in deliverable D2.1 [36] and analysed in [3] and [4]. In this deliverable, we present new variants of the DiCE algorithm which have been developed within the iJOIN project. The reduced overhead DiCE (RO-DiCE) introduced in [5] avoids an exchange of Lagrange

multipliers among the iSCs and thus shows a reduced communication overhead. The Fast-DiCE proposed in [38] shows an accelerated convergence compared to DiCE by introducing predictors into the update equations. The priority-based augmented Lagrangian cooperative estimation (PALCE) aims to reduce overhead using no auxiliary variable and less exchange of multipliers among iSCs [86]. The different DiCE algorithms are compared by means of error rate and throughput performance using an LTE Release 10 compliant uplink level simulation chain. Additionally, the corresponding computational complexity and the backhaul traffic with specific backhaul topologies are investigated with corresponding results published in [39] and [82].

CT2.2 Multi-Point Turbo Detection (MPTD) evaluates the benefit of applying the turbo detection principle in an uplink multi-user scenario with dense small cell deployment. On the opposite side of the traditional frequency partitioning scheme that relies on orthogonal allocation, the main idea here is to allow cell edge users to be scheduled on the exact same resources. By processing iteratively the received signals from various small cells, the joint detection of the user flows can be performed either centrally or locally. Preliminary results on the worst case scenario where two edge users are attached to two different small cells, but are close enough to have similar propagation conditions toward each small cell (same received power), show a significant gain when using the MPTD centralized approach compared to advanced linear receivers, traditionally applied locally. The gains are evaluated in terms of frame error rate (FER) and number of retransmission needed (HARQ) using link level simulations with the LTE Release 10 compliant modulation and coding scheme over a well-accepted MIMO channel model (ITU-R InH). We can observe that for a given radio condition, the MPTD approach gives a lower FER and requires less retransmission, leading to a potential throughput increase when applied.

CT2.3 investigates Joint Network-Channel Coding (JNCC) solutions for uplink transmission in a dense small cell deployment. The small cell deployment is separated in a number of multiple-access relay channels (MARC), each one comprising two users that use an intermediate (relay) small cell to communicate with a final (destination) small cell. This CT focuses on the joint optimization of users' channel codes and the network coding (NC) technique, which allows combining relayed messages prior to relay transmission. The main goal is to maximize the user throughput, while minimizing the amount of data to be transmitted on the backhaul (relay-to-destination) link. In this deliverable, we propose and evaluate several bit-interleaved NC designs, including (1) different constructions of the bit-interleaver placed between channel and network codes, and (2) different structures of the NC graph (e.g. using binary or non-binary codes, possibly coupled with a repeat-accumulate structure). The proposed NC designs are further combined with different channel codes at users' side, in order to evaluate the overall performance of the joint NC and channel coding scheme. This investigation highlights a trade-off between the error-correction performance at relay and destination nodes. Finally, we investigate the impact of the JNCC technique on the backhaul traffic load, and show that significant load reduction (30 to 90 %) can be achieved, depending on the quality of user-to-destination links.

CT2.4 Sum-Rate and Energy-Efficiency metrics of DL CoMP with backhaul constraints (CoMP) considers Coordinated Multi-Point (CoMP) techniques for downlink transmission taking backhaul constraints into account. First, the energy consumption of access and backhaul for joint transmission (JT) and for coordinated beamforming (CB) in the presence of backhaul capacity constraints was investigated as reported in deliverable D2.1 [36]. In order to forward precoded signals over quantized signals to the users, a distributed Wyner-Ziv coding scheme was proposed and examined in [99]. In this report, we investigate the optimal number of iSCs allocated to a veNB for joint downlink transmission in terms of network sum-rate in the presence of latency that is mainly caused by the RANaaS processing delay and channel state information (CSI) feedback delay. Theoretically, increasing the number of iSCs that are jointly processed would improve the interference cancellation gains leading to higher spectral efficiency. However, as the latency increases with the number of cooperating iSCs, the mismatch between the channel used for designing the precoder and the actual transmit channel causes performance degradations. In order to trade off the interference and the delay-caused mismatch and maximizing throughput, the whole network is divided into separate veNBs and the veNB size is treated as the optimization objective. In contrary to traditional cellular systems which are usually modelled by placing base stations (BSs) on a grid, we consider a random placement of the iSCs and the users and both large scale and fast fading factors of the wireless channels are taken into account. The output signal to interference and noise ratio (SINR) of the network is derived based on the zero forcing precoding scheme. Then the sum-rate optimization problem in terms of the cluster size is formulated and solved. Simulation results show the effectiveness of proposed algorithms.

CT2.5 Partially Centralized Inter-Cell Interference Coordination (ICIC) is focused on the problem of designing downlink cooperation schemes exploiting efficiently the backhaul architecture. Indeed, the heterogeneous backhaul makes the assumption of fully centralized or perfect CSI at the iSCs unrealistic and requires studying more advanced imperfect CSI configurations. In this deliverable, we focus on a particular CSI configuration, denoted as "hierarchical", where the iSCs can be ordered by increasing quality of CSI. This configuration is practically interesting and widely used (Master-Slave is a particularly simple hierarchical configuration). For that CSI configuration, we derive a new precoding scheme being robust to this imperfect knowledge of the CSI at the iSCs and outperforming conventional robust schemes from the literature. The precoding algorithm described in this deliverable has led to the publication [24].

CT2.6 Data Compression for Radio over Fibre (RoF) analyses possible PHY functional splits between RANaaS and iSCs with the corresponding impact on the backhaul load and control information required at the iSCs. Particular focus is given to frequency domain (FD) transmission in order to keep the iSCs as simple as possible, reduce data rate without losing precision (as in compression methods) and enable downlink precoding at the iSCs on a per-user basis. The backhaul load is calculated analytically in terms of message size and frequency for both payload and signalling, analysing also the relation between backhaul load, system parameters (e.g. system bandwidth, number of quantization bits) and radio interface load. These dependencies are important in the selection of the most effective PHY layer functional splits as a function of the backhaul characteristics and deployment scenarios.

CT2.7 Millimetre Wave Backhauling (mmWave) investigates the use of wireless links in the 60-90 GHz range for a high-capacity, low-latency alternative to fibre. Since these wireless links need to cope with varying outdoor channel conditions, the CT focuses on an adaptive joint coding and decoding of radio access and backhaul links. To reduce latency, the possibility to use uncoded backhaul links has been investigated [12]. To mitigate the performance loss as compared to the more common approach of an additional adaptive coding on the backhaul link, we propose two novel schemes: an error resilient decoder [81] and a soft-input/soft-output dequantizer (SISODQ) [34]. These approaches improve the bit error rate performance while introducing far less latency than an additional en/decoder for the backhaul. We show that combining the uncoded backhaul with the novel schemes can reduce the latency on wireless backhaul links and, at the same time, improve the performance compared to conventional approaches of using two separate codes for radio access and backhaul.

In addition to the development of specific approaches in the CTs, WP2 investigates in detail several common topics. In particular, an extensive analysis of possible functional splits between the iSCs and the RANaaS has been performed, investigating different options regarding backhaul data rate requirements, additional signalling and the centralized processing that the splits enabled. This is of fundamental importance to a flexible cloud-based architecture implementation as it will have a considerable impact on current networks. WP2 and WP3 partners jointly published a paper investigating requirements, constraints and technologies to enable such a flexible functional split [88]. In addition, the characteristics of different backhaul techniques for have been analysed in [13]. As iJOIN aims to exploit the benefits of cloud processing for the centralized RANaaS platform, the possibility of implementing PHY functionality on commodity hardware needs to be addressed. As the precoder is the main functional block for downlink transmission and the FEC decoder is the most complex block for uplink transmission, the impacts of implementing these basic processing blocks on a RANaaS platform are investigated. The implementation study of different FEC decoders on commodity hardware and corresponding performance evaluations are presented in [80].

The benefits that cloud-computing offers for 5G mobile networks and the implications on signal processing algorithms has been discussed by several WP2 partners in the special issue "Signal Processing for the 5G Revolution" of the IEEE Signal Processing Magazine [37].

3 Functional Split and Virtual eNB

3.1 iJOIN Architecture

In the RAN-as-a-service (RANaaS) architecture a RANaaS platform builds together with one or several connected iSCs a virtual evolved Node B (veNB). This veNB represents the evolution of the LTE eNB in the iJOIN architecture, but allows for a flexible distribution of radio access network (RAN) functionalities between the RANaaS platform and the iSCs. The actual shift of functionality is controlled by the iJOIN virtual eNB Controller (iveC) located in the RANaaS. Neighbouring iSCs can exchange information with each other directly by the logical J2 interface and are connected to the RANaaS platform by the logical J1 interface.

The dense deployment of iSCs leads to strong interference between neighbouring cells which needs to be considered within PHY processing approaches for downlink (DL) transmission and uplink (UL) reception. Using the distributed nature of the iJOIN architecture, joint transmission and detection schemes are required, exploiting either distributed processing among iSCs or centralized processing within the RANaaS platform or a mixture of both. The actual functional split will depend on numerous factors as outlined in the following subsections.

In the envisaged iJOIN architecture, the exchange of information bearing signals among the distributed RAN entities has to be realized over different kind of backhaul links. In order to design the PHY processing, the transmission properties of the different backhaul technologies, as discussed in Section 4 of deliverable D4.2 [84], have to be taken into account. In particular, supported data rate and the inherently introduced latency will play an important role for the feasibility of implementing PHY processing schemes.

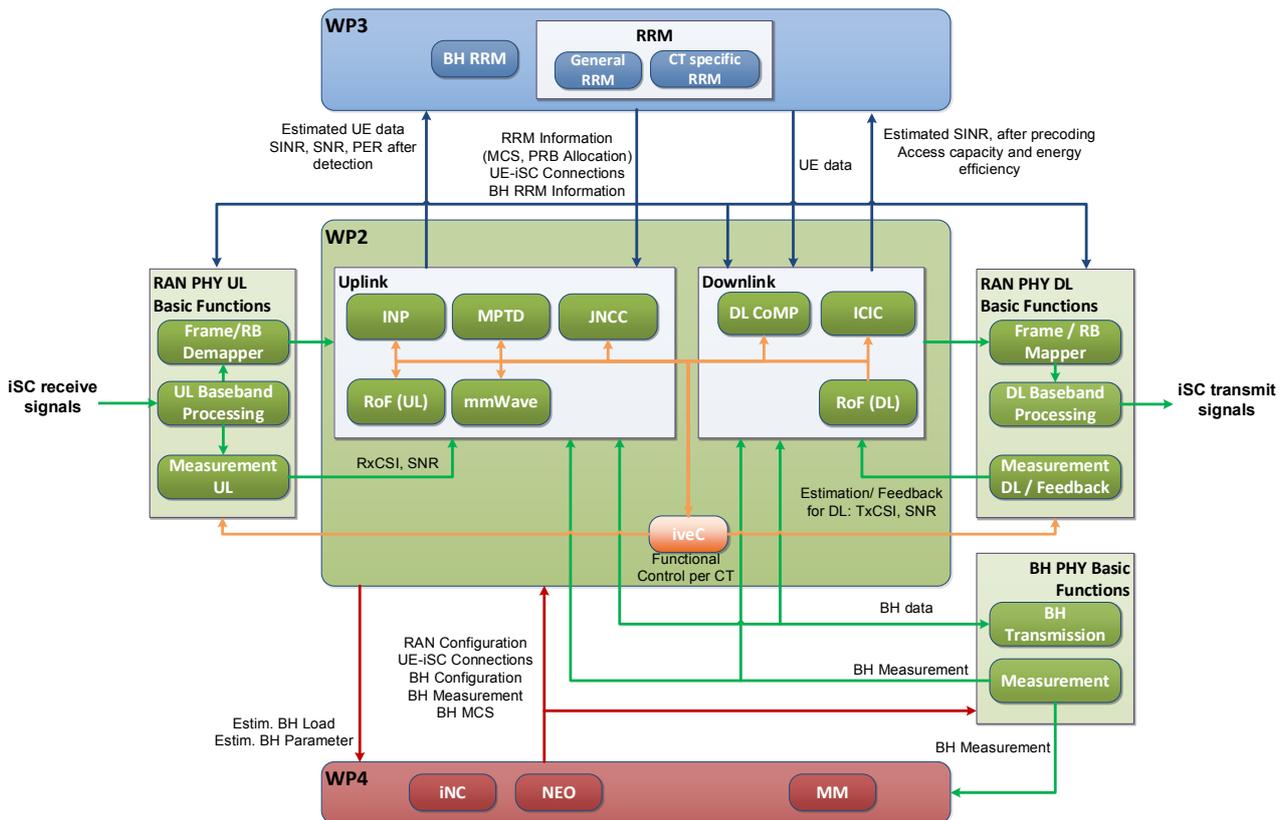


Figure 3-1: Functional Architecture of PHY processing

In deliverable D2.1 [36], the principle interfaces between WP2 CTs with basic PHY functions and with other iJOIN CTs were defined by specifying the requested input information and the provided output information per CT. For reference, the input and output messages are summarized in Appendix I. Furthermore, Figure 3-1 depicts the overall functional architecture of PHY processing approaches by means of interactions with PHY basic functions on the RAN and the BH, WP3, and WP4 as derived in D2.1 [36]. A grouping of WP2 CTs with respect to UL or DL transmission has been introduced to indicate the main objective of each CT.

On the UL, the different iSCs have received the UE signals over different channels. The antenna signals are processed in the “UL Baseband Processing” block yielding the I/Q receive signals in frequency domain. After an optional resource block demapping based on the UL radio resource management (RRM) information provided by WP3, the inphase/quadrature (I/Q) signals are forwarded to the CTs for further processing. Additionally, the block “Measurement UL” estimates the receive CSI and the receive signal to noise ratio (SNR). It provides this information to CTs as well. Note that the actual location of the measurement block and the execution of functions in the RAN architecture depend on the chosen functional split option, which is controlled by the iveC. The required exchange of signals over the backhaul links is organized by the iJOIN network controller (iNC) in cooperation with the iveC. This cooperative control also requires BH information provided by the “BH Measurement” block. Based on the available UL signals, the CTs perform an estimation of the transmitted information signals and provide the estimate to WP3. For the RRM either common LTE functions or CT specific functions are used.

The processing on the DL is executed accordingly. For time division duplex (TDD) transmission mode, the DL CSI can be estimated from the corresponding UL channel. In case of frequency division duplex (FDD), feedback based schemes are used. Based on this measurement and the RRM information from WP3, the user data is processed. After mapping to resource blocks, the DL baseband processed signal is transmitted over the iSCs. Again, either general LTE RRM functions or CT-specific RRM functions are used.

The kind of information to be exchanged within one CT between the RAN entities over the J1 and J2 interfaces will vary among the functional split options per CT. In addition, the localisation of channel measurements for the RAN links and the location where this information is needed may change with the functional split option. Thus, based on the principle discussion of functional split options applicable for PHY processing, we derive for each split the corresponding message size in Section 3.2.1 and elaborate the principle advantages and disadvantages per split. Based on these general relations, each CT will specify its messages in the corresponding part of Section 4 in order to measure the required backhaul rate.

3.2 Functional Split

One of the main benefits of the network architecture proposed by iJOIN is the ability to flexibly assign functionality to either the iSCs or the RANaaS within a virtual eNodeB. This functional split can be different for the various veNBs depending on both location and time according to, e.g. the traffic demand or the deployment scenario. As discussed in D2.1 [36], the actual split has some implications on the processing needs for iSCs and RANaaS, the reliability and latency requirements of the backhaul links (i.e. iSC-iSC and iSC-RANaaS links), and the backhaul load (associated to the payload and the signalling information) that needs to be transmitted on the backhaul connection between the iSCs and the RANaaS. These implications results in a number of decision factors that have to be taken into account to decide the actual split for each veNB as discussed in more detail in Section 3.2.1 of D3.2 [87]. In this section, we analyse the functional split options that are most relevant from WP2 perspective for both downlink and uplink and derive formulas for the analytical calculation of backhaul load at different functional splits.

3.2.1 Functional Split Options

Within iJOIN, four main functional split options A, B and C and D have been defined that split between local processing in the iSC and central processing in the RANaaS. These splits are depicted in Figure 3-2 and are further described in D5.2 [97]. From the PHY-layer perspective, split A, split B, and split C are most interesting, as for all splits above C the PHY layer is terminated in the iSC and thus it forms the transition to the MAC layer being considered in D3.2 [87]. Thus, all splits on MAC layer will be interpreted as split C from PHY layer perspective.

In WP2 we additionally investigate more detailed variants on how to split the PHY layer processing. Figure 3-3 shows these variants and the main functional blocks of the LTE DL/UL baseband signal processing chain. In the DL, the information bits coming from the higher layers are first subject to FEC encoding, rate matching and scrambling. For simplicity, these last two operations are not represented by separate blocks and thus it can be assumed that they are executed in the FEC encoding block. The encoded bits are then converted to complex symbols in the modulation block and subsequently precoded through the multiplication with a specific precoding vector or matrix. The complex signals after precoding are then mapped over the transmission resources (i.e. the OFDM subcarriers) of the different transmit antennas in the mapping block. After that, the signals are converted from frequency to time domain by means of an IFFT operation. The

IFFT output after oversampling and low pass filtering, not indicated for simplicity in the figure, is then converted to the analogue domain by the digital to analogue (D/A) conversion block.

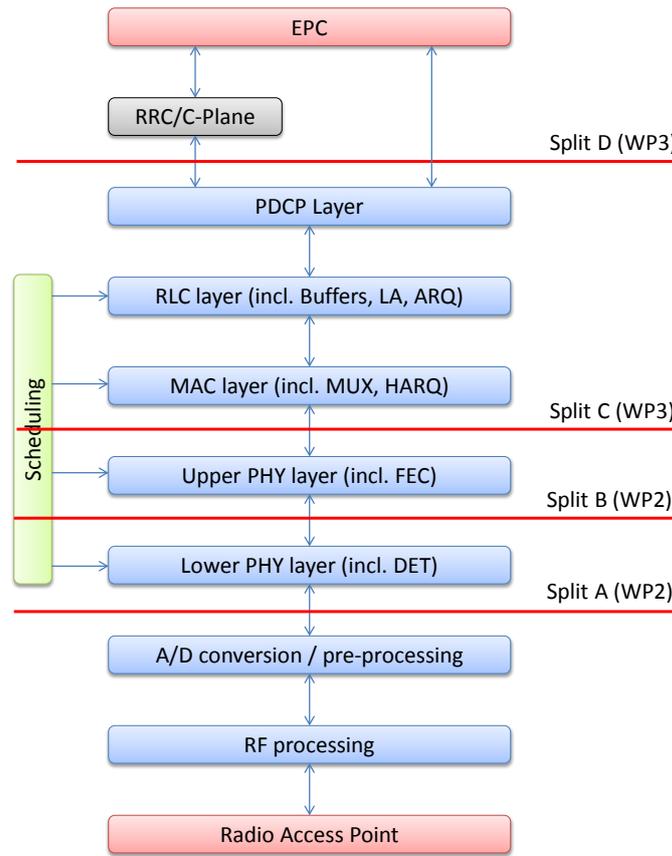


Figure 3-2: Project-wide functional split options

The UL signal processing operations are basically complementary to the DL ones. The only difference is the equalization, which is specifically done only in UL to compensate the frequency selectivity of the radio channel. In general the equalization requires also performing channel estimation that, still for simplicity, is considered to be executed in the equalization block.

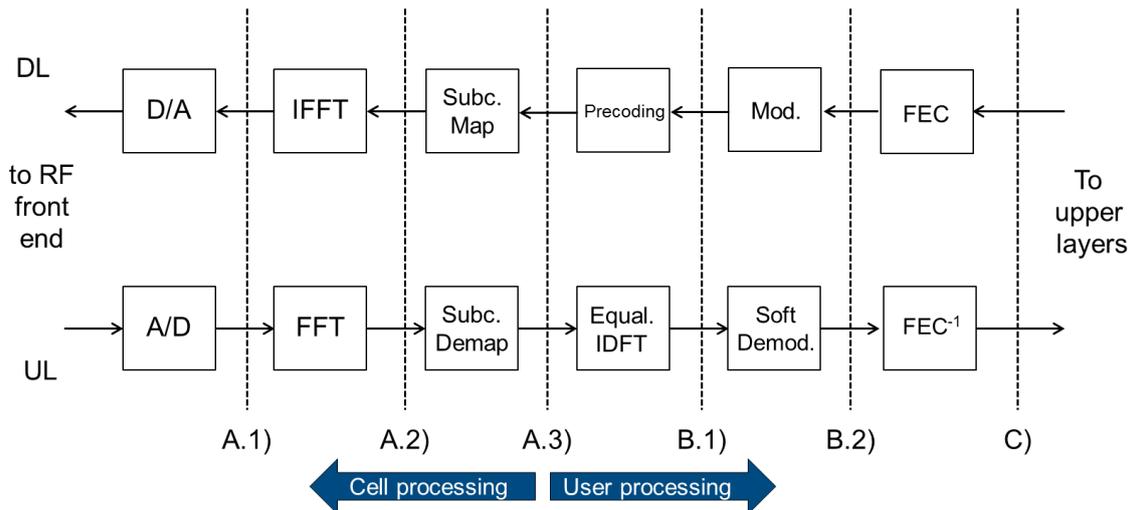


Figure 3-3: PHY-layer detailed functional split options

The choice of the actual functional split shown in Figure 3-3 depends upon the BH data rates, the timing constraints within the protocol stack and the possible centralization gains. For sake of simplicity the functional splits for one iSC-RANaaS link are analysed in the following. However, this analysis is also

important for more general systems with distributed processing among iSCs as applied by some CTs. The corresponding split options per CT are described in detail in Section 4.

The system parameters used in the subsequent calculations in terms of bandwidth requirements, together with the related meaning and exemplary values for LTE /LTE-A system, are summarized in Table 3-1.

Table 3-1: System parameters for BH load calculations

Parameter	Meaning	Exemplary Value	Unit
B	System bandwidth	{1.4, 3, 5, 10, 15, 20}	MHz
f_s	Sampling frequency at FFT/IFFT output	{1.92, 3.84, 7.68, 15.36, 23.04, 30.72}	MHz
OF	Oversampling factor	{1, 2}	-
N_{FFT}	FFT/IFFT size	{128, 256, 512, 1024, 1536, 2048}	Samples
N_{CP}	CP size	{9, 18, 36, 72, 108, 144}	Samples
N_{sc}	Number of available subcarriers for transmission (scales with system bandwidth B)	{72, 180, 300, 600, 900, 1200}	-
$N_{sc,u}^j$	Number of subcarriers allocated to the u -th user served by the j -th iSC in one OFDM symbol	variable	-
N_{RB}	Number of available PRBs (scales with system bandwidth B)	{6, 15, 25, 50, 75, 100}	-
$n_{RB,u}^j$	Number of PRBs allocated to the u -th user served by the j -th iSC	variable	-
N_{SYMB}^{SUB}	Number of OFDM symbols per subframe	{14}	-
T_{SUB}	Subframe duration	1	ms
T_s	OFDM symbol duration (w/o CP)	66.6	μ s
N_T^{iSC}	Number of transmit antennas at the iSC	{1, 2, 4, 8}	-
N_R^{iSC}	Number of receive antennas at the iSC	{1, 2, 4, 8}	-
Q_m	Number of bits per symbol carried by the modulation	{2, 4, 6}	-
N_L	Number of spatial layers transmitted to or received by a user	{1, 2}	-
N_Q	Number of quantization bits for each signal component	Actual value depends on the specific functional split	-
γ	Overhead introduced by the BH protection coding ($\gamma > 1$)	(For example $\gamma = 4/3$ for line coding and CPRI overhead in split option A.1)	-

3.2.1.1 Downlink

In the following the formulas for the analytical calculation of the message size for the payload D_p and the signalling information D_s transmitted on the backhaul for the different functional splits shown in Figure 3-3 are provided. The analysis considers the load between the RANaaS and one iSC with N_T^{iSC} antennas serving N_{UE} users.

In the equations derived in the following the size of the messages transmitted on the BH are calculated without including the overhead of the coding techniques that may be used for error detection or synchronization purposes. For example, for the split option A.1 based on common public radio interface

(CPRI) a typical 8b/10b line code [30] is used, so that the factor γ is equal to 10/8 that corresponds to a rate increase of 1.25 over the backhaul.

The value N_Q represents the number of quantization bits used for each signal (e.g. I/Q component, soft bits) transmitted on the BH, e.g. for each real and each imaginary part of a complex signal sample. For simplicity of notation, the same symbol N_Q is used in all the considered cases, even though the actual value depends on the considered functional split. Some indication about the number of quantization bits is provided in [33]. For example in case of option A.1, where orthogonal frequency division multiplexing (OFDM) or single carrier frequency division multiple access (SC-FDMA) signals in the time domain are transmitted on the BH, a rather high resolution (15 bits per dimension per sample) is necessary to keep quantization noise at a tolerable level [33]. Conversely, for split option A.2 where the OFDM/SC-FDMA signal in the frequency domain is transmitted on the BH, a lower resolution can be used compared to split option A.1, due to the smaller amplitude fluctuations of the signals in the frequency domain. In particular, the authors in [33] postulate a minimum of 5 to 6 bit depending on the actual SINR. A further margin of 1 or 2 bits may be necessary in presence of strong interfering signals and thus a total number of about 7 to 8 quantization bits for each signal component can be considered reasonable for the split options A.2 to B.1. Concerning the split option B.2, in the uplink case “soft bits” (e.g. log likelihood ratios (LLRs) after symbol-to-bit demapping or their expected values) are transmitted to the channel decoder on the BH. Conversely, in the downlink the channel coded bits after rate matching are transmitted to the subcarrier mapping function located in the iSC. In the downlink case a 1 bit resolution is clearly used, while for the uplink a LLR resolution over 4 to 5 bits leads to a negligible quantization loss [75]. Finally, concerning the split option C, where the MAC bits are transmitted on the BH, a 1 bit resolution is applied on both DL and UL.

Split option A.1: Time-Domain I/Q

With this split option the I/Q signal samples in the time domain, i.e. after IFFT operation and cyclic prefix (CP) insertion, are transmitted from the RANaaS to the iSCs. This split option is used in centralized RAN (C-RAN) architectures where the remote radio heads (RRHs) are connected to the central unit by means of CPRI interfaces [85].

The signal transmitted on the backhaul is the multiplexing of the user signals served by the same iSC and of the control channels that carry signalling information. It follows that specific in band signalling is not required by this split option because this information is already embedded in the signal. The backhaul load with this split option is independent on the actual radio interface load, thus preventing any form of statistical multiplexing of the BH load generated by different iSCs. This option is clearly the most promising in terms of centralization gain as all digital processing is performed in the RANaaS, enabling any kind of joint processing. This also means that iSCs potentially require very little digital hardware, making them very small and low-cost. On the other hand very high backhaul rates are required, putting a very high load on the BH network.

The message size for the payload can be calculated with the formula

$$D_P^{A.1} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot OF \cdot (N_{FFT} + N_{CP}) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-1)$$

where N_{FFT} and N_{CP} are the FFT and CP lengths expressed in samples, N_{SYMB}^{SUB} is the number of OFDM symbols in one subframe of duration T_{SUB} and OF is the oversampling factor. The multiplying factor 2 accounts for the I/Q components of the OFDM signal. The transmission frequency of the message is equal to $1/T_{SUB}$. The meaning of the symbols together with a set of typical values specific for the LTE system are provided in Table 3-1.

The factor OF is introduced to take into account the oversampling and low-pass filtering operations, which have to be executed before the digital to analogue (D/A) conversion. If these operations are performed in the RANaaS, then an oversampling factor larger than one must be considered, otherwise if the oversampling operation is executed at the iSCs no oversampling is required, i.e. $OF=1$.

The backhaul load associated to in-band signalling is equal to zero for this split, according to the explanation provided above.

$$D_S^{A.1} = 0 \quad (3-2)$$

The backhaul load of this split is very high and does not depend on the actual traffic load of the network. On the other hand, it is the most versatile, as all PHY-layer functionalities are executed in the RANaaS, enabling many types of joint processing.

Split option A.2: Frequency-Domain I/Q before IFFT

With this split option the I/Q signal samples in the frequency domain, i.e. before the IFFT operation, are transmitted from the RANaaS to the iSCs. The signal transmitted on the backhaul is the multiplexing of the frequency domain user signals served by the same iSC and of the control channels. It follows that specific in band signalling is not required by this split option because this information is already embedded in the signal. Considering that the iSCs do not know any resource allocation information, all the available subcarriers need to be quantized and transmitted over the backhaul, including those not allocated by the scheduler. The backhaul load with this split option is independent of the actual radio interface load, thus preventing any form of statistical multiplexing of the BH load generated by different iSCs.

The message size associated to the payload can be calculated with the formula

$$D_P^{A.2} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot N_{sc} \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-3)$$

where N_{sc} is the number of subcarriers available for transmission. The transmission frequency of the message equals to $1/T_{SUB}$.

The reduction of the backhaul load of this split option compared to the split option A.1, is mainly related to the fact that both CP and guard subcarriers at the edge of the channel are not transmitted over the backhaul (comparing equation (3-1) and (3-3)) where $N_{sc} < (N_{FFT} + N_{CP})$ and by the absence of the oversampling factor (i.e. $OF = 1$). Furthermore, the number of quantization bits N_Q can be reduced, e.g. $N_Q = 7$ was suggested in [76] for DL transmission.

The backhaul load associated to in-band signalling is equal to zero for this split, according to the explanation provided above.

$$D_S^{A.2} = 0 \quad (3-4)$$

Split option A.3: Frequency-Domain I/Q before resource mapping

With this split option the I/Q signal samples in the frequency domain before the resource element (RE) mapping block are transmitted from the RANaaS to the iSCs. Compared to option A.2, only the physical resource blocks (PRBs) allocated by the scheduler in a given subframe are quantized and transmitted over the BH. However, in order to perform the mapping on the transmission resources at the iSC, the resource allocation information for every subframe has to be provided to the iSC. The backhaul load with this split option is proportional to the actual radio interface load, thus enabling the statistical multiplexing of the BH load generated by different iSCs.

In case of orthogonal access of a group of N_{UE} users served by a given iSC, the message size is obtained as sum of the message sizes associated to the different users

$$D_P^{A.3} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-5)$$

where $N_{sc,u}^j$ is the number of subcarriers allocated to the u -th user served the j -th iSC in one OFDM symbol.

As for the previous split options, the transmission frequency of the message is equal to $1/T_{SUB}$.

In the non-orthogonal access case (e.g. MU-MIMO), where the users are allocated over sets of overlapping subcarriers, precoding and user signal superposition is performed in the RANaaS and thus the load depends only on the total number of allocated subcarriers, independently that a part of them are shared among two or more users. In this case the equation (3-5) modifies as follows

$$D_P^{A.3} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot \left(\bigcup_{u=1}^{N_{UE}} S_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-6)$$

where $S_{sc,u}^j$ is the set of subcarriers allocated to u -th user served the j -th iSC containing $N_{sc,u}^j = |S_{sc,u}^j|$ subcarriers. Thus, the union set

$$S_{sc}^j = \bigcup_{u=1}^{N_{UE}} S_{sc,u}^j \quad (3-7)$$

contains all subcarriers allocated to the users served by iSC j .

In case of LTE/LTE-A, the number of allocated subcarriers is equal to the number of allocated PRBs ($n_{RB,u}^j$) multiplied by 12 (i.e. the number of subcarriers per PRB)

$$N_{sc,u}^j = n_{RB,u}^j \cdot 12 \quad (3-8)$$

It should be noted that the variable $N_{sc,u}^j$ is always lower or equal than the number of available subcarriers for the considered system bandwidth B (i.e. $N_{sc,u}^j \leq N_{sc}$). Besides, in the equations (3-5) and (3-6) the load associated to the transmission of the common reference signals (RS) in the PRBs not allocated for data transmission is not considered. However, in the exemplary case of a LTE MIMO 2x2 system, this factor accounts for only about the 5% of the backhaul load with respect to the case of full resource allocation.

Concerning the backhaul load associated to signalling, the information to be transmitted to the iSCs is the resource allocation decided by the MAC scheduler in terms of position of the allocated PRBs in each subframe. Several ways to encode this information could be envisaged. A very simple but not particularly efficient protocol consists in transmitting to each iSC a bitmap composed by N_{RB} bits, where the PRBs allocated by the scheduler are associated to a bit value '1' and the other ones to '0'. The PRBs marked by '1' will have full size as they carry the physical downlink shared channel (PDSCH) channel, whereas the RBs marked with '0' will be shortened as they only carry the REs used for transmission of the RSs and of the control channels located in the first part of the subframe, i.e. the physical downlink control channel (PDCCH), physical control format indicator channel (PCFICH) and physical HARQ indicator channel (PHICH) channel in the specific LTE case. The message size necessary to transmit this bitmap corresponds to

$$D_S^{A.3} = N_{RB} \quad [\text{bit/iSC}] \quad (3-9)$$

with the transmission frequency of this message equal to $1/T_{SUB}$, where T_{SUB} is the subframe period defined in Table 3-1.

Split option B.1: Frequency-Domain I/Q before precoding

In this option, precoding is performed at the iSC instead of the RANaaS. For this purpose, channel state information has to be available at the iSC. Alternatively, if the precoding matrix is calculated at the RANaaS, the iSC does not necessarily need CSI as well. However, the precoding matrix then has to be transmitted to the iSC separately. If users share a certain group of subcarriers (e.g. in the MU-MIMO case) their signals have to be transmitted separately as the superposition is performed in the iSC. As compared to option A.3, the load is now dependent on the number of users served by an iSC and not only on the number of scheduled subcarriers. Also, the number of layers is now relevant instead of the number of transmit antennas, giving the message size as

$$D_P^{B.1} = 2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-10)$$

with the same transmission frequency $1/T_{SUB}$. If the precoding matrix is calculated in the RANaaS and assuming that the same number of quantization bits is used for it, the additional signalling message size can be calculated as

$$D_S^{B.1} = 2 \cdot N_L \cdot N_T^{iSC} \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-11)$$

Split option B.2: Frequency Domain after FEC encoding

With this option the RANaaS-iSC interface is located at the output of the FEC encoder so that the channel coded bits after rate matching (i.e. puncturing or repetition) are transmitted on the backhaul link. The iSC executes the modulation, spatial precoding, RE mapping, IFFT, CP insertion, up-sampling, filtering, D/A conversion, and radio frequency (RF) processing. The message size associated to the payload of one user u can be calculated with the following formula

$$D_{P,u}^{B.2} = N_{sc,u}^j \cdot Q_m \cdot N_L \cdot N_{SYMB}^{SUB} \quad [\text{bit/user}] \quad (3-12)$$

where N_L is the number of spatial layers and Q_m is the number of bits per modulation symbol of the u -th user (assumed equal for all the users for simplicity of notation). The transmission frequency of the message is equal to $1/T_{SUB}$.

The message size associated to a group of N_{UE} users served by a given iSC is therefore obtained as the sum of the load associated to each user

$$D_P^{B.2} = \sum_{u=1}^{N_{UE}} N_{sc,u}^j \cdot Q_m \cdot N_L \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-13)$$

With this functional split some specific signalling information is required at the iSC in order to execute the modulation, precoding, RE mapping and power adjustment operations. In the following it is provided an analytical estimate of this signalling load on the BH where some characteristics of the radio interface (e.g. modulation schemes, precoding codebook size, downlink power scaling range, etc.) are defined taking exemplary the Release 10 Specification of the LTE system. However, the same principle and formulas can be used for other radio interfaces by changing the relevant parameters. In the calculation it is considered the case of an LTE cell with bandwidth of $B=20$ MHz equipped with $N_T^{iSC} = 2$ transmit antennas.

The first information that must be transmitted for each scheduled user to the iSCs is the used modulation scheme. In case of LTE, three modulation schemes are currently foreseen in downlink for data channels (i.e. QPSK, 16-QAM and 64-QAM) [15] and therefore two bits per user ($N_{MOD} = 2$ bits) are sufficient to convey this information. Because the assigned modulation scheme is applied for all the PRBs allocated to a given user in one subframe, this signalling information does not depend on the actual number of allocated PRBs. The message size related to the modulation operation for each scheduled user is then calculated as

$$D_{S,MOD}^{B.2} = N_{MOD} \quad [\text{bit/user}] \quad (3-14)$$

where the transmission frequency of this message is equal to $1/T_{SUB}$.

The second information to be transmitted for each scheduled user is the actual resource allocation decided by the MAC scheduler in terms of number and position of the allocated PRBs. This information is necessary in order to perform the resource mapping operation for each user at the iSC side. In case of LTE, the downlink and uplink resource allocation of each user is conveyed by the PDCCH channel. In order to minimize the signalling overhead and at the same time provide sufficient flexibility for the scheduler, the LTE standard foresees three types of resource allocations denoted as Type 0, 1 and 2 [16].

The Type 0 and 1 are used for the downlink only, while Type 2 is used for both downlink and uplink allocations. The Type 0 resource allocation addresses resource block groups (RBGs), where the group size P (with $P=2, 3$ or 4) depends on the system bandwidth B . The number of bits N_{RA} required to encode the allocation information is equal to $\lceil N_{RB} / P \rceil$, where N_{RB} is the total number of PRBs available in the channel bandwidth B . The Type 1 resource allocation addresses individual PRBs in a subset of RBGs. The number of subsets (2, 3, or 4) depends on the system bandwidth B . As for Type 0 the number of bits N_{RA} required to encode the allocation information is equal to $\lceil N_{RB} / P \rceil$.

Finally the Type 2 allocation addresses any possible arrangement of contiguous PRB allocations. The type 2 allocation is signalled to each user in terms of starting position and number of allocated PRBs. The number of bits N_{RA} required to encode the allocation information is equal to $\lceil \log_2(N_{RB} \cdot (N_{RB} + 1)) \rceil$. Assuming for example the Type 2 allocation, the message size related to the resource allocation operation for each scheduled user is then calculated as

$$D_{S,RA}^{B,2} = \lceil \log_2(N_{RB} \cdot (N_{RB} + 1)) \rceil \quad [\text{bit/user}] \quad (3-15)$$

where the transmission frequency of this message is still equal to $1/T_{SUB}$.

The third information to be transmitted is related to the execution of the multi-antenna signal processing operations at the iSCs. The LTE standard foresees several transmission modes (TM) which are based on both open loop and closed loop MIMO techniques. In particular in the Release 10 of the standard are foreseen 9 possible transmission modes ranging from single antenna transmission (TM1) up to eight layer transmission (TM9) [16]. It follows that for each scheduled user 4 bits ($N_{TM}=4$ bits) are sufficient to convey the information of the used transmission mode to the iSC. In addition, considering the case of transmission modes that use closed loop precoding, it is necessary to convey also the information of the used precoding element. Assuming wideband (i.e. equal for all the allocated PRBs) codebook based precoding, formed by N_{CB} elements (i.e. N_{CB} vectors or matrixes depending on the number of transmitted layers), the number of bits to be transmitted for each scheduled user in one subframe for the precoding operations is equal to $\log_2(N_{CB})$. At the end, the message size on the backhaul related to the multi-antenna operations for one user can be calculated as:

$$D_{S,MIMO}^{B,2} = N_{TM} + \log_2(N_{CB}) \quad [\text{bit/user}] \quad (3-16)$$

which is transmitted every subframe (i.e. with frequency $1/T_{SUB}$).

Notice that in case of frequency selective precoding the related overhead $\log_2(N_{CB})$ will increase by a factor equal to the number of allocated PRBs (or to the number of sub-bands in which the overall transmission bandwidth may be divided).

The fourth information to be transmitted is related to the execution of the power adjustment operation at the iSC for each user transmission. For LTE the energy per resource element (EPRE) of the data subcarriers allocated to a given user is defined with respect to the energy of the common reference signals (RS-EPRE). The downlink cell-specific RS-EPRE is constant across the downlink system bandwidth and constant across all subframes, until different cell-specific RS power information is received through higher layer signalling. For LTE two parameters ρ_A and ρ_B are defined [16], which represent the data to RS EPRE ratio in the OFDM symbols that do not contain RS and in the OFDM symbols that contain RS, respectively

$$\rho_A = \frac{E_A}{E_{RS}} \quad \rho_B = \frac{E_B}{E_{RS}} \quad (3-17)$$

where E_A and E_B is the EPRE of the data subcarrier and E_{RS} is the EPRE of the RSs, as schematically shown in Figure 3-4.

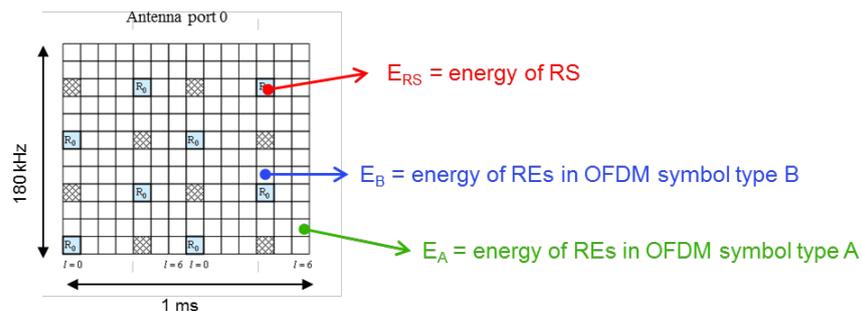


Figure 3-4: Power adjustment operation at the iSC in the LTE case

In case of LTE, it is possible to derive that $N_{PA} = 5$ bits are sufficient to encode the possible values of the parameters ρ_A and ρ_B [16]. More in general assuming that the power adjustment control information can be represented with N_{PA} bits, the message size related to the power adjustment operations for each scheduled user can be calculated as

$$D_{S,PA}^{B.2} = N_{PA} \quad [\text{bit/user}] \quad (3-18)$$

where it is assumed that the EPRE is kept the same for all the RBs allocated to a given user in a certain subframe (as it is in the current design of the LTE system). Clearly, if some form of power loading (e.g. water filling) is possible, the related overhead increases by a factor equal to the number of allocated PRBs.

The total message size related to signalling can then be calculated by adding the four contributions given above for modulation, resource allocation, MIMO processing and power adjustment

$$D_S^{B.2} = D_{S,MOD}^{B.2} + D_{S,RA}^{B.2} + D_{S,MIMO}^{B.2} + D_{S,PA}^{B.2} \quad [\text{bit/user}] \quad (3-19)$$

and by substituting the related equations we finally obtain

$$D_S^{B.2} = (N_{MOD} + \lceil \log_2(N_{RB}^{DL} \cdot (N_{RB}^{DL} + 1)) \rceil + N_{TM} + \log_2(N_{CB}) + N_{PA}) \quad [\text{bit/user}] \quad (3-20)$$

Split option C: Frequency Domain before FEC encoding

With this split option the data at the MAC layer output are transmitted on the backhaul. This also means that if PHY-layer functionalities are to be performed in a joint fashion, this has to be done in the iSCs as it cannot be done in the RANaaS. The resulting BH load for the payload largely depends on the used modulation and coding scheme (MCS). By defining as $TBS_{1,u}$ and $TBS_{2,u}$ the transport block size (TBS) for the first and second codeword assigned to the u -th user in a subframe, the corresponding message size associated to the payload can be calculated with the following formula

$$D_{P,u}^C = (TBS_{1,u} + TBS_{2,u}) \quad [\text{bit/user}] \quad (3-21)$$

where the transmission frequency of this message depends on the scheduling decisions and can be equal up to $1/T_{SUB}$, when the user is scheduled in every subframe.

The message size associated to a group of N_{UE} active users (i.e. simultaneously scheduled in the same subframe) served by a given iSC is therefore obtained as sum of the message size associated to each user

$$D_P^C = \sum_{u=1}^{N_{UE}} (TBS_{1,u} + TBS_{2,u}) \quad [\text{bit/iSC}] \quad (3-22)$$

In case of single codeword transmission, the value of $TBS_{2,u}$ is equal to zero, whereas in case of double codeword transmission both values will be larger than zero. Like for split option A.3, the equation (3-22) provides the message size in a given subframe.

In [76] an alternative method for the calculation of the average backhaul load for this split option is described. This method is based on the average spectral efficiency S of the radio interface, expressed in bit/s/Hz. Assuming that S is known, the average backhaul load generated by one iSC is simply the product of the system bandwidth B in Hz and the spectrum efficiency S . Clearly this way of calculation provides the average values, while the peak load values can be higher. However, also considering the peak values, the split option C remains the less critical in terms of required backhaul capacity.

Concerning the signalling load, also the FEC coding operation is executed at the iSC and therefore for each user it must be provided the assigned MCS in the subframe. As up to 32 MCSs are defined in the standard for LTE [16], five bits are sufficient to encode this information. By exploiting equation (3-19), defined for split option B.2, and replacing N_{MOD} with the new field N_{MCS} (as it includes also the modulation information in addition to the TBS), the message size can be now calculated with the following formula

$$D_{S,MCS}^C = N_{MCS} \quad [\text{bit/user}] \quad (3-23)$$

and thus the total message size per each scheduled user becomes

$$D_S^C = \left(D_{S,MCS}^C + D_{S,RA}^C + D_{S,MIMO}^C + D_{S,PA}^C \right) \quad [\text{bit/user}] \quad (3-24)$$

where the specific message elements $D_{S,RA}^C$, $D_{S,MIMO}^C$ and $D_{S,PA}^C$ for the split option C are identical to those calculated for the split option B.2.

3.2.1.2 Uplink

In the following the equations for estimating the uplink message size, related to data payload and signalling, on the backhaul connection between the iSCs and the RANaaS are derived. Due to the multi-carrier nature of the SC-FDMA signal, the equations derived for the downlink can be easily extended to the uplink by changing the related parameters.

Split option A.1: Time-Domain I/Q

With this split option the I/Q SC-FDMA signal samples in the time domain, i.e. before CP removal and FFT operation, are transmitted from the iSCs to the RANaaS. This split option is typically adopted in centralized architectures based on remote radio heads (RRHs), where all the physical layer operations are centralized in a baseband pool.

The message size can therefore be derived from the downlink equation (3-1) by introducing the number of receiving antennas at the iSC

$$D_P^{A.1} = 2 \cdot N_R^{iSC} \cdot N_Q \cdot OF \cdot (N_{FFT} + N_{CP}) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-25)$$

where the transmission frequency of the message is equal to $1/T_{SUB}$.

As all signalling information is included in the quantized time domain SC-FDMA received signal, the message for in-band signalling is of size zero

$$D_S^{A.1} = 0 \quad (3-26)$$

An example of BH load calculation for this functional split is provided in [76]. In this case the value assumed for the number of quantization bits of each signal component is $N_Q = 15$ bits, while the overhead γ is assumed equal to $\gamma = (16/15) \cdot (10/8) = 4/3$, where the factor 16/15 accounts for the CPRI overhead and the factor 10/8 accounts for the 8B/10B line coding. In case of an iSC equipped with $N_R^{iSC} = 2$ receive antennas and bandwidth $B = 20$ MHz, this leads to a backhaul rate of 2.46 Gbit/s. In [76] it is also mentioned the option of using compression techniques that are capable of reducing the load by a factor of up 3-5 times, leading to solutions denoted in literature as Compressed CPRI [77], [78]. In any case, the backhaul load of this split is very high and does not depend on the actual traffic load of the network. On the other hand, it is the most versatile, as all PHY-layer functionalities are executed in the RANaaS, enabling many types of joint processing.

Split option A.2: Frequency-Domain I/Q after FFT

With this split option the I/Q SC-FDMA signal samples in the frequency domain, i.e. after the FFT operation, are transmitted from the iSCs to the RANaaS. The signal transmitted on the backhaul contains the user signals served by the particular iSC and the corresponding control channels.

The message size can be derived from the downlink equation (3-3) by introducing the number of receiving antennas at the iSC

$$D_P^{A.2} = 2 \cdot N_R^{iSC} \cdot N_Q \cdot N_{sc} \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-27)$$

where the transmission frequency of the message is equal to $1/T_{SUB}$.

The backhaul load associated to in-band signalling is equal to zero for this split

$$D_S^{A.2} = 0 \quad (3-28)$$

An example of BH load calculation for this functional split is provided in [76]. In this case the value assumed for the number of quantization bits of each signal component is $N_Q = 9$ or $N_Q = 10$ bits. Considering a LTE configuration with $B = 20$ MHz bandwidth and 2 receive antennas, the analysis in [76] leads to a BH rate of 0.66 Gbit/s in the case of $N_Q = 9$ sample resolution when including the overhead associated to the physical random access channel (PRACH).

Split option A.3: Frequency-Domain I/Q after resource demapping

With this split option the I/Q signal samples in the frequency domain, i.e. after the RE demapping block, are transmitted from the iSCs to the RANaaS. The backhaul load is therefore proportional to the actual radio interface load (unlike split options A.1 and A.2).

The message size associated to a group of N_{UE} users served by a given iSC can therefore be calculated with equation (3-5) defined for the downlink split option A.3 in case of orthogonal access

$$D_P^{A.3} = 2 \cdot N_R^{iSC} \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-29)$$

and by (3-6) for non-orthogonal access (i.e. MU-MIMO case):

$$D_P^{A.3} = 2 \cdot N_R^{iSC} \cdot N_Q \cdot \left(\bigcup_{u=1}^{N_{UE}} S_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-30)$$

As for the downlink, $S_{sc,u}^j$ is the set of subcarriers with cardinality $N_{sc,u}^j$ allocated to u -th user served the j -th iSC.

The frequency of the message is also in this case the reciprocal of the subframe period, namely $1/T_{SUB}$. The equations (3-29) and (3-30) provide an estimate of the backhaul load considering only the subcarriers allocated to the physical uplink shared channel (PUSCH). In general this represents the larger part of the uplink load. In the uplink of the LTE system there are also other control channels and signals that contribute to the backhaul load, in particular the physical uplink control channel (PUCCH), the PRACH and the sounding reference signals (SRS). The frequency location and periodicity of these channels is decided by the RRC radio resource control (RRC) function in the RANaaS and must be signalled to the iSCs so that these channels can be demapped and quantized. In [33] it is shown that the load associated to these channels can be still expressed as the product of the number of occupied subcarriers by the quantization resolution, divided by the specific time periodicity of the considered channel or signal.

Concerning signalling, the iSC must be provided with the resource allocation information every subframe in order to perform the demapping of the uplink physical channels. An estimate of the related load can be derived with the equation (3-9) for the PUSCH allocation part.

Split option B.1: I/Q after equalization and IDFT

In this option, the signals have already been equalized, MIMO processing has been performed and the signals have been transformed back into the time domain. As compared to split A.3, the message size now depends on the number of transmitted layers and can be calculated as

$$D_P^{B.1} = 2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-31)$$

with the transmission frequency $1/T_{SUB}$. This option has the advantage that no channel information or reference signals have to be forwarded to the RANaaS for equalization. However, since equalization and MIMO processing can be a complex task, it requires a lot more digital processing power in the iSC than split options A.3.

Split option B.2: quantized soft demodulator output

With this option the quantized output of the soft demodulator is transmitted on the BH link between the iSCs and the RANaaS. The message size, associated to a group of N_{UE} users served by a given iSC, is calculated

as in the equation (3-13) derived for the downlink case, with the only modification of the factor N_Q to take into account the quantization

$$D_P^{B.2} = \sum_{u=1}^{N_{UE}} N_{sc,u}^j \cdot Q_m \cdot N_Q \cdot N_L \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-32)$$

where the frequency of this message is equal to $1/T_{SUB}$.

The number of quantization bits N_Q is equal to 1 in case of hard bit transmission. In case of soft bit transmission a number of quantization bits N_Q equal to 4-5 bits is sufficient to obtain a negligible quantization error, as shown by the analyses provided in [75] or in [79].

Concerning the signalling information, the downlink equation (3-20) for the split option B.2 can be used also for the uplink.

Split option C: FEC decoder output

With this split option the data at the FEC decoder output are transmitted on the backhaul. As this means that all PHY-layer functionality is located in the iSCs, joint processing can only be performed among the iSCs and not in the RANaaS. The resulting BH load for the payload largely depends on the used MCS. The equation (3-22), defined for the downlink split option C, can then be applied also for the LTE Release 10 uplink with double codeword transmission

$$D_P^C = \sum_{u=1}^{N_{UE}} (TBS_{1,u} + TBS_{2,u}) \quad [\text{bit/iSC}] \quad (3-33)$$

The equation (3-33) provides the size of the message transmitted in one subframe and thus it follows that the message frequency is $1/T_{SUB}$.

Concerning the signalling information the equation (3-24) can be applied also for the uplink.

3.2.1.3 Numerical Example and Summary

For an exemplary evaluation of the previously derived payload calculations we consider the downlink of a LTE cell equipped with MIMO 2x2 and system bandwidth $B = 20$ MHz. The values of the parameters used in the calculations are summarized in Table 3-2. The different functional splits are compared in terms of message size related to the payload. As the message frequency is the same for all the functional splits (i.e. $1/T_{SUB} = 1$ kHz), the message size is also indicative for the BH throughput without including the BH coding overhead γ .

Table 3-2: System parameters used in the numerical example

Parameter	Exemplary Value	Unit	Parameter	Exemplary Value	Unit	Parameter	Exemplary Value	Unit
B	20	MHz	$N_{sc,1}^j$	600	-	N_T^{iSC}	2	-
f_s	30.72	MHz	N_{RB}	100		N_R^{iSC}	2	-
OF	1	-	$n_{RB,1}^j$	50	-	Q_m	6	-
N_{FFT}	2048	Samples	N_{SYMB}^{SUB}	14	-	N_L	2	-
N_{CP}	144	Samples	T_{SUB}	1	ms	N_Q	15 (A.1) 7 (A.2, A.3, B1) 1 (B.2, C)	-
N_{sc}	1200	-	T_s	66.6	μ s			

The message size for the split option A.1 is calculated with equation (3-1)

$$D_p^{A.1} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot OF \cdot (N_{FFT} + N_{CP}) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-34)$$

leading to

$$D_p^{A.1} = 2 \cdot 2 \cdot 15 \cdot 1 \cdot (2048 + 144) \cdot 14 = 1841280 \quad [\text{bit/iSC}] \quad (3-35)$$

As pointed out before, the message size provided by equation (3-35) does not include the BH protection overhead. However, by including this factor (i.e. $\gamma = 4/3$) the corresponding transmission size is equivalent to a throughput of 2.46 Gbit/s as derived in [76].

Similarly, for the split option A.2 the message size is calculated with the equation (3-3)

$$D_p^{A.2} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot N_{sc} \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-36)$$

leading to

$$D_p^{A.2} = 2 \cdot 2 \cdot 7 \cdot 1200 \cdot 14 = 470400 \quad [\text{bit/iSC}] \quad (3-37)$$

For the split option A.3 the message size is instead calculated with the equation (3-5)

$$D_p^{A.3} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-38)$$

and, assuming for example that only half of the transmission resources are allocated (i.e. $\sum_u N_{sc,u}^j = 600$), it leads to

$$D_p^{A.3} = 2 \cdot 2 \cdot 7 \cdot 600 \cdot 14 = 235200 \quad [\text{bit/iSC}] \quad (3-39)$$

For the split option B.1 and assuming that the precoding matrix is calculated at the iSC and has not to be transmitted on the BH the message size is calculated using equation (3-10)

$$D_p^{B.1} = 2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (3-40)$$

which yields

$$D_p^{B.1} = 2 \cdot 2 \cdot 7 \cdot 600 \cdot 14 = 235200 \quad [\text{bit/iSC}] \quad (3-41)$$

being identical to $D_p^{A.3}$ because 2 layers and 2 transmit antennas are assumed. If the precoding matrix is calculated in the RANaaS, it has to be transmitted on the BH leading to an increase of the BH message size.

The message size for the split option B.2 is calculated with the equation (3-13)

$$D_p^{B.2} = \sum_{u=1}^{N_{UE}} N_{sc,u}^j \cdot Q_m \cdot N_L \cdot N_{SYMB}^{SUB} \quad (3-42)$$

and still assuming a half occupancy of the transmission resources it provides

$$D_p^{B.2} = 600 \cdot 6 \cdot 2 \cdot 14 = 100800 \quad [\text{bit/iSC}] \quad (3-43)$$

Finally, for the split option C the message size is calculated using the equation (3-22)

$$D_p^C = \sum_{u=1}^{N_{UE}} (TBS_{1,u} + TBS_{2,u}) \quad [\text{bit/iSC}] \quad (3-44)$$

The TBS for the first and second codeword is derived from [16] considering the case of a single user $N_{UE} = 1$, which is scheduled over half of the available PRBs (i.e. $n_{RB,1}^j = 50$) and performs spatial multiplexing with $N_L = 2$ double layer transmission. In particular assuming that the 64-QAM modulation and the highest coding rate are used, which correspond to MCS 28, the TBS sizes are both equal to $TBS_{1,1} = TBS_{2,1} = 36696$ bits. The message size is thus equal to

$$D_p^C = 36696 + 36696 = 73392 \quad [\text{bit/iSC}] \quad (3-45)$$

The example calculation shows, as expected, that the message size and thus the BH rate requirements decrease as the split section is progressively moved towards the higher layers, i.e. from split option A.1 to C. Furthermore, only for the split options A.3 to C the message size is proportional to the fraction of used resources on the radio interface, opening the possibility to further exploit this characteristic for reducing the BH requirements.

A factor not considered so far is that usually additional overhead is introduced on the BH due to protocols and error protection. An example of this is provided in [76]. The overhead γ is assumed to be equal to $\gamma = (16/15) \cdot (10/8) = 4/3$, where the factor 16/15 accounts for the CPRI protocol overhead and the factor 10/8 accounts for the 8B/10B line coding. This increases the backhaul rate of split A.1 to 2.46 Gbit/s. In [76] it is also mentioned the option of using compression techniques that are capable of reducing the load by a factor of up 3-5 times, leading to solutions denoted in literature as compressed CPRI [77], [78].

The findings of this section are compactly summarized in Table 3-3.

Table 3-3: Summary of functional split options

Split option	Lowest functionality centralized	Impact on LTE	BH Latency requirements	Exemplary BH Bandwidth	Centralization scheme	Advantage	Disadvantage
A.1	FFT/IFFT	small	5 μ s (CPRI)	1,841 Mbps	Fully centralized PHY processing	- Very simple iSCs - All centralized processing options	- Very high BH bandwidth - Strict latency requirements
A.2	Subcarrier mapping	small	CSI constrained	470 Mbps	Fully centralized PHY processing	- Reduced BH load - All centralized processing options	- High BH bandwidth
A.3	Precoding/Equalization	- Sync. signalling between iSC/RANaaS req. - Subc. mapping signalling req.	CSI constrained	235 Mbps	Fully centralized PHY processing of payload	- Reduced BH load - All centralized processing options	- Additional signalling required
B.1	Modulation/demodulation	Precoding signalling req.	4 ms (HARQ)	235 Mbps	Centralized modulation/detection	- Reduced latency requirements - Joint detection possible	- No centralized precoding - Turbo Equalization infeasible - Additional signalling required
B.2	FEC	small	4 ms (HARQ)	101 Mbps	Centralized coding	- Joint decoding/ NC possible	- No centralized precoding - Turbo Equalization unfeasible
C	MAC	small	4 ms (HARQ)	73 Mbps	Centralized MAC	- Greatly reduced BH load	- No joint PHY processing

3.2.2 Functional Split Investigation for Main Blocks for DL and UL

When performing the functional split on the physical layer, some lower layer PHY functions are executed in the iSCs, whereas some higher layer functions are centrally executed in the RANaaS. Although centralized baseband pools have already been investigated for the RAN to allow for efficient resource usage and advanced multi-cell algorithms, these technologies still require dedicated hardware and do not offer the same characteristics as cloud-computing platforms, i.e. on-demand provisioning, virtualization, resource pooling, elasticity, service metering, and multi-tenancy [37]. However, these properties of cloud computing are key enablers for the iJOIN architecture supporting an ultra-dense deployment of iSCs leading to severe multi-cell interference in combination with a significant increase of the number of access nodes and huge fluctuations of the rate requirements over time. The main difficulty of implementing RAN functionality in a cloud-platform lies in the tight constraints caused by the 3GPP LTE protocol stack. This implies that individual tasks need to finish within a predefined time window as discussed in D3.2 [87]. In order to evaluate the principal possibility of such a cloud based implementation, we investigate in this section the main challenges

for DL and UL transmissions by elaborating two fundamental components of both processing chains: precoding in DL and FEC decoding in UL.

Precoding is a generalization of beamforming to support multi-stream (or multi-layer) transmission in multi-antenna wireless communications and, thus, it is the key building block in the DL processing chain for coordinated multi-point (CoMP) systems. After FEC encoding, rate matching and modulation, the signal streams of one or several users are processed to construct the joint transmit signal of the cooperating iSCs. By mapping the data stream(s) to physical antennas with appropriate weighting two kinds of gains are realized: beamforming/diversity gain to overcome the interference and noise contamination on desired signals and multiplexing gain to maximize the throughput. Centralized precoding can significantly facilitate to achieve these gains by multi-site coordination mechanisms and releases the demanding computational resource requirement to iSCs. However, centralized processing introduces also extra delay and backhaul constraints as discussed subsequently in Section 3.2.2.1.

On the UL, the key building block is the FEC decoder being responsible for error correction at the receiver side. It is the last functional component of the PHY layer and consumes the main computational effort of the UL processing chain. Due to the heavy demand of computational resources and the strict timing requirements, FEC decoders are usually implemented on specialized hardware and the implementation on general purpose hardware (GP-HW) based cloud platform is a challenging approach as discussed in Section 3.2.2.2.

3.2.2.1 Precoder @ Cloud

The strong interference between neighbouring iSCs, caused by the ultra-dense deployment of iSCs in the iJOIN architecture, needs to be taken into account and managed if possible for DL transmissions. As reported in D2.1 [36], several CoMP approaches are known which utilize the inherent interference for increasing the system throughput. In particular, for the highly promising joint transmission scheme in downlink, the iSCs cooperatively transmit pre-processed signals to the UEs in order to improve SINR at the receiver side. Commonly considered precoding methods include linear processing such as maximum ratio transmission (MRT), zero forcing (ZF) or minimum mean square error (MMSE) and nonlinear methods such as dirty paper coding (DPC) [90], [92], Tomlinson Harashima precoding (THP) [93], etc. Among them, global ZF precoding is widely used due to its good trade-off between computational complexity and performance [90].

Generally, the precoding process includes the calculation of the precoder matrix based on the available CSI and the calculation of the transmit signals (i.e. application of the precoder). To this end, the precoding matrix is multiplied with the modulated I/Q signals and the output signals are mapped to the corresponding physical antennas. As can be realised, precoding is generally a highly demanding process; huge processing capability may be required and low delays should be introduced within the whole process in order to achieve the theoretical promised benefits. Thus, while centralizing part or whole of the process may benefit from the global obtained view of the system channels and high computational capability network entities, distributed processing may involve less delays (leading to CSI aging) and less backhauling workload.

In the iJOIN architecture, different options for the precoder implementation (i.e. for calculating the precoder matrix and for determining the transmit signals) exist.

- Option a1: *Fully centralized precoding at the RANaaS*. In that option, global CSI (i.e. CSI for all channels in the cooperation set) is fed back to the RANaaS over the J1 link and precoding process is completely performed at the RANaaS. The precoded signals are transmitted to iSCs via the J1 link. Note that no J2 link is required in this case.
- Option a2: *Centralized calculation of precoding matrix at RANaaS*. In this option, global CSI is fed back to the RANaaS over J1 links, the calculated precoding weights along with the UE signals (e.g. info bits, codebits, or modulated symbols) are sent to the corresponding iSCs via J1 link and application of precoder is performed at iSCs. No J2 link is required in this option as well.
- Option b: *Distributed calculation of joint precoder*. In that case, the J2 link is used for global CSI exchange between iSCs and distributed precoding matrix calculation is performed at iSCs, while the J1 link is used only for sending UE signals from RANaaS to iSCs.

- Option c: *Local precoding*. This option considers no exchange of information among iSCs. J1 links are used for the UE signals and no J2 link is required in that case of no cooperation between iSCs.

Of course, compared to option c, options a (a1 and a2) and b can take advantage of multi-site coordination mechanisms to improve the network spectrum efficiency. However, the actual choice of functional split depends on several system factors such as the properties of the backhaul network, the processing capability of the involved nodes as well as the latency constraints imposed (e.g. due high UE mobility).

In particular, for option b, constraints related to the deployment of iSCs itself may limit the local implementation of joint processing. Firstly, iSCs are normally limited in terms of processing capability, especially when a high number of total system antennas are involved (e.g. multiple iSCs cooperation) [90], the computational demands may make the calculation of precoding matrix at iSCs impractical. Secondly, the limited (in terms of availability, bandwidth and latency) backhaul links between iSCs may pose a serious challenge for collecting global CSI at the iSCs.

A promising solution is to move the precoding matrix calculation at RANaaS (option a1 and a2), where computational capabilities are much higher and global CSI can generally be more easily collected from each individual iSC through J1 links. However, there are also some constraints for implementing precoding at the RANaaS. The backhaul capacity is one critical factor; the payload to backhaul for the DL functional split after precoding may be significantly higher compared to the one before precoding (see Section 3.2.1.1 for details). To this end, CT2.5 focuses on the critical issue of timely exchange of the CSI (or the precoded signals) between cooperating nodes and investigates methods of precoding which adapt to any backhaul topology and exploit the available backhaul in the most efficient way. Another critical issue has to deal with is the accuracy of CSI estimation; the move of precoding matrix calculation to the RANaaS will cause extra latency and CSI quantisation errors. These factors will cause a mismatch between the CSI estimate used for precoding and the actual transmission channels. Consequently, extra interference arises that may significantly reduce the performance and throughput. Among the channel estimation error, the channel quantization error and the latency, the last one is particularly of interest for joint transmission. This latency is mainly caused by the RANaaS processing and CSI feedback delay and its value will depends on the number cooperating iSCs (e.g. the number of iSCs within one veNB working on the same time-frequency resources to jointly transmit to a group of UEs). This is one of the most critical constraints for implementing centralised precoding at the RANaaS. In this section we will model the cluster size related latency including RANaaS processing and CSI feedback delay. The model will be used for veNB size optimization in CT2.4.

Delay model for centralized precoding

Let us consider a veNB composed by a RANaaS and N_{iSC} iSCs each equipped with N_T^{iSC} antennas serving N_{UE} UEs, each UE with N_R^{UE} antennas. In general, the total delay of the precoding operation is caused by several factors such as (see Figure 3-5):

- pilot estimation and processing delay at the UE (step 1 in Figure 3-5)
- two-way propagation delay from UE to iSC and from iSC to RANaaS (step 2, 4, 6, 7 in Figure 3-5)
- UE feedback scheduling and channel waiting delay (step 1 in Figure 3-5)
- iSC and RANaaS processing delay(step 3 and 5 in Figure 3-5)

These delay factors are depicted together in Figure 3-5 and the overall delay can be expressed as

$$\Delta t = N_{ant_tot} \Delta t_{chan_est_UE} + N_{fb_time} (\Delta t_{fb} + \Delta t_{prop_total}) + \Delta t_{process_RANaaS} + \Delta t_{process_iSC} \quad (3-46)$$

In the following we describe each delay factor in detail.

Channel estimation delay

The first term $\Delta t_{chan_est_UE}$ in (3-46) denotes the channel estimation delay at the UE. Traditionally, least square (LS) channel estimation is implemented at the UE with rather low complexity [91]. Assuming non-frequency selective channels for simplify, the number of channel coefficients to be estimated by each UE is

$$N_{ant_tot} = N_T^{iSC} N_{iSC} N_R^{UE} \quad (3-47)$$

Note also that more advanced estimation algorithms such as MMSE may be adopted to improve the estimation accuracy. However, the complexity will increase by a cubic order with the estimation dimension [91]. Nevertheless, due to the rather low overall complexity, this impact is neglected in the subsequent analysis.

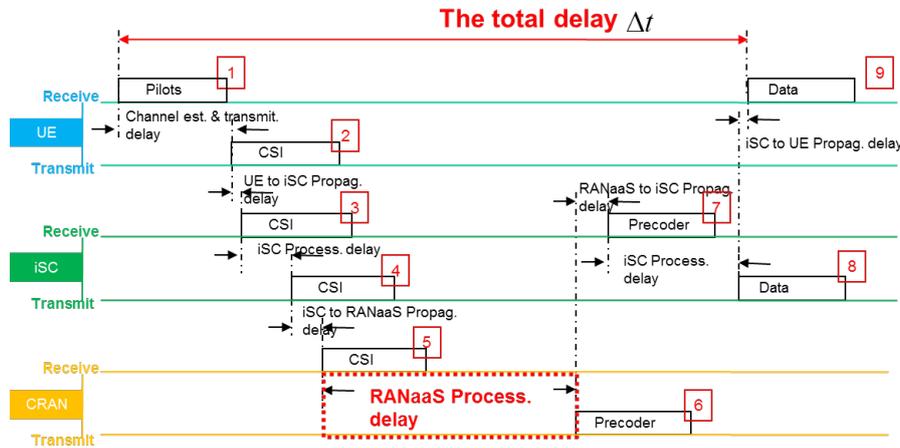


Figure 3-5: The total CSI delay of the precoding operation

CSI feedback and propagation delay

The second term in (3-46) contains two items; the first parameter Δt_{fb} is the channel feedback delay due to the channel waiting and scheduling, while the second parameter Δt_{prop_total} denotes the two-way propagation delay per feedback. Assuming no multiplexing used for CSI feedback and limited feedback resources per time interval, the total number of time intervals required to feed back all of the channel coefficients to the iSCs is given by

$$N_{fb_time} = \frac{N_{ant_tot} N_{UE}}{N_{CSI_per_fb}} = \frac{N_T^{iSC} N_{iSC} N_R^{UE} N_{UE}}{N_{CSI_per_fb}} \quad (3-48)$$

with $N_{CSI_per_fb}$ specifying the number of channel coefficients that can be fed back at each time interval. To make it easier to understand the meaning of $N_{CSI_per_fb}$, we give the following example: suppose that we need feedback 50 (i.e. $N_{iSC} N_T^{iSC} N_{UE} N_R^{UE} = 50$) channel coefficients in total for precoding, due to the UL resource limitation (and no multiplexing used), it is impossible to feedback all of the coefficients in one scheduling time, suppose that each time we can only feedback 10 coefficients (i.e. $N_{CSI_per_fb} = 10$), then we need 5 times (i.e. $N_{fb_time} = 5$) in total (serially) to feedback all of the coefficients. Apparently, the more iSCs and UEs are contained in a vNB, the more channel coefficients are required.

Suppose that the CSI is quantized by N_{Q_CSI} bits and the data rate for transmitting the CSI feedback from one UE to one iSC is given by C_{fb} , then the feedback delay can be calculated as:

$$N_{fb_time} \Delta t_{fb} = \frac{N_{iSC} N_T^{iSC} N_{UE} N_R^{UE} N_{Q_CSI}}{C_{fb} N_{CSI_per_fb}} \quad (3-49)$$

Moreover, the total propagation delay can be expressed as:

$$N_{fb_time} \Delta t_{prop_total} = \frac{2N_{iSC} N_T^{iSC} N_{UE} N_R^{UE} (d_{ue_to_iSC} + d_{iSC_to_RANaaS})}{c_0 N_{CSI_per_fb}} \quad (3-50)$$

where c_0 is the speed of light and $d_{ue_to_iSC}$ and $d_{iSC_to_RANaaS}$ denote the average distances from UE to iSC and from iSC to RANaaS, respectively. Note that even though the distances from each iSC to RANaaS and from each UE to iSCs are different due to the randomness of the placement. However, the propagation delay is negligible in the small cell networks and here we have used an average distances $d_{ue_to_iSC}$ and $d_{iSC_to_RANaaS}$ as an approximation.

RANaaS processing delay

The third term $\Delta t_{process_RANaaS}$ in (3-46) denotes the RANaaS processing delay and is composed of two parts

$$\Delta t_{process_RANaaS} = \Delta t_{Tx1} + \Delta t_{precoder_cal} \quad (3-51)$$

The first part Δt_{Tx1} is attributed to the baseband processing (such as coding, modulation, precoding, etc.) at the RANaaS. The more functionality is implemented in the RANaaS, the smaller will be the corresponding processing delay Δt_{Tx2} at the iSCs. However, the total baseband processing delay for transmission $\Delta t_{Tx} = \Delta t_{Tx1} + \Delta t_{Tx2}$ can be considered roughly constant. In LTE-A, the worst case of Tx processing time is around 2.3 ms, i.e. $\Delta t_{Tx} = 2.33$ ms can be assumed here as well [15].

The second part of the RANaaS processing delay $\Delta t_{precoder_cal}$ is the precoder calculation delay. This may become a dominant factor when the number of iSCs per veNB is extremely huge and the RANaaS computational capability is limited. Moreover, different precoding algorithms lead to dramatically different computational complexity. For example, ZF precoding has significant larger complexity than MRT beamforming. Considering ZF as an example, the total number of real-valued multiplications K_{multi} and real-valued additions K_{add} for the matrix calculation can be expressed as:

$$K_{multi} = 8(N_{UE} N_R^{UE})^2 N_{iSC} N_T^{iSC} + O[4(N_{UE} N_R^{UE})^3] + 2N_{UE} N_R^{UE} N_{iSC} N_T^{iSC} \quad (3-52)$$

and

$$K_{add} = 8(N_{UE} N_R^{UE})^2 N_{iSC} N_T^{iSC} - 2(N_{UE} N_R^{UE})^2 - 2N_{UE} N_R^{UE} - 2\zeta_1 N_{UE} N_R^{UE} N_{iSC} N_T^{iSC} + O[4(N_{UE} N_R^{UE})^3] \quad (3-53)$$

The term $O[4(N_{UE} N_R^{UE})^3]$ arises from the matrix inversion and its complexity depends on the specific algorithm but generally follows the cubic order of matrix dimension. For simplicity, the parameter ζ_1 represents the complexity of one division by equivalent addition operations and its typical value takes from 15-25 in modern processors [100]. Similarly, parameter ζ_2 is the equivalent number addition operation times for each multiplication operation and its typical value takes 1 in the modern processor [100].

Thus, the delay caused by the precoder matrix calculation can be written as

$$\Delta t_{precoder_cal} = \frac{K_{add} + \zeta_2 K_{multi}}{N_{FLOPS} q_c} \quad (3-54)$$

where F_{FLOPS} represents the total RANaaS computational capability and q_c is a resource division factor. The purpose of the resource division factor is the following: since the computational resources could be shared by multi-tasks (coding/decoding etc.), joint transmit processing of different groups of UEs within a veNB (i.e. precoding is performed only for UEs assigned to same PRBs) and the implementation of several veNB on one cloud computing platform, only the fraction $1/q_c$ of the total computational resources will be allocated to the precoding matrix calculation. Note that q_c could be set smaller than 1, which corresponds to

the case that multiple processors contribute to the precoding matrix computation in parallel. Therefore, $F_{FLOPS}q_c$ denotes the available computational capability allocated to the precoding matrix calculation.

Note here we have not considered the granularity of the precoding, i.e. how many subcarriers in frequency domain and how many subframes in time domain use the same precoding matrix. Generally speaking, smaller granularity leads to an accurate precoding, but requires more computational resources. With fixed computational resources, the optimal granularity is depending on the correlation in time and frequency. Fast moving UEs (with high frequency selective channel) may need denser precoding in the frequency domain and fast changing channel may results in smaller precoding granularity in time domain.

iSC processing delay

The last term $\Delta t_{process_iSC}$ in (3-46) represents the iSC processing delay which is also composed of two parts:

$\Delta t_{process_iSC} = \Delta t_{Tx2} + \Delta t_{CSI_fd}$. The first part Δt_{Tx2} is attributed (as already mentioned above at the RANaaS processing delay part) to the transmitter chain functionalities (such as IFFT, etc.) at the iSCs. The second part Δt_{CSI_fd} is due to uplink CSI feedback where iSCs forward the received CSI to the RANaaS. Since this part is minimal compared to other delay factors, it can be treated as a constant.

The total processing delay at the RANaaS and iSC can be written as $\Delta t_{Tx} + \Delta t_{process_iSC} + \Delta t_{process_RANaaS}$. Then the total delay (3-46) can be rewritten as

$$\Delta t = N_{ant_tot} \Delta t_{chan_est_UE} + N_{fb_time} (\Delta t_{fb} + \Delta t_{prop_total}) + \Delta t_{precoder_cal} + \Delta t_{Tx} + \Delta t_{CSI_fd} \quad (3-55)$$

According to our derivation in equation (3-55), the total delay is increasing with the antenna number in a veNB that composes of multiple cooperative iSCs, especially for RANaaS processing delay and CSI feedback delay. The RANaaS processing delay is cubic order of the total receiving antenna number, which may lead to a significant delay when the cooperation number of iSCs is large, then may make the centralised precoding at RANaaS infeasible. However, the delay at RANaaS also critically depends on the available computational resources for precoding. Sufficient such resources may result in a negligible additional delay for the system. On the other hand, the CSI feedback delay is second order of the antenna number but also depends on the CSI feedback capacity from UE to iSC. Large capacity may lead to a very small delay to the system. Therefore, the decision on whether centralised precoding at RANaaS is feasible or not will not only depend on the total number of cooperating antennas in a veNB, but also depend on the network and cloud configurations (CSI feedback capacity and available computational resource, etc.). In Section 4.4, based on the developed delay model and specific configurations, we will investigate the optimal veNB size (number of cooperating iSCs in a veNB) for maximising network sum-rate.

3.2.2.2 Decoder @ Cloud

In the context of a flexible cloud-RAN, the FEC decoding may be performed either within the small cell (local processing) or within the cloud platform (central processing). However, in the uplink direction, if the functional split between local and central processing is done within the physical layer, the FEC decoding will necessarily have to be performed within the cloud platform. This poses a significant challenge to the design of a flexible cloud-RAN architecture, since FEC decoders are usually implemented in specialized hardware, such as application-specific integrated circuit (ASIC) or field-programmable gate array (FPGA). The use of specialized hardware makes it possible to cope with the computational complexity of FEC decoders, while meeting the increasing throughput requirements of wireless communication systems. Nonetheless, the advent of many-core architectures provides outstanding computing power and opens new perspectives for massively parallel implementations. In order to meet stringent requirements on data rate, cloud-based FEC decoders will need to fully exploit the high level of parallelism available in the computing platform. In this context, Low Density Parity Check (LDPC) [45] and Turbo codes [46] appear to be the two most promising candidates, due to the intrinsic capacity of their decoding algorithms to accommodate various degrees of parallelization.

As illustrated in Figure 3-6, two main approaches can be used to exploit parallelism in multiprocessor/multicore platforms. First, fully or partially *parallel decoders* can be implemented through the use of concurrent threads, with every processor or core executing a separate thread. The efficiency of such a parallel implementation depends also on the degree of parallelism allowed by the implemented algorithm. The

second approach consists of using *multi-codeword decoders*, with each processor or core running a separate image that decodes a different codeword. It is worth noting that both approaches allow increasing the decoder throughput, although in two different ways: the first by reducing the latency per decoded codeword, the second by increasing the number of codewords decoded within the same latency period. In addition, both approaches can be combined, depending on the implemented algorithm and the specific characteristic of the computing platform.

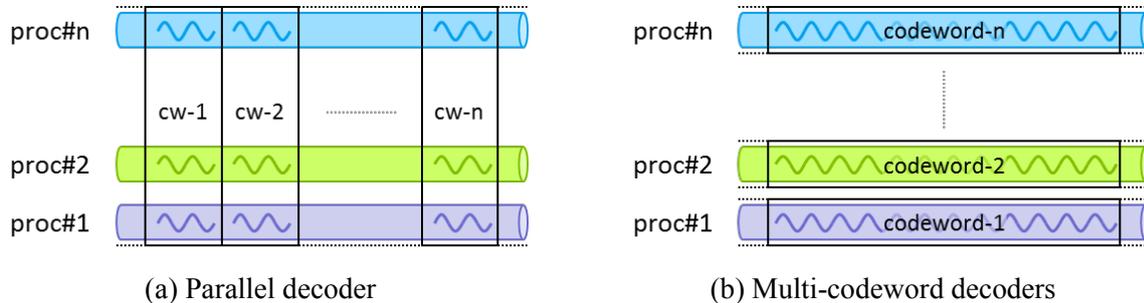


Figure 3-6: Exploiting parallelism in multiprocessor/multicore platforms

For high throughput (e.g. multi-gigabit/sec) applications, LDPC are known to compare favourably against Turbo codes, due to the fact that LDPC decoding allows a higher degree of parallelism [47]-[51]. This also makes LDPC codes more suitable with regard to aforementioned parallel decoder approach, and explains the surge of interest on flexible software-based implementations of LDPC decoders on various parallel computing platforms [52]-[59]. However, software-based parallel LDPC decoders barely achieve throughputs of a few tens of Mbits/s, as reported in [57] for General Purpose Graphics Processing Units (GPUs), or in [52] for the Signal-processing On-Demand Architecture (SODA). This is mainly due to large overheads required to exchange information between threads (e.g. through an interconnection network or by using a shared memory), thereby resulting in scalability issues [57], [59].

In contrast with the parallel decoder approach, the multi-codeword decoder approach does not rely on the parallelization of the decoding algorithm, thus eliminating the need for exchanging information between threads. This results in better scalability properties and the throughput of the multi-codeword decoder is known to increase almost linearly with the number of cores [59]. Another advantage of this approach is that it can easily accommodate other families of codes, such as convolutional or Turbo codes, which is of particular interest since these codes are used in a large number of cellular wireless standards, including 3GPP-LTE. Multi-codeword LDPC decoders have been reported to achieve throughputs up to 80 Mbits/s on the IBM CELL Broadband Engine [53], [57], with 24 to 96 codewords decoded in parallel. Recently, CPU and GPU implementations of multi-codeword Turbo decoders have also been reported in [60], with a peak throughput increasing from 55 Mbits/s to 122 Mbits/s, as the number of decoding iterations decreases from 8 to 4.

In conclusion, the cloud-based implementation of modern FEC decoders, such as LDPC and Turbo codes, is a challenging task. To reach the increasing throughput requirements of 5G cellular systems will require the simultaneous optimization both at the algorithmic and the architectural abstraction levels. At the best of our knowledge, such an algorithm-architecture co-design for cloud-based decoders has not been addressed so far in the literature.

3.2.3 Decoder Implementation on Cloud Platform

This section reports preliminary results concerning the implementation of LDPC and Turbo decoders on general purpose hardware. The goal is to assess the achievable throughput of such implementations and to investigate the scalability with respect to the number of processors and/or cores of the computing platform. Results presented in this section have been published in [80].

Implementation of Message Passing Algorithm

LDPC codes [45] are a class of linear codes that can be advantageously represented by sparse bipartite graphs. The bipartite-graph of an LDPC code contains two types of nodes – variable-nodes, corresponding to codeword bits, and check-nodes, corresponding to parity-check equations – which are connected according to the non-zero entries of the parity check matrix defining the code. LDPC decoding is performed by *message-passing* (MP) algorithms, which consist of an exchange of messages along the edges of the bipartite

graph, taking place in several rounds or iterations. Thus, variable-nodes collect more and more information with each new decoding iteration, which gradually improves the estimation of the sent codeword. The most effective MP decoding algorithms are the *belief-propagation* (BP) decoding – also referred to as *sum-product* (SP) – and the *min-sum* (MS) decoding.

The BP decoding is known to be optimal on cycle-free graphs, but can also be successfully applied to decode linear codes defined by graphs with cycles, which is actually the case of all practical codes. However, in practical applications the BP algorithm is disadvantaged by its computational complexity, numerical instability, and the fact that it is highly sensitive to the SNR estimation, which may lack accuracy in practical situations.

One way to deal with complexity and numerical instability issues is to simplify the computation of messages exchanged within the BP decoding. The most complex step of the BP decoding is the computation of check-node messages, which makes use of computationally intensive hyperbolic tangent functions. The MS algorithm is aimed at reducing the computational complexity of the BP, by using max-log approximations of these messages [64]. The only computations required by the MS decoding are additions and comparisons, which solves the complexity and numerical instability problems. The performance of the MS decoding is also known to be independent of the knowledge of the channel parameter (e.g. SNR), for most of the usual channel models.

Practical implementations of MP decoders may also deal with different *scheduling strategies*, indicating the order in which variable and check-node messages are updated during the MP iterative process. The classical convention is that, at each iteration, all check-nodes and subsequently all variable-nodes pass new messages to their neighbours. This message-passing schedule is usually referred to as *flooding scheduling* [65]. A different approach is to process check or variable nodes in a serial manner [66], [67]. Applied to check-nodes, the *serial scheduling* enables them to be processed sequentially, with an immediate propagation of their messages to the neighbour variable-nodes. The decoder updates the neighbour variable-nodes, so as to profit from the propagated messages, and then proceeds to the next check-node. The main advantage of the serial schedule is that it propagates information faster and converges in about half the number of iterations compared to the flooding schedule. For applications requiring a small number of decoding iterations, the serial schedule provides superior decoding performance, thanks to its faster convergence. Finally, the *layered scheduling* [68] allows processing sequentially groups (layers) of check-nodes, while check-nodes within the same group are processed by using a flooding schedule strategy. This scheduling technique is also known to converge about twice faster than the flooding scheduling [69]. Actually, if the parity-check matrix is split into horizontal layers such that any variable-node is connected to at most one check-node in each layer (e.g. Quasi-Cyclic LDPC codes [70]), then both serial and layered scheduling have the same performance and convergence speed. It is also worth noting that the number of operations per decoding iteration is the same for all flooding, serial, and layered schedules.

On the basis of the previous discussion, the Min-Sum decoder with layered scheduling has been chosen for our investigation. In order to assess its achievable throughput on general purpose hardware, a C++ implementation of a multi-codeword decoder (Section 3.2.2.2) has been carried out. Multithreading is implemented by using Message Passing Interface (MPI) and Open Multi-Processing (OpenMP) directives [71]. The maximum number of decoding iterations has been set to 20. The decoder stops when whether a codeword have been found (syndrome equal zero) or the maximum number of iterations is reached. The FER performance and the average number of decoding iterations for the WiMAX LDPC codes [10] with rate 1/2 and 5/6 are presented in Figure 3-7 (for QPSK modulated signal corrupted with additive white Gaussian noise). It can be seen that the average number of decoding iterations decreases with the FER. For FER = 10^{-2} , the average number of iterations is about 6.9 for rate 1/2 (SNR = 2dB), and about 4 for rate 5/6 (SNR = 5.9dB).

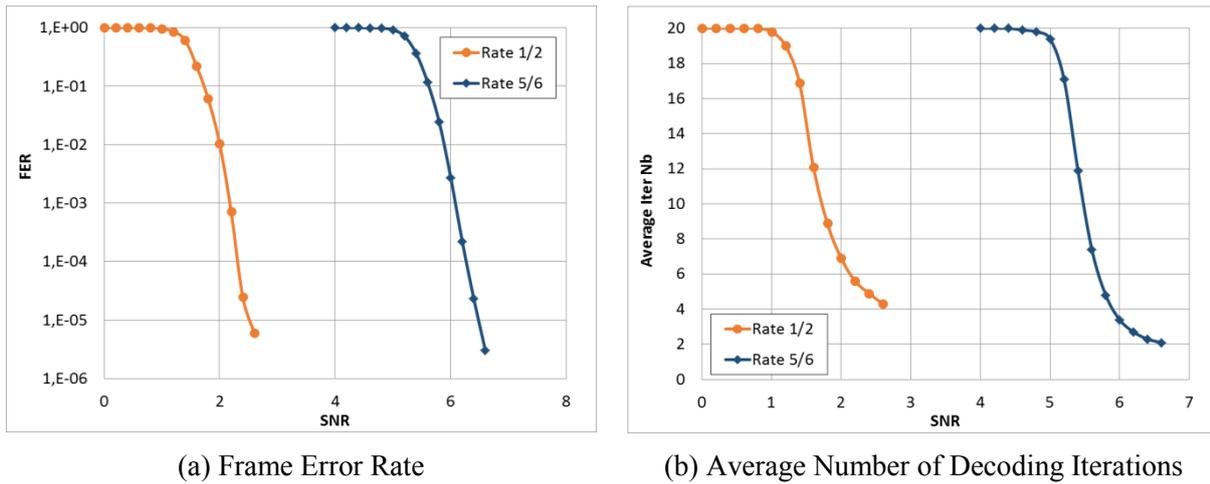


Figure 3-7: FER and average number of decoding iterations, WiMAX LDPC with rate 1/2 and 5/6

For the multi-codeword decoder, codewords decoded by different threads can take a different number of decoding iterations. However, as illustrated in Figure 3-8, threads decoding different codewords are synchronized, in the sense that they all start decoding in same time and wait for the slowest one to complete decoding.

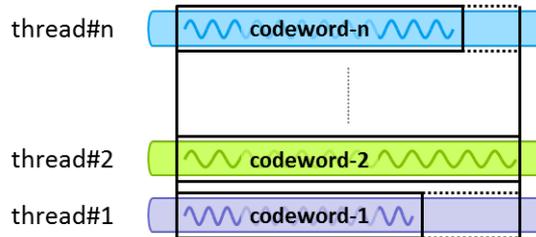


Figure 3-8: Multi-codeword decoder with synchronized threads

The Min-Sum multi-codeword decoder has been run on two Intel Xeon x5650 @2.67GHz processors, each one composed of 6 physical cores. Each physical core is further composed of 2 logical cores, which leads to a total of 24 logical cores divided in two processors. The average throughput, as function of the number of cores, is presented in Figure 3-9, for SNR values corresponding to FER values of 10^{-2} and 10^{-4} . The throughput is expressed as number of useful (information) bits per second. The difference between rate-1/2 and rate-5/6 throughput is explained by (i) the faster convergence of the rate-5/6 decoder, which results in a reduced average number of decoding iterations, and (ii) the increased number of information bits for each decoded codeword. It can also be observed that the achieved throughput doubles when the FER improves from 10^{-2} to 10^{-4} , corresponding to an SNR increase of about 0.3-0.4 dB. For an SNR of 6.3 dB ($FER = 10^{-4}$), the rate-5/6 decoder achieves an average throughput of 140 Mbits/s, by decoding 22 codewords in parallel (22 logical cores are used).

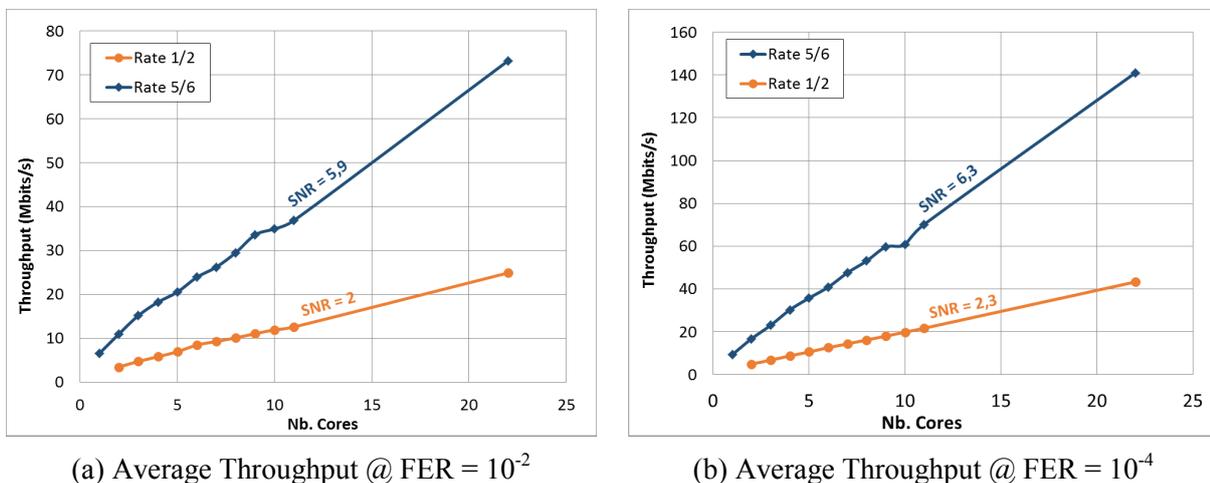


Figure 3-9: Average Throughput as function of the number of cores, WiMAX LDPC with rate 1/2 and 5/6

Implementation of LTE Turbo Decoder

The implementation of the channel decoder on a cloud platform promises computational diversity, since the required computational power depends on the current access link qualities and on the rate demand of the user. To assess this issue, a multi-threaded C++ implementation of the LTE turbo decoder on general purpose hardware has been carried out. The performance of this decoder was evaluated using a basic link-level simulation employing the LTE uplink MCSs 6 through 28 with 4QAM, 16QAM and 64QAM modulation and code rates larger than 1/3. The simulation was performed under the Ubuntu Linux 12.04 host operating system of the UoB RANaaS testbed. Initial results on a VMware ESXi system with fewer MCSs were originally presented in [37], while these latest results have been presented in [80]. The left plot in Figure 3-10 shows the achievable spectral efficiency for a target block error rate of 1% depicted over the SNR. For each SNR, the highest MCS which fulfilled the target block error rate was selected for a given number of 8 or 2 maximum iterations. It can be seen that the restriction to 2 iterations causes a small penalty on the spectral efficiency and thus on the throughput. In the right plot in Figure 3-10, the resulting computational complexity, denoted by number of required CPU cores is shown. Here it can be observed that the restriction to a maximum of 2 iterations will reduce the computational load severely. Furthermore, it can be seen that it does not show monotonic behaviour, but shows significant variance, in particular if the throughput is maximized. This implies that through pooling on a cloud platform, significant gains through computational diversity can be achieved. The impact of the SNR on the decoding time has initially been assessed in the companion deliverable D6.1 [89].

In the literature, there exists a row of publications investigating the latency of turbo decoders or aiming for a low latency implementation [72], [73], [74]. In order to allow real-time processing on general purpose hardware, these approaches will be considered for future implementations.

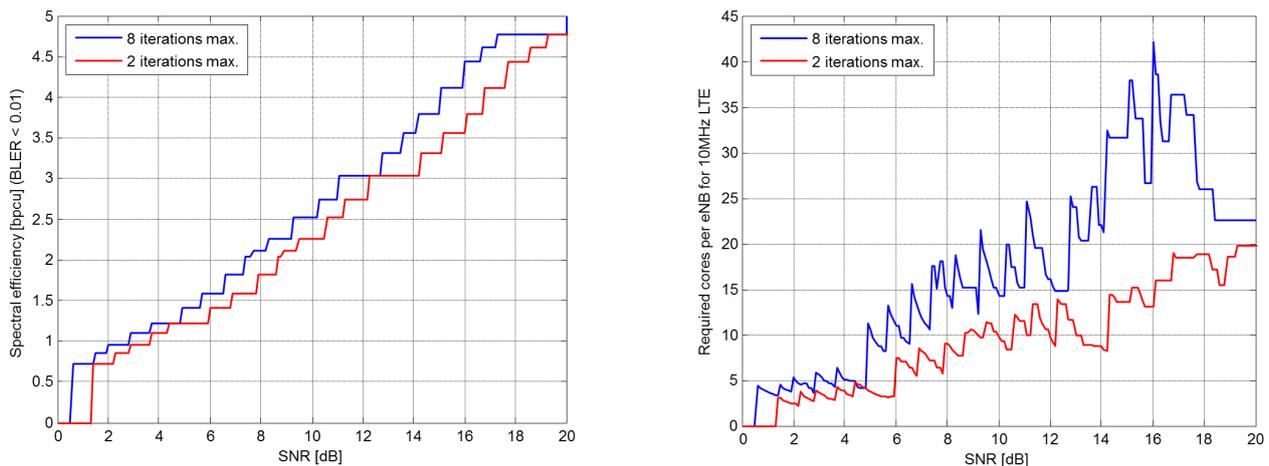


Figure 3-10: Spectral efficiency and computational complexity results for turbo decoding on the UoB RANaaS testbed

3.3 Virtual eNB Implementation

3.3.1 veNB and 3GPP compliance

In order to facilitate the adoption of the RANaaS concept, its system architecture should provide the required interfaces without disruptive changes to the existing deployments. The Figure 3-11 illustrates the integration of the RANaaS concept into the 3GPP LTE system architecture. It does not imply changes to existing interfaces but introduces the concept of a virtual eNodeB (veNB). A veNB is composed of one or more iSCs, a cloud based RANaaS platform, and the necessary BH links between these nodes. It maintains the same interfaces as a 3GPP LTE eNB in order to maximize backward-compatibility, i.e., X2 between eNBs and veNBs as well as S1 between veNBs and EPC. It is expected that these standard interfaces terminate at the RANaaS platform. The iSCs communicate with the RANaaS platform through the logical J1 interface, while neighbouring iSCs can exchange information with each other directly by the logical J2 interface.

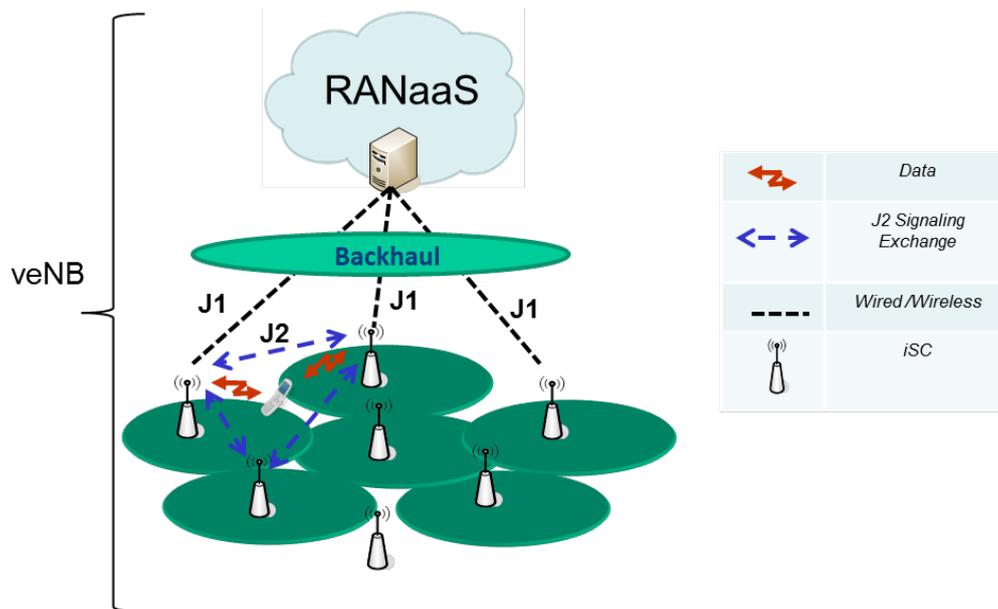


Figure 3-11 System architecture with virtual eNodeB (veNB) consisting of a central RANaaS platform and several iSCs

In the classical LTE architecture, one eNB can host several cells and the most known example being the tri-sector site in a macro deployment, while a Home eNB (HeNB) only supports one cell. Within the WP2 scope, a cell is essentially defined by its physical cell identity (PCI). As a reminder, the PCI is made of two values broadcasted to help cell synchronisation/frame timing:

- Primary Synchronization Signal (PSS): a 64-length sequence made of Zadoff-Chu sequence; 3 sequences are defined and identified by $N_{ID}^{(2)} = 0, \dots, 2$.
- Secondary Synchronization Signal (SSS): a 64-length sequence made of two binary sequences interleaved and concatenated, scrambled by a PSS-dependent sequence; 168 sequences are defined and identified by $N_{ID}^{(1)} = 0, \dots, 167$.

The PCI is given by the equation below; more details are given in [15].

$$PCI = N_{ID}^{CELL} = 3N_{ID}^{(1)} + N_{ID}^{(2)} \quad (3-56)$$

The PCI can take 504 different values and dictates the position of the other reference signals. For instance, the cell reference signals (CRSs) in the downlink are positioned in the resource grid based on the PCI value modulo 6. Therefore, the PCI is of primary importance for the physical layer.

Since a veNB should ideally support the same features as a classical eNB, it should also support the hosting of multiple cells. If an iSC can only support the broadcasting of one PCI, which is the most likely scenario, then:

- When the veNB is made of RANaaS and one iSC, the veNB would be able to support only one PCI/cell (similar to a femtocell since it is unlikely to deploy iSC with directive antennas).
- When the veNB is made of RANaaS and several iSCs, the veNB would be able to support multiple cell hosting, the maximum number of cells being equal to 256 as per 3GPP specifications regarding the cell addressing.

Usually PCIs are chosen such that each neighbouring cell has a different value, facilitating the cell discovery at the UE side. However, if the same PCI is used within the iSCs of a single veNB, then the system should be seen as a distributed antenna system to avoid this PCI confusion. In this case, a specific care should be taken in the way the antenna ports are assigned to the antenna. Indeed, LTE defines antenna port as a way to link data to be transmitted to the antennas. If the iSCs have only one antenna then it can only transmit one antenna port. If it has more than one, then either the same ports could be used over the whole iSCs of

different port could be used, increasing the maximal number of spatial streams being able to be transmitted (the maximum being 8 in the current specifications).

3.3.2 Further description of veNB

The iJOIN system architecture described above requires that the functionality at the eNB can be decomposed into re-assignable functions, and each function can be assigned either to the central RANaaS or local iSCs. The actual shift of functionality is controlled by the iJOIN virtual eNB Controller (iveC) located in the RANaaS. There are four main characteristics of a veNB:

- a) Functional decomposition: The functionality executed by a veNB can be decomposed in modules. Each module implements a functionality of the RAN protocol stack such as PHY layer, MAC layer, or RRC layer procedures
- b) Functional split: Each of the modules may be located either locally at the iSCs or centrally at the RANaaS entity. Furthermore, only a subset of all modules may be executed at the iSCs while another subset is implemented at the RANaaS entity. The interaction between modules, the placement of modules, and the interaction of modules with other logical entities is controlled by the iveC
- c) Functional interworking (Control Plane): iJOIN defines interfaces that are used for the coordination and control of functional split in order to transparently configure which modules are executed at the iSC and which modules are executed in the RANaaS entity
- d) Logical unity: On a black-box view, e.g. from the core network perspective, the veNB appears as a standard eNB in order to ensure compatibility with 3GPP specifications.

3.4 Physical layer constraints

The implementation of PHY layer processing approaches in the iJOIN architecture is subject to several practical impairments caused by the distributed placement of the cooperating nodes and physical layer aspect of the transmission schemes. In this section, we identify the main PHY layer constraints and discuss their impact of joint UL reception and for joint DL transmission schemes. However, as the impact will highly depend on the applied UL or DL processing approach and the chosen functional split, the discussion is restricted on the principle influences only without distinguishing all possible centralized or distributed implementation options. Furthermore, as the investigation of impairments is not in the main focus of iJOIN, but has been thoroughly been investigated in national projects like EASY-C [120] or FP7 projects like ARTIST-4G [121] in the context of CoMP, we restrict our self to provide an overview here. A general detailed overview of the different impacts using also results from CoMP field trials are provided in [109] and [119].

Hence, we start by presenting briefly the main approaches being used or discussed in LTE and LTE-A to feedback the CSI obtained at the UEs to the cooperating iSCs. This channel feedback is necessarily imperfect as it inevitably introduces some delay and only finite feedback resources are available. Therefore, we then discuss the main principles driving the delay requirements. In order to implement cooperative transmission and reception techniques in the iJOIN architecture, messages like CSI, received signals, or precoded signals have to be exchanged among the nodes via the J1 or the J2 interfaces. As a consequence, the signals have to be quantized in order to efficiently use the backhaul network. Some of the main results available in the literature regarding the impact of this quantization or its optimization are presented. Finally, an efficient cooperative transmission requires an accurate synchronization between the iSCs. Thus, the main synchronization techniques are presented and their advantages and drawbacks are compared. The optimization of the spectral shape used in the transmission is also shortly described, as it represents one of the most promising candidates for reducing these synchronizations requirements and improving the performance in future 5G networks.

CSI Feedback Schemes

In LTE release 8, the feedback consists of a link adaptation and channel quality indicator (CQI) feedback which is used to adapt the modulation and coding to the current channel state. In addition to that feedback, from release 10 onwards, it is possible for the UEs to feedback some precoding information to efficiently use the multiple-antennas. This feedback consists in the index of an element in the precoding codebook. This

feedback is said to be “implicit” in the sense that the CSI is not directly feedback, but a precoding decision is fed back [126]. This feedback is not adapted to an efficient transmission in a multi-user MIMO scenario which requires a precoder design depending on the channels of all the users. Therefore, the feedback design is one of the topics being heavily investigated in future LTE networks [125].

Explicit feedback has then appeared as the privileged solution for future LTE releases as it allows more advanced and flexible precoding which justifies its additional complexity compared to implicit feedback. Many quantization schemes can be used depending on which figure of merit is minimized. Grassmannian quantization is however a very efficient approach which has been the focus of the major part of the literature dealing with transmission with imperfect CSI and is the most used quantization scheme [127]. It consists in quantizing a unit-norm channel vector \mathbf{h} by its estimate $\hat{\mathbf{h}}$ such that

$$\hat{\mathbf{h}} = \underset{\mathbf{c} \in C}{\operatorname{argmax}} \quad |\mathbf{h}^H \mathbf{c}| \quad (3-57)$$

where C is the quantization codebook and is formed from unit-norm vectors \mathbf{c} . This quantization scheme has also the advantage of being well understood and theoretically analysed (see [127] and references therein). Although very efficient, this approach also comes with some drawbacks, particularly in the case of cooperative transmissions. Indeed, the norm of the channel is not feedback since only its unit-norm value is quantized. Furthermore, the figure of merit is invariant by multiplication by any unit-norm complex number and has to be modified so as to adapt to the case of multiple cooperating iSCs [22].

The quantization is also the key in efficiently reducing the impact of delay, the limited feedback resources, or the limited backhaul resources. Hence, the feedback design in future networks is an active research area both in the academic field and in the industrial world (e.g. [128] and references therein). It is also currently discussed inside the 3GPP.

CSI Delay Requirements

In order to perform coherent detection in the uplink and coherent transmission in the downlink, the channel coefficients have to be estimated. In LTE-A the channel is estimated at the receiver side using the demodulation reference signals (DMRS). These pilot signals are placed in a dedicated symbol in every slot, spanning the whole bandwidth used for the corresponding transmission. In order to distinguish different spatial layers orthogonal reference signal are used allowing separating the antennas of the UEs in the uplink and the iSCs in the downlink. The quality of the channel estimates could be improved by adding more pilot symbols [123] at the cost of the possible payload size which results in a corresponding trade-off [124].

In case of centralized processing in the RANaaS, the individual channel estimates need to be quantized by the iSC before forwarding over the J1 link. Thus, in addition to the channel estimation error also the quantization error will affect the overall quality of the channel estimate. In addition to DMRS uplink sounding reference signals (SRS) are transmitted periodically, but less often or even on UE request. This SRS is used to acquire CSI only for scheduling and link adaptation. Their design also allows the separation of several UEs. Since SRS might be split over several subframes, an iSC will most likely obtain CSI locally and forward the quantized CSI to the RANaaS.

For the CoMP uplink different approaches for joint channel estimation and the impact of estimation errors were investigated in [118] and [122] considering the LTE-A pilot structure. However, the impact of channel estimation errors is more critical for the downlink.

In case of FDD, pilot signals are inserted in the downlink transmit signal such that the UEs estimate the CSI and feed them back in the next uplink transmission, before the precoder can be calculated based on this CSIT (channel state information at the transmitter). A mismatch between the CSIT and the actual transmit channel may cause severe interference. In order to provide the nodes (iSCs and/or RANaaS) with an accurate CSI estimate, it is necessary to be able to exchange this information in a timely manner. Specifically, this means that the total delay between the channel estimation at the UE and the transmissions from the iSC in the DL should lead only to acceptable performance degradation. It is unavoidable that each step of the CSI sharing and of the precoding introduces some delay. Hence, a critical problem consists in evaluating this delay to ensure a limited degradation of the performance. A precise analysis of the delay generated by every step of the precoding is done in Section 3.2.2.1. Comparing the delay incurred by the transmission with the corresponding transmission scheme and the rate of channel variation allows gaining a good understanding of the impact of CSI aging. Therefore, we will now roughly evaluate the time scale of the channel change for

two commonly used scenarios. Even though the exact calculation depends on many parameters of the system, the orders of magnitude obtained provides a useful guideline for the system design and spans most of the practical mobile scenarios.

In order to consider the variation of the channel over time, the rate of change is defined by

$$\rho_{T_{50}} = \frac{E[h(t)h^*(t + T_{50})]}{E[|h(t)h^*(t)|^2]} \quad (3-58)$$

where T_{50} denotes the 50% coherency time. This 50% coherency time T_{50} can then be used as a measure for the time at which there is a significant change of the channel state. The system design should be done such that the total precoding delay Δt as defined in Section 3.2.2.1, remains much smaller than T_{50} , i.e. $\Delta t \ll T_{50}$. The 50% coherence time T_{50} can be approximated by the rule of thumb [42]

$$T_{50} = \frac{0.423}{f_d} \quad (3-59)$$

where the maximum Doppler shift f_d is given by

$$f_d = f_c \frac{v}{c_0} \quad (3-60)$$

Here, f_c is the carrier frequency, $c_0 = 300000$ km/s is the speed of light and v is the speed of the mobile in km/s. Using typical parameters of LTE mobile communications as $f_c = 2$ GHz and pedestrian speed of 3 km/h, we obtain

$$T_{50} \approx 80 \text{ ms} \quad (3-61)$$

while for the vehicular speed 50 km/h, it is reduced to

$$T_{50} \approx 5 \text{ ms} \quad (3-62)$$

In both cases, achieving a round trip lower than the latency requirements seems to be possible. Although, depending on the implementation of the precoder and the network topology, it might be challenging and requires a special care when designing the overall system. Indeed, in heterogeneous networks with low backhaul capability, the delay due to the transmitter cooperation could degrade strongly the performance if not properly taken into account in the system design.

Quantization of exchanged payload signals

When performing joint detection at the RANaaS for UL transmission, the signals received from the iSCs need to be forwarded via the backhaul. Consequently, quantization is required not only for the CSI but also for payload.

For centralized linear precoding in the DL, either the resulting precoding matrix is forwarded to the iSCs individually from the actual user data or filtering is already performed at the RANaaS. The first case requires quantization of the precoding matrix, while in the latter case the resulting precoded transmit signals need to be quantized. Evaluations of the required overhead for quantizing precoded data are provided for example in [101]. Furthermore, the backhaul load can be reduced by limiting the number of iSCs involved in joint precoding as discussed in [102].

Due to the different nature of signals transmitted via the backhaul for different functional splits, the required quantization resolution can differ. For example, CPRI (split A.1) requires up to 15 bits per dimension, while soft-bits (split B.2) are commonly quantized with three bits. In general the resolution needs to be high enough that quantization noise does not become the limiting factor of the transmission. Some indication about the required number of quantization bits is given in [33].

In general, the optimal design of quantizers is a widely discussed research topic. Beginning with the classical papers of Max [107] and Lloyd [108], many methods to reduce the number of required quantization bits while keeping the quantization noise to a minimum have been investigated. Compression techniques that

utilize such an optimized quantizer design and exploit the properties of OFDM signals in addition are described in [105] and [106]. It is shown, that the required number of quantization bits can be reduced by a factor of 1/2 to 1/3 compared to CPRI. Compression schemes for joint detection have also been investigated in [103] and evaluated via field trail measurements in [104]. In [19], the correlation between the signals at the different iSCs is optimized to maximize the sum-rate. It is also shown how this approach can be made robust to an imperfect knowledge of the channel statistics. In [21], the precoding is optimized jointly with the quantization error. In [20], it is studied how to optimize the allocation of the backhaul links so as to maximize the sum rate in the uplink.

Synchronization of iSCs

Joint transmission among multiple cells can provide significantly large throughput gains at the users. Nevertheless, [109] points out the price to pay to achieve an efficient CoMP transmission with OFDM waveform. One of these identified drawbacks is that the multiple transmitters must be well synchronized to prevent mutual interference. There are different local oscillators in each base station and mobile terminal that lead to deviations in the carrier frequency according to its nominal value. Hence, the orthogonality between subcarriers collapses at the receiver and the performance is degraded. In addition, the beamforming gain will be affected due to the mismatch between the precoder and the actual channel. To avoid a significant performance loss, in the current LTE/LTE-A systems (without cooperation), the frequency accuracy requirements are set to ± 50 ppb (parts per billion), ± 100 ppb, ± 250 ppb for wire area, local area and home scenarios, respectively [111], [112]. Determining the required frequency accuracy is an active research area as the performance loss due to the carrier frequency offset (CFO) for CoMP depends in a complicated manner of the number of antennas, oscillators and UEs and larger values of those parameters make the system more sensitive to CFO [116], [117].

Another source of imperfection comes from the variations in the symbol timing between each transmitter and receiver station, which may lead to inter symbol interference (ISI) and performance reduction. The time accuracy requirements in the current LTE/LTE-A for CoMP transmission vary as a function of the specific scheme. For example joint transmission/processing and dynamic point blanking (DPB) require the distributed base stations to have time accuracy as low as $1.5 \mu\text{s}$. In contrast, coordinated scheduling and DL non-coherent transmission can relax the requirements to $5 \mu\text{s}$ [111], [112].

In general, several different synchronization approaches can be used: satellite-based synchronization using global position system (GPS), network-based synchronization and over the air (OTA)-based synchronization [109]. The satellite-based synchronization may provide higher accurate frequency and time synchronization for multiple iSCs. It relies however on the line of sight (LOS) link between the iSCs and the satellites, thus limiting its use in the indoor small cell deployments where high penetration losses may hinder the reception of the weak GPS signals. In addition, the need for GPS or global navigation satellite system (GNSS) modules in small cells for the satellite-based synchronization may increase the iSC deployment cost.

An alternative solution is the precision time protocol (PTP), also known as IEEE 1588 [113]. It is one of the most widely used network-based synchronization method and can attain up to sub-microsecond accuracy when implemented in hardware. By moving the clock synchronization as close to the physical layer as possible, sources of jitter and processing delay introduced in higher layers can be mitigated. In the iJOIN architecture, the reference clock and oscillator could be located at the RANaaS data centre to which all of the cooperative iSCs are connected by a J1 link. Products already available on the market claim clock offsets within $1 \mu\text{s}$ and frequency offsets better than ± 10 ppb [114].

Different to the satellite-based or network-based synchronization approaches, the OTA-based synchronization approaches are based on the principle of exchanging beacons between the small cells and synchronizing carriers based on estimates of the phase and frequency of these beacons. [115] proposed a so-called two-way synchronization protocol with which a clock offset standard deviation better than 100 ps and a frequency offset standard deviation better than 0.4 ppb can be obtained with moderate SINR and re-synchronization interval.

The different synchronization approaches presented have been designed so as to answer the stringent needs for synchronization in OFDM systems. Another approach to reduce the synchronization requirements currently discussed for 5G mobile communication systems is the modification of the spectral shapes used for modulation. In particular, FBMC (filter bank multi carrier) waveform with its spectrally well shaped prototype filters and overlapped time symbols has some inherent features which make it a natural choice for

some of the anticipated 5G application scenarios. First of all, it does not require a cyclic prefix and has an almost perfect separation of frequency subbands without the need for strict synchronization. Consequently its properties make it especially suited for fragmented spectrum and CoMP transmission and reception. In addition, it is shown in [110] that both the CFO estimation and its compensation can be realized at the UE side, in the frequency domain. These operations require no feedback from the UE to the BSs. Reference [110] also demonstrated that FBMC allows an easier and more accurate estimation and compensation of the CFO than OFDM. Thus, the synchronization constraints currently imposed by OFDM in LTE-A could be relaxed in future 5G systems.

Another issue is the necessary feedback for time synchronization at the UEs. In the case of large cooperating areas, the size of the cyclic prefix can reach high values, leading to high spectral efficiency loss for OFDM systems. In [110], FBMC modulation was demonstrated to be very resistant to time propagation differences between signals from two base stations, due to its overlapping structure. Hence, delays up to 120 samples, i.e. 7.8 μ s (2340 meters), can be tolerated at the UE without any estimation or correction, thus reducing the cost of the feedback. Again, this shows an alternative solution when implementing FBMC in 5G.

4 iJOIN Physical Layer Candidate Technologies

In this chapter, promising PHY CTs for a dense deployment of small cells and a joint optimization of access and backhaul networks are described. The basic concept of each CT is introduced together with updates from D2.1 [36] and future studies for the next iJOIN phase. For each CT the “evaluation method” that indicates how the CT will be evaluated is also defined.

Cooperative multi-user detection techniques for the uplink are investigated by CT2.1 “In-Network Processing” and CT2.2 “Multi-Point Turbo Detection”. The joint optimization of the FEC codes for the access links and the backhaul link applying network coding is considered by CT2.3’s “Joint Network-Channel Coding”. For the CoMP transmission CT2.4 “Sum-Rate and Energy-Efficiency metrics of DL CoMP with backhaul constraints” and CT2.5 “Partially Centralized Inter-Cell Interference Coordination” analyse different precoder and interference coordination techniques with backhaul constraints. Techniques to reduce the backhaul throughput for radio over fibre (RoF) links are the topics of CT2.6 “Data Compression over RoF”, whereas CT2.7 “Millimetre Wave Backhauling” addresses wireless backhaul links and joint access and backhaul FEC. The list of iJOIN CTs is provided in Table 4-1.

Table 4-1: List of iJOIN PHY Candidate Technologies (CTs)

CT	Topic	Abbreviation
2.1	In-Network-Processing	INP
2.2	Multi-Point Turbo Detection	MPTD
2.3	Joint Network-Channel Coding	JNCC
2.4	Sum-Rate and Energy-Efficiency Metrics of DL COMP with backhaul constraints	CoMP
2.5	Partially Centralized Inter-Cell Interference Coordination	ICIC
2.6	Data Compression over RoF	RoF
2.7	Millimetre wave backhauling	mmWave

4.1 CT2.1: In Network Processing

4.1.1 Technical description

Scenario

In this CT, joint multi-user detection (MUD) in a very dense deployment of iSCs via In-Network-Processing (INP) is investigated and corresponding results were published in [4]. The MUD shall be implemented in a distributed fashion by an exchange of information among multiple iSCs observing several users in their coverage area. This can be accomplished by the so-called Distributed Consensus-Based Estimation (DiCE) algorithm [2] which was described in detail in D2.1 [36] and which was analysed in [3] for erroneous backhaul links. Depending on the interference scenario, this distributed detection will achieve improved performance compared to local detection per iSC. Centralized detection, where all receive signals are forwarded to the central processing node, acts as a benchmark for distributed procedures as all observations are incorporated into the detection process. The distributed detection schemes aim to achieve the central solution iteratively, but limiting the main backhaul traffic to the area of the iSCs. Thus, the required rate on the backhaul link to the RANaaS is reduced significantly.

In the following, the system model together with a brief description of the DiCE algorithm and the noise-free exchange DiCE (nfX-DiCE) algorithm are introduced. The nfX-DiCE has the same algorithmic procedure as the DiCE but it only suited for perfect backhaul links among the iSCs. Subsequently, two recently developed variants of the DiCE algorithm are presented, namely the Reduced Overhead DiCE (RO-DiCE) [5] and the Fast-DiCE algorithm [38]. These algorithms are modified versions of the DiCE algorithm for reducing the communication overhead among iSCs and improving the convergence speed of the algorithm, respectively. Additionally, the Priority-based Augmented Lagrangian Cooperative Estimation (PALCE) algorithm is described as another approach for distributed MUD. Algorithmic descriptions regarding the DiCE algorithms are corroborated by simulation results using an LTE uplink transmission chain based on Release 10

regarding the bit error rate (BER), frame error rate (FER) and throughput performance. In order to integrate the DiCE algorithms into the LTE uplink transmission chain the common equalizer is substituted by the different DiCE algorithms in each iSC.

System Model

Figure 4-1 depicts an exemplary multi-user scenario with $N_{UE} = 2$ UEs and $N_{iSC} = 4$ iSCs. Each user u transmits a vector of complex symbols $\mathbf{x}^{(u)}$ on each subcarrier by spatial multiplexing using N_T^{UE} transmit antennas to all N_{iSC} iSCs. The symbol vectors $\mathbf{x}^{(u)}$ of all UEs can be stacked into a vector \mathbf{x} because users are assumed to use the same physical resource blocks (PRBs):

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \vdots \\ \mathbf{x}^{(N_{UE})} \end{bmatrix}. \quad (4-1)$$

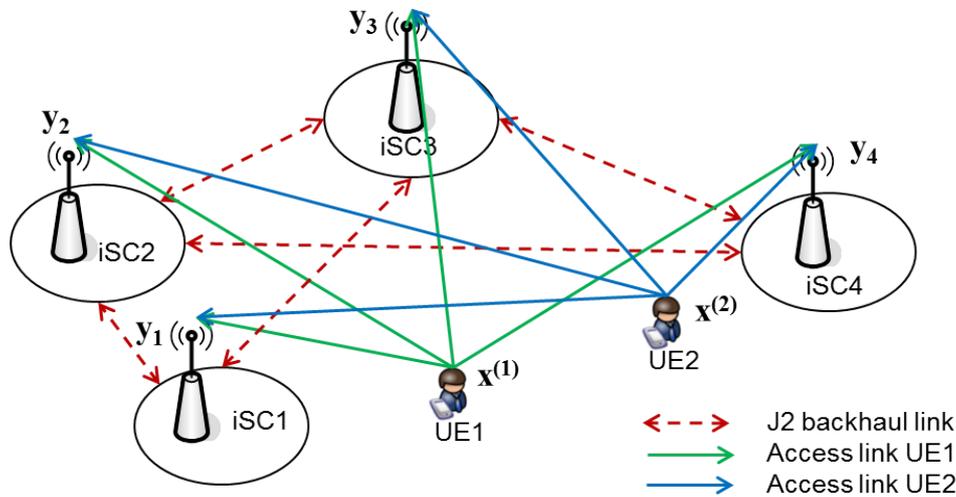


Figure 4-1: Multi-user scenario in a dense deployment of iSCs

Then the stacked symbol vector \mathbf{x} has dimension $N_I \times 1$ with $N_I = N_{UE} N_T^{UE}$ denoting the total number of input signals. Each iSC is equipped with N_R^{iSC} receive antennas resulting in the definition of the total number of output signals $N_O = N_{iSC} N_R^{iSC}$. The transmission from all UEs to a single iSC j is stacked into the symbol vector \mathbf{x} which is distorted by individual complex-valued MIMO channel matrices \mathbf{H}_j of dimension $N_R^{iSC} \times N_I$ and by complex white Gaussian noise \mathbf{n}_j of dimension $N_R^{iSC} \times 1$:

$$\mathbf{y}_j = \mathbf{H}_j \mathbf{x} + \mathbf{n}_j. \quad (4-2)$$

Based on these observations, local, centralized and distributed detection schemes can be applied. In order to illustrate the basic ideas of the following algorithms we consider a transmission on one subcarrier only. For numerical simulations based on LTE the algorithms can be extended appropriately for multiple subcarriers.

Approach

DiCE Algorithm

In the following, we briefly recall the Distributed Consensus-Based Estimation (DiCE) algorithm from [2] as a distributed approach for multi-user detection. A detailed description of the algorithm is contained in the deliverable D2.1 [36]. The following equations constitute the DiCE algorithm for each iSC j :

$$\mathbf{z}_j(k) = \frac{\mu}{|\mathcal{N}_j^+|} \sum_{i \in \mathcal{N}_j^+} \left(\frac{\mathbf{x}_i(k-1)}{\mu} - \boldsymbol{\lambda}_{ij}(k-1) \right) \quad (4-3)$$

$$\boldsymbol{\lambda}_{ji}(k) = \boldsymbol{\lambda}_{ji}(k-1) - \frac{1}{\mu} (\mathbf{x}_j(k) - \mathbf{z}_i(k)) \quad \forall i \in \mathcal{N}_j^+ \quad (4-4)$$

$$\mathbf{x}_j(k) = \left(\mathbf{H}_j^H \mathbf{H}_j + \frac{|\mathcal{N}_j^+|}{\mu} \mathbf{I}_{N_j} \right)^{-1} \cdot \left[\mathbf{H}_j^H \mathbf{y}_j + \sum_{i \in \mathcal{N}_j^+} \left(\boldsymbol{\lambda}_{ji}(k) + \frac{\mathbf{z}_i(k)}{\mu} \right) \right]. \quad (4-5)$$

The set $\mathcal{N}_j^+ = \mathcal{N}_j \cup \{j\}$ denotes the augmented neighbourhood of iSC j which contains all neighbouring iSCs and the iSC j itself. Each iSC j transmits its auxiliary variable $\mathbf{z}_j(k)$, its Lagrange multipliers $\boldsymbol{\lambda}_{ji}(k)$ and its local estimate $\mathbf{x}_j(k)$ per iteration to its neighbouring iSCs over the J2 link. Note that an exchange of the Lagrange multipliers $\boldsymbol{\lambda}_{ji}(k)$ has to be done in a unicast fashion since it differs depending on the transmitting iSC j and receiving iSC i . Consequently, each iSC j receives $\mathbf{z}_i(k)$, $\boldsymbol{\lambda}_{ij}(k)$ and $\mathbf{x}_i(k)$ from its neighbours $i \in \mathcal{N}_j$.

Noise-free exchange DiCE (nfX-DiCE) Algorithm

In case of noise-free backhaul links (e.g. ideal backhaul connections) the Lagrange multipliers $\boldsymbol{\lambda}_{ji}(k)$ and $\boldsymbol{\lambda}_{ij}(k)$ can be calculated locally at iSC j . This is due to the assumption that the local variables $\mathbf{x}_j(k)$ and $\mathbf{z}_j(k)$ are not disturbed by noise during a transmission to neighbouring iSCs and thus, the Lagrange multipliers will be the same independent of which iSC they are calculated at. Therefore, an exchange of Lagrange multipliers is not necessary anymore resulting in the so-called noise-free exchange DiCE (nfX-DiCE). Compared to the DiCE algorithm, only the calculation of the multipliers in (4-4) is split into two calculations:

$$\boldsymbol{\lambda}_{ij}(k) = \boldsymbol{\lambda}_{ij}(k-1) - \frac{1}{\mu} (\mathbf{x}_i(k-1) - \mathbf{z}_j(k)), \quad \forall i \in \mathcal{N}_j^+ \quad (4-6)$$

$$\boldsymbol{\lambda}_{ji}(k) = \boldsymbol{\lambda}_{ji}(k-1) - \frac{1}{\mu} (\mathbf{x}_j(k-1) - \mathbf{z}_i(k)), \quad \forall i \in \mathcal{N}_j^+ \quad (4-7)$$

The overall communication overhead can thus be significantly reduced by avoiding an exchange of the multipliers. Note that the nfX-DiCE is only applicable for noise-free backhaul links.

Reduced Overhead DiCE (RO-DiCE) Algorithm

It could be seen that for the DiCE algorithm an additional transmission of the Lagrange multipliers $\boldsymbol{\lambda}_{ji}$ for each iSC j is necessary. This transmission can be very costly e.g. in a wireless transmission context because of its unicast nature. This means that each iSC needs to transmit the multiplier $\boldsymbol{\lambda}_{ji}$ separately to each of its neighbours which obviously increases the overall communication overhead among the iSCs significantly. The RO-DiCE algorithm avoids such an exchange of multipliers by using locally available multipliers only. The basic idea is to approximate the average of multipliers by the local multipliers $\boldsymbol{\lambda}_{jj}$. This results in the following approximations of the Lagrange multipliers:

$$\frac{1}{|\mathcal{N}_j^+|} \sum_{i \in \mathcal{N}_j^+} \boldsymbol{\lambda}_{ij}(k) \approx \boldsymbol{\lambda}_{jj}(k) \approx \frac{1}{|\mathcal{N}_j^+|} \sum_{i \in \mathcal{N}_j^+} \boldsymbol{\lambda}_{ji}(k) \quad (4-8)$$

By this approximation an exchange of multipliers can be avoided and the DiCE algorithm is modified as follows:

$$\mathbf{z}_j(k) = \frac{1}{|\mathcal{N}_j^+|} \left(\sum_{i \in \mathcal{N}_j^+} \mathbf{x}_i(k-1) \right) - \mu \cdot \boldsymbol{\lambda}_{jj}^j(k-1) \quad (4-9)$$

$$\boldsymbol{\lambda}_{jj}(k) = \boldsymbol{\lambda}_{jj}(k-1) - \frac{1}{\mu} (\mathbf{x}_j(k) - \mathbf{z}_j(k)) \quad (4-10)$$

$$\mathbf{x}_j(k) = \left(\mathbf{H}_j^H \mathbf{H}_j + \frac{|\mathcal{N}_j^+|}{\mu} \mathbf{I}_{N_j} \right)^{-1} \cdot \left[\mathbf{H}_j^H \mathbf{y}_j + |\mathcal{N}_j^+| \cdot \boldsymbol{\lambda}_{jj}(k) + \frac{1}{\mu} \sum_{i \in \mathcal{N}_j^+} \mathbf{z}_i(k) \right]. \quad (4-11)$$

As can be seen the RO-DiCE algorithm uses locally available Lagrange multipliers only. Nevertheless, by applying (4-8) a systematic error is introduced into the estimation process which will decrease the overall estimation performance of the algorithm.

Fast-DiCE Algorithm

According to the DiCE algorithm each iSC j needs to transmit the local estimates to its neighbouring iSCs in each iteration. Thus, the number of iterations required for an acceptable estimation performance will lead to a considerable communication overhead over the J2 link causing latency for the processing at the iSCs. In order to accelerate the convergence speed, and thus reduce the communication overhead, we proposed a novel algorithm called Fast-DiCE [38] as a modification of the DiCE algorithm by adopting the optimal gradient descent method [40] and the accelerated alternating direction method of multipliers (ADMM) approach [41]. The main idea of the Fast-DiCE is to optimize the update function of the local estimates with two prediction-type variables taking roles of ‘predictors’ for certain estimates. Both predictors are calculated with respect to the estimates from the two latest consecutive iteration steps. Then each new estimate will be extended to a promising value in a direction indicated by the difference between these two consecutive estimates. Those newly predicted estimates are then applied to the next update. Here, we consider a non-ideal backhaul link and the respective update equations for the Fast-DiCE algorithm are given by

$$\mathbf{z}_j(k) = \frac{\mu}{|\mathcal{N}_j^+|} \sum_{i \in \mathcal{N}_j^+} \left(\frac{\mathbf{x}_i(k-1)}{\mu} - \tilde{\boldsymbol{\lambda}}_{ji}(k-1) \right) \quad (4-12)$$

$$\boldsymbol{\lambda}_{ji}(k) = \tilde{\boldsymbol{\lambda}}_{ji}(k-1) - \frac{1}{\mu} (\mathbf{x}_j(k) - \mathbf{z}_i(k)), \quad \forall i \in \mathcal{N}_j^+ \quad (4-13)$$

$$\tilde{\mathbf{z}}_{ji}(k) = \mathbf{z}_i(k) + \gamma(k) (\mathbf{z}_i(k) - \mathbf{z}_i(k-1)), \quad \forall i \in \mathcal{N}_j^+ \quad (4-14)$$

$$\tilde{\boldsymbol{\lambda}}_{ji}(k) = \boldsymbol{\lambda}_{ji}(k) + \gamma(k) (\boldsymbol{\lambda}_{ji}(k) - \boldsymbol{\lambda}_{ji}(k-1)), \quad \forall i \in \mathcal{N}_j^+ \quad (4-15)$$

$$\mathbf{x}_j(k) = \left(\mathbf{H}_j^H \mathbf{H}_j + \frac{|\mathcal{N}_j^+|}{\mu} \mathbf{I}_N \right)^{-1} \cdot \left[\mathbf{H}_j^H \mathbf{y}_j + \sum_{i \in \mathcal{N}_j^+} \left(\tilde{\boldsymbol{\lambda}}_{ji}(k) + \frac{\tilde{\mathbf{z}}_{ji}(k)}{\mu} \right) \right]. \quad (4-16)$$

The predictors are initialized as $\tilde{\mathbf{z}}_{ji}(0) = \tilde{\boldsymbol{\lambda}}_{ji}(0) = 0$. The step size parameter is given by

$$\gamma(k) = \frac{\alpha(k-1) - 1}{\alpha(k)} \quad \text{with} \quad \alpha(k) = \frac{1 + \sqrt{1 + 4\alpha^2(k-1)}}{2} \quad (4-17)$$

with $\alpha(0) = 1$. Similar to the DiCE algorithm, variables $\mathbf{z}_j(k)$, $\boldsymbol{\lambda}_{ji}(k)$ and $\mathbf{x}_j(k)$ are exchanged among the iSCs. Additionally, at each iSC j a local calculation of the predictors $\tilde{\mathbf{z}}_{ji}(k)$ and $\tilde{\boldsymbol{\lambda}}_{ji}(k)$ is performed. These predictors are not exchanged among neighbouring iSCs so that the total communication overhead per iteration remains the same for Fast-DiCE and DiCE. Nevertheless, the Fast-DiCE algorithm requires less iterations to achieve the same estimation quality in comparison to the DiCE algorithm, therefore, leading to a reduction of communication effort over the J2 backhaul link and a lower latency.

Priority-based Augmented Lagrangian Cooperative Estimation (PALCE) Algorithm

A new algorithm named PALCE has also been proposed for the distributed MUD, which is based on the distributed Augmented Lagrangian Multipliers (ALM) method employing a priority-based processing scheme. Different from the DiCE algorithm and its variants, PALCE adopts a new approach to decouple the network consensus constraint $\mathbf{x}_j = \mathbf{x}_i$ [2] for parallel processing among all iSCs. Instead of using additional auxiliary variable \mathbf{z}_j , the constraint is decoupled into two parts based on the priority principle (e.g., iSC j has higher priority than iSC i if $j < i$): the instant estimate $\mathbf{x}_j(k)$ of iSC j intends to gather the estimate $\mathbf{x}_i(k-1)$ of neighbouring iSC i with lower priority; and the instant estimate $\mathbf{x}_i(k)$ of neighbouring iSC i with higher priority has to collect the estimate $\mathbf{x}_j(k-1)$ of iSC j . Then following the similar derivation of the DiCE algorithm, the update equations for PALCE can be obtained:

$$\mathbf{x}_j(k) = \left(\mathbf{H}_j^H \mathbf{H}_j + \frac{|\mathcal{N}_{j^+}| + |\mathcal{N}_{j^-}|}{\mu} \mathbf{I}_N \right)^{-1} \cdot \left[\mathbf{H}_j^H \mathbf{y}_j + \sum_{i \in \mathcal{N}_{j^+}} \left(\frac{\mathbf{x}_i(k-1)}{\mu} + \boldsymbol{\lambda}_{ji}(k-1) \right) + \sum_{i \in \mathcal{N}_{j^-}} \left(\frac{\mathbf{x}_i(k-1)}{\mu} - \boldsymbol{\lambda}_{ij}(k-1) \right) \right] \quad (4-18)$$

$$\boldsymbol{\lambda}_{ji}(k) = \boldsymbol{\lambda}_{ji}(k-1) - \frac{1}{\mu} (\mathbf{x}_j(k) - \mathbf{x}_i(k-1)), \quad \forall i \in \mathcal{N}_{j^+} \quad (4-19)$$

with the set \mathcal{N}_{j^+} denoting the neighbouring iSCs i of iSC j satisfying $i > j$ and set \mathcal{N}_{j^-} denoting the neighbouring iSCs i of iSC j when $i < j$. Like other distributed algorithms, during the iterative processing of PALCE, each iSC j updates the estimate $\mathbf{x}_j(k)$ based on the received estimates $\mathbf{x}_i(k-1)$ from neighbouring iSCs $i \in \mathcal{N}_j$ over the J2 backhaul links. However, the multipliers $\boldsymbol{\lambda}_{ij}(k-1)$ are sent from those iSCs with higher priority to lower priority iSCs, i.e., only the multipliers $\boldsymbol{\lambda}_{ji}(k)$ for $j < i$ need to be updated at iSC j and to be delivered to its neighbouring iSCs $i \in \mathcal{N}_{j^+}$. Therefore PALCE shows a great advantage of saving communication overhead compared to DiCE, since there is no exchange of auxiliary variables among iSCs, and only part of all iSCs need to send the multipliers to the neighbouring iSCs, which can achieve a considerable reduction in overhead and latency.

Table 4-2 summarizes the algorithmic steps of the presented variants of the DiCE algorithm and the PALCE algorithm within one iteration.

Table 4-2: Overview of algorithmic steps of presented DiCE variants and PALCE per iteration

Steps	DiCE	nfX-DiCE	RO-DiCE	Fast-DiCE	PALCE
Calculation	$\mathbf{z}_j = f_z(\mathbf{x}_i, \boldsymbol{\lambda}_{ij})$	$\mathbf{z}_j = f_z(\mathbf{x}_i, \boldsymbol{\lambda}_{ij})$	$\mathbf{z}_j = f_z(\mathbf{x}_i, \boldsymbol{\lambda}_{ij})$	$\mathbf{z}_j = f_z(\mathbf{x}_i, \tilde{\boldsymbol{\lambda}}_{ij})$	$\mathbf{x}_j = f_x(\boldsymbol{\lambda}_{ji}, \mathbf{x}_i)$
Exchange	\mathbf{z}_j	\mathbf{z}_j	\mathbf{z}_j	\mathbf{z}_j	\mathbf{x}_j
Calculation	$\boldsymbol{\lambda}_{ji} = f_\lambda(\boldsymbol{\lambda}_{ji}, \mathbf{x}_j, \mathbf{z}_i)$	$\boldsymbol{\lambda}_{ji} = f_\lambda(\boldsymbol{\lambda}_{ji}, \mathbf{x}_j, \mathbf{z}_i)$ $\boldsymbol{\lambda}_{ij} = f_\lambda(\boldsymbol{\lambda}_{ij}, \mathbf{x}_i, \mathbf{z}_j)$	$\boldsymbol{\lambda}_{jj} = f_\lambda(\boldsymbol{\lambda}_{jj}, \mathbf{x}_j, \mathbf{z}_j)$	$\tilde{\boldsymbol{\lambda}}_{ji} = f_{\tilde{\lambda}}(\mathbf{z}_i)$ $\boldsymbol{\lambda}_{ji} = f_\lambda(\tilde{\boldsymbol{\lambda}}_{ji}, \mathbf{x}_j, \mathbf{z}_i)$	$\boldsymbol{\lambda}_{ji} = f_\lambda(\boldsymbol{\lambda}_{ji}, \mathbf{x}_i)$
Exchange	$\boldsymbol{\lambda}_{ji}$	-	-	$\boldsymbol{\lambda}_{ji}$	$\boldsymbol{\lambda}_{ji}$
Calculation	$\mathbf{x}_j = f_x(\boldsymbol{\lambda}_{ji}, \mathbf{z}_i)$	$\mathbf{x}_j = f_x(\boldsymbol{\lambda}_{ji}, \mathbf{z}_i)$	$\mathbf{x}_j = f_x(\boldsymbol{\lambda}_{ji}, \mathbf{z}_i)$	$\tilde{\boldsymbol{\lambda}}_{ji} = f_{\tilde{\lambda}}(\boldsymbol{\lambda}_{ji})$ $\mathbf{x}_j = f_x(\tilde{\boldsymbol{\lambda}}_{ji}, \mathbf{z}_i)$	-
Exchange	\mathbf{x}_j	\mathbf{x}_j	\mathbf{x}_j	\mathbf{x}_j	-

4.1.2 Implementation in the iJOIN architecture

In order to implement the INP approach in the LTE environment, the common equalizer is substituted by one of the DiCE algorithms for the detection of UE messages located within the same PRBs. Thus, the DiCE algorithm is applied to PRBs which are shared by multiple UEs. Those iSCs which do not receive relevant data for a joint detection of the UE messages are not considered for a cooperation based on the INP approach. For orthogonally scheduled UE messages the receiving iSCs can perform local detection and decoding of the UE messages. Then, only the iSC with a successful cyclic redundancy check (CRC) forwards its data to the RANaaS. Alternatively, the receiving iSCs can also cooperate to detect the common received UE messages by using the DiCE algorithm.

Following two system functional splits, using the DiCE algorithm the iSCs detect non-orthogonally scheduled UE data cooperatively. This causes rather high data traffic on the J2 links due the iterative exchange of messages between iSCs. However compared to the J2 links, lower data traffic is expected on the long distance backhaul between the iSCs and the RANaaS, i.e. the J1 link, since only one iSC needs to forward its estimates to the RANaaS.

Functional split B.1: Joint detection among iSCs, decoding in RANaaS

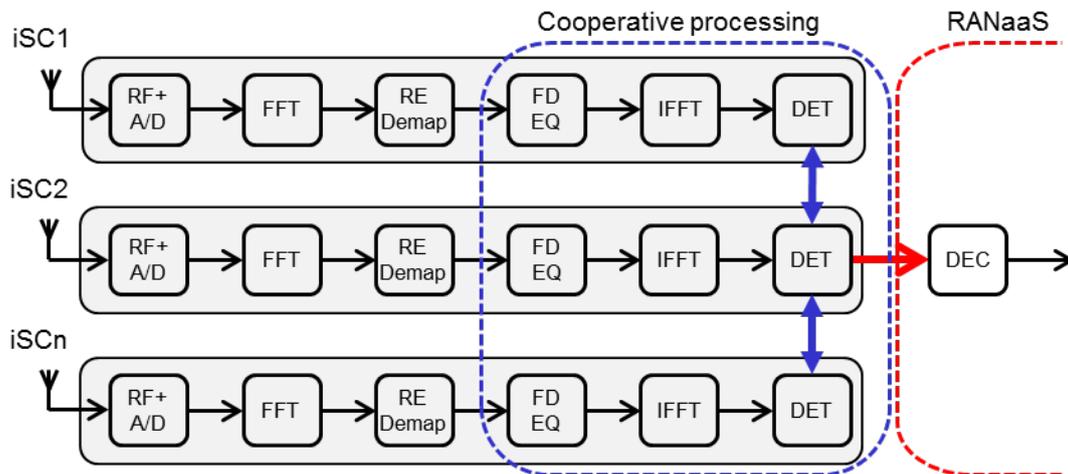


Figure 4-2: Functional split B.1 for CT2.1

In this variant, the iSCs perform joint detection of the UE UL signals using the DiCE algorithm as fundamental approach, and the estimated soft symbols are then forwarded to the RANaaS over a J1 link by one iSC (in general, the one with the momentarily most suitable backhaul connection). Equivalently, we may also forward quantized LLRs corresponding to split option B.2, where the number of quantization bits is roughly independent of the modulation order. Subsequently, we discuss the J1 payload for split option B.1 only and note that the corresponding J1 payloads for split B.2 is directly given by (3-32).

The J2 links which facilitate cooperation among the iSCs are located within a relatively short range such that long distance connections with high rate are not required. Vice versa, the properties of the BH links, e.g. error rate, rate and latency, but also physical and logical topology, will affect the distributed detection algorithm and its performance. The RANaaS will perform channel decoding on the soft symbols transmitted on the J1 link, again an interface within the veNB. As the decoder is implemented in the RANaaS on general purpose hardware (GP-HW) major challenges are imposed as discussed in Section 3.2.2.2. In particular, such decoder implementation has to meet timing constraints imposed by the 3GPP standard which require a low-latency implementation of the channel decoder.

According to the CT description in Section 4.1.1, the local estimate \mathbf{x}_j , the auxiliary variable \mathbf{z}_j and the multiplier λ_{ji} need to be quantized and exchanged among iSCs via J2 links within one subframe duration T_{SUB} . In the following we assume that all considered UEs are allocated on the same PRBs. As the dimension of the exchanged variables equals the dimension of the filter output signals of the system functional split B.1, the message size of the J2 payload per iSC-iSC link in each iteration for each variable is determined based on (3-31) as

$$2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (4-20)$$

For our evaluations, we assume that the number of layers N_L is equal to the number of transmit antennas N_T^{UE} per UE. In practice not all UEs will usually be completely scheduled on the same PRBs but, instead, only part of the available PRBs might be shared by several UEs. For such a scenario the DiCE algorithm is only applied to the detection of those PRBs which are shared by several UEs.

As the distributed processing of the receive signals is the same for all functional splits, the same signals are exchanged between the iSCs leading to the same J2 payload per iteration. For the symbols to be estimated within one subframe the three messages of the same size have to be exchanged for each iteration of the DiCE algorithm. Therefore, the iSC-iSC J2 payload per iteration within one subframe is given by

$$D_{P,J2} = 3 \cdot 2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (4-21)$$

Obviously, the J2 payload per iSC-iSC then scales linearly with the total number of iterations N_{It} used for the DiCE algorithm. For a sufficient number of iterations the local estimates of all iSCs converge to the same estimate. Thus, only one iSC j with a good J1 link condition needs to forward the consensus estimate \mathbf{x}_j to the RANaaS for decoding. In case a consensus is not achieved among the iSCs due to an insufficient number of iterations, several iSCs may forward their local estimates to the RANaaS in order to improve the decoding performance. The corresponding total message size of the payload on J1 interface in case only one iSC forwards its consensus estimate is given by (3-31) as

$$D_{P,J1}^{B.1} = 2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (4-22)$$

Functional split C: Joint detection among iSCs, decoding at iSC

In this functional split variant, each iSC will perform channel decoding and the one which is first able to decode correctly, verified by a CRC, forwards the decoded bits to the RANaaS. The iterative detection can be aborted as soon as one CRC is successful, which allows saving computational effort.

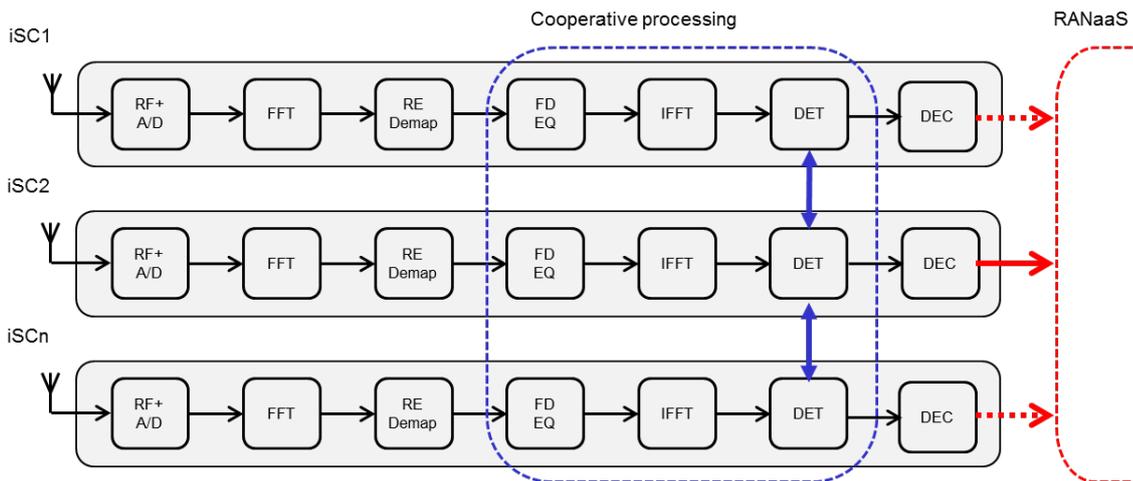


Figure 4-3: Functional split C for CT2.1

As the iterative approach for joint detection is not affected by the functional split, the variables exchanged between iSCs do not change leading to the same J2 payload (4-21) per iteration as in split B.1. However, as decoding is performed at each iSC, it is likely that less iterations are required for this split option to achieve comparable performance results. In contrast to split B.1 only the estimated information bits are delivered to the RANaaS by one iSC. The BH load on the J1 link can then be expressed by the transport block sizes $TBS_{1,u}$ and $TBS_{2,u}$ for each codeword and each user during one subframe. In accordance to (3-33) the message size of the payload on J1 is given by

$$D_{p,J1}^C = \sum_{u=1}^{N_{UE}} (TBS_{1,u} + TBS_{2,u}) \quad [\text{bit/iSC}] \quad (4-23)$$

Backhaul Load

In this CT, each iSC has to perform signal detection and demodulation using the local CSI information as well as the local SNR per UE for each functional split. As the measurements of RxCSI and SNR are not shared among iSCs, no backhaul is caused as summarized in Table 4-3.

Table 4-3: PHY Measurement placement and exchange links for CT2.1

Functional Split Approach	Measurement	Measurement Location	Exchange link
Functional split B.1	RxCSI, RxSNR per UE	iSC	-
Functional split C	RxCSI, RxSNR per UE	iSC	-

In order to implement the distributed detection approach, several control signals as defined in defined in Table 6-1 and Table 6-2 of Appendix I have to be forwarded to the iSCs via the J1 links. In addition, for functional split C the CRC check can be exchanged among iSCs via the J2 links. The respective control information has been summarized into Table 4-4.

Table 4-4: Backhaul control information of CT 2.1

Link	Description
iSC → iSC	CRC check (functional split C)
iSC → RANaaS	RRM Information <ul style="list-style-type: none"> - O2.3: Effective SINR per UE - O2.4: Effective SINR per UE-iSC link - O2.5: Estimated packet error rate (PER)
	PHY Information (access link) <ul style="list-style-type: none"> - O2.7: ACK/NACK, HARQ process - O2.8: User data to upper layers, if ACK
RANaaS → iSC	Network parameters <ul style="list-style-type: none"> - I2.1: System Parameter - I2.3: RAN connection table of iSCs and UEs - I2.4: BH connection table for iSC-iSC links and iSCs-RANaaS links
	PHY information (access and BH link) <ul style="list-style-type: none"> - I2.8: RxCSI, SNR, noise variance of UE-iSC links - I2.9: Received signal at iSC - I2.12: RxCSI, SNR, noise variance of iSC-RANaaS link - I2.13: Parameter of iSC-RANaaS - I2.17: Parameter of iSC-iSC link
	CT specific control information <ul style="list-style-type: none"> - I2.5: Functional control for shifting/splitting functions - I2.6: RRM information per UE

The information of the BH payload for different functional splits discussed above has been summarized in Table 4-5. The message frequency depends linearly on the number of iterations N_{it} within one subframe duration T_{SUB} . For a satisfying detection performance several iterations (e.g. 10) might be needed for the DiCE algorithm. The maximum latency is restricted by the ACK/NACK time of 3ms. Within this time frame the iterative information exchange and estimation by the DiCE algorithm and the decoding procedure need to be performed. Thus, approximately 1ms remains for the iterative processing by the DiCE algorithm. If we assume 10 iterations for the iterative processing the corresponding maximum latency will roughly be 0.1ms.

Table 4-5: Backhaul payload of CT 2.1

Link	Description	Msg Size	Msg Frequency	Latency
Functional split B				
iSC → iSC	Estimate \mathbf{x}_j per iSC j	$2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB}$	$N_H \cdot T_{SUB}^{-1}$	< 0.1ms
	Auxiliary variable \mathbf{z}_j per iSC j	$2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB}$	$N_H \cdot T_{SUB}^{-1}$	< 0.1ms
	Multiplier λ_{ji} for iSC j to iSC i	$2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB}$	$N_H \cdot T_{SUB}^{-1}$	< 0.1ms
iSC → RANaaS	Consensus based estimate	$2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB}$	T_{SUB}^{-1}	< 0.1ms
RANaaS → iSC	-			
Functional split C				
iSC → iSC	Estimate \mathbf{x}_j per iSC j	$2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB}$	$N_H \cdot T_{SUB}^{-1}$	< 0.1ms
	Auxiliary variable \mathbf{z}_j per iSC j	$2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB}$	$N_H \cdot T_{SUB}^{-1}$	< 0.1ms
	Multiplier λ_{ji} for iSC j to iSC i	$2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB}$	$N_H \cdot T_{SUB}^{-1}$	< 0.1ms
iSC → RANaaS	Hard decisions of information bits	$\sum_{u=1}^{N_{UE}} (TBS_{1,u} + TBS_{2,u})$	T_{SUB}^{-1}	< 0.1ms
RANaaS → iSC	-			

4.1.3 Evaluation of the CT

Compliance with iJOIN objectives

This CT performs joint MUD by cooperatively exploiting the receive signals at the iSCs. Thus, different UEs can use the same physical resources to transmit their information leading to an increased overall throughput as long as the MUD algorithm is able to estimate the UE signals correctly. Correspondingly, CT2.1 addresses the objective “area throughput”.

The distributed detection algorithm leads to low rate J1 (link between iSC and RANaaS) communication at the expense of demanding J2 (link between iSC and iSC) traffic. However, as the distances between the iSCs is relatively small compared to the distance to the RANaaS, the main backhaul traffic occurs in the restricted local area of the cooperating iSCs. Thus, the objective “utilisation efficiency” is addressed.

Description of the baseline used for the evaluation

For evaluation the baseline is an LTE Release 10 configuration, where local receive processing is executed in an eNB with varying number of receive antennas (e.g. 2 or 4 antennas) and the eNB covers the same geometrical area as the veNB. The eNB is equipped with the same number of receive antennas as the total number of transmit antennas and applies linear MMSE detection. Furthermore, a certain part of the available bandwidth can be shared by several UEs.

Discussion of results of the CT

In the following, numerical results regarding the BER, FER and throughput of the various DiCE algorithms integrated into an LTE uplink chain will be presented. For the evaluations, split option C is implemented. The basic system parameters as the number of UEs and iSCs are chosen according to the indoor small cell deployment defined in Table 6-7. The total bandwidth is chosen to 1.4 MHz in order to illustrate initial performance gains by this CT. The different DiCE algorithms are performed on each subcarrier in the frequency domain. For the symbol demapping noise power estimation based on QPSK symbols is currently used. For all evaluations coded systems are considered using the standard LTE turbo decoder with 8 iterations. We assume a fully meshed iSC network, i.e. each iSC is connected to all other iSCs over the J2 link, and perfect channel knowledge at each iSC. As channel model between UEs and iSCs a flat Rayleigh-fading channel which is constant over one subframe is used. Currently, no geometrical model for the path losses between UEs and iSCs is included so that each UE has the same average channel gain but a different Rayleigh variable. Furthermore, all UEs are allocated on the same PRBs with a bandwidth of 1.4 MHz and use the MCS 7 defined in the LTE standards. The effective code rate will then be $R_c = 712/1584 \approx 0.45$ for a TBS of 712 and a codeword size of 1584. The antenna configuration between each UE and iSC will be a 2x2 MIMO setup with spatial multiplexing at each UE.

Error rate performance and throughput

Figure 4-4 shows the error rate performance versus SNR of an LTE uplink transmission for a system with $N_{iSC} = 6$ iSCs in a full mesh configuration, $N_R^{iSC} = 2$ receive antennas per iSC, $N_{UE} = 4$ UEs and $N_T^{UE} = 2$ transmit antennas per UE. The BER and the FER are averaged over all iSCs and UEs. However, depending on the number of iterations the detection performances of the iSCs can be different. In this case, the iSC with the best error rate performance reflects the overall performance of the network in contrast to the averaged error rate depicted here. Both figures show the performances for the baseline system, for the central processing and for the distributed MMSE equalization using the DiCE, the RO-DiCE and the Fast-DiCE algorithm. For the baseline system, one eNB with $N_R = 4$ receive antennas and $N_{UE} = 4$ UEs each with $N_T^{UE} = 2$ transmit antennas are assumed where the UEs share 50% of the total bandwidth of 1.4 MHz. The performance of the DiCE variants is shown after 10 iterations. Obviously, MUD in the baseline system does not achieve a satisfying performance. Comparing centralized and distributed processing it can be observed, that in the low SNR region the DiCE algorithms achieve a performance close to the centralized MMSE equalization which serves as an upper performance bound. However, in the high SNR region an error floor can be observed for all DiCE variants since they are stopped after 10 iterations. Furthermore, the Fast-DiCE shows the best performance among all DiCE variants due to its accelerated convergence. The RO-DiCE algorithm performs slightly worse compared to the DiCE algorithm since no multipliers are exchanged and only local multipliers are used. In general, a better performance for the DiCE variants can be achieved by increasing the number of iterations reducing the error floors.

Figure 4-5 shows the corresponding average throughput in Mbps per UE for the baseline system, for the central and for distributed processing. Again, it can be seen that detection by the baseline system is not satisfactory. The maximum achievable throughput is 0.7 Mbps which is approximately half of the peak data rate of the DiCE algorithms. The DiCE algorithms show a throughput performance close to the central processing. In particular, the Fast-DiCE algorithm achieves the same peak data rate as the central detector. The peak data rates of the DiCE algorithm approximately differs by only 0.1Mbps compared to the central detector. It should be mentioned, that a better performance for the distributed processing based on the DiCE algorithm can be expected if the exact filter matrix is used for symbol demodulation. As described before, currently an estimation of the noise power is used at the iSCs in order to perform a soft demodulation of the symbols. The integration of the filter matrix into the LTE simulation chain is currently under investigation.

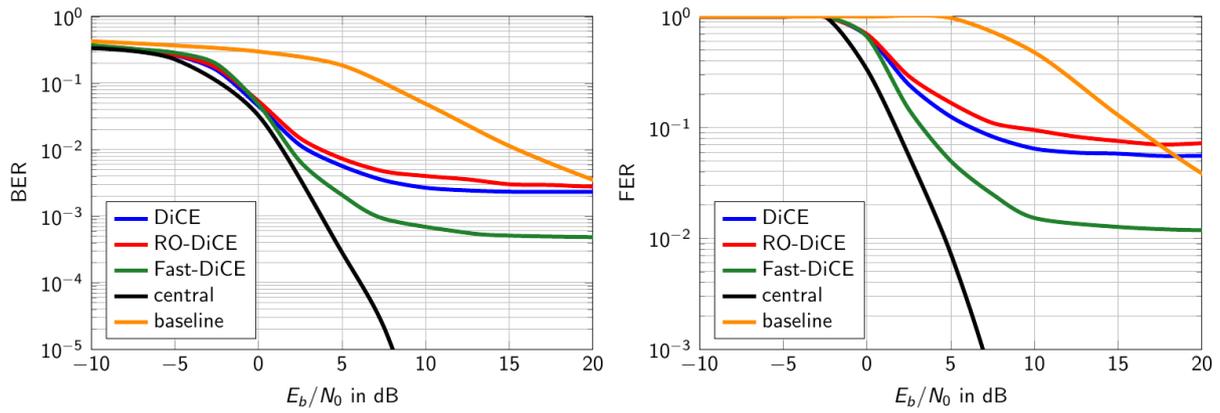


Figure 4-4: BER and FER of an LTE uplink transmission with MCS 7 and a bandwidth of 1.4 MHz

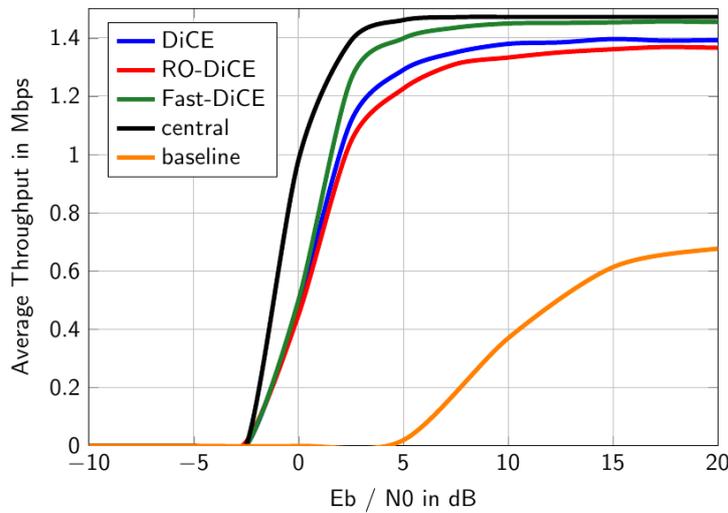


Figure 4-5: Throughput of an LTE uplink transmission with MCS 7 and a bandwidth of 1.4 MHz

Noisy backhaul connections

Figure 4-6 shows the BER and FER for ideal and noisy backhaul connections among the iSCs with an SNR of 20 dB on the backhaul connections. The noisy backhaul connections are used to model quantization noise for the information exchange among iSCs which would appear in a more realistic setup. It can be seen that for noisy backhaul connections a small loss in performance is present for the DiCE algorithm while the loss is highest for the Fast-DiCE and the RO-DiCE. In the future, the effects of a real quantization of the exchanged information among iSCs depending on the number of quantization bits will be investigated.

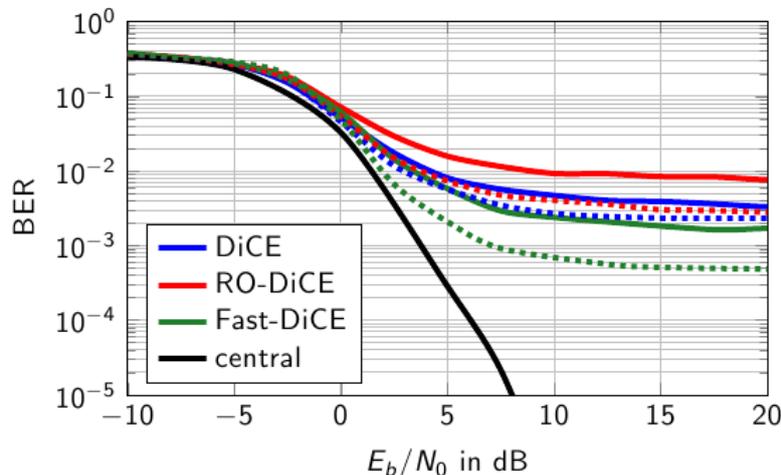


Figure 4-6: BER of an LTE uplink transmission with MCS 7 and a bandwidth of 1.4 MHz with perfect (dotted) and noisy (solid) backhaul links

Computational complexity

Table 4-6 lists the number floating point operations (FLOPS) per subcarrier of central and local linear equalization methods as well as of the DiCE, Fast-DiCE and RO-DiCE algorithms [82].

Table 4-6: Number of floating point operations per subcarrier and per iSC

Scheme	FLOPS
cZF	$\frac{2}{3}N_I^3 + \left(6N_O - \frac{1}{2}\right)N_I^2 + \left(\frac{4}{3} - \frac{3}{2}N_O\right)N_I + 4N_I N_O N_{SYMB}^{SUB}$
cMMSE	$2N_I^3 + (6N_O - 2)N_I^2 + \left(3 - \frac{3}{2}N_O\right)N_I + 4N_I N_O N_{SYMB}^{SUB}$
ZF	$\frac{2}{3}N_I^3 + \left(6N_R^{iSC} - \frac{1}{2}\right)N_I^2 + \left(\frac{4}{3} - \frac{3}{2}N_R^{iSC}\right)N_I + 4N_I N_R^{iSC} N_{SYMB}^{SUB}$
MMSE	$2N_I^3 + (6N_R^{iSC} - 2)N_I^2 + \left(3 - \frac{3}{2}N_R^{iSC}\right)N_I + 4N_I N_R^{iSC} N_{SYMB}^{SUB}$
DiCE	$\frac{4}{3}(N_R^{iSC})^3 + (6N_I + 5)(N_R^{iSC})^2 + \left(2N_I^2 - \frac{3}{2}N_I + 4N_I N_{SYMB}^{SUB} - N_{SYMB}^{SUB} - \frac{17}{6}\right)N_R^{iSC} +$ $\left(4N_{SYMB}^{SUB}N_{It} - 4N_{SYMB}^{SUB} - \frac{1}{2}\right)N_I^2 + (6 \mathcal{N}_j + 6)N_{SYMB}^{SUB}N_{It}N_I - (8 \mathcal{N}_j + 8)N_{SYMB}^{SUB}N_I + 1$
Fast-DiCE	$\frac{4}{3}(N_R^{iSC})^3 + (6N_I + 5)(N_R^{iSC})^2 + \left(2N_I^2 - \frac{3}{2}N_I + 4N_I N_{SYMB}^{SUB} - N_{SYMB}^{SUB} - \frac{17}{6}\right)N_R^{iSC} +$ $\left(4LN_{It} - 4L - \frac{1}{2}\right)N_I^2 + (8 \mathcal{N}_j + 12)N_{SYMB}^{SUB}N_{It}N_I - (12 \mathcal{N}_j + 20)N_{SYMB}^{SUB}N_I + 1$
RO-DiCE	$\frac{4}{3}(N_R^{iSC})^3 + (6N_I + 5)(N_R^{iSC})^2 + \left(2N_I^2 - \frac{3}{2}N_I + 4N_I N_{SYMB}^{SUB} - N_{SYMB}^{SUB} - \frac{17}{6}\right)N_R^{iSC} +$ $\left(4N_{SYMB}^{SUB}N_{It} - 4N_{SYMB}^{SUB} - \frac{1}{2}\right)N_I^2 + (2 \mathcal{N}_j + 6)N_{SYMB}^{SUB}N_{It}N_I - (2 \mathcal{N}_j + 7)N_{SYMB}^{SUB}N_I + 1$

The complexity of central schemes highly depends on the total number of input signals N_I , the total number of output signals N_O and the number of SC-FDMA symbols per subframe N_{SYMB}^{SUB} . For local schemes the number of output signals reduces to the number of receive antennas N_R^{iSC} per iSC. Additionally, the complexities for the DiCE schemes depend on the number of iterations N_{It} and the number of neighbours $|\mathcal{N}_j|$ per iSC.

Figure 4-7 shows the number of FLOPS for the same system which has been used for BER investigations. The complexity is depicted over the number of iterations for the various DiCE approaches. For reference purposes the complexity of local and central linear equalization schemes are depicted as well. In the first iteration the complexity of the DiCE schemes is comparable to that of the local equalization schemes. With an increasing number of iterations the complexity of the DiCE approaches grows linearly and reaches the same number of FLOPS as the central schemes after approximately 8 to 14 iterations. Compared to the DiCE algorithm the RO-DiCE has a lower complexity due to the introduced approximation. The Fast-DiCE algorithm shows the highest complexity per iteration due to the additional prediction step. However, as discussed subsequently, it requires less iterations to achieve the same detection performance.

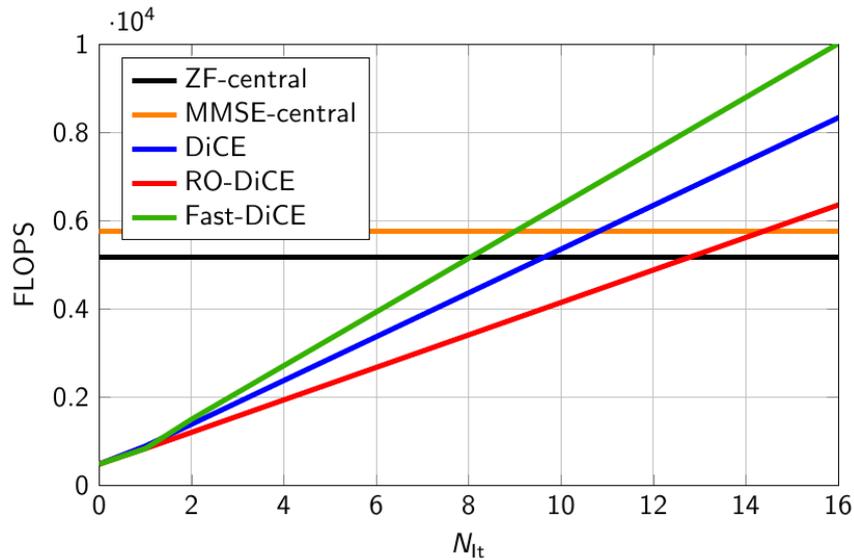


Figure 4-7: Floating point operations per subcarrier and per iSC for a system with $N_{SC} = 6$, $N_R^{SC} = 2$, $N_{UE} = 4$, $N_T^{UE} = 2$

Backhaul traffic

For the evaluation of the backhaul traffic we consider three general physical topology cases with a logical backhaul topology defined by the set of edges \mathcal{E} (see Figure 4-8) [82]. The first case considers a topology with point-to-point (P2P) links between the iSCs, e.g. via millimetre Wave (mmWave), fibre or copper links. In such a topology only unicast transmissions among the iSCs are possible such that each variable requires a separate transmission depending on the neighbouring iSC. In the DiCE algorithm, variables \mathbf{z}_j , $\boldsymbol{\lambda}_{ji}$ and \mathbf{x}_j need to be transmitted by each iSC j to its neighbours. Consequently, six variables are exchanged for each logical connection in the network resulting in a total communication overhead of $6|\mathcal{E}|N_I N_{SYMB}^{SUB}$ per iteration with $|\mathcal{E}|$ being the total number of logical links in the network. For the RO-DiCE only \mathbf{z}_j and \mathbf{x}_j need to be exchanged reducing the overhead per iteration to $4|\mathcal{E}|N_I N_{SYMB}^{SUB}$.

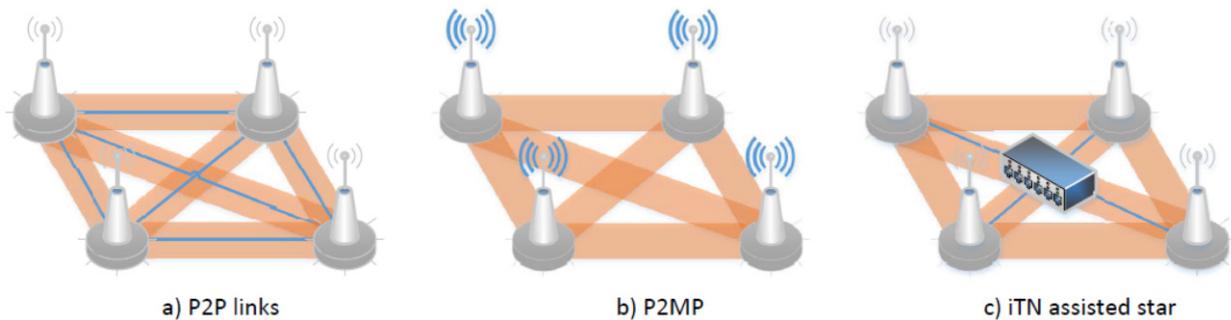


Figure 4-8: Considered backhaul topologies for an example of $N_{SC} = 4$ iSCs

The second case considers point-to-multi-point (P2MP) connections among the iSCs, e.g. by omnidirectional wireless connections. In such a scenario broadcast transmissions among the iSCs are possible such that the exchange of the variables \mathbf{z}_j and \mathbf{x}_j only requires one transmission each and thus, does not depend on the number of logical edges anymore. The exchange of $\boldsymbol{\lambda}_{ji}$ still has to be done in a unicast fashion since these are specific to on logical edge in the network. In contrast, the RO-DiCE avoids an exchange of the multipliers $\boldsymbol{\lambda}_{ji}$ which reduces the overhead in a point-to-multi-point scenario significantly compared to the DiCE.

As a third case we consider the integration of the iTN (iJOIN Transport Node) into the backhaul topology. For this, we assume that all iSCs are connected to the iTN in a physical star topology. Here, the iTN controls the communication among the iSCs, i.e. packets are transmitted to the iTN first which directs them to the desired iSCs. For the iTN we assume that it is able to multicast packets to multiple destinations. Therefore, variables \mathbf{z}_j and \mathbf{x}_j need to be transmitted to the iTN only once. The iTN duplicates the variables in order to forward these to the logical neighbours.

Table 4-7 lists the total communication overhead per iteration and per subcarrier for the DiCE and RO-DiCE algorithm for the mentioned topology scenarios. It should be noted that the overhead of the DiCE algorithm is identical to the overhead of the Fast-DiCE algorithm.

Table 4-7: Communication overhead per iteration and subcarrier for different topologies

PHY Topology	Links	D_{DiCE}	$D_{\text{RO-DiCE}}$
Arbitrary	P2P	$6 \mathcal{E} N_I N_{\text{SYMB}}^{\text{SUB}}$	$4 \mathcal{E} N_I N_{\text{SYMB}}^{\text{SUB}}$
Arbitrary	P2MP	$2(\mathcal{E} + N_{\text{iSC}})N_I N_{\text{SYMB}}^{\text{SUB}}$	$2N_{\text{iSC}}N_I N_{\text{SYMB}}^{\text{SUB}}$
iTN	P2P	$(2N_{\text{iSC}} + 8 \mathcal{E})N_I N_{\text{SYMB}}^{\text{SUB}}$	$(2N_{\text{iSC}} + 4 \mathcal{E})N_I N_{\text{SYMB}}^{\text{SUB}}$

Figure 4-9 depicts the FER and the corresponding throughput, respectively, over the number of transmitted complex scalars per subcarrier for the described backhaul scenarios at $E_b/N_0 = 10$. We consider a full mesh topology of $N_{\text{iSC}} = 6$ iSCs, $N_R^{\text{iSC}} = 2$ receive antennas per iSC, $N_{\text{UE}} = 4$ UEs and $N_T^{\text{UE}} = 2$ transmit antennas per UE with noisy backhaul links with an $\text{SNR}_{\text{BH}} = 20$ dB. The number of transmitted complex scalars can be calculated from Table 4-7 and is linearly growing with the number of iterations used for the different DiCE algorithms. It is obvious that in general a better performance can be achieved with more iterations which means with a higher total communication overhead in the network. The system setup is identical to the setup used to evaluate the BER, FER and throughput in the LTE uplink. According to these results Fast-DiCE achieves the same FER performance with a lower number of iterations and thus with a lower communication effort. In case of point-to-multi-point transmissions significant benefits can be observed for the RO-DiCE since it avoids unicast transmissions of multipliers at all. For an iTN-assisted backhaul topology the overall communication overhead is increased especially for the DiCE and RO-DiCE because variables have to be transmitted to the iTN first before being routed to the desired destinations. Nevertheless the increase in communication overhead comes with a reduction of the number of physical backhaul links.

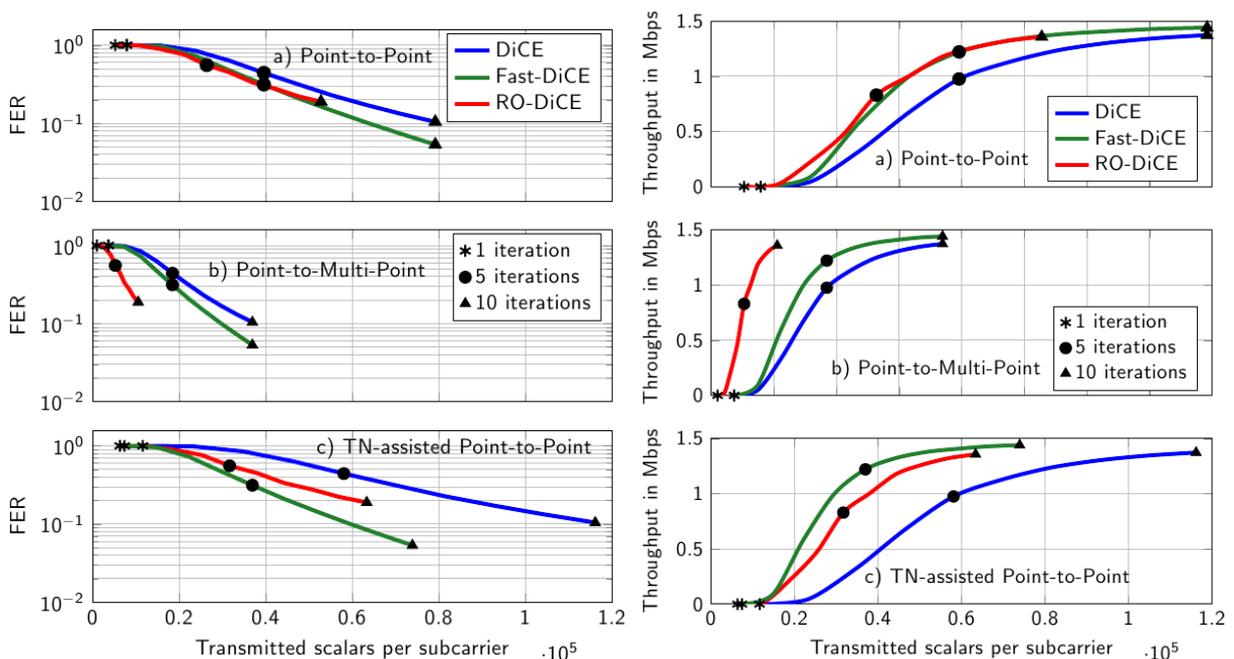


Figure 4-9: FER and throughput over total communication overhead for considered backhaul topologies

4.2 CT2.2: Multi-Point Turbo Detection

4.2.1 Technical description

Scenario

The CT2.2 investigates the benefit of relying on the turbo detection principle to increase the aggregated user throughput in the uplink direction. In a dense small cell deployment, one user can more easily see other small cells in addition to its serving one, especially if he is at the edge of the cell. By scheduling the (edge) users on the same resources and exploiting the created interference as a source of information through the turbo detection principle in each concerned small cell, the aggregated throughput of the system should be improved, as “more” spectrum and diversity are made available.

We will use turbo detection processing to perform the multi-user detection. This processing can either be done:

- centrally at the RANaaS data centre: we call this approach **Multi-Point Turbo Detection (MPTD)**, or,
- locally in each iSC: we call this approach **Single-Point Turbo Detection (SPTD)**.

In theory, this CT could be applied to any iJOIN Common Scenario (CS) as long as the UEs are not moving too fast. In practice, this CT will be assessed in the context of CS4: Shopping Mall/Airport.

System Model

To evaluate the benefit of MPTD/SPTD solutions compared to the previous baseline, link-level simulations will be carried out. The setup depicted in Figure 4-10 will be assumed. In this setup, two iSCs are connected to the RANaaS platform through the J1 interface and to each other through the J2/X2 interface. One UE is connected to one iSC. For simplicity, UE₁ is attached to iSC₁ and UE₂ is attached to iSC₂. Both UEs are scheduled on the same resources such that they interfere with each other at their respective serving cell. All equations and an in-deep technical description associated to the MPTD and the SPTD without cooperation strategies are already provided in D2.1 [36].

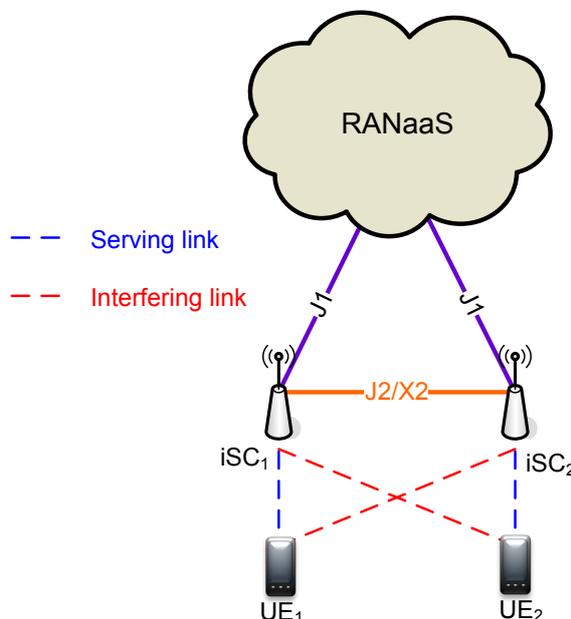


Figure 4-10: MPTD Simulation Set-up

So far, the symmetric setup (i.e., same MCS, same receive power) over a Rayleigh channel model has been evaluated in D2.1 [36], highlighting the tremendous gain offered by MPTD over an enhanced linear minimum mean square error receiver with interference rejection combining (E-LMMSE-IRC). With this model for the MIMO channel, the same realisation (a random Rayleigh-based matrix) was used for all the subcarriers and for all the OFDM symbols of the allocated Resource Block (RB). To rely on a more

“realistic” modelling, the channel model used for the MIMO fast fading part has been updated to the Indoor/Hotspot (InH) one from the ITU-R [44].

Assuming an LTE system on the 2.6 GHz frequency band using a 10 MHz bandwidth for transmission, 50 RBs are available. Using 1 RB at each extremity for the PUCCH channel leaves 48 RBs for the PUSCH channel. Amongst these 48 available RBs, one RB is randomly picked and used by both UE₁ and UE₂ as shown in Figure 4-11. Within the Monte-Carlo framework, a new fast fading (i.e., a MIMO matrix per subcarrier) following the InH profile is generated at each transmission. A speed of 3 km/h and non-line of sight conditions are assumed. The channel is supposed to be constant during the whole subframe.

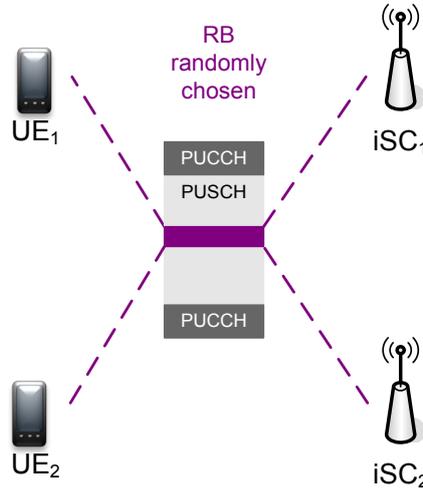


Figure 4-11: MPTD System Model

Figure 4-12 shows how the LTE-compliant Hybrid Automatic Repeat Request (HARQ) with Incremental Redundancy (IR) is captured in our Monte-Carlo framework. At a given SNR, the two users send their frame toward their respective serving iSC. If one frame cannot be correctly decoded (check with the Cyclic Redundancy Check - CRC), then it is retransmitted 8ms later with a different Redundancy Version (RV). The statistics are collected only on the two first frames and the number of retransmissions needed to convey them is stored. We allow up to four transmissions just like in the LTE standard. If one user has its frame successfully decoded and the other not, the first user will still send dummy frames which will not be accounted in the metrics computation as shown in Figure 4-12 where the frame of user 2 has needed 4 transmissions.

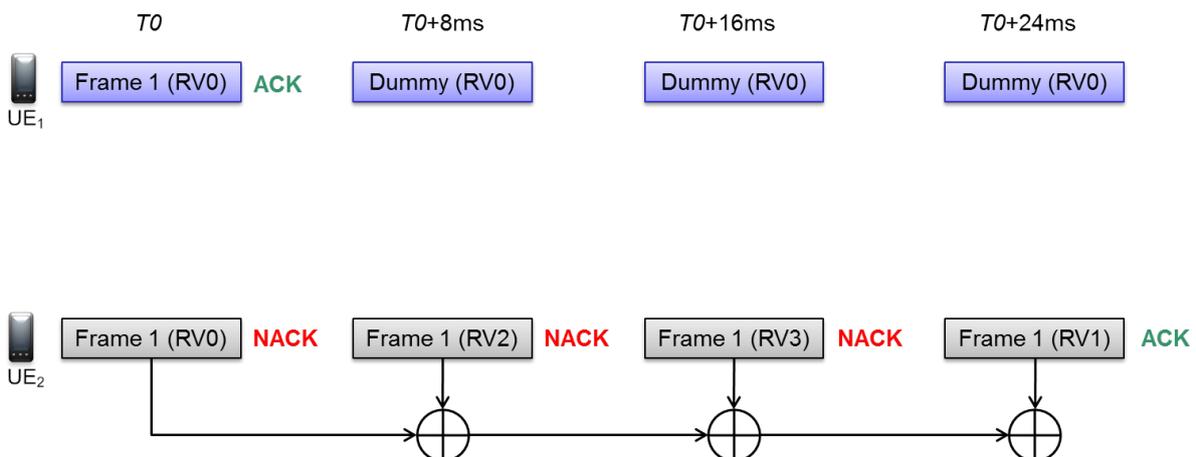


Figure 4-12: HARQ modelling in the Monte-Carlo framework

Approach

The principle of the turbo detection is to use the detection of one user’s stream, reconstruct its contribution and subtract it from the received signal. By doing this iteratively, the detection of each user’s streams is improved compared to an attempt to detect them independently (classical approach). Figure 4-13 describes this iterative signal processing when *U* users are being detected.

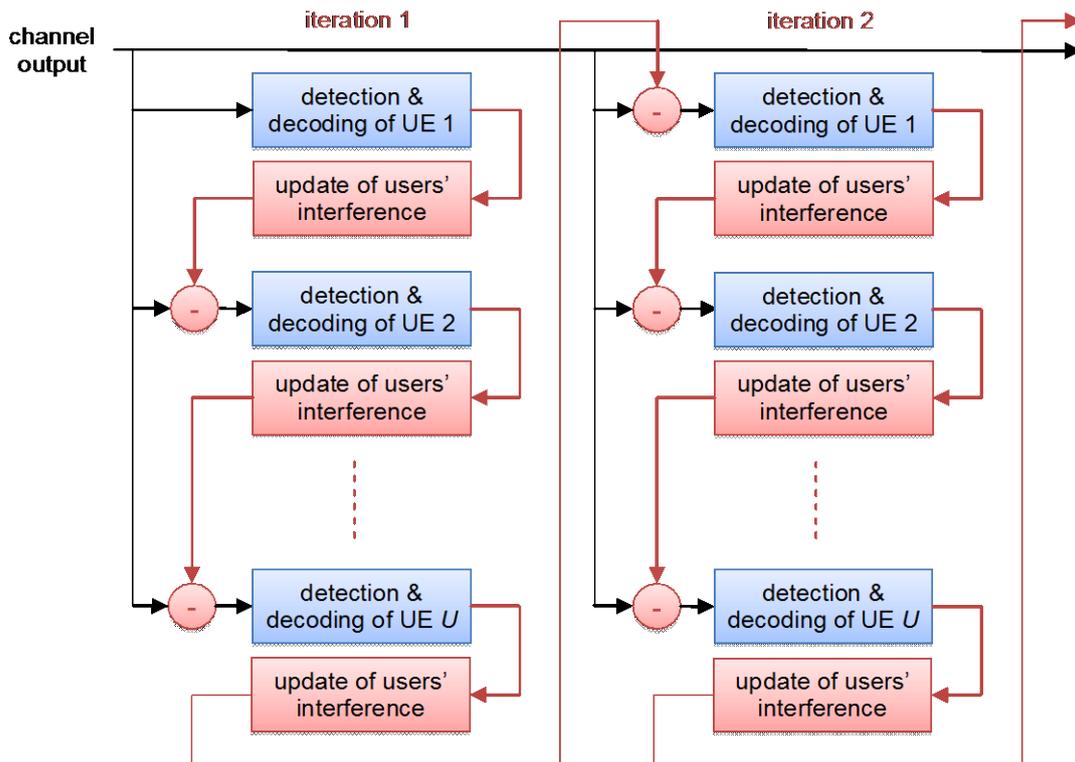


Figure 4-13: Turbo-detection principle

Using LTE Release 10 compliant MCSs coupled with SC-FDMA symbol mapping, a Monte-Carlo approach is used to assess the performance of the different receiver strategies (central/distributed processing) taking into account the HARQ procedure. In particular, FER against SNR of each UE is evaluated through link-level simulations. For a given UE, the interference coming from the other UE is not taken into account into the SNR. It will be considered by the receiver. The MPTD/SPTD solutions should provide lower FER allowing higher data rate to be used in the same radio link conditions, addressing Objective 1 on the “area throughput”.

4.2.2 Implementation in the iJOIN architecture

In this CT, a veNB can be made of RANaaS and one iSC or of RANaaS and several iSCs. In the first veNB case, this means that the functional split envisaged in both variants described in section 4.2.1, fully centralized (MPTD) or distributed with or without cooperation (SPTD), will involve two veNBs. As a consequence, the logical J2 interface may not be used for cooperation, but an extension to the logical X2 interface should be investigated. In the second veNB case, i.e. veNB made of RANaaS and several iSCs, the cooperation between iSCs is made within the same veNB, meaning that the logical J2 interface will be used.

There is no real advantage to use one or the other veNB definition. With the first definition, we do not need to add the J2 interface but rather extend the X2 one, opening the turbo-detection processing to “traditional” eNBs.

As described in D2.1 [36], we envisaged two cases for the turbo detection implementation:

Functional Split A.2: Multi-Point Turbo Detection (MPTD)

In this configuration, the physical processing is performed at the RANaaS platform as shown in Figure 4-14 (split option A.2 in section 3.2.1). Raw quantized I/Q samples after the first FFT stage are sent to the RANaaS, meaning that a high bandwidth, but more importantly, a low latency is required for the J1 interface.

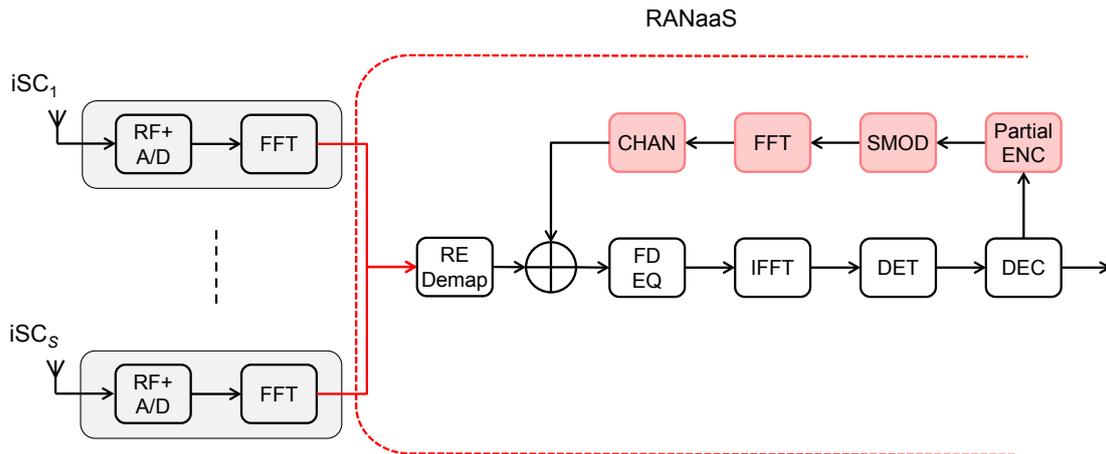


Figure 4-14: MPTD performed at the RANaaS platform using the J1 interface

Functional Split above C: Single-Point Turbo Detection (SPTD)

Single-Point Turbo Detection (SPTD): In this configuration, the turbo detection is performed locally in each iSC as shown in Figure 4-15. Cooperation through a direct interface may be envisaged as well. To stick with the 3GPP constraints and the veNB concept, this interface (J2 or extended X2 based on the veNB definition) should provide high bandwidth but more importantly low latency which is more likely to be achievable in a dense deployment where iSCs will be closely deployed.

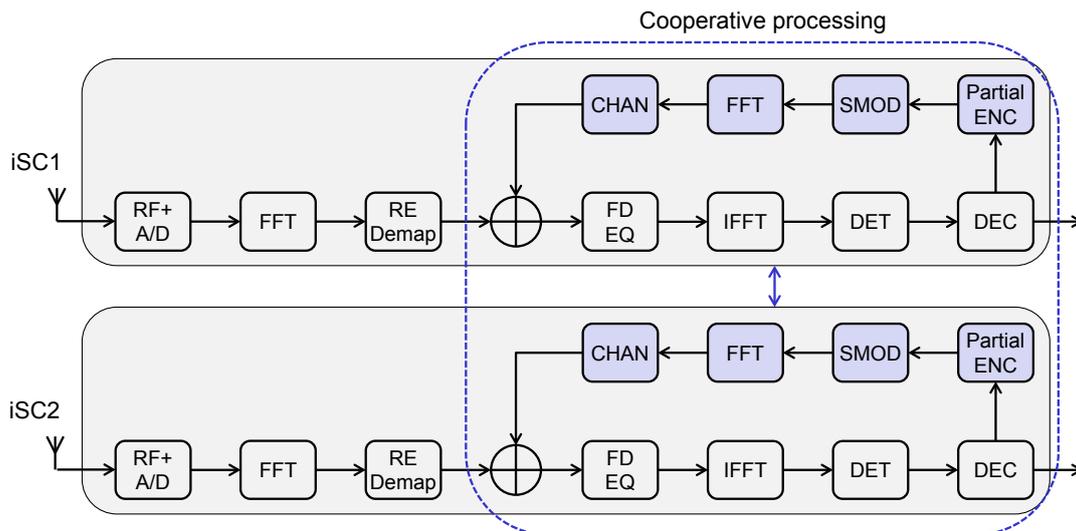


Figure 4-15: SPTD performed in each iSC with possible cooperation using the X2 interface

Backhaul Load per Functional split

In this CT, decision to choose either MPTD or SPTD is taken at higher layer than the physical one. The pairing of users to be involved in a turbo-processing is based so far only on the downlink measurement on the reference signals (Reference Signal Received Power – RSRP) performed by the UE, which is then fed back to the iSC in a standardised way. This measurement goes up to the RANaaS only if for a given UE, the RSRP toward a neighbouring iSC is within a given range from the RSRP of the serving iSC (see CT3.7 in D3.2 [87] for more details).

Table 4-8: PHY Measurement placement and exchange links for CT2.2

Functional Split Approach	Measurement	Measurement Location	Exchange link
A.2 (MPTD)	RSRP	UE	iSC -> RANaaS (J1)
C (SPTD)	RSRP	UE	iSC -> RANaaS (J1)

When MPTD has to be performed, I/Q samples on each subcarrier allocated to a paired UE have to be sent through the J1 interface toward the RANaaS data centre. Demodulation reference signal (DMRS) used to derive channel state information are part from I/Q samples (the OFDM symbol in the middle of a slot is dedicated to host DMRS, one subframe being made of two slots). In the downlink, the acknowledgement (1 bit) needs to be sent back from the RANaaS to the iSC.

When SPTD is chosen, there is no exchange on the J1 interface regarding the physical layer. There must be some exchange to coordinate the scheduling/decoding of both iSCs either dynamically through the J2 interface or configured by RANaaS through the J1 interface, but the content of these messages is higher layer related (WP3). Regarding the WP2 scope, we may envisage exchanging information during the turbo-processing between the iSCs (e.g., LLRs as they both decode the two paired users locally) but this is left for further studies.

Table 4-9: Backhaul control information of CT 2.2

Link	Description
iSC → iSC	FFS
iSC → RANaaS	RRM Information <ul style="list-style-type: none"> - O2.3: Effective SINR per UE - O2.4: Effective SINR per UE-iSC link - O2.5: Estimated PER
	PHY Information (access link) <ul style="list-style-type: none"> - O2.7: ACK/NACK, HARQ process - O2.8: User data to upper layers, if ACK
RANaaS → iSC	CT specific control information <ul style="list-style-type: none"> - I2.5: Functional control for shifting/splitting functions I2.6: RRM information per UE

Table 4-10: Backhaul Load of CT 2.2

Link	Description	Msg Size	Msg Frequency	Latency
Functional split A.2: MPTD				
iSC → iSC	None			
iSC → RANaaS	I/Q signals after FFT (uplink)	see split option A.2 in 3.2.1	Every TTIs when a UE is scheduled	<0.1 ms
RANaaS → iSC	ACK (downlink)	1bit	4ms after when a UE is scheduled	<0.1 ms
	Scheduling information (RB, MCS)	See CT3.7 in WP3	RRM framework	<50 ms
Functional split above C: SPTD				
iSC → iSC	Scheduling synchronisation (e.g., MCS)	FFS	When a paired UE is scheduled	< 0.1ms
	Codeword information (e.g. LLR)	FFS	During a paired UE decoding	< 0.1ms
iSC → RANaaS	None at PHY layer			
RANaaS → iSC	Scheduling information (RB, ...)	See CT3.7 in WP3	RRM framework	<50 ms

4.2.3 Evaluation of the CT

Compliance with iJOIN objectives

By scheduling users on the same resources and ensuring that the physical processing performed allows the decoding of their streams, the overall throughput of the system should be improved, meaning that the “area throughput” (objective 1) will go up as a consequence.

By moving the processing part to the RANaaS platform where more computational resources are available, the “utilisation efficiency” (objective 4) may also be improved as more traffic will transit through the backhaul since higher transmission data rate on the air interface may be used.

Description of the baseline used for the evaluation

The baseline used for the evaluation is an LTE Release 10 configuration. Small-cells will be equipped with two receive antennas, while UEs will only support one transmit antenna. At the small cell side, the advanced receiver E-LMMSE-IRC will be assumed as the baseline receiver. Such receiving strategy is currently evaluated at the UE side in the Release 12 Work Item on Network-Assisted Interference Cancellation (NAIC, [43]), so it makes sense to incorporate it as a baseline at the base station side.

Discussion of results of the CT

In this deliverable, we only evaluate the performance of the MPTD algorithm against the baseline receiver. The SPTD approach, which could include cooperation during the turbo detection signal processing, will be investigated on a later stage. We consider 4 iterations for the turbo detection. We also add the so-called “Genie” curve corresponding to a perfect user interference cancellation which represents the lower bound of the turbo detection. We still consider a symmetric setup, i.e., that each iSC received both users with the same power on average, the fast fading will however unbalance the instantaneous received power.

Three LTE-compliant MCSs are used, with a transmission spanning over one RB. As a reminder one RB has 12 subcarriers lasting over 14 OFDM symbols. Two of these OFDM symbols are dedicated to the DMRS, leading to 144 modulation symbols available for data per RB.

The three MCSs are the following:

- QPSK and a coding rate of 0.5: Transport Block Size (TBS) of 144 bits for a Codeword Size (CWS) of 288 bits
- 16QAM and a target coding rate of 0.5: TBS of 280 bits, CWS of 576 bits
- 64QAM and a target coding rate of 0.5: TBS of 408 bits, CWS of 864 bits

Figure 4-16 to Figure 4-18 show (a) the FER against the SNR and (b) the average number of needed transmissions against the SNR of the first user for the three previous MCSs. Since we have a symmetric setup, the results of user1 or user 2 are equivalent. On a general comment, compared to the results of D2.1, we can notice on the figures (a) that the HARQ procedure manages to reduce the gap between the baseline receiver (E-LMMSE-IRC, solid black curve) and our advanced one (MPTD, solid red curve). However, MPTD still outperforms E-LMMSE-IRC for any given FER target. In addition, the gap may have been a bit reduced thanks to HARQ, but it comes with a cost when observing the figures (b), where more repetitions are needed, leading to an overall lower throughput.

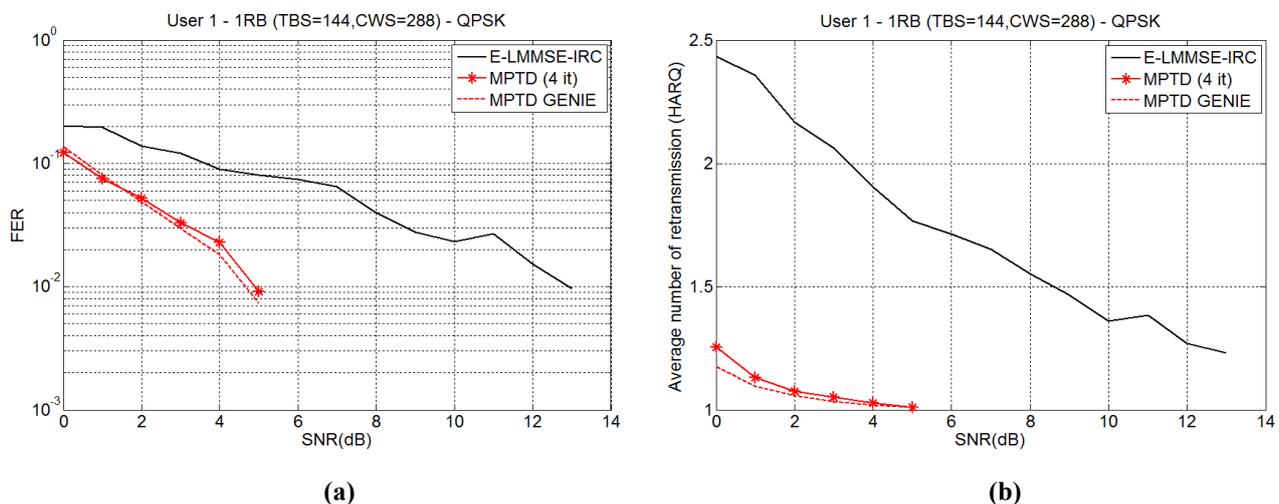


Figure 4-16: FER vs SNR (a) and Average number of transmissions vs SNR (b) for the QPSK MCS

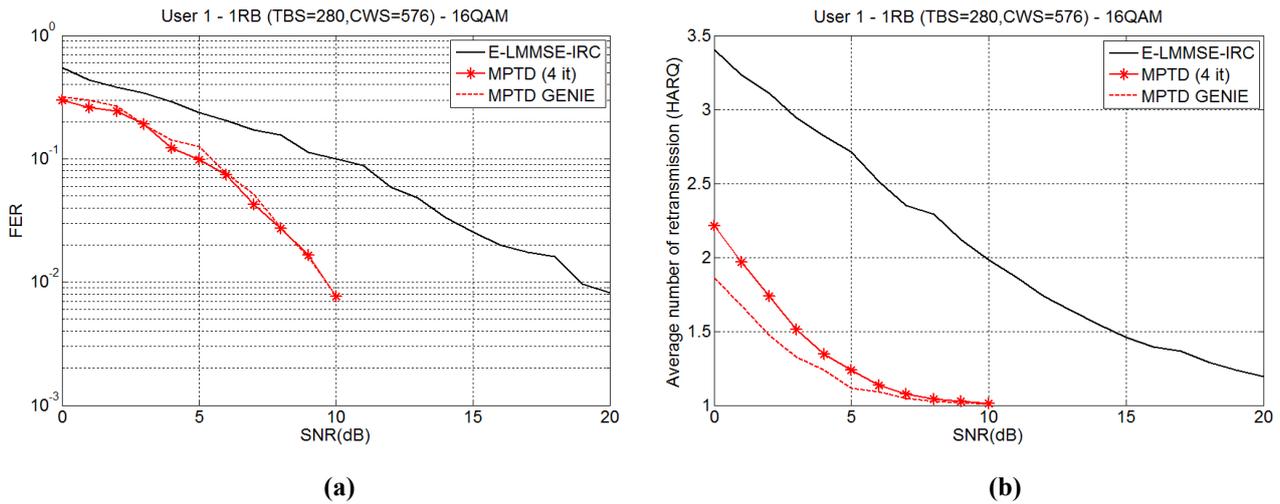


Figure 4-17: FER vs SNR (a) and Average number of transmissions vs SNR (b) for the 16QAM MCS

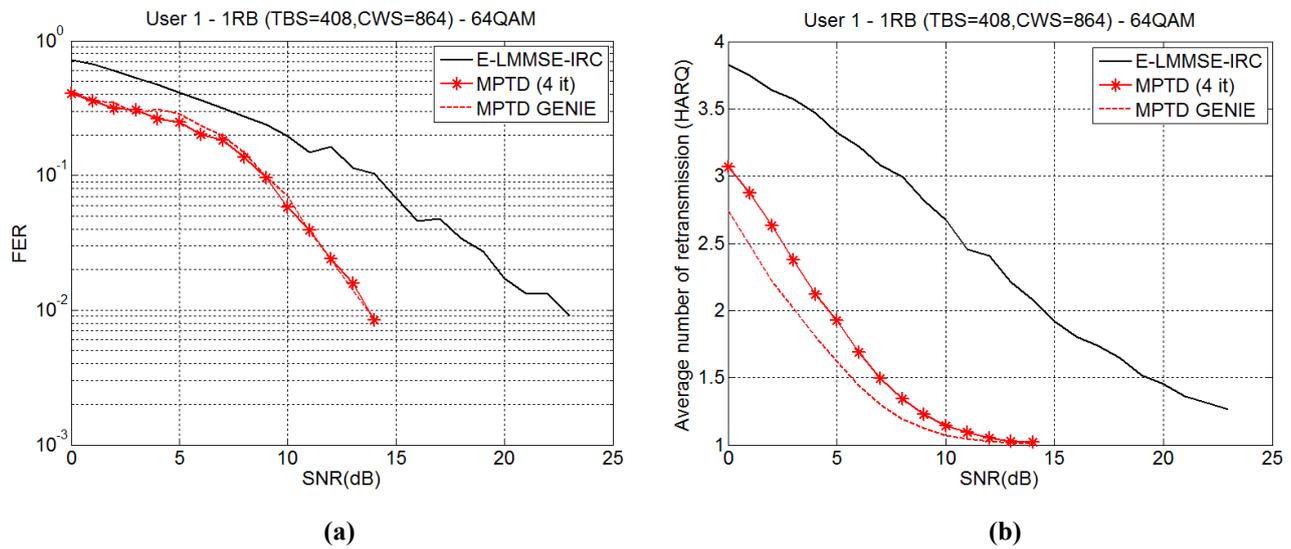


Figure 4-18: FER vs SNR (a) and Average number of transmissions vs SNR (b) for the 64QAM MCS

We can also notice that the MPTD curves converge to the Genie curves in terms of FER for all three MCSs. However, MPTD usually needs a bit more of retransmission attempts for the high FER area (e.g., 10^{-1}) compared to the Genie, while almost the same number of retransmissions both is needed in the low FER area (e.g., 10^{-2}).

As a conclusion, MPTD improves the detection quality both in terms of FER and retransmission need. As a consequence the area throughput should be increase for the same transmitted power, which will also benefit to the air interface utilisation efficiency. For the future work, the SPTD solution will be investigated as the MPTD one needs a very low latency backhaul. In terms of FER and average number of retransmissions, the MPTD curves for various MCSs are sufficiently close enough to the Genie ones, which assume a perfect interference cancellation of the other user. Therefore, such physical abstraction of the MPTD will be used for system-level simulations without losing in accuracy.

4.3 CT2.3: Joint Network-Channel Coding

4.3.1 Technical description

Scenario

The CT2.3 investigates the benefits of the JNCC technique for increasing the aggregated user throughput in the uplink direction. The proposed scenario assumes that the veNB domain contains two classes of iSCs: a first class of iSCs with virtually unlimited backhaul capacity, and a second class of iSCs with limited backhaul capacity. For instance, one could include in the first class the iSCs connected to the RANaaS via fibre links, and in the second class those connected via wireless links. Another, but similar situation is depicted in Figure 4-19(a): the first class of iSCs communicate directly to the RANaaS (grey lines), while iSCs of the second class have no direct connection with RANaaS, but communicate with iSCs of the first class via wireless interface (green lines). The baseline case is represented by users being connected to only one iSC (solid orange lines), with the consequence of possible congestion over the iSC-to-iSC connections. To avoid traffic congestion, we allow some users to communicate with a second iSC (dashed orange lines), in such a way that the veNB domain is separated in a number of multiple-access relay channels (MARC). The MARC model is illustrated in Figure 4-19(b), and consists of two UEs that use a common intermediate iSC₁, which acts as a relay, in order to communicate with a final destination iSC₂.

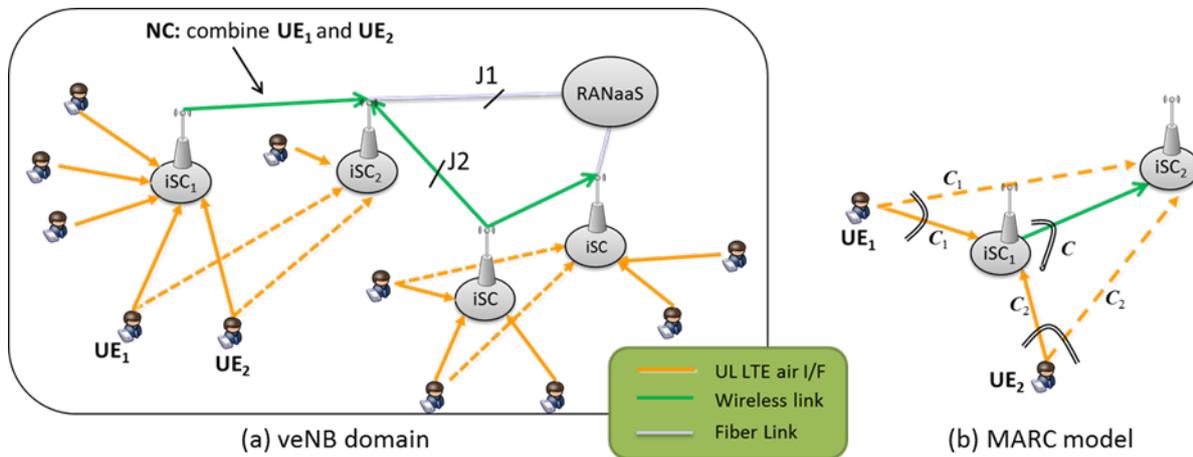


Figure 4-19: Proposed scenario for joint network-channel coding

System Model

We assume that UEs' transmissions are encoded by two channel codes \mathcal{E}_1 and \mathcal{E}_2 , while a network code \mathcal{E} is used at the intermediate iSC₁. Transmitted codewords are denoted by $\mathbf{c}_1 \in \mathcal{E}_1$, $\mathbf{c}_2 \in \mathcal{E}_2$, and $\mathbf{c} \in \mathcal{E}$, and are assumed to be of same length denoted by N_{block} .

The intermediate iSC₁ first decodes \mathbf{c}_1 and \mathbf{c}_2 , and then generates a NC codeword \mathbf{c} that is transmitted to iSC₂. Classical NC strategy assumes that \mathbf{c} is the bit-wise XOR of \mathbf{c}_1 and \mathbf{c}_2 , which is denoted by $\mathbf{c} = \mathbf{c}_1 \oplus \mathbf{c}_2$. Alternative NC strategies are discussed in the following paragraph. The use of network coding (NC) at the intermediate iSC₁ allows reducing the traffic load on the iSC₁-iSC₂ link, hence avoiding traffic congestion and therefore increasing area throughput in the uplink direction. Finally, the destination iSC₂ decodes the received signals by using knowledge of \mathcal{E}_1 , \mathcal{E}_2 , and \mathcal{E} , which are assumed to be transmitted over orthogonal channels.

Approach

For the proposed CT, users' transmissions are encoded by using Low-Density Parity-Check (LDPC) codes. Such codes may be represented by bipartite graphs, with a first set of nodes corresponding to coded bits, and a second set of nodes corresponding to parity check equations. The former are referred to as bit-nodes (represented by white circles) and the latter as check-nodes (represented by squares). These graphs are depicted in Figure 4-20, where they are denoted by LDPC@UE₁ for UE₁ and LDPC@UE₂ for UE₂,

respectively. A third bipartite graph is also represented, denoted by NC@Relay (in blue), which is associated with the network code at the relay. The NC-bit-nodes represent the bits generated by the network code, while NC-check-nodes represent parity check equations between UE₁-bit-nodes, UE₂-bit-nodes, and NC-bit-nodes.

The graph in Figure 4-20, consisting of LDPC@UE₁, LDPC@UE₂ and NC@Relay sub-graphs taken together, is referred to as the joint network-channel graph. This graph is used to perform decoding at the destination.

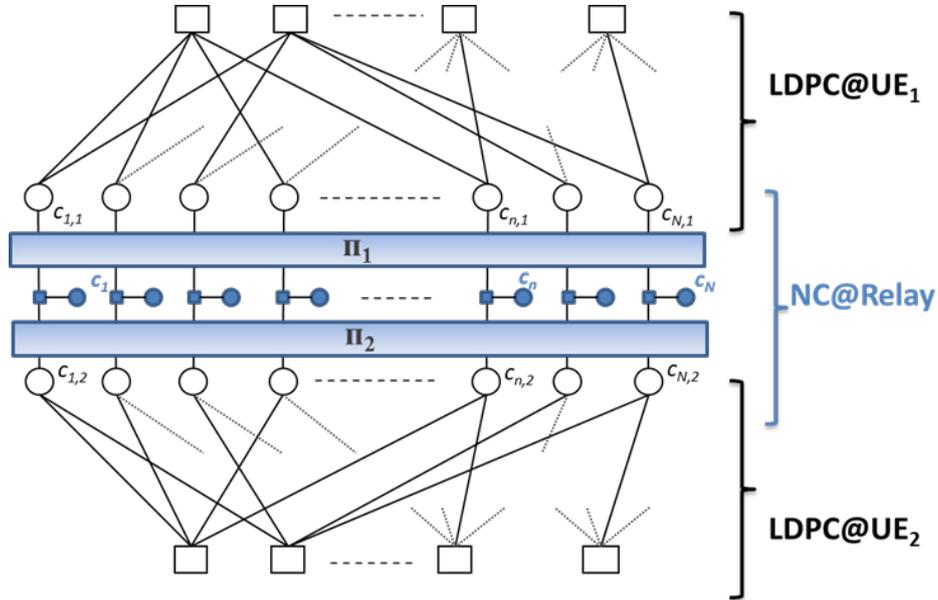


Figure 4-20: Joint network-channel graph

In case that no interleaving is used between users' bit-nodes and NC-check nodes, each bit c_n generated by the network-code is simply the XOR of the corresponding bits of users UE₁ and UE₂, that is $c_n = c_{n,1} \oplus c_{n,2}$. This corresponds to the traditional network coding strategy.

As an alternative NC strategy, in Figure 4-20 we have also represented the possibility to include an interleaver Π_u between each user's bit-nodes and the NC-check nodes. LDPC codes are known to have specific topologies, as short cycles, stopping sets, trapping sets [9], [11], which are responsible for decoding failures. Roughly speaking, with a message passing type algorithm, cycles lead to auto-confirmation of exchanged messages, and then degrade the decoder performance. Figure 4-21 shows a case of mirroring of a cycle of length 4 that amplifies the degradation effects. The goal of the interleaver is to avoid connecting together such harmful topologies, which would be detrimental to the performance of the joint channel-network code.

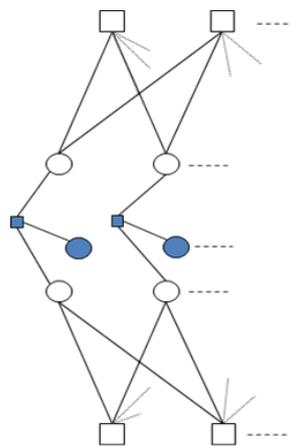


Figure 4-21: Example of mirroring of a cycle of length 4

Two designs of channel-network interleavers have been investigated:

1. A girth-optimized interleaver, inspired by the Progressive Edge-Growth (PEG) algorithm [6], in which each user's bit-nodes and NC-check-nodes are connected in a progressive manner, so that to maximize the girth of the JNCC graph. This girth-optimized interleaver is dependent on the LDPC codes of the users.
2. A random interleaver, in which the connections between opposite bit-nodes are avoided but the construction of a JNCC graph with large girth is not assured.

The goal of the analysis is to demonstrate that the network coding design influences the decoding performance. Moreover, we note that with these constructions, all the network NC-bit-nodes have degree equal to one, which generally produces bad performance since the degree-1 bit-nodes cannot provide extrinsic information during the iterative decoding process. For these reasons we also propose a Repeat-Accumulate (RA) [7] like construction, in which each NC-bit-node is further connected to the next network check-node. In Figure 4-22 a joint channel-network graph is depicted, in which the network coding part presents the RA structure. Hence, the first NC bit is generated by $c_1 = c_{1,1} \oplus c_{1,2}$, while $c_n = c_{n,1} \oplus c_{n,2} \oplus c_{n-1}$, for $n > 1$. Eventually, while not represented in this figure, an interleaver can also be interposed between the bit-nodes of each user and the NC-check-nodes.

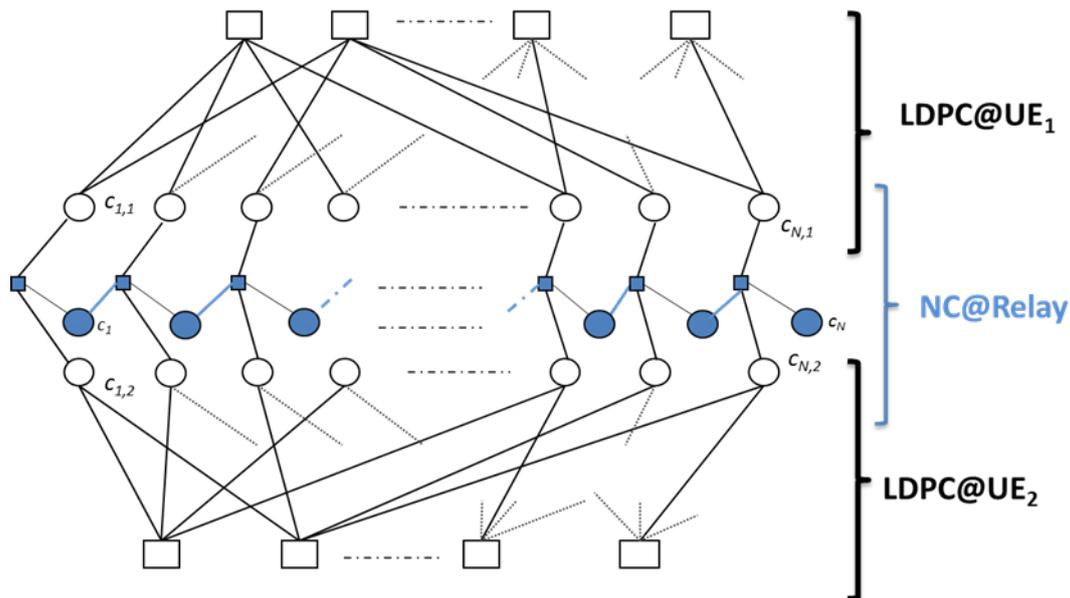


Figure 4-22: Joint network-channel graph with Repeat Accumulate (RA) structure

4.3.2 Implementation in the iJOIN architecture

In this CT, a veNB comprises a RANaaS and two or more iSCs. The veNB domain is further divided into a number of MARCs, each one consisting of two users, an intermediate (or relaying) iSC, and a destination iSC. The functional split allows distributing part of functionalities within the relaying and/or the destination small cell, or centralizing them in the RANaaS. Two functional split strategies have been presented in Section 6.2.3 of deliverable D2.1 [36]. They are briefly discussed and also slightly modified in the following.

Functional Split C: Joint Network-Channel Decoding within iSC₂

The first functional split implements joint network-channel decoding within iSCs (Figure 4-23). Cooperative processing is represented by the blue arrow between the two iSCs, indicating that network encoded data is transmitted from iSC₁ to iSC₂ by the J2 link. The destination node iSC₂ jointly decodes the received signals from UEs and iSC₁. Decoded data is then transmitted to the RANaaS through the J1 link.

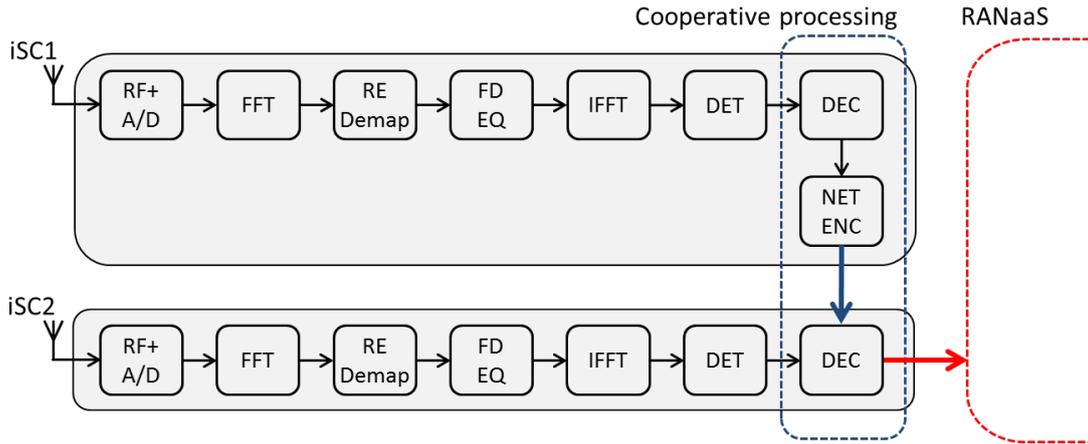


Figure 4-23: Functional Split C for CT2.3

Let N_R denote the number of NC bits transmitted from the relaying small cell iSC_1 to the destination small cell iSC_2 . The total number of NC bits that can be generated at iSC_1 is equal to $N_{\text{block},u}$, which is the common length of channel codes used by UE_1 and UE_2 . However, N_R can be dynamically adapted within the range $0, \dots, N_{\text{block},u}$, according to the quality of the links between users and the destination small cell iSC_2 . Therefore, we have $N_R = \eta_u \cdot N_{\text{block},u}$, with parameter $\eta_u \leq 1$ (note also that NC bits transmitted on the iSC_1 - iSC_2 link do not need further channel coding protection). Finally, assuming that all users served by iSC_1 are coupled in pairs within different MARCs, the message size of J2 payload for iSC_1 is given by:

$$D_{P,J2}^C = \frac{1}{2} \sum_{u=1}^{N_{UE}} \eta_u \cdot N_{\text{block},u} \text{ [bit/iSC]} \quad (4-24)$$

After JNCC decoding at the destination small cell iSC_2 , information bits of both users are transmitted from iSC_2 to the RANaaS (note that transmission on the iSC_2 -RANaaS link should further be protected by channel coding, according to the quality of the link). Hence, according to Section 3.2.1, the message size of J1 payload for iSC_2 is given by:

$$D_{P,J1}^C = \sum_{u=1}^{N_{UE}} (TBS_u) \text{ [bit/iSC]} \quad (4-25)$$

Functional Split B.2: Joint Network-Channel Decoding within RANaaS

The second functional split implements centralized decoding within RANaaS as depicted in Figure 4-24. The intermediate node iSC_1 transmits the network encoded data directly to the RANaaS. As iSC_2 does not receive the network encoded data, it is not able to decode the received user signals. Hence, decoding is centralized on the RANaaS, and iSC_2 forwards soft LLR values to RANaaS. Successively, RANaaS performs JNCC decoding using the information received from both iSC_1 and iSC_2 .

More in general, we note that the functional split at iSC_2 could be performed anywhere between the RF and the detection (DET) modules. However, we only consider the case when the functional split is performed between the detection and decoding modules (as illustrated in Figure 4-24), since the decoding module is the only that needs information from both iSC_1 and iSC_2 .

Within this variant, the number of bits transmitted from iSC_1 to the RANaaS is exactly the same as the number of bits transmitted from iSC_1 to iSC_2 in variant C. The number of bits transmitted from iSC_2 to RANaaS can be computed by the formula derived in Section 3.2.1 for functional split B.2 in uplink, and is given by:

$$D_{P,J1}^{B,2} = \sum_{u=1}^{N_{UE}} N_{sc,u}^{iSC_2} \cdot Q_m \cdot N_Q \cdot N_{CW} \text{ [bit/iSC]} \tag{4-26}$$

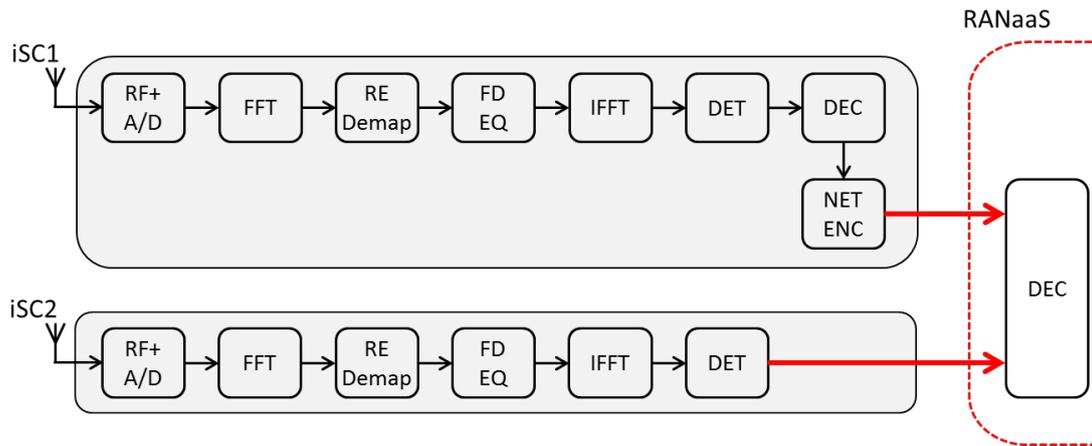


Figure 4-24: Functional Split B.2 for CT2.3

Backhaul Load per Functional split

The two variants do not affect the baseband processing operations within the relaying small cell iSC₁. The number of NC bits transmitted by iSC₁ does also remain constant; the only difference is that NC bits are transmitted either to iSC₂ for variant C or to the RANaaS for variant B.2.

However, the above discussed variants do affect the baseband processing at iSC₂ and the load of the backhaul link iSC₂-RANaaS. Indeed, for the variant C, all baseband functionalities are processed at the iSC₂, while variant B.2 allows offloading the iSC₂ node of part of the baseband processing (e.g., decoding), with an increasing load of the backhaul link. The BH payload for the two functional splits has been discussed in the corresponding paragraphs above, and is further summarized in Table 4-13 below.

Table 4-11: PHY Measurement placement and exchange links for CT2.3

Functional Split Approach	Measurement	Measurement Location	Exchange link
B.2 and C	CSI _{RAN} , SNR _{RAN}	iSC	iSC-RANaaS
B.2 and C	CSI _{BH} , SNR _{BH}	RANaaS	N/A

For all functional splits, some additional control signals, as detailed in the Input/Output tables Table 6-1 and Table 6-2 of Appendix I for the CT and system functions, have also to be exchanged over J1 link. The respective control information has been summarized in Table 4-12.

Table 4-12: Backhaul control information of CT 2.3

Link	Description
iSC → RANaaS	RRM Information <ul style="list-style-type: none"> - O2.5: Estimated PER
	PHY Information (access link) <ul style="list-style-type: none"> - I2.8: RxCSI, SNR, noise variance of UE-iSC links (functional split B.2) - O2.7: ACK/NACK, HARQ process - O2.8: User data to upper layers, if ACK
	Network parameters <ul style="list-style-type: none"> - I2.3: RAN connection table of iSCs and UEs - I2.4: BH connection table for iSC-iSC links and iSCs-RANaaS links
RANaaS → iSC	PHY information (BH link) <ul style="list-style-type: none"> - I2.17: Parameter of iSC-iSC link
	RRM Information <ul style="list-style-type: none"> - I2.6: RRM information per UE - I2.7: MCS for backhaul link
	CT specific control information <ul style="list-style-type: none"> - I2.5: Functional control for shifting/splitting functions

Table 4-13: Backhaul Load of CT 2.3

Link	Description	Msg Size	Msg Frequency	Latency
Functional Split C: Joint Network-Channel Decoding within iSC₂				
iSC ₁ → iSC ₂	NC bits	$\frac{1}{2} \sum_{u=1}^{N_{UE}} \eta_u \cdot N_{\text{block},u}$	T_{SUB}^{-1}	< 0.1 ms
iSC ₂ → RANaaS	Estimated messages (information bits) of users UE1 and UE2	$\sum_{u=1}^{N_{UE}} (TBS_u)$	T_{SUB}^{-1}	< 0.1 ms
Functional Split B.2: Joint Network-Channel Decoding within RANaaS				
iSC ₁ → RANaaS	NC bits	$\frac{1}{2} \sum_{u=1}^{N_{UE}} \eta_u \cdot N_{\text{block},u}$	T_{SUB}^{-1}	< 0.1 ms
iSC ₂ → RANaaS	LLR values for UE1 and UE2	$\sum_{u=1}^{N_{UE}} N_{sc,u}^{iSC_2} \cdot Q_m \cdot N_Q \cdot N_{CW}$	T_{SUB}^{-1}	< 0.1 ms

4.3.3 Evaluation of the CT

Compliance with iJOIN objectives

The use of JNCC is aimed at increasing the users’ throughput in the uplink direction. This technique is particularly suitable when the iSC₁-iSC₂ link has limited capacity (due to, e.g., limited bandwidth) for serving all the users connected to iSC₁, and the wireless links UEs-iSC₂ have degraded capacity due to increased distance. In this case, iSC₁ can provide iSC₂ with additional data pertaining to the user transmissions, hence improving the error correction capability at the destination. The additional data is actually provided by the network code used at iSC₁, and the whole coding scheme can be optimized such as to minimize the amount of data to be transmitted on the iSC₁-iSC₂ link, subject to a target throughput for users UE₁ and UE₂. Finally, as each veNB contains several iSCs serving many users, our CT can be applied on different MARCs within the same veNB, which allows increasing the area throughput in the uplink direction.

Description of the baseline used for the evaluation

The baseline scenario is depicted in Figure 4-25. We consider that users are connected to one single base station, namely iSC₁, which has no direct link to the core network. Then iSC₁ is connected to the core network through a second small/macro cell iSC₂/eNB. Since users are connected to only one small cell, there are no direct links between the users and the iSC₂/eNB. Then iSC₁ acts as a relay node that separately decodes and forwards user messages to iSC₂, which establishes the connection with the core network. The link between iSC₁ and iSC₂ is assumed to be noisy, and therefore user messages are assumed to be transmitted from iSC₁ to iSC₂ by using an appropriate modulation and coding scheme.

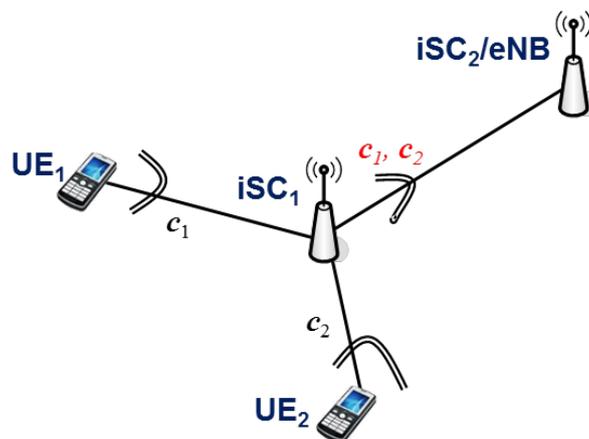


Figure 4-25: Baseline scenario, no direct UEs-iSC₂ links.

In order to assess the benefits of the proposed CT with respect to the baseline scenario, we evaluate the load on the backhaul link iSC_1 - iSC_2 for a target quality of services (QoS), expressed in terms of FER at iSC_2 .

Description of the evaluation methodology

For the baseline scenario, the amount of data that needs to be transmitted on the iSC_1 - iSC_2 link (expressed in number of transmitted bits) is given by the ratio between the number of source (useful) bits of the two users and the rate of the error correcting code used to encode the transmission on the iSC_1 - iSC_2 link.

For the proposed CT, the signals sent by the users are also received at iSC_2 by direct links, even though in a noisy version that could not be correctly decoded. In this case, NC can be advantageously applied at iSC_1 in order to reduce the load on the iSC_1 - iSC_2 link. Instead of separately relaying data packets for each of the users accessing the channel, the NC technique allows combining them together for transmission. Thus, let \mathbf{c}_1 , \mathbf{c}_2 be the codewords of UE_1 and UE_2 respectively, iSC_1 transmits only additional information $\mathbf{c} = \text{NC}(\mathbf{c}_1, \mathbf{c}_2)$, in order to help iSC_2 decoding the received signals from the UEs. The goal is then to evaluate the minimum amount of data that iSC_1 has to transmit, in order to ensure a FER at iSC_2 .

Discussion of results of the CT

Two main directions have been investigated. The first investigation is related to the design of the NC. First of all we analyse the behaviour of the joint channel-network coding when a bit interleaver is used in order to spread the users' bits before the network encoding. We also study specific structures based on the RA technique in order to obtain better coding gains. Moreover, binary or non-binary (NB) NC designs are also analysed.

The second investigation is related to the backhaul offload allowed by the use of NC. NC can reduce the number of relayed bits for each couple of users and ensure at the same time a target QoS at the destination. Then the MARC model together with the JNCC yields a double benefit: better error-correcting performance is obtained and traffic load can be reduced. In addition, there is a trade-off between error correcting performance and traffic load on the backhaul. This trade-off is not possible for the baseline scenario.

Simulation scenarios and results

We consider different symmetric or asymmetric scenarios. The symmetric scenario is characterized by the same quality of UEs- iSC_2 and iSC_1 - iSC_2 links. The asymmetric scenario is characterized by the same SNR for the UEs- iSC_2 links, but different SNR on the iSC_1 - iSC_2 link.

All links are modelled as additive white Gaussian noise (AWGN) channels, and transmitted symbols are QPSK modulated. We assume that the intermediate iSC_1 node successfully decodes the received user signals, and then transmits the network-encoded vector to the destination node iSC_2 . Signals transmitted by UE_1 , UE_2 , and iSC_1 are assumed to be orthogonal (e.g., in either frequency or time). At the destination iSC_2 node, Belief-Propagation (BP) [8] decoding is performed on the JNCC graph.

We focus on two families of binary LDPC codes:

- 1) WiMax codes specified by the IEEE 802.16e standard [10]. These codes are quasi-cyclic LDPC codes with irregular degree profile.
- 2) Regular (2,4)-LDPC codes, with bit-nodes of degree $d_v = 2$ and check-nodes of degree $d_c = 4$.

Binary Network-coding

Simulation results for the symmetric and asymmetric scenarios are shown in Figure 4-26 and Figure 4-27, respectively. The abscissa represents SNR values between the users and the destination small cell iSC_2 , denoted by SNR_{USD} . SNR values between the relaying small cell iSC_1 and the destination are denoted by SNR_{RD} . Blue curves correspond to JNCC with WiMax channel codes, while all the others correspond to JNCC with (2,4)-LDPC channel codes. Channel codes of both UE_1 and UE_2 are of length $N_{\text{block}} = 600$ coded bits, which is also equal to the number of network encoded bits transmitted by the relaying small cell iSC_1 .

The solid lines correspond to the case when no interleaving is used between the channel and the network graphs, whereas the dashed lines correspond to the network interleaving; we differentiate two cases in the figure legend: the first corresponding to random interleaving ("*W rand interl.*"), and the second corresponding to optimized interleaving ("*W girth-optim.*"). From these figures, it can be seen that in case a

network interleaver module is placed before the network coding, the obtained performance is always better compared to the case in which no interleaver is used.

The performance of the JNCC with *girth-optimized* interleaver is only shown in case of RA structure and (2,4) LDPC codes; this is because for the other cases we have not observed any significant gain with respect to random interleaving. Note also that the RA structure yields good results only when the (2,4)-LDPC codes are used. Hence, up to 0.5dB of gain is observed in Figure 4-26 and Figure 4-27(a) if the *girth-optimized* interleaving is used. When the iSC_1 - iSC_2 link is perfect, NC-bit-nodes are correctly received at iSC_2 and therefore the RA structure does not produce any advantage (Figure 4-27(b)). We remark also that, for different scenarios, the behaviour of the JNCC varies with the used channel codes. However, it can be observed that for a targeted FER = 10^{-2} , using regular (2,4)-LDPC channel codes with interleaved network coding and RA structure proves to be the best transmission strategy.

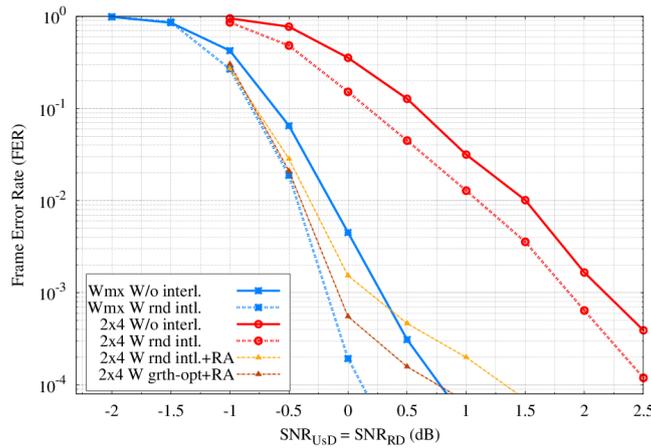
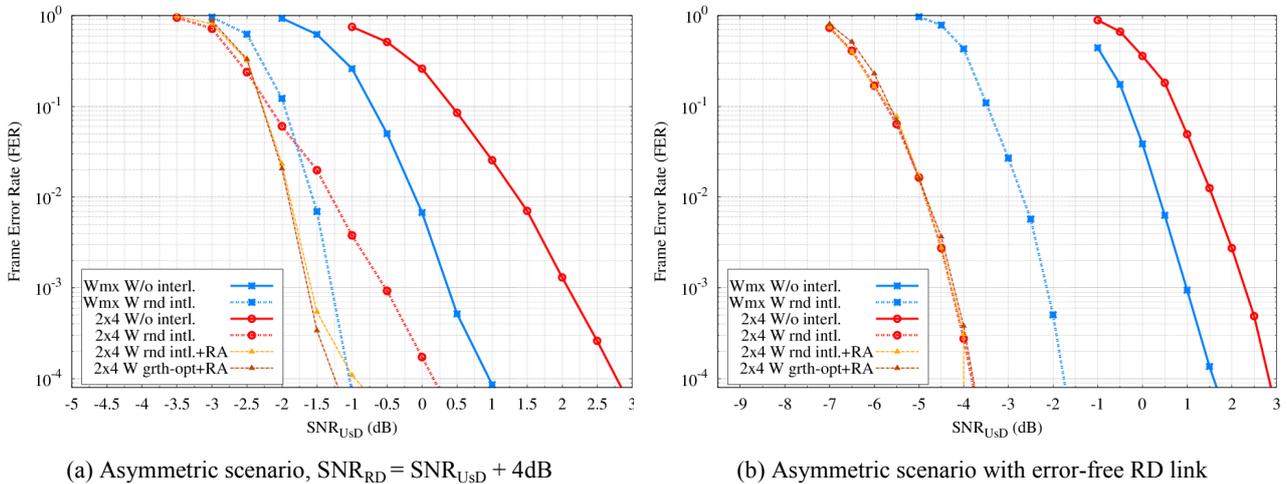


Figure 4-26: FER for WiMax and (2,4)-LDPC codes with different network graph constructions (symmetric scenario)



(a) Asymmetric scenario, $SNR_{RD} = SNR_{USD} + 4dB$

(b) Asymmetric scenario with error-free RD link

Figure 4-27: FER for WiMax and (2,4)-LDPC codes with different network graph constructions (asymmetric scenario)

Non-binary Network-Coding

This paragraphs aims at evaluating the performance of non-binary network coding, in which the network coded symbols are defined over the Galois Field with $q = 2^p$ elements, denoted by $GF(q)$. The use of non-binary NC with binary channel codes allows taking advantage of the coding gain brought by non-binary NC, while limiting the complexity increase at both relay and destination.

Figure 4-28 shows an example of network check-node defined over $GF(16)$ – hence a non-binary symbol corresponds to $p = 4$ bits. Each check-node is connected to $p = 4$ bit-nodes $c_{i,u}$ for each UE_u with $u = \{1, 2\}$ and $i \in \{1, 2, \dots, N_{block}\}$, where N_{block} represents the number of bit-nodes. The NC-symbol-node c_k (with k

$\in \{1, 2, \dots, N_{block}/p\}$) is defined over GF(16). The coefficients h are the columns of the binary images of the non-binary coefficients. We use this design with optimized coefficients reported in Section 4.3.3 of D2.1 [36]. For the sake of clarity, we represent only the case of no interleaving between the users' graphs and the relay's graph, but the use of an interleaver is not precluded and, as in the binary NC, it yields better results.

In Figure 4-29 we compare the FER performance for the two families of channel codes with binary (solid curves) and non-binary (dashed curves) network coding and random interleaver. We present only the symmetric scenario for which the non-binary NC performs better than the binary NC. For these simulations we use the same parameters as mentioned above, hence the number of network symbol-nodes is $N_{block} / 4 = 150$.

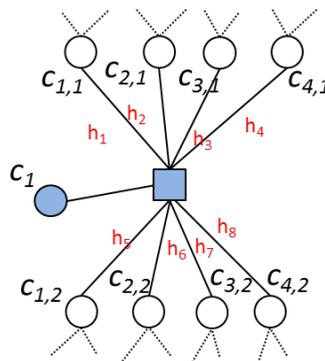


Figure 4-28: Example of non-binary network check-node

We can observe small gains with NB network coding with respect to binary network coding. However, this is not what we have expected from this technique, and we consider that these gains are not sufficient to justify the complexity increase induced by the use of NB network coding.

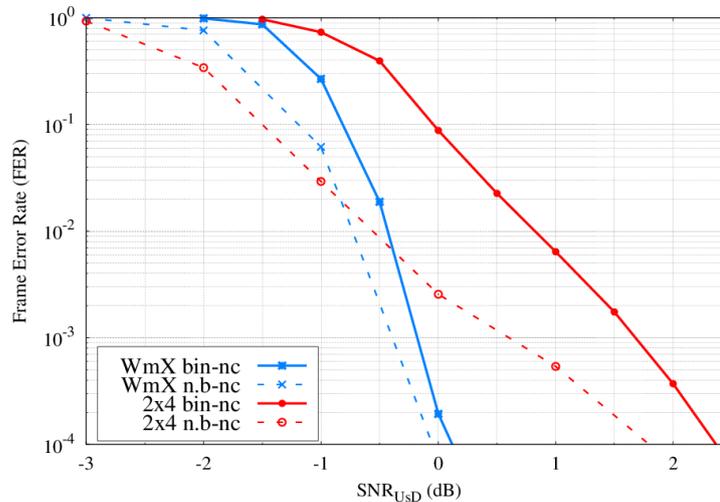


Figure 4-29: FER for WiMax and (2,4)-LDPC codes with binary and NB network coding and random interleaving

Backhaul off-load

One of the aims of this CT is to reduce the traffic load on the backhaul. Therefore, we analyse the case in which the number of bits transmitted by the relaying small cell iSC_1 is dynamically adapted to obtain a target FER of 10^{-2} at the destination.

We restrict our analysis to the case when both UE_1 and UE_2 use the WiMax LDPC code with rate $1/2$ and length $N_{block} = 2304$ coded bits (which is the standard coded length considered in the WiMax protocol [10]), then the number of information bits for each user is $N_{info} = 1152$. Users' transmissions are assumed to be perfectly decoded at the relaying small cell iSC_1 . Throughout this paragraph, the SNR on the iSC_1 - iSC_2 link is fixed to $SNR_{RD} = 3.5$ dB.

For the baseline scenario (illustrated in Figure 4-25), the relaying small cell iSC_1 transmits the information bits of both users towards the destination small cell iSC_2 . According to the SNR_{RD} value, the transmission on the iSC_1 - iSC_2 link needs to be encoded by a code with rate $2/3$, in order to achieve a target $FER = 10^{-2}$ at iSC_2 . Thus in the baseline scenario, the relay first decodes the users' messages and obtains the $2N_{info}$ information bits (N_{info} information bits for each user). These bits are then transmitted from iSC_1 to iSC_2 . Since the iSC_1 - iSC_2 transmission needs to be protected by a code with rate $2/3$, the number of bits transmitted on the iSC_1 - iSC_2 link is given by $N_R = 3N_{info}$ bits.

We focus on the case of MARC in which direct links assure a connection with the destination small cell and, even if they are noisy, they can be advantageously exploited. In order to compare our CT with the baseline, we consider two possible strategies at the relay: users' messages are either combined (NC strategy) or sent separately (pure relaying strategy).

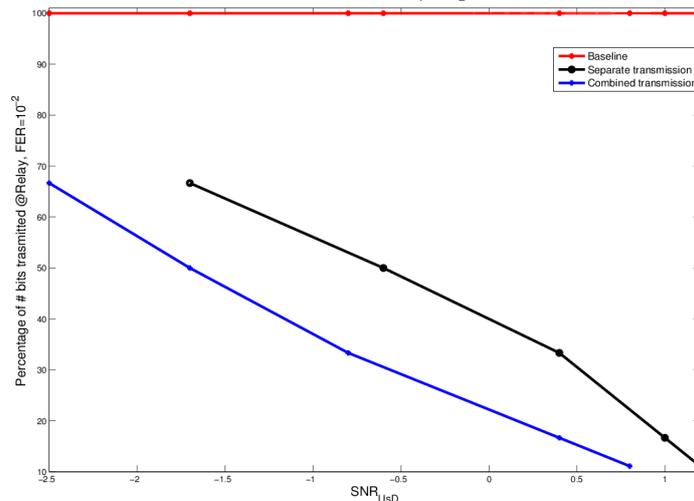


Figure 4-30: Analysis of the percentage of transmitted bits from the relay

In Figure 4-30 we plot the percentage of transmitted bits (with respect to the baseline case) sent by the relay within the above strategies, in order to achieve a target $FER = 10^{-2}$. The red line corresponds to the baseline (no link from users to iSC_2) in which the relay always sends $3N_{info}$ bits. The MARC configuration allows dynamically adapting the number of bits transmitted by the relay, N_R , according to the quality of the links between users and the destination small cell iSC_2 . In our simulations, N_R varies in the range $0, \dots, N_{block} = 2N_{info}$, since N_{block} is the maximum number of NC bits that can be generated at the relay.

The black curve represents the case in which users' messages are sent separately by the relay (pure relaying strategy). In this case, $N_R/2$ bits are sent for each user, priority being given to the N_{info} information bits.

For example, in case $N_R = 1728$, the relay transmits 864 bits for each user. Since $864 < N_{info} = 1152$, all the bits transmitted are information bits. The blue curve represents the case of a combination of N_R bits of the first user with N_R bits of the second user (NC strategy). In this case, a random interleaver is used before the network coding module, which is constructed such that the following constraint is fulfilled: the NC combines information bits of UE₁ with information bits of UE₂ (and, consequently, parity bits of UE₁ with parity bits of UE₂). The N_{info} NC-bits generated by information bits of UE₁ and UE₂ are referred to as NC-information bits, while the remaining NC-bits are referred to as NC-parity bits. Similar to the pure relaying case, the relay transmits N_R NC-bits, priority being given to the N_{info} NC-information bits. For example, for $N_R = 1728$, the relay transmits $N_{info} = 1152$ NC-information bits and 576 NC-parity bits. As we can observe in this figure, the NC strategy allows to reduce the number of bits sent by the relay, for the same target $FER = 10^{-2}$, compared to the baseline case and the pure relaying strategy. As a consequence, one can reduce the system load and determine the number of transmitted bits from the relay according to the SNR of the UEs- iSC_2 links.

4.4 CT2.4: Sum-Rate and Energy-Efficiency metrics of DL CoMP with backhaul constraints

Sum-rate and energy-efficiency are two critical metrics to characterize the performance of the DL CoMP algorithms. In deliverable D2.1 [36], energy efficiency for two of the most representative CoMP schemes joint transmission (JT) and coordinated beamforming (CB) have been analysed in the presence of backhaul capacity constraints. The results indicated that the different CoMP schemes should be selected in different scenarios to optimise the system energy efficiency. Furthermore, a distributed Wyner-Ziv coding schemes was investigated to forward compressed signals from the iSCs to the UEs. We mapped this architecture to the multiple-relay compress-and-forward (CF) problem, derive the achievable rate of such a system, and demonstrated that the achievable rate can be maximized by optimising the distributed compression rate at each individual coordinating point in a joint manner [99]. For both approaches the number of iSCs per veNB, i.e. veNB size, was assumed to be fixed. However, the veNB size can also be treated as an optimization objective to optimising the network performance. In the next, we will investigate the optimal number of iSCs in a veNB for joint DL transmission in terms of network sum-rate in the presence of latency that is mainly caused by the RANaaS processing delay and CSI feedback delay.

4.4.1 Technical description

Scenario

We consider the precoding matrix calculation at RANaaS, to benefit from its strong computational capability and from the fact that global CSI can be collected from each individual iSC through J1 links. Theoretically, larger joint processing veNB size (i.e. number of cooperating iSCs), gives the potential for better interference cancelation, thus, higher spectral efficiency gain can be achieved. However, this is generally not true in practice when factors such as latency in the network are taken into account. As discussed in Section 3.2.2.1, the complexity (therefore, RANaaS processing delay) for calculating the precoder matrix depends on the total number of transmit antennas of cooperating iSCs and also the number of users. This is especially true for more complicated channel estimators (e.g. MMSE) and precoders (e.g. ZF) whose complexity is at cubic order of the number of transmitting/receiving antennas. In addition, the amount of channel coefficients required for joint processing increases with the number of cooperating antennas and introduces a significant CSI feedback delay. All of this additional latency can cause a severe mismatch between the actual transmit channel and the channel coefficients used for the precoder calculation, consequently leading to performance and throughput degradation.

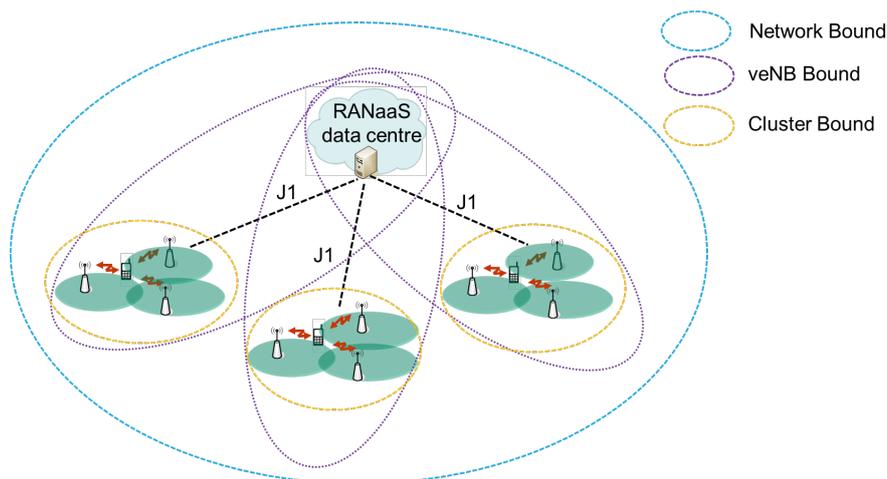


Figure 4-31: Grouping of iSCs to veNBs operated on a same RANaaS data centre

One feasible solution is to divide a network comprising a large number of iSCs into a set of veNBs with each one composing reasonable number of cooperative iSCs. The considered system scenario is depicted in Figure 4-31, where several cooperative iSCs along with a RANaaS constitute a veNB that operates joint DL transmission, while multiple veNBs share the same RANaaS data centre. In that case, each veNB only requires CSI between the iSCs and the UEs within the veNB and ZF precoding could be performed in the RANaaS separately for each veNB. Therefore, the computational complexity will be reduced. Meanwhile, the feedback amount will drop since low dimension ZF requires less CSI. Consequently, the introduced delay

could be mitigated; however, extra inter-veNB interference due to small veNB size may reduce system performance and throughput. Therefore, there must be an optimal veNB size that is not so small (to mitigate interference into a reasonable level) and not so large either (to save the performance loss due to the delay caused channel mismatch), which optimises system performance.

System Model

Let us consider a planar deployed network with an area of A_{net} (note for a network shaped as a circle, the volume $A_{net} = 2\pi R_t^2$ with R_t being the radius of the network), served by a RANaaS data centre. The network is evenly divided into N_{veNB} veNB domains, each with an area of:

$$a_{veNB} = \frac{A_{net}}{N_{veNB}} \quad (4-27)$$

Each veNB will implement joint transmission to mitigate the intra-veNB interference, while the transmissions among veNBs are independent and each UE will still experience interference from adjacent veNBs. In order to model the statistical dropping of UEs and iSCs within the covered area, we introduce the UE density ρ_{UE} and the iSC density ρ_{iSC} , with each UE and iSC being equipped with a single antenna.

Thus, the number of UEs $n_{UE,i}$ ($i = 1, 2, \dots, N_{veNB}$) and the number of iSC $n_{iSC,i}$ ($i = 1, 2, \dots, N_{veNB}$) per veNB is random with expected values given by [94]:

$$\begin{aligned} N_{UE} &= E\{n_{UE,i}\} = a_{veNB}\rho_{UE} \\ N_{iSC} &= E\{n_{iSC,i}\} = a_{veNB}\rho_{iSC} \end{aligned} \quad (4-28)$$

Inside each veNB, we suppose that each iSC is connected with the RANaaS by fibre or another link with sufficient capacity, while no direct link exists between the iSCs. Therefore each iSC in the veNB can only acquire a local CSI and the veNB global CSI is accumulated at the RANaaS by iSC feedback via the J1 links.

By considering both path-loss and fast fading effects of the wireless channel between iSC and UE, the channel from the j -th veNB to the u -th UE in the i -th veNB is given by:

$$\tilde{\mathbf{h}}_{i,j,u}[t] = \mathbf{h}_{i,j,u}[t]\mathbf{G}_{i,j,u} \quad (4-29)$$

$\mathbf{G}_{i,j,u}$ is a real-values diagonal matrix with dimension $n_{iSC,i} \times n_{iSC,i}$ denoting the path-loss and $\mathbf{h}_{i,j,u}[t]$ is a complex-valued matrix with dimension $1 \times n_{iSC,i}$ describes the fast fading factor of the channel with complex Gaussian distribution $CN(0,1)$. Similarly, the channel at $t - \Delta t$ can be expressed as:

$$\tilde{\mathbf{h}}_{i,j,u}[t - \Delta t] = \mathbf{h}_{i,j,u}[t - \Delta t]\mathbf{G}_{i,j,u} \quad (4-30)$$

Note that t and Δt denote the current time index and total CSI delay of the precoding operation, respectively. It is supposed that the channels are spatially uncorrelated and temporally correlated as follows [95]:

$$\mathbf{h}_{i,j,u}[t] = \lambda \mathbf{h}_{i,j,u}[t - \Delta t] + \mathbf{v}_{i,j,u} \quad (4-31)$$

where $\lambda = J_0(2\pi f_D \Delta t)$ is the temporal correlation factor with the zero-order Bessel function $J_0(\cdot)$ of first kind and f_D denoting the Doppler frequency. Here we assume that the temporal correlation factors are the same for different UEs and iSCs within one veNB. $\mathbf{v}_{i,j,u}$ is the channel mismatch error vector with each element having a complex Gaussian distribution of $CN(0, (1 - \lambda^2))$ [95]. Note for increasing delay Δt , λ decreases, therefore, the variance $1 - \lambda^2$ of the channel error elements increases. Consequently, the mismatch between the actual and observed channels increases. Substituting (4-31) into (4-30), we get:

$$\tilde{\mathbf{h}}_{i,j,u}[t] = \lambda \mathbf{h}_{i,j,u}[t-\Delta t] \mathbf{G}_{i,j,u} + \mathbf{v}_{i,j,u} \mathbf{G}_{i,j,u} \quad (4-32)$$

where $\tilde{\mathbf{h}}_{i,j,u}[t]$ and $\tilde{\mathbf{h}}_{i,j,u}[t-\Delta t]$ are the actual and the observed channels. Note that the calculation for the precoder is based on $\tilde{\mathbf{h}}_{i,j,u}[t-\Delta t]$, whereas $\tilde{\mathbf{h}}_{i,j,u}[t]$ describes the actual transmit channel. In the next, we will drop the time indexes $[t]$ and $[t-\Delta t]$ for brevity. The received signal of the u -th UE within the i -th veNB can be written as:

$$y_{i,u} = \tilde{\mathbf{h}}_{i,i,u} \mathbf{w}_{i,u} x_{i,u} + \sum_{l=1, l \neq u}^{n_{UE}} \tilde{\mathbf{h}}_{i,i,u} \mathbf{w}_{i,l} x_{i,l} + \sum_{j=1, j \neq i}^{N_C} \sum_{k=1}^{n_{UE}} \tilde{\mathbf{h}}_{i,j,k} \mathbf{w}_{j,k} x_{j,k} + n_{i,u} \quad (4-33)$$

where $\mathbf{w}_{i,u}$ is the complex-value precoding vector of dimension $n_{iSC,i} \times 1$ and $x_{i,u}$ denotes the signal for the u -th UE in the i -th veNB. Obviously, the precoding matrix $\mathbf{W}_i = [\mathbf{w}_{i,1}, \mathbf{w}_{i,2}, \dots, \mathbf{w}_{i,n_{UE,i}}]$ for all of the iSCs in the i -th veNB calculation is based on the observed channel $\tilde{\mathbf{h}}_{i,j,u}[t-\Delta t]$ for $u = 1, 2, \dots, n_{iSC,i}$. Here we have assumed that the power is evenly allocated to each UE in a veNB. All transmitted signals have unit power and the input SNR is varied by the noise power, i.e. the noise has a distribution of $n_{i,u} \sim CN(0, 1/SNR)$, where SNR stands for the input SNR.

In (4-33), the first item in the right-hand side of the equation is the desired signal for the u -th UE in the i -th veNB while the second and third items represent the intra-veNB and inter-veNB interference, respectively. From this equation, it becomes apparent that a mismatch arises since the precoding matrix is based on the observed delay channels in (4-31) instead of the actual channel in (4-29). Then, the output SINR of the u -th UE in the i -th veNB can be expressed as

$$SINR_{i,u} = E \left\{ \frac{|\tilde{\mathbf{h}}_{i,i,u} \mathbf{w}_{i,u} x_{i,u}|^2}{\left| \sum_{l=1, l \neq u}^{n_{UE}} \tilde{\mathbf{h}}_{i,i,u} \mathbf{w}_{i,l} x_{i,l} \right|^2 + \left| \sum_{j=1, j \neq i}^{N_C} \sum_{k=1}^{n_{UE}} \tilde{\mathbf{h}}_{i,j,k} \mathbf{w}_{j,k} x_{j,k} \right|^2 + \frac{1}{SNR}} \right\} \quad (4-34)$$

Approach

Signal and interference power without precoding

Before we consider any specific precoding algorithm and channel fast fading factor, we first derive the desired signal power and interference power without fast fading channel and precoding consideration. Thus it is assumed that all iSCs transmit signals without precoding (i.e. the precoding coefficient is set to be 1 for each iSC). In addition, the mismatch caused by temporal change and feedback delay is not taken into account yet. However, we will consider all of these factors in the next and we will show that the fast fading, delay caused error and specific precoding coefficients could be treated independently.

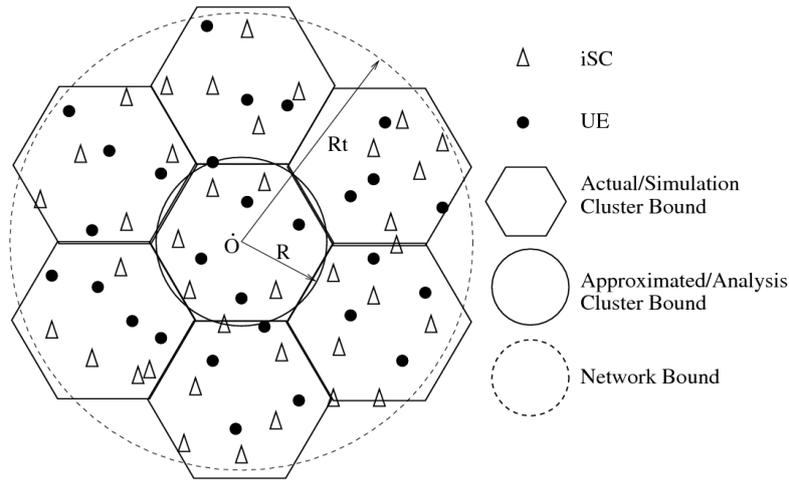


Figure 4-32: veNB division for planar deployment and each veNB is shaped as a hexagon

The network is divided into N_{veNB} veNBs with the same shape and an area of a_{veNB} . One of the practical shapes of a veNB is a hexagon in order to avoid adjacent veNBs overlap. However, as the hexagon boundary is difficult to analyse, the veNB boundary will be approximated by a circle at the same central point and the same area as shown in Figure 4-32. In addition, to approximate the circle bounded network of the veNBs, we will consider a network comprising full-tiers of veNBs, i.e. the number of veNBs in a network can only be $1+3n(n-1)$, where n is the number of tiers. Figure 4-32 gives an example of a 2-tier network that is composed of 7 veNBs. Note that even in this case, there will be an approximation error since the circular-area network (see the network bound by dot line in Figure 4-32) is not precisely overlapped by the hexagonal veNBs at the edge of the network. Nevertheless, networks with more veNB tiers will lead to a better approximation.

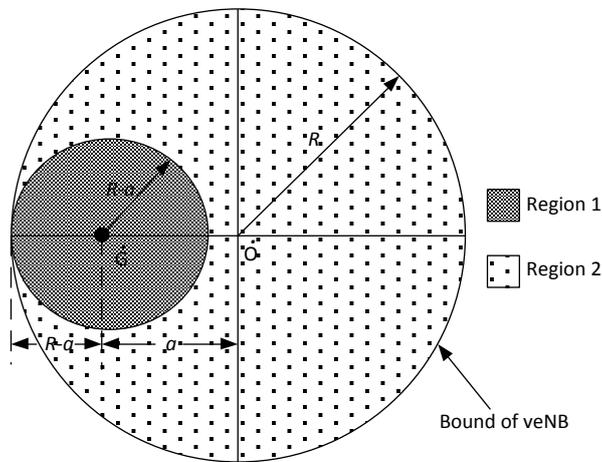


Figure 4-33: veNB area containing a single UE at \dot{G} and two contribution regions of the desired signal power

In the next, we will first propose a new approach for the desired signal and interference power calculation based on randomly distributed iSCs and UE. Let us first consider the desired signal power contributed by the cooperative iSCs in the circle (\dot{O}, R) shown in Figure 4-33 where R is the radius of the circle and \dot{O} is the circle centre. Without loss of generality, we assume an UE located at a point with distance $a \in [0, R]$ from the centre \dot{O} of the veNB, as depicted in Figure 4-33. For the iSCs that are randomly distributed in the circle (\dot{O}, R) with density ρ_{iSC} , the desired signal power at this UE can be found by integrating all power contributions from the iSCs. These contributions can be divided into two regions as shown in Figure 4-33 and is given by:

$$\bar{P}_x(a) = \bar{P}_{x1}(a) + \bar{P}_{x2}(a) = 2\pi\rho_{iSC} \int_0^{R-a} P_L(r, \kappa) r dr + 2\rho_{iSC} \int_{R-a}^{R+a} P_L(r, \kappa) r \arccos \frac{r^2 + a^2 - R^2}{2ra} dr \quad (4-35)$$

The first part in (4-35), $\bar{P}_{x1}(a)$, is attributed to the circular region $(\dot{G}, R-a)$ and the power could be conducted straightforwardly since it is a centrally symmetric region. All points that are at the same distance from the centre of circle \dot{G} (where the UE is located) will contribute the same power. The second part in (4-35), $\bar{P}_{x2}(a)$, is attributed to the remaining circular region of (\dot{O}, R) subtracting $(\dot{G}, R-a)$. Furthermore, $P_L(r, \kappa)$ is the path-loss with values given by

$$P_L(r, \kappa) = \begin{cases} R_0^{-\kappa} & r \leq R_0 \\ r^{-\kappa} & r > R_0 \end{cases} \quad (4-36)$$

where κ is the path-loss exponent and its typical value varies between 2 and 3.5 depending on the deployment scenarios. The micro-cellular model suggests a smaller value of $\kappa = 2$, and the macro-cellular model suggests a much larger value of $\kappa = 3.5$ [96]. Moreover, R_0 is a minimum distance between the UE and iSC for bounding the path-loss, i.e. the path-loss within the distance $[0, R_0]$ keeps constant. In some other channel path-loss modelling, an exclusive zero around the iSC may be introduced to bound the path-loss or a uniform closed-form expression is used [94]. Note here $P_L(r, \kappa)$ is another expression of the channel large scale factor $g_{i,j,u,k}$, which is the k -th diagonal element of the $\mathbf{G}_{i,j,u}$ defined in equation (4-29).

If we treat the UE location a as a random variable that is evenly distributed within the circle (\dot{O}, R) , the desired signal within the veNB radius can be expressed as

$$\bar{P}_x = \bar{P}_{x1} + \bar{P}_{x2} = \int_0^R \varphi(a) [\bar{P}_{x1}(a) + \bar{P}_{x2}(a)] da = \int_0^R \frac{2a}{R^2} [\bar{P}_{x1}(a) + \bar{P}_{x2}(a)] da \quad (4-37)$$

where we define

$$\varphi(a) = \frac{2a}{R^2}, \quad 0 \leq a \leq R \quad (4-38)$$

as the probability density function of a .

Equation (4-37) gives the desired signal power expression. On the other hand, the expression for the realised interference power cannot be straightforwardly derived since the integral region is irregular and, in addition, its value depends not only on the UE location in a veNB, but also on the location of the veNB in the network. However, the interference power can be approximated by evaluating the power received from the whole network and subtracting the desired signal power. Thus, we supposed that all of the veNBs are in the centre of the network and suffer similar inter-veNB interference since all of the networks are surrounding by other networks. On the other hand, when the network contains many tiers of veNBs, the UE is roughly in the centre of the network. Then the total receive power of the UE is

$$\bar{P}_{tot} = 2\pi\rho_{iSC} \int_0^{R_t} P_L(r, \kappa) r dr = \pi\rho_{iSC} R_0^2 + 2\pi\rho_{iSC} \frac{R_t^2 - R_0^2}{2 - \kappa} \quad (4-39)$$

and the interference power contributed by the area outside of the veNB can be expressed as:

$$\bar{P}_I = \bar{P}_{tot} - \bar{P}_x \quad (4-40)$$

Output SINR with delay and precoding

So far, we have evaluated the desired signal and interference powers for the case where all of the iSCs transmit omni-directionally without precoding. Both of them have been expressed as a function of the veNB size. Subsequently, a specific precoding matrix to combat the fast fading factor of the channel in the presence of CSI latency will be proposed.

Based on the delayed version of channel vector (4-32), the precoding vectors will be calculated on the RANaaS separately. Considering ZF precoding, the precoding matrix for the u -th UE in the i -th veNB can be given as

$$\mathbf{w}_{i,u} = \frac{\hat{\mathbf{H}}_i^H (\hat{\mathbf{H}}_i \hat{\mathbf{H}}_i^H)^{-1} \mathbf{e}_u}{\|\hat{\mathbf{H}}_i^H (\hat{\mathbf{H}}_i \hat{\mathbf{H}}_i^H)^{-1} \mathbf{e}_u\|} \quad (4-41)$$

where $\hat{\mathbf{H}}_i = [\hat{\mathbf{h}}_{i,i,1}, \hat{\mathbf{h}}_{i,i,2}, \dots, \hat{\mathbf{h}}_{i,i,n_{UE}}]$ contains all considered channel vectors $\hat{\mathbf{h}}_{i,j,u}$ considered for calculating the precoders. For brevity the vector $\hat{\mathbf{h}}_{i,j,u}$ was introduced to denote the delayed version of $\tilde{\mathbf{h}}_{i,j,u}$, i.e. $\hat{\mathbf{h}}_{i,j,u}[t] = \tilde{\mathbf{h}}_{i,j,u}[t - \Delta t]$. Furthermore, $\mathbf{e}_u = [0, \dots, 1, \dots, 0]^T$ denotes a $n_{UE,i} \times 1$ vector with its u -th element being 1 and the rest being zero. Substituting (4-41) and (4-33) into (4-34) and applying some stochastic analysis and derivations, the ergodic output SINR based on ZF precoding can be expressed as a function of the veNB size as following:

$$SINR \approx \frac{\frac{[N_{iSC} - N_{UE}] \lambda^2 + 1}{N_{iSC}} \bar{P}_x}{\frac{(N_{UE} - 1)(1 - \lambda^2)}{N_{iSC}} \bar{P}_x + \frac{N_{UE}}{N_{iSC}} \bar{P}_I + \sigma^2} \quad (4-42)$$

Note that this SINR depends not on the actual channel realization, but only on the temporal correlation factor λ being a function of the CSI delay Δt .

In equation (3-55) of Section 3.2.2.1 this delay was derived for the case of only one cooperative veNB with given number of iSCs and UEs. In order to extend the analysis for multiple veNBs with a stochastic dropping of UEs and iSCs, equation (4-28) is used to express the total number of UEs and iSCs as a function of the veNB size. As the multiple veNBs share the same computational resources as well as time-frequency resources for data transmissions the CSI delay for the precoding processing Δt in equation (3-55) can be rewritten as

$$\Delta t = N_{ant_tot} \Delta t_{chan_est_UE} + N_{fb_time} (\Delta t_{fb} + \Delta t_{prop_total}) + \Delta t_{precoder_cal} + \Delta t_{Tx} + \Delta t_{CSI_fd} \quad (4-43)$$

where the five items reflect the channel estimation delay at the UE, the CSI feedback and propagation delay, the processing delay at the RANaaS (precoding matrix calculation delay), the total transmission chain delay, and the CSI forward delay at the iSC, respectively. Using the assumption $N_R^{UE} = N_T^{iSC} = 1$ for simplicity, the total number of channels coefficients (3-47) that one UE needs to estimate due to the specified iSC density ρ_{iSC} equals

$$N_{ant_tot} = \pi R^2 \rho_{iSC} \quad (4-44)$$

The number of CSI feedbacks (3-48) becomes

$$N_{fb_time} = \frac{\pi R^2 \rho_{iSC} A_{net} \rho_{UE}}{N_{CSI_per_fb}} \quad (4-45)$$

with $N_{CSI_per_fb}$ specifying the number of channels coefficients that can be fed back every time and with UE density being ρ_{UE} . Finally, the delay caused by calculating the precoder matrix at the RANaaS defined in (3-54) is changed to

$$\Delta t_{precoder_cal} = \frac{(K_{add} + \zeta_2 K_{multi})}{F_{FLOPS} q_c} \quad (4-46)$$

with

$$K_{multi} = 8\pi^2 R^4 A_{net} \rho_{iSC} \rho_{UE}^2 + O(4\pi^2 R^4 A_{net} \rho_{UE}^2) + 2\pi A_{net} R^2 \rho_{UE} \rho_{iSC} \quad (4-47)$$

and

$$K_{add} = 8\pi^2 R^4 A_{net} \rho_{iSC} \rho_{UE}^2 - 2\pi A_{net} R^2 \rho_{UE}^2 - 2A_{net} \rho_{UE} - 2\zeta_1 \pi R^2 A_{net} \rho_{UE} \rho_{iSC} + O(4\pi^2 R^4 A_{net} \rho_{UE}^2) \quad (4-48)$$

describing the number of real-valued operations for addition and multiplication. Again, ζ_1 and ζ_2 are the equivalent addition operation times for each division and multiplication operations, respectively. Furthermore, F_{FLOPS} and q_c are the RANaaS computational capability and computational resource division factor as introduced in Section 3.2.2.1. Thus, the expressions derived in Section 3.2.2.1 have been modified to contain multiple veNBs, where the number of iSCs and UEs per veNB is a function of the UE and iSC density. Since λ is a function of the CSI delay Δt and Δt could be expressed by the veNB size, we use $\lambda(R)$ in the sequel. Substituting (4-28) into (4-42), the expected SINR per UE can be written in terms of the veNB size as

$$SINR \approx \frac{[\pi R^2 (\rho_{iSC} - \rho_{UE}) N_{iSC}] \lambda(R)^2 + 1}{\pi R^2 \rho_{iSC}} \frac{\bar{P}_x}{(\pi R^2 \rho_{UE} - 1)(1 - \lambda(R)^2) \bar{P}_x + \frac{\pi R^2 \rho_{UE}}{\pi R^2 \rho_{iSC}} \bar{P}_l + \sigma^2} \quad (4-49)$$

Thus, the optimization problem for maximizing the sum-rate of the network in terms of veNB size can be expressed as

$$\max_R \eta_s, \quad \text{subject to } 0 < R < R_t \quad (4-50)$$

where R_t is the radius of the network and the sum-rate η_s is based on the Gaussian transmitting signal, it could be written as

$$\begin{aligned} \eta_s &= E \left\{ \sum_{i=1}^{N_c} \sum_{u=1}^{N_{UE}} \log_2(1 + SINR_{i,u}) \right\} \\ &= N_c E \left\{ \sum_{u=1}^{N_{UE}} \log_2(1 + SINR_u) \right\} \\ &= \rho_{UE} A_{net} E \left\{ \log_2(1 + SINR) \right\} \\ &\leq \rho_{UE} A_{net} \log_2(1 + E \{ SINR \}) \end{aligned} \quad (4-51)$$

The second equation in (4-51) is based on the assumption that the network is surrounded by other non-overlapping networks with the same configuration. In other words, it implies that each veNB could be treated as the central veNB in a network, therefore, equally contributing to the sum-rate. The third equation in (4-51) holds when each UE is randomly distributed in the network and independently dropped, where the total number of UEs is calculated in terms of the density and network area, i.e. $E \{ N_{veNB} N_{UE} \} = N_{veNB} E \{ N_{UE} \} = \rho_{UE} A_{net}$. Since it is very difficult to optimize the ergodic sum-rate directly due to the expectation operation being outside of the logarithm, the well-known Jensen's inequality is used in the fourth step of (4-51), i.e. the optimization will be based on the upper bound of the ergodic sum-rate. Although the upper bound of sum-rate is a loose bound in some cases the optimization problem solutions, match very well as will be seen at the respective evaluation section.

The optimization problem in (4-51) is approximately equivalent to

$$\max_R \text{SINR}, \quad \text{subject to } 0 < R < R_t \quad (4-52)$$

Since the cost function at the aforementioned optimisation is extremely complex in terms of veNB radius R , it is difficult to obtain the optimal solution in close form. Thus, numerical methods will be adopted in Section 4.4.3 to verify its effectiveness.

4.4.2 Implementation in the iJOIN architecture

The proposed CT relies on centralized precoding at the RANaaS to maximise the sum-rate in terms of veNB size in the presence of delays caused by RANaaS processing and CSI feedback. This implies that only functional split A.1, A.2 and A.3 described in Section 3.2.1 are practical. Otherwise, not only the data, but also the precoding matrix shall be transmitted to the individual iSCs from the RANaaS, which may significantly increase the load of backhaul in fast fading channels. Therefore, we present the backhaul load only for the appropriate A.1, A.2 and A.3 functional split defined in Section 3.2.1.1. However, as the three function split options will result in the same cooperative processing gain, we will focus on the functional split A.3 subsequently.

Functional split A.3

The functional split A.3 for CT2.4 is shown in Figure 4-34, where the precoding matrix and precoded signal calculation are operated in RANaaS, and then the precoded signals are transmitted to the iSCs for further RE mapping and IFFT operation via J1 link.

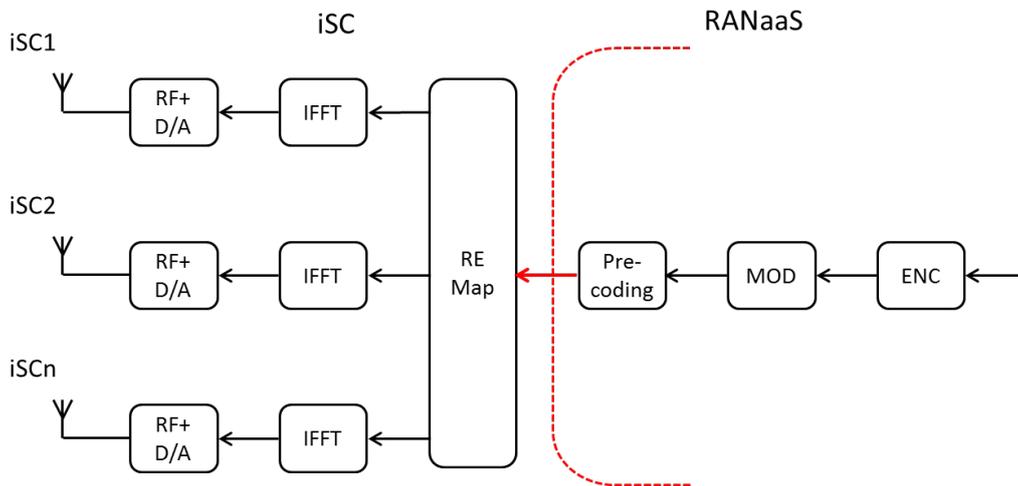


Figure 4-34: functional split A.3 for CT2.4

To implement the centralized precoding at RANaaS, the estimated CSI will be first fed back from the UEs to the iSCs via the radio access links, and then will be forwarded to the RANaaS via J1 links. With those CSIs and network configuration parameters (i.e. available computational resources, CSI feedback rate, etc.), the optimal veNB size will be calculated based on the approach proposed in Section 4.4.1. The ZF-based precoding on the optimal veNB size will be implemented in RANaaS. The precoded signal will be sent over the backhaul link and accords to the principle calculation approach proposed in equation (3-6) in Section 3.2.1.1, the backhaul load can be expressed as

$$D_P^{A.3} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot N_{sc}^i \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (4-53)$$

where $N_{sc}^i = N_{sc,u}^i$ is the number of PRBs allocated to all users served by the i -th veNB. Notice that it is currently assumed, that for joint transmission all UEs in one veNB are served by the same set of subcarriers $S_{sc,u}^i$ to implement multi-user transmission. Thus the total number of subcarriers is given by $N_{sc}^i = N_{sc,u}^i = |S_{sc,u}^i|$. The payload transmission from RANaaS to iSC over J1 link is assumed subframe-based, which means that the message transmission frequency is T_{SUB}^{-1} .

Similarly, following the functional split A.3 in Section 3.2.1.1, the signal load for the backhaul can be given as

$$D_S^{A.3} = N_{RB} \quad [\text{bit/iSC}] \quad (4-54)$$

The backhaul loads (both data and signal) of functional split A.1 and A.2 in our CT are exactly the same as the counterparts expressed in Section 3.2.1.1, respectively. Thus, the equations are omitted here.

Backhaul Load per Functional split

In each functional split, the RANaaS implements the global ZF precoding to mitigate the intra-veNB interference based on the feedback CSI from iSCs. This information has to be estimated by the UEs, fed back from the individual UEs to the iSCs, and finally forwarded over the J1 links to the RANaaS as summarized in Table 4-14.

Table 4-14: PHY Measurement placement and exchange links for CT2.4

Functional Split Approach	Measurement	Measurement Location	Exchange link
A.1	CSI	UE	UE → iSC → RANaaS
A.2	CSI	UE	UE → iSC → RANaaS
A.3	CSI	UE	UE → iSC → RANaaS

For all functional splits, some additional control signals and system parameters, as detailed in the Input/Output tables Table 6-1 and Table 6-2 of Appendix I, have also to be exchanged over the J1 links. The respective control information has been summarized into Table 4-15.

Table 4-15: Backhaul control information of CT 2.4

Link	Description
iSC → RANaaS	PHY Information (access link) <ul style="list-style-type: none"> - I2.10: TxCSI, noise variance, and path-loss of each iSC-UE link - O2.6: Effective SNR/SINR of each UE (with precoding) - O2.10: Access link capacity and energy efficiency
RANaaS → iSC	Network parameters <ul style="list-style-type: none"> - I2.3: RAN connection table of iSCs and UEs - I2.4: BH connection table for iSCs-RANaaS links
	RRM Information <ul style="list-style-type: none"> - I2.6: RRM information per UE - I2.7: MCS for backhaul link
	Network parameters <ul style="list-style-type: none"> - I2.1: System Parameter

The information of BH payload for different functional splits discussed above has been summarized in Table 4-16. Note that the total CSI (note that we only consider channel coefficients) exchange between each iSC and RANaaS is depending on the total number of channels $N_T^{iSC} N_R^{UE} N_{UE}$ between the iSC and corresponding UEs, more channels leads to larger message size. It is also depending on the granularity of the precoding, i.e. how many subframes/symbols in time domain and how many subcarriers in frequency domain use the same precoding matrix and the granularity for precoding is also discussed in Section 3.2.2.2. Here we assume that the estimation and feedback is done in every symbol and subcarrier. Other factors that affect the number of CSI feedback amount are the total allocated subcarriers $N_{sc,u}^i$ for those UEs and number of quantization bits N_{Q_CSI} for each channel coefficient. Here the CSI exchange frequency and latency are set the same as the payload for consistence. The CSI feedback message size from each iSC to RANaaS has been shown in Table 4-15.

Table 4-16: Backhaul Load of CT 2.4

Link	Description	Msg Size	Msg Frequency	Latency
Functional Split A.1				
iSC → RANaaS	CSI	$N_T^{iSC} N_R^{UE} N_{UE} N_{Q_CSI} N_{sc,u}^i N_{SYMB}^{SUB}$	T_{SUB}^{-1}	<0.1ms
RANaaS → iSC	TD I/Q samples	$2 \cdot N_T^{iSC} \cdot N_Q \cdot OF \cdot (N_{FFT} + N_{CP}) \cdot N_{SYMB}^{SUB}$	T_{SUB}^{-1}	<0.1ms
Functional Split A.2				
iSC → RANaaS	CSI	$N_T^{iSC} N_{iSC} N_R^{UE} N_{UE} N_{Q_CSI} N_{sc,u}^i N_{SYMB}^{SUB}$	T_{SUB}^{-1}	<0.1ms
RANaaS → iSC	FD I/Q samples before IFFT	$D_P^{A,2} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot N_{sc} \cdot N_{SYMB}^{SUB}$	T_{SUB}^{-1}	<0.1ms
Functional Split A.3				
iSC → RANaaS	CSI	$N_T^{iSC} N_{iSC} N_R^{UE} N_{UE} N_{Q_CSI} N_{sc,u}^i N_{SYMB}^{SUB}$	T_{SUB}^{-1}	<0.1ms
RANaaS → iSC	FD I/Q samples before resource mapping	$2 \cdot N_{sc,u}^i \cdot N_T^{iSC} \cdot N_Q$	T_{SUB}^{-1}	<0.1ms

4.4.3 Evaluation of the CT

Compliance with iJOIN objectives

In the presence of RANaaS processing and CSI feedback delay, the CT formulated a sum-rate optimisation problem in terms of the cluster size to trade off the channel mismatch and interference. Specifically, the optimal number of iSCs within a veNB (or veNB size) for cooperative DL transmission is solved to achieve the optimal sum-rate. Correspondingly, this CT addresses the objective “area throughput” and can potentially address the objective “energy efficiency”.

Description of the baseline used for the evaluation

We investigate the proposed optimization problem and the analytical results, which will be verified by simulations. The maximum sum-rate achieved by the optimal veNB size will be compared to the sum-rate with other non-optimal veNB sizes. The baseline corresponds to the case without cooperation among iSCs. However, the performance of the baseline is absent at the current simulations and will be provide along with the evaluations on iJOIN testbed in the next stage.

Description of the evaluation methodology

The evaluation for the proposed algorithms, in the first instance, is assumed based on a server consisting of an Intel Xeon Processor E5-2680, where the delay caused by the RANaaS processing is replaced by the theoretical values in equation (4-43). However, by implementing the proposed algorithm for the RANaaS testbed at UoB, the actual delay will be evaluated in the future.

In our MATLAB simulations, 2000 iSCs and 1000 active UEs are uniformly distributed in a circle with a network radius $R_t = 500$ meters, which corresponding to $\rho_{iSC} = 0.0025$ iSC/ m² and $\rho_{UE} = 0.00125$ UE/m². All of the cooperative veNBs in the network will share the same resource elements. We assume that all UEs experience the same noise power and the input SNR is set to $SNR = 30$ dB. The path-loss exponent equals $\kappa = 2.2$. A Rayleigh fast fading factor is considered for the channel between iSC and UE by equation (4-30). The temporal correlation of the channel is modelled by (4-31) with the Doppler spread $f_D = 10$ Hz for all iSC to UE links.

For the delay model derived in the Section 3.2.2.1, we use the worse-case transmit processing time defined in 3GPP, i.e. $\Delta t_{TX} = 2.33$ ms [15]. The channel estimation delay and small cell processing delay are set to zero since they are negligible compared to the processing and feedback delay, i.e. $\Delta t_{chan_est_UE} = 0$. For $\Delta t_{process_cal}$, we take the typical values $\zeta_1 = 15$ and $\zeta_2 = 1$ [100], respectively. The distance between UEs

and iSCs plus the distance between iSCs and RANaaS data centre is considered 1000 meters. The RANaaS platform consists of an Intel Xeon Processor E5-2680 with computational capability $F_{FLOPS}=344\text{GFLOPS}$, which corresponds to 172 double-precision DB-GFLOPS (i.e. $1.72E+11$ FLOPS). The value is calculated as follows: the 8-core Intel Xeon Processor E5-2680 has a CPU frequency $2.7E+9$ Hz and 2 operations per clock period, it supports 256-bit advanced vector extensions. Therefore, the computational capability is $8 \times 2.7 \times 10^7 \times 256 / 64 = 1.72 \times 10^{11}$ DP-FLOPS. In the simulations, we assume the addition operation is double-precision floating point (64-bit). In addition, the computational resources are shared by multi-tasks and multiple sub-carriers of the networks. Therefore the available computational capability for precoding calculation is $1.72E+11/q_c$ addition operations per second. The factor q_c is set to be 10, otherwise specified. The feedback rate of the back-haul link from iSC to RANaaS is $C_{fb} = 10^7$ bits per second and $N_{CSI_per_fb} = 1$ to avoid interference.

Discussion of results

Based on different cloud processing capabilities (q_c changing from 1/2, 10, 40, to 80), Figure 4-35 and Figure 4-36 depict the output SINR and the sum-rate of the ZF-based precoding in terms of different veNB sizes. From the figures we can see that each curve calculated by the proposed theoretic method roughly matches the corresponding simulation results and the peak point (optimal veNB size) are strictly overlapped to each other. One should notice that the peak points tend to move towards to the right as the latency reduces (computational capacity factor changing from 1/2 to 80). This could be explained by the fact that when q_c goes up, the system is affected more by the latency and smaller veNB size is needed to improve the delay-caused channel mismatch and achieve the optimal solution.

On the other hand, the peaks in Figure 4-35 and Figure 4-36 show the optimal veNB sizes for maximizing the output SINR and maximizing sum-rate are overlapped. This essentially means that although the Jensen's inequality gives the upper bound of the sum-rate, the optimal veNB size is not affected by the approximation in (4-49), i.e. the optimization in equations (4-51) is accurate. Obviously, the theoretic sum-rate for each q_c is higher than the simulation counterparts due to the fact that Jensen's inequality gives an upper bound of the sum-rate.

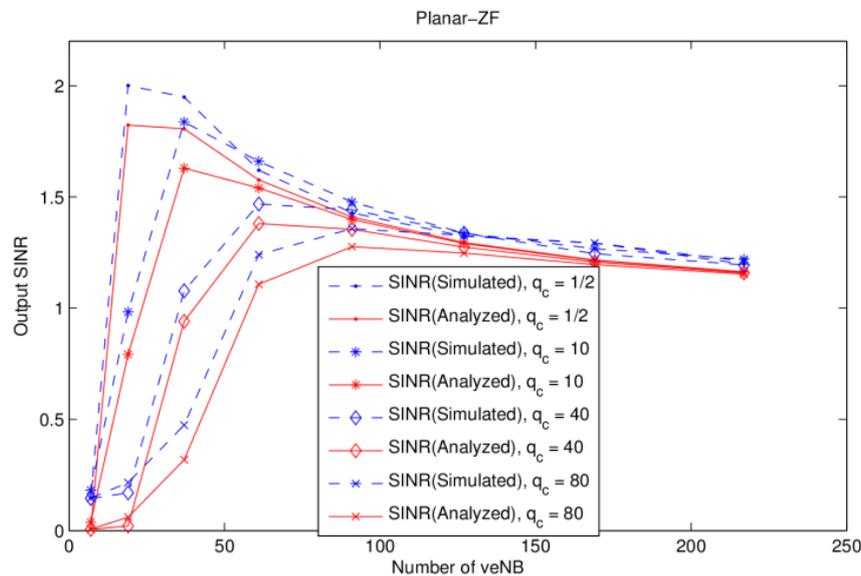


Figure 4-35: Output SINR for ZF-based planar deployment versus available computational resource factor q_c

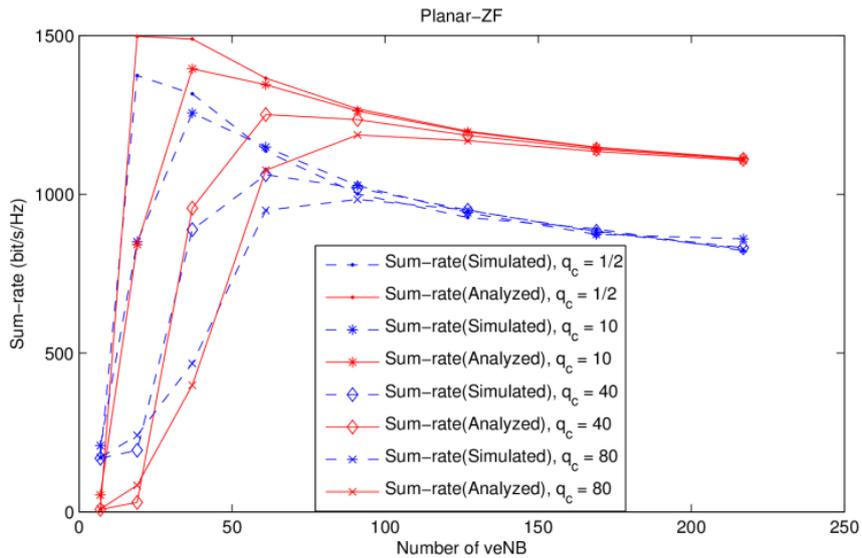


Figure 4-36: Sum-rate for ZF-based planar deployment versus available computational resource factor q_c

The time consumption at each optimal configuration (veNB size) and the contribution by each component are shown in Figure 4-37. We can see that in the optimal value the contribution from the feedback delay decreases sharply. However, the total delay is not monotonically reducing with q_c ; from left to right, the optimal veNB size goes down, which will release the required cloud computational resource according to equation (4-43), leading to a non-monotonic delay.

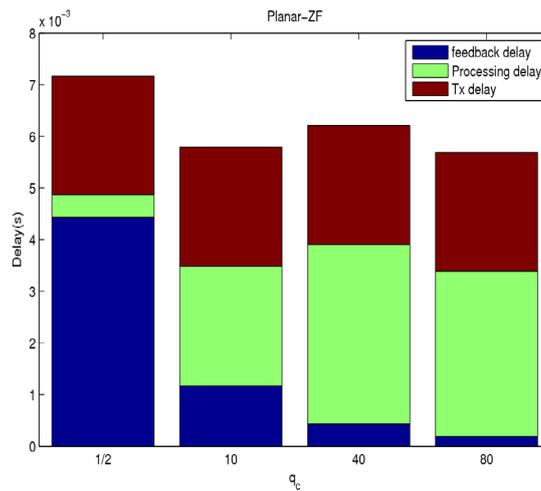


Figure 4-37: The total delay and their component versus available computational resource factor q_c

4.5 CT2.5: Partially Centralized Inter-Cell Interference Coordination

4.5.1 Technical description

A flexible and rich backhaul topology with both J1 and J2 links as shown in Figure 4-38 provides many opportunities for improving the efficiency of the joint transmission CoMP in the downlink, in particular by exploiting the cooperation of the iSCs. However, it also opens many interesting and difficult questions.

Indeed, when a central node is available, it is commonly assumed that centralized precoding is realized, and hence the numerous results from the extensive literature on that topic can be applied (see [1] and references therein). When only limited J2 links are available for exchanging the CSI, it is then only possible to rely on distributed precoding. Due to the imperfection of the J2 links (quantization, delay, ...), this gives rise to a so-called distributed CSI configuration [22], [23] where each iSC has to decide its transmit parameters based on its own local information.

In the case of a general and flexible backhaul topology, finding the optimal precoding strategy, or even an efficient one, which adapts to any scenario is a completely open and challenging question. In particular, it is not known whether the backhaul links should be exploited to carry the channel estimates (explicit channel state information at transmitter (CSIT)) or to carry the precoded signals (implicit CSIT), or a mix of the two.

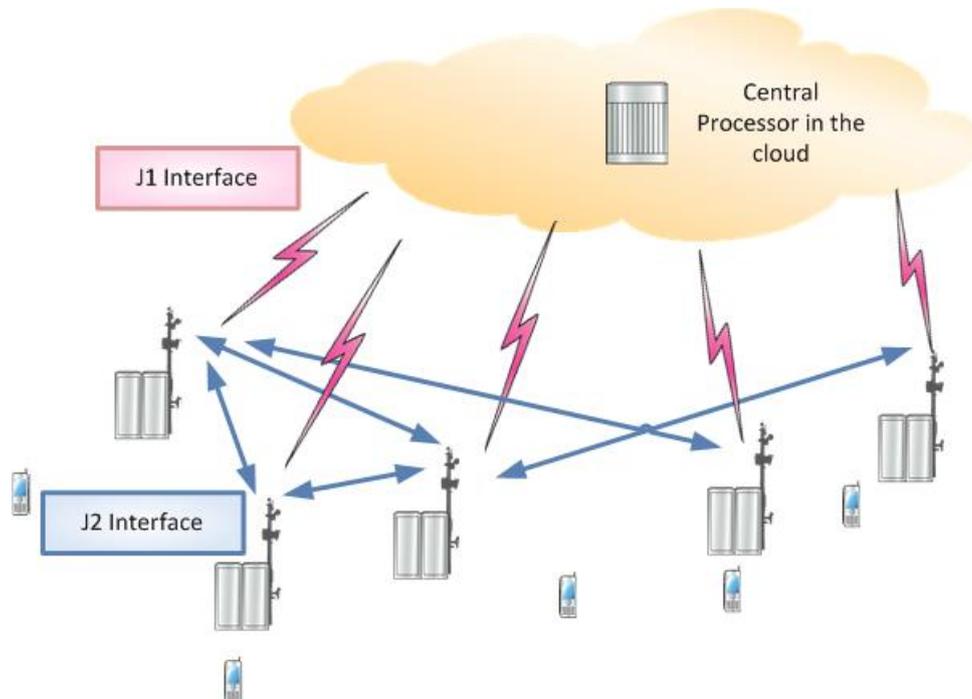


Figure 4-38: iJOIN architecture with both J1 links between the RANaaS and the iSCs and J2 links between the iSCs

The general goal of our approach is to investigate precoding methods which adapt to any backhaul topology and which exploit in the most efficient way the available backhaul resources. As the discussion of physical layer constraints in Section 3.4 illustrates, the most critical issue is the timely exchange of the CSI (or the precoded signals) between the cooperating nodes, and we will therefore focus on that issue.

System Model

In this deliverable, we focus on a CSI configuration, which is denoted as the “hierarchical CSI configuration” [24], and is an interesting candidate for enforcing efficient coordination between iSCs. In this CSIT configuration, it is possible to order the iSCs by increasing level of CSIT accuracy. In practice, this means that according to this order, an iSC is able to reconstitute the CSI estimate is available at the preceding iSCs and, hence, to do the same signal processing as this iSC. This CSI configuration is well known in the team decision community [25] where different actors aim at maximizing a common objective on the basis of individual information, and also denoted as “nested CSI configuration” [27]. This CSI

configuration is practically interesting as it has the benefit of enforcing coordination at a lower cost than the general CSI configuration.

This setting is in fact very similar to the scenario where an iSC simply receives through the J2 link the output of the processing done at the preceding iSCs. Indeed, in both cases, the important aspect is that an iSC is aware of the signal processing done at the preceding iSC. The two possible scenarios are presented in Figure 4-39.

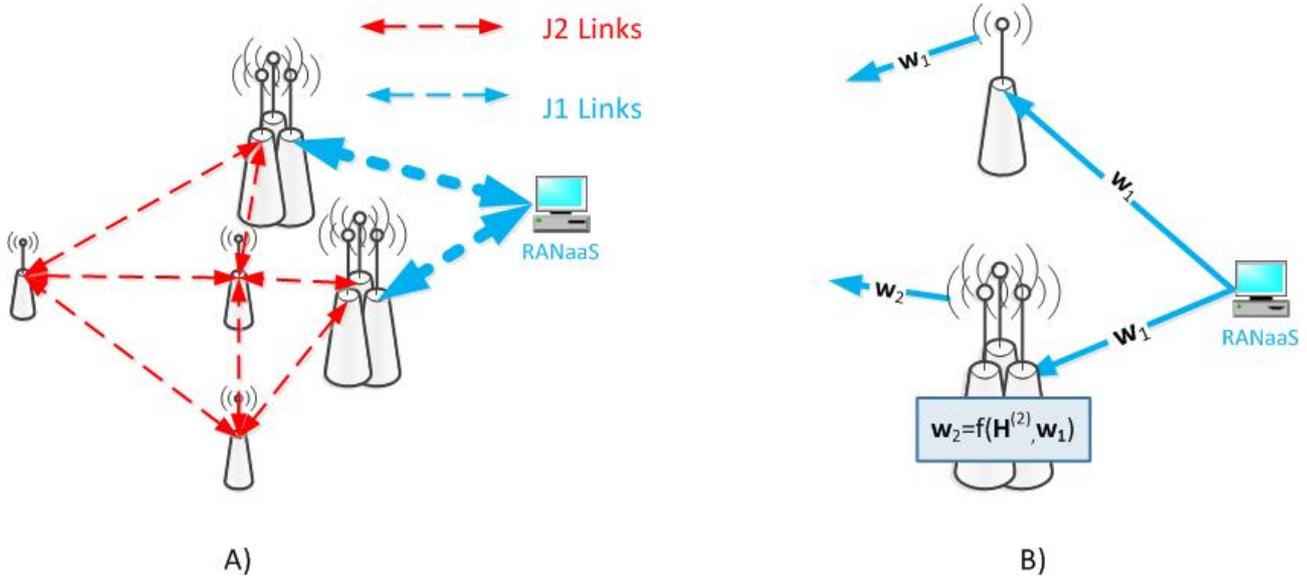


Figure 4-39: Two possible practical scenarios leading to a hierarchical CSIT configuration.

This CSI configuration is used to model the limited exchange of information through the J2 links. Indeed, the multi-user channel estimate available locally is obtained after an exchange protocol during which the iSCs share the CSI feedback from the UEs. With infinite backhaul resources, the iSCs are able to exchange perfectly the CSI and there is no need for robust precoding. Considering imperfect backhaul links, it becomes then necessary to design a robust precoder. More precisely, we consider a setting with N_{iSC} iSCs and N_{UE} UEs. We assume for the sake of clarity that all the iSCs have the same number of antennas which is denoted by N_T^{iSC} . Similarly, we consider that each UE has a number of antennas equal to N_R^{UE} . The multi-user channel matrix from all the iSCs to all the UEs is then denoted by \mathbf{H}^H and written as

$$\mathbf{H}^H = \begin{bmatrix} \mathbf{H}_1^H \\ \vdots \\ \mathbf{H}_{N_{UE}}^H \end{bmatrix}. \quad (4-55)$$

with the channel from all the iSCs to UE u equal to \mathbf{H}_u^H of size $N_R^{UE} \times (N_T^{iSC} \cdot N_{iSC})$. The received signals at the UEs is then written as

$$\mathbf{y} = \mathbf{H}^H \mathbf{x} + \mathbf{n}. \quad (4-56)$$

where \mathbf{x} is the signal vector transmitted jointly by all the iSCs, \mathbf{y} is the vector made of the stacked received signals at all the UEs, and \mathbf{n} is a vector containing the additive white Gaussian noise of unit variance at all the UEs.

The multi-user transmitted signal \mathbf{x} is then obtained from the multi-user symbol vector \mathbf{s} after multiplication by the multi-user precoder \mathbf{W} . Following distributed precoding, the signal emitted at iSC j , denoted as \mathbf{x}_j is written as

$$\mathbf{x}_j = \mathbf{W}_j^H \mathbf{s} \quad (4-57)$$

such that the multi-user precoder \mathbf{W} is given by

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1^H \\ \vdots \\ \mathbf{W}_{N_{iSC}}^H \end{bmatrix}. \quad (4-58)$$

We consider in the rest of this contribution a distributed CSIT configuration such that iSC j receives its own estimate of the multi-user channel state, denoted by $\mathbf{H}^{(j)}$. Based on this multi-user CSI estimate, iSC j then designs its transmit coefficients \mathbf{W}_j . As described above, we focus in fact on a particular so-called hierarchical CSI configuration where the iSCs can be ordered such that one iSC has more information relative to the channel state than the preceding one. We assume from now on that the iSCs are ordered according to this particular order such that iSC j has received a more accurate CSI than iSC k if $k < j$.

In this CSI configuration, the iSCs aim at cooperatively maximizing the expected sum rate *without any further exchange of information* between the iSCs. The team decision optimization problem can then be written as

$$(w_1^*, \dots, w_{N_{iSC}}^*) = \arg \max_{(w_1, \dots, w_{N_{iSC}})} E[R(w_1(\mathbf{H}^{(1)}), \dots, w_{j-1}(\mathbf{H}^{(j-1)}), w_j(\mathbf{H}^{(j)}), w_{j+1}(\mathbf{H}^{(j+1)}), \dots, w_{N_{iSC}}(\mathbf{H}^{(N_{iSC})}))] \quad (4-59)$$

where w_j is a precoding function and R denotes the sum rate function when the transmit coefficients taken in argument as used for the transmission.

Approach

This optimization problem considered is a very intricate optimization problem which falls in the category of the team decision problems [25]. A usual approach to deal with team decision problems is to consider the best-response optimization which allows attaining a locally optimal solution.

The best response optimization consists in finding precoding functions satisfying [26]

$$w_j^{BR} = \arg \max_{w_j} E[R(w_1^{BR}(\mathbf{H}^{(1)}), \dots, w_{j-1}^{BR}(\mathbf{H}^{(j-1)}), w_j(\mathbf{H}^{(j)}), w_{j+1}^{BR}(\mathbf{H}^{(j+1)}), \dots, w_{N_{iSC}}^{BR}(\mathbf{H}^{(N_{iSC})}))] \quad (4-60)$$

In the case of hierarchical CSI, this problem can then be further simplified since the signal transmitted by iSC l to iSC $j-l$ can be obtained at iSC j and hence are already given in the point of view of the optimization problem considered. It remains then only to take the expectation over the precoding functions of iSC $j+1$ to iSC N_{iSC} .

Although this problem has already been simplified significantly, it remains a difficult optimization problem as it is still a stochastic optimization problem.

As an answer, we propose in this CT a heuristic solution consisting in letting iSC j approximate the optimization problem by considering that it can in fact decide what is transmitted by the remaining iSCs. The optimization problem is then approximated as

$$w_j^{BR} = \arg \max_{w_j, \dots, w_{N_{iSC}}} E[R(w_1^{BR}(\mathbf{H}^{(1)}), \dots, w_{j-1}^{BR}(\mathbf{H}^{(j-1)}), w_j(\mathbf{H}^{(j)}), w_{j+1}^{BR}(\mathbf{H}^{(j)}), \dots, w_{N_{iSC}}^{BR}(\mathbf{H}^{(j)}))] \quad (4-61)$$

The optimization problem becomes then very similar to the conventional robust precoder design as studied in [29]. The only difference being that part of the precoder is already fixed. Indeed, when considering the precoding at iSC j , the precoding coefficients at iSC k , with $k < j$ are *already* fixed since following the hierarchical CSIT configuration, it can be seen as a received signal to iSC j .

We can then apply the precoding algorithm from [28] with only little change. The details of the precoding algorithm can be found in [24].

4.5.2 Implementation in the iJOIN architecture

The proposed CT relies on distributed precoding at the iSCs to enforce more robustness with respect to imperfect CSIT and delay in the backhaul. Hence, this CT requires the precoding to be done at the iSCs. This

means that only functional splits B.1, B.2, and C as described in Section 3.2.1 are adapted to the use of the proposed hierarchical precoding scheme.

Table 4-17: PHY Measurement placement and exchange links for CT2.5

Functional Split Approach	Measurement	Measurement Location	Exchange link
Functional Split B.1	Downlink CSI, eff. SNR/SINR per UE	UE	UE → iSC via feedback channel iSC → iSC via J2 links
Functional Split B.2	Downlink CSI, eff. SNR/SINR per UE	UE	UE → iSC via feedback channel iSC → iSC via J2 links
Functional Split X	Downlink CSI, eff. SNR/SINR per UE	UE	UE → iSC via feedback channel iSC → iSC via J2 links

Table 4-18: Backhaul control information of CT 2.5

Link	Description
iSC → RANaaS	PHY Information (access link) <ul style="list-style-type: none"> - I2.10: TxCSI, noise variance, and path-loss of each iSC-UE link - O2.6: Effective SNR/SINR of each UE (with precoding) - O2.10: Access link capacity and energy efficiency
RANaaS → iSC	Network parameters <ul style="list-style-type: none"> - I2.3: RAN connection table of iSCs and UEs - I2.4: BH connection table for iSCs-RANaaS links
	RRM Information <ul style="list-style-type: none"> - I2.6: RRM information per UE - I2.7: MCS for backhaul link
	Network parameters <ul style="list-style-type: none"> - I2.1: System Parameter

The physical measurements and their exchange through the backhaul links are shown in Table 4-17. The most important exchange of information consists in the exchange of the timely CSI estimates between the iSCs. This exchange is important to compute the precoder distributedly at the iSCs. The CSI has to be exchanged to reach a hierarchical CSI configuration, or it has to be complemented by the exchange of the computed precoder in an iterated manner between the iSCs. The other measurement which is done consists in the measurement of the general informations of the networks which are then feedback to the RANaaS in order to do the RRM. We also show in Table 4-18 the information exchanged between the iSCs and the RANaaS. Following the same pattern, the iSC receives all the information obtained as output of the RRM at the RANaaS while the iSCs send back the necessary information for this RRM step.

As mentioned before, this CT requires distributed precoding to be done at the iSCs. Thus, only the functional splits B.1, B.2, and C can be applied. The detailed calculations for the payload signals are given in Table 4-19. However, this CT is focused on the sharing of the CSI such that the most important parameter will be the exchange of the CSI. Hence, each iSC requires obtaining the full multi-user channel for each of the frequency slot in which the iSC is transmitting. This corresponds to $\sum_{u=1}^{N_{UE}} N_{sc,u}^J$ by definition. This is then

multiplied by the number of OFDM symbols N_{SYMB}^{SUB} . Finally, the quantization has to be done for each scalar element of the channel matrix and both for the complex and the real component, which leads to the remaining coefficients.

The message frequency and the maximum delay depend on the channel delay. It has to be done frequently enough and with a delay such that the estimate used for the precoding is timely and accurate. We recommend therefore as a rule of thumb to transmit the message every $0.1 T_{50}$ seconds since T_{50} corresponds by definition to the delay after which the channel correlation is below 50%, i.e., after which the channel has significantly changed. Such an update allows to always keep a sufficiently accurate estimate. However, if this is frequency is possible, the proposed CT is not really necessary as it exactly aims at coping with less accurate and less timely CSI estimates. In LTE systems with users moving at pedestrian speed, this typically corresponds to a frequency of sharing in the order of several ms and the same holds for the maximum delay.

As the focus of this CT, we do not detail the calculation of the other informations exchanged. This detailed description can be found along with the detailed description of the different functional splits in Section 3.2.1.1. The frequency of the message exchange follows directly from the calculation in Section 3.2.1.1 and is therefore not further discussed. The latency however depends on the protocol and the transmission scheme used. Determining the latency requirements and study how to relax them is a topic of current investigation.

Table 4-19: Backhaul Load of CT 2.5

Link	Description	Msg Size	Msg Frequency	Latency
Functional Split B.1				
iSC → iSC	CSI exchange between iSCs	$2 N_R^{iSC} N_{iSC} N_Q N_{SYMB}^{SUB} \sum_{u=1}^{N_{UE}} N_{sc,u}^j$	$0.1/T_{50}$	$0.1 \cdot T_{50}$
RANAAS → iSC	Data Load	$D_P^{B.1} = 2 \cdot N_L \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB}$ [bit/iSC]	T_{SUB}^{-1}	Under investigation
Functional Split B.2				
iSC → iSC	CSI exchange between iSCs	$2 N_R^{iSC} N_{iSC} N_Q N_{SYMB}^{SUB} \sum_{u=1}^{N_{UE}} N_{sc,u}^j$	$0.1/T_{50}$	$0.1 \cdot T_{50}$
RANAAS → iSC	Data Load	$D_P^{B.2} = \sum_{u=1}^{N_{UE}} N_{sc,u}^j \cdot Q_m \cdot N_L \cdot N_{SYMB}^{SUB}$ [bit/iSC]	T_{SUB}^{-1}	Under investigation
	Modulation information	$D_{S,MOD}^{B.2} = N_{MOD}$ [bit/user]	T_{SUB}^{-1}	Under investigation
	RRM information	$D_{S,RA}^{B.2} = \lceil \log_2(N_{RB} \cdot (N_{RB} + 1)) \rceil$ [bit/user]	T_{SUB}^{-1}	Under investigation
Functional Split C				
iSC → iSC	CSI exchange between iSCs	$2 N_R^{iSC} N_{iSC} N_Q N_{SYMB}^{SUB} \sum_{u=1}^{N_{UE}} N_{sc,u}^j$	$0.1/T_{50}$	$0.1 \cdot T_{50}$
RANAAS → iSC	Data Load	$D_P^{B.2} = \sum_{u=1}^{N_{UE}} N_{sc,u}^j \cdot Q_m \cdot N_L \cdot N_{SYMB}^{SUB}$ [bit/iSC]	T_{SUB}^{-1}	Under investigation
	RRM information	$D_{S,RA}^{B.2} = \lceil \log_2(N_{RB} \cdot (N_{RB} + 1)) \rceil$ [bit/user]	T_{SUB}^{-1}	Under investigation
	MCS information	$D_{S,MCS}^C = N_{MCS}$ [bit/user]	T_{SUB}^{-1}	Under investigation

4.5.3 Evaluation of the CT

Compliance with iJOIN objectives

We study in this CT the ergodic throughput of the multi-user setting in the CSI configuration described above. In particular, the method proposed aims at improving the robustness of the transmission when the backhaul links are too limited for centralization. In our approach the backhaul resources available are used to exchange CSI through the logical J2 links, thus allowing for cooperation between the iSCs via the joint precoder. Thus, this CT provides an improvement of the area throughput for the case in which the backhaul links have a significant delay or are too weak. This improvement of the area throughput is also related to an improvement of the energy efficiency in the sense that it is possible to reduce the power used to achieve given area throughput requirements.

Description of the baseline used for the evaluation

The baseline is obtained when using conventional sum-rate maximizing algorithms [28] from the literature applied in a distributed manner at the iSCs. Hence, the partial centralization is not taken into account in the precoding scheme used which leads to performance degradations. The area throughput achieved will also be compared to the area throughput achieved in a fully centralized setting with infinite backhaul. This will help us evaluate the cost of having a limited backhaul.

Discussion of results of the CT

To evaluate the performance of the proposed hierarchical precoding algorithm, we show the ergodic rate achieved in a simple wireless configuration with $N_{iSC} = N_{UE} = 4$ and $N_R^{UE} = N_T^{iSC} = 1$. We furthermore assume a Rayleigh fading environment with all the wireless links having a unit-variance. One of our main task for the coming months will be to improve the simulation environment to obtain an evaluation of the performance in a more realistic scenario and in agreement with the common scenario parameters described in the Appendix II.

We consider a simple CSI configuration where two iSCs receive a perfect estimate of the full multi-user channel \mathbf{H} while the two other iSCs receive an estimate corrupted with an additive white Gaussian noise of variance 0.25. To evaluate the efficiency of the robust hierarchical precoding proposed, we compare its performance to the sum rate achieved when perfect instantaneous CSI is available at all iSCs. This corresponds solely to a theoretical upper bound as providing perfect instantaneous CSIT at every iSC or at a central node is not possible. We also compare our algorithm to the performance obtained when conventional robust precoding from the literature is applied at each iSC.

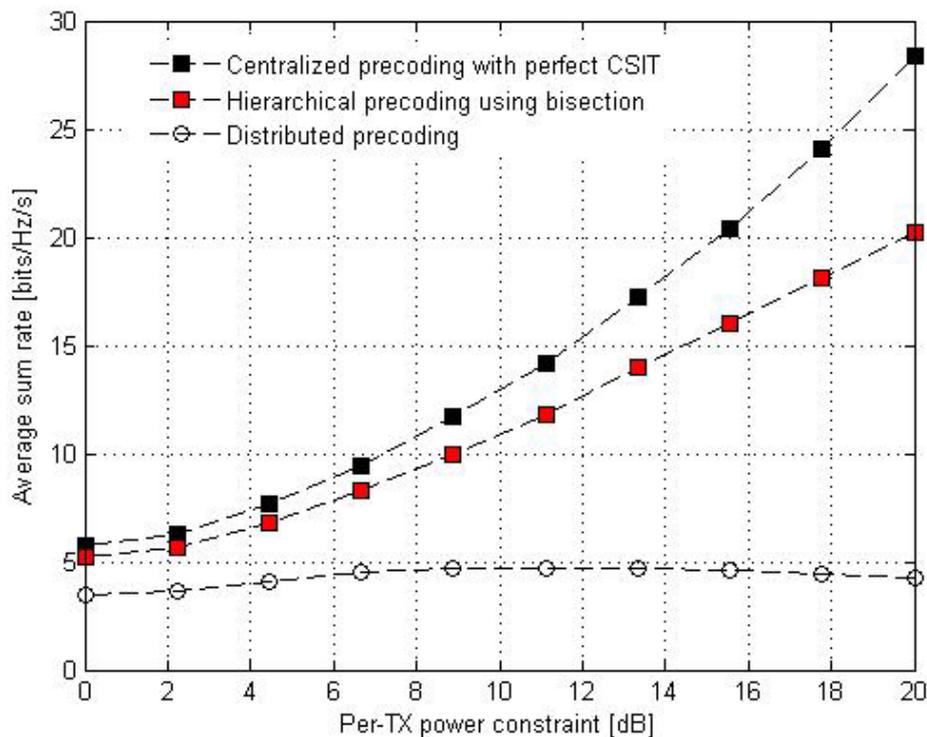


Figure 4-40: Ergodic sum rate achieved in a wireless network with 4 single-antennas iSCs and 4 single-antennas UEs with only two iSCs having access to a perfect CSI

It can be seen in Figure 4-40 that the proposed hierarchical precoding algorithm outperforms significantly the conventional transmission scheme with distributed use of the sum rate algorithm [29]. In particular, the strong improvement comes from the fact that the proposed hierarchical precoding algorithm is able to achieve an unlimited increase of the sum rate with the transmit SNR. This is due to the fact that two of the iSCs have access to a perfect CSI and are able to exploit this information in a coordinated manner.

4.6 CT2.6: Data Compression over RoF

4.6.1 Technical description

The CT2.6 provides a general investigation about the possible functional splits at the physical layer between RANaaS and iSCs and thus also supports the other CTs for the analysis of the possible DL and UL approaches. In particular, the main goal of CT2.6 is to reduce the amount of data that must be transmitted on the BH. The considered functional splits are placed at an intermediate level between the fully centralized case, commonly denoted in literature as centralized RAN (C-RAN) [85] and the fully distributed solution, where all the physical layer functionalities are implemented at the iSCs side. A detailed analysis of all the considered functional split options is provided in section 3.2.1.

Scenario

The CT2.6 primarily addresses the problem of the high capacity required by the backhaul network in a dense small cell deployment scenario. In particular this problem is addressed by analysing the functional splits that leveraging on the structure of the transmitted signals permit a reduction of the backhaul load.

Approach

The used approach aims to identify by means of an analytical investigation the functional splits that best combine the characteristics listed in the following:

- backhaul load reduction compared to the fully centralized solution
- possibility to perform statistical multiplexing of the backhaul load generated by different iSCs (characteristic not applicable in the fully centralized case with time domain transmission)
- possibility to implement centralized coordination algorithms

4.6.2 Implementation in the iJOIN architecture

The impact of CT2.6 on the vNB concept is essentially related to the physical layer functional split. The frequency domain transmission, characterized by the execution of the FFT/IFFT operations at the iSC side, basically represents an internal interface of the vNB that is also applicable to backhaul technologies different than fibre (e.g. wireless). In the following one possible functional split in the frequency domain that combines the characteristics discussed before is analysed.

Functional Split A.3

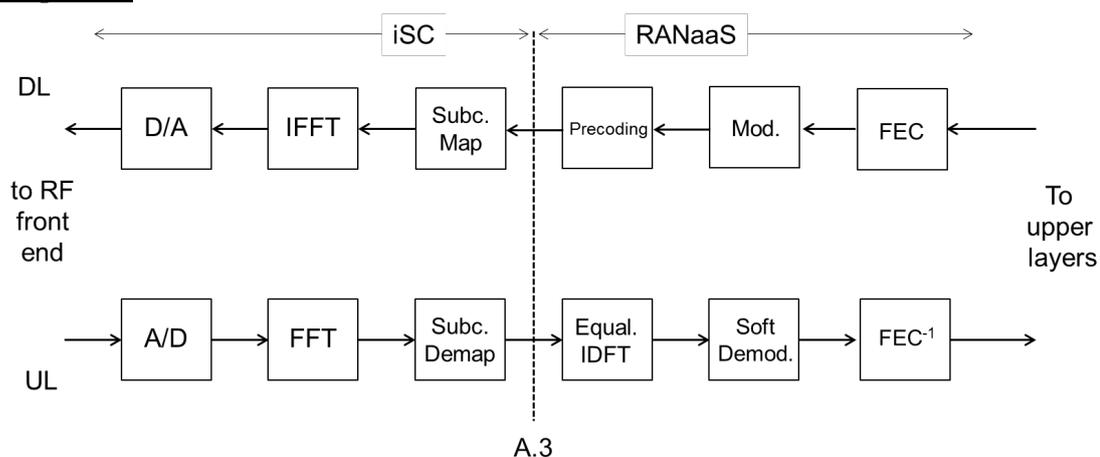


Figure 4-41: Functional split A.3

The functional split shown in Figure 4-41 places the RANaaS/iSC boundary at the input of the resource mapping block in the downlink. Symmetrically the boundary is placed at the resource demapping output in the uplink. The resource mapping/demapping and FFT/IFFT operations are therefore executed at the iSC side so that only the allocated transmission resources need to be transmitted on the backhaul. This split provides a significant reduction of the backhaul load compared to the functional split A.1 with I/Q time domain

transmission and enables also the statistical multiplexing among multiple iSCs that share one or several hops of the backhaul network. Besides, it entails the centralization in the cloud of some uplink physical layer functions that are most computationally intensive, like channel estimation, multi-user detection and FEC decoding. The downlink precoding weights are calculated and applied centrally at the RANaaS, allowing the implementation of centralized interference coordination algorithms.

A further aspect related to the functional split in Figure 4-41 is the generation and multiplexing directly at the iSC side of specific downlink channels or signals that carry content with no (or infrequent) variation in time, thus providing a further reduction of the backhaul load.

The functional split in Figure 4-41 where the backhaul load is proportional to the radio interface load opens also the possibility to design the backhaul capacity below the full load case (i.e. all iSCs with full resource utilization at the same time). Clearly, the design of the backhaul capacity below the full load case brings advantages in terms of scalability and costs but is paid by a certain level of blocking probability as a function of the expected traffic distribution in space and time.

Backhaul Load per Functional split

For the functional split option A.3 the downlink precoding and the uplink detection/decoding are executed in the RANaaS. The CSI measurement is therefore located in the RANaaS as summarized in Table 4-20.

Table 4-20: PHY Measurement placement and exchange links for CT2.6

Functional Split Approach	Measurement	Measurement Location	Exchange link
A.3: IFFT/FFT and mapping demapping at iSC	Downlink CSI	RANaaS	N/A
	Uplink CSI	RANaaS	N/A

The backhaul control information and the backhaul load for the payload part are provided in Table 4-22 and Table 4-22, respectively.

Table 4-21: Backhaul control information of CT 2.6

Link	Description
iSC → RANaaS	RRM Information <ul style="list-style-type: none"> - O2.2: Estimated BH parameter (load)
	PHY Information (access link) <ul style="list-style-type: none"> - O2.8: User data to upper layers, if ACK
RANaaS → iSC	Network parameters <ul style="list-style-type: none"> - I2.1: System Parameter - I2.3: RAN connection table of iSCs and UEs
	PHY information (access and BH link) <ul style="list-style-type: none"> - I2.8: RxCSI, SNR, noise variance of UE-iSC links - I2.9: Received signal at iSC - I2.11: User data to be transmitted in DL - I2.13: Parameter of iSC-RANaaS link - I2.15: Parameter of RANaaS-iSC link
	CT specific control information <ul style="list-style-type: none"> - I2.5: Functional control for shifting/splitting functions - I2.6: RRM information per UE

Table 4-22: Backhaul Load of CT 2.6 for payload

Link	Description	Msg Size	Msg Frequency	Latency
Functional Split A.3				
iSC → iSC	N/A	N/A	N/A	N/A
iSC → RANaaS	I/Q in the frequency domain after resource demapping	$D_P^{A.3} = 2 \cdot N_R^{iSC} \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB}$	$1/T_{SUB}$	<0.1 ms (to satisfy HARQ timing in the LTE case)
RANaaS → iSC	I/Q signal in the frequency domain before resource mapping	$D_P^{A.3} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB}$	$1/T_{SUB}$	<0.1 ms (to satisfy HARQ timing in the LTE case)

4.6.3 Evaluation of the CT

Compliance with iJOIN objectives

The Data Compression over RoF aims at reducing the throughput requirements on the backhaul links by means of a specific partitioning of the baseband modem functionalities between RANaaS and iSCs keeping at the same time some advantages related to the centralization of specific physical layer functions. The data rate reduction is obtained without loss of information. Correspondingly, CT2.6 addresses the objective “area throughput” and can potentially address the “cost efficiency”, considering that BH capacity and related cost are in general correlated. Besides, also the “utilization efficiency” can be positively affected when considering the possibility of performing the statistical multiplexing of the traffic generated by different iSCs over BH links.

Description of the baseline used for the evaluation

The baseline used for evaluating the backhaul load reduction provided by the CT2.6 is a fully centralized architecture based on Common Public Radio Interface (CPRI) [14], where the RANaaS performs all the physical layer operations up to the generation of the composite digital baseband OFDMA signal in the time domain and that corresponds to the split option A.1.

Discussion of results of the CT

The backhaul load for the functional split A.3 shown in Figure 4-41 is calculated with the formulas given in section 3.2.1. A LTE cell equipped with MIMO 2x2 and using a B=20 MHz channel is considered. The downlink payload message size is calculated using the equation (3-5), reported below for simplicity

$$D_P^{A.3} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot \left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (4-62)$$

Considering the case of a fully loaded cell, we then have the number of used subcarriers

$$\left(\sum_{u=1}^{N_{UE}} N_{sc,u}^j \right) = 1200 \quad (4-63)$$

and assuming a number of quantization bits $N_Q = 7$, $N_T^{iSC} = 2$ iSC transmit antennas the downlink BH payload message size is equal to

$$D_P^{A.3} = 2 \cdot 2 \cdot 7 \cdot 1200 \cdot 14 = 470400 \quad [\text{bit/iSC}] \quad (4-64)$$

that corresponds to a BH throughput of about $\eta = 470$ Mbit/s without including the BH protection overhead γ . For the uplink the equation (3-29), symmetrical to the downlink, provides the same value of 470400 bit/iSC, under the same assumption of $N_Q = 7$.

Concerning signalling, the resource allocation information is required at the iSC to perform the resource mapping/demapping operations. By using the equation (3-9) it is possible to calculate the signalling message size

$$D_S^{A.3} = N_{RB} = 100 \quad [\text{bit/iSC}] \quad (4-65)$$

Considering that the transmission rate is equal to the reciprocal of the subframe period T_{SUB} , in case of B=20 MHz bandwidth the BH signalling throughput is then equal to

$$\eta_S^{A.3} = D_S^{A.3} \cdot \frac{1}{T_{SUB}} = \underbrace{100}_{\text{message size}} \cdot \underbrace{\frac{1}{0.001}}_{\text{frequency}} = 100 \text{ kbit/s} \quad (4-66)$$

indicating that the resource allocation information requires only a small fraction of the backhaul capacity compared to the payload while, on the other hand, it enables the statistical multiplexing of the load generated by different iSCs.

The backhaul load of the functional split A.3 can be compared with the fully centralized solution, corresponding to the split option A.1 of Figure 3-3. Using the equation (3-1), together with the parameter values discussed in section 3.2.1

$$D_P^{A.1} = 2 \cdot N_T^{iSC} \cdot N_Q \cdot OF \cdot (N_{FFT} + N_{CP}) \cdot N_{SYMB}^{SUB} \quad [\text{bit/iSC}] \quad (4-67)$$

namely $N_T^{iSC} = 2$, $N_Q = 15$, $OF = 1$, $N_{FFT} = 2048$, $N_{CP} = 144$, we then obtain

$$D_P^{A.1} = 2 \cdot 2 \cdot 15 \cdot 1 \cdot (2048 + 144) \cdot 14 = 1841280 \quad [\text{bit/iSC}] \quad (4-68)$$

that corresponds to a BH throughput of about $\eta = 1.84$ Gbit/s without including the BH protection overhead γ . The BH throughput reduction factor, in full load condition, of the functional split A.3 is therefore equal to 3.9 times and increases as the radio interface load decreases, due to the fact that the BH load of the split option A.1 is constant and does not depend on the radio interface load.

The impact of the considered functional split on the area throughput metric [17] can be estimated using the model described in [83]. In particular, making an assumption of the available BH capacity per area C_{BH} expressed in Gbit/s/km² and given the required backhaul capacity per iSC $\eta \cdot \gamma$, it is possible to determine the maximum iSC density (ρ_{iSC}), expressed in iSCs/km²

$$\rho_{iSC} = \frac{C_{BH}}{\eta \cdot \gamma} \quad [\text{iSCs/km}^2] \quad (4-69)$$

The assumption of a limited BH capacity per area holds for example in case of wireless backhaul with specific network topologies (e.g. tree, ring, chain) or also for other BH technologies, like fibre, when is it difficult or even not possible to add further capacity.

The spatially averaged rate per user R in bit/s/Hz is then determined using the analytical model given in [83], which is derived on the base of the Shannon capacity, assuming iSCs equipped with omnidirectional antennas that transmit a power P and using a bandwidth B

$$R = \frac{\pi^{5/2}}{2} \sqrt{\frac{\rho_{UE} \rho_{iSC} P}{\sigma^2}} \cdot \text{erfc} \left(\frac{\pi^2 \rho_{UE}}{4} \sqrt{\frac{P}{\sigma^2}} \right) \cdot \exp \left(\frac{\pi^4 \rho_{UE}^2 P}{16 \sigma^2} \right) \quad [\text{bit/s/Hz}] \quad (4-70)$$

and where ρ_{UE} is the assumed user density expressed in UE/km² and σ^2 is the thermal noise power. The exponential term in equation (4-70) usually dominates the other terms. If, however $\rho_{UE} \cdot \sqrt{P/\sigma^2} > 4/\pi^2$, as it occurs in an interference limited system, the spatially averaged rate R can be approximated by

$$R \approx 2 \cdot \sqrt{\frac{\rho_{iSC}}{\rho_{UE}}} \quad [\text{bit/s/Hz}] \quad (4-71)$$

Finally the area throughput R_A is calculated using the equation below

$$R_A = \rho_{UE} \cdot R \cdot B \quad [\text{bit/s/km}^2] \tag{4-72}$$

The Figure 4-42 and Figure 4-43 show the results of the application of the model described above. The functional split A.3 is compared with the fully centralized physical layer, corresponding to split option A.1 of Figure 3-3. A uniform small cell deployment, with the iSCs placed at the centre of hexagonal cells is assumed in the model. In particular the area throughput R_A and the iSC density ρ_{iSC} are calculated as a function of the available backhaul capacity C_{BH} per area. The iSC density for both split options is calculated assuming that the backhaul capacity provided to each iSC is sufficient to support the full load condition. In order to make a fair comparison the same BH protection overhead $\gamma = 4/3$ is assumed for both split options.

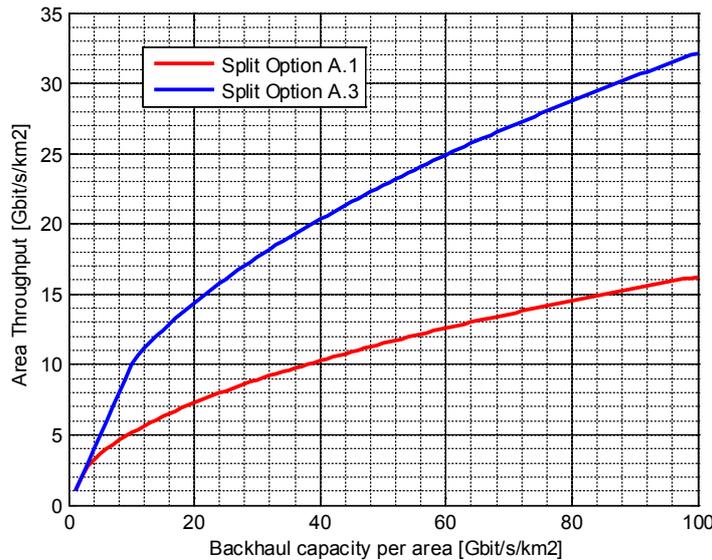


Figure 4-42: Area throughput versus available BH capacity per area in Gbit/s/km²

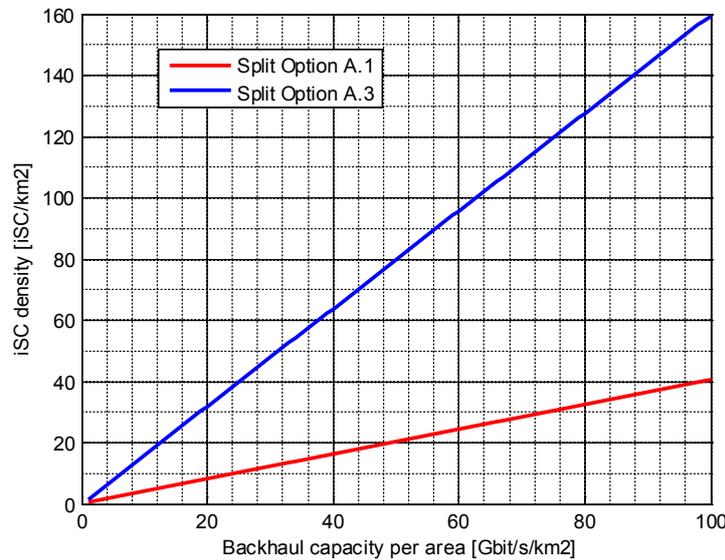


Figure 4-43: iSC density versus available BH capacity per area in Gbit/s/km²

The user density ρ_{UE} is assumed uniform and equal to 5000 UE/km² in the two cases [63]. The transmit power P of each iSC is set to 30 dBm and the equivalent noise bandwidth is 18 MHz. The thermal noise power σ^2 at the UE side is calculated assuming a receiver noise figure F equal to 9 dB [63].

The results show that the backhaul data rate reduction associated to the functional split option A.3 allows a increase of the iSC density for a given BH capacity per area, with respect to the split option A.1. The network densification achievable with the split option A.3 turns therefore into a corresponding increase of

the area throughput metric. As the available BH capacity per area increases, also the number of deployable iSC increases and in parallel the number of served users by each iSC reduces (assuming a uniform user distribution). It can be also observed that when overall BH capacity per area becomes lower than a given threshold, it occurs that the BH capacity provided to each iSC becomes lower than the aggregated data rate on the radio interface (i.e. the BH becomes the bottleneck for the radio interface), and this effect is captured by the slope change of the area throughput curves in Figure 4-42.

4.7 CT2.7: Millimetre Wave Backhauling

4.7.1 Technical description

Scenario

Backhaul technologies play an important role in the iJOIN architecture. Centralized processing requires the BH links to provide high capacity, high reliability and low latency. At the same time, they have to be cheap and energy efficient. Fibre is the technology currently used for that purpose. However, it is expensive to deploy and might not even be deployable at every location [13]. Therefore, CT2.7 investigates mmWave wireless links as an alternative.

Wireless links in the mmWave bands above 60 GHz offer very high capacity, but at a limited range, so they are only suitable for the “last mile” of the backhaul network [12]. Long distance backhaul of about 10 km is currently also possible but requires large antennas that are not suitable for small cell deployments. We therefore investigate an architecture as depicted in the bottom part of Figure 4-44, featuring a RANaaS that is directly connected to an iSC via a mmWave link. However, other architectures like multi-hop links or additional iTNs where the backhaul switches from wireless to fibre-based links, can be also supported by most approaches discussed by CT2.7. Remarks on this can be found in the corresponding sections below.

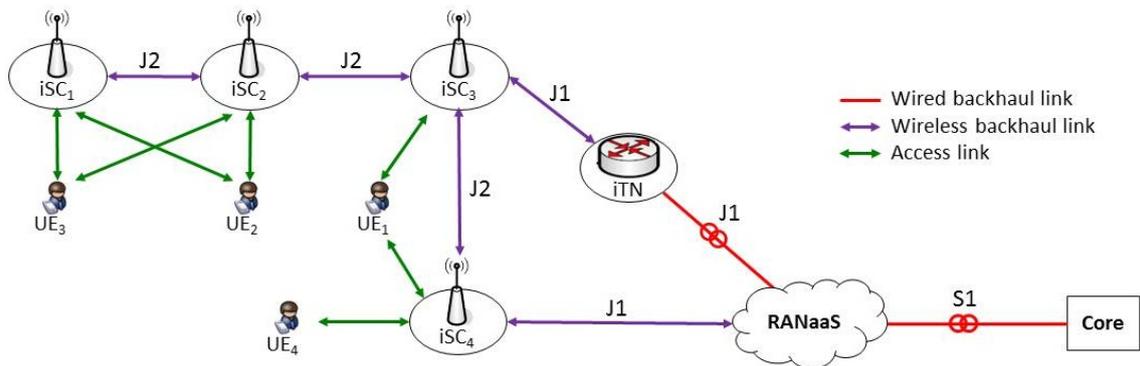


Figure 4-44: Wireless BH architecture

In principle, mmWave links can carry any digital information, meaning that they can be used to enable any functional split as described in Section 3.1 in either the UL or the DL. However, this CT focuses firstly, on the UL, and secondly, on a lower layer BH where the decoding is located in the RANaaS. This means that (pre-processed) digital I/Q samples are forwarded from the iSCs to the RANaaS.

System Model

For the joint RAN/BH investigations, we model a single link system comprising of one UE, one iSC and the RANaaS instance. This is depicted in Figure 4-45

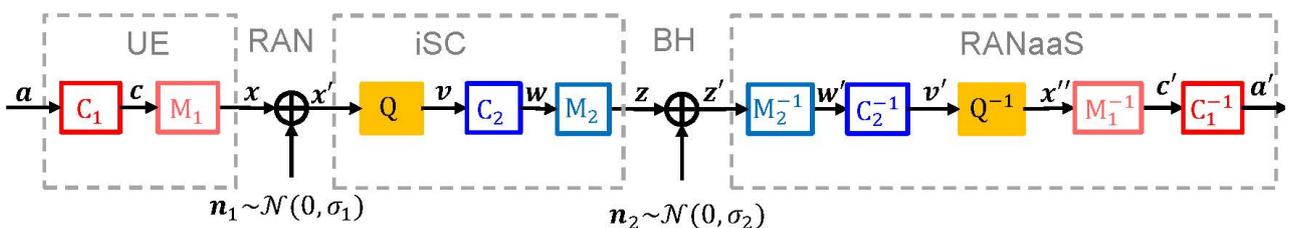


Figure 4-45: CT2.7 system model

The UE encodes its user bits \mathbf{a} using encoder C_1 with rate R_c^{RAN} to a codeword \mathbf{c} which is then modulated by modulator M_1 to one or multiple symbols \mathbf{x} . These symbols are sent over the RAN link to the iSC and are received as \mathbf{x}' . The RAN is depicted as an AWGN channel of a certain $SNR_{\text{RAN}} = \text{var}(\mathbf{x}) / \sigma_1^2$. However, in cases where we use a frequency selective Rayleigh channel, a MMSE equalizer will be located either before \mathbf{x}' or before \mathbf{x}'' , depending on the functional split options discussed below.

In the iSC, the received signal is sampled and the samples are digitalized by a quantizer Q to a bit-vector \mathbf{v} according to a codebook. These bits can optionally be encoded by a second encoder C_2 with rate R_c^{BH} to a codeword \mathbf{w} that is modulated by M_2 to one or multiple symbols \mathbf{z} .

These symbols are sent over the BH link and are received as \mathbf{z}' . The BH link is modelled as an AWGN channel of a certain $SNR_{\text{BH}} = \text{var}(\mathbf{z}) / \sigma_2^2$, which is a valid assumption for highly directive, LOS mmWave links.

The received symbols are demapped and optionally decoded to receive the bit vector \mathbf{v}' . This bit-vector is used by dequantizer Q^{-1} to reconstruct \mathbf{x}'' by choosing the corresponding entries from the quantization codebook. The reconstructed received symbols are then demapped by M_1^{-1} and decoded by C_1^{-1} to result in the received user information \mathbf{a}' .

Approach

The unique challenge arising in mmWave links is that, in contrast to fibre links, the channel conditions can vary, as mmWave links are attenuated by, e.g., rain. This problem can be solved by adaptive coding. An additional encoder can encode the I/Q samples of a received UE signal in case of degrading channel conditions. However, encoding and especially decoding the BH data is a task that introduces additional latency, which is very undesirable in cloud architectures. Furthermore, the implementation of complex decoders at the very high throughput faced when using in I/Q backhauling would require expensive, energy-hungry hardware.

To avoid this, CT2.7 investigates how to omit a BH code and still be able to mitigate bad channel conditions. For this we follow two approaches, depicted in Figure 4-46:

- Joint encoding: the encoder at the UE can take the BH channel into account, lowering the code rate R_c^{RAN} in case of unfavourable BH conditions
- Joint decoding: the decoder of the RAN-code can be adapted to take an erroneous BH channel into account

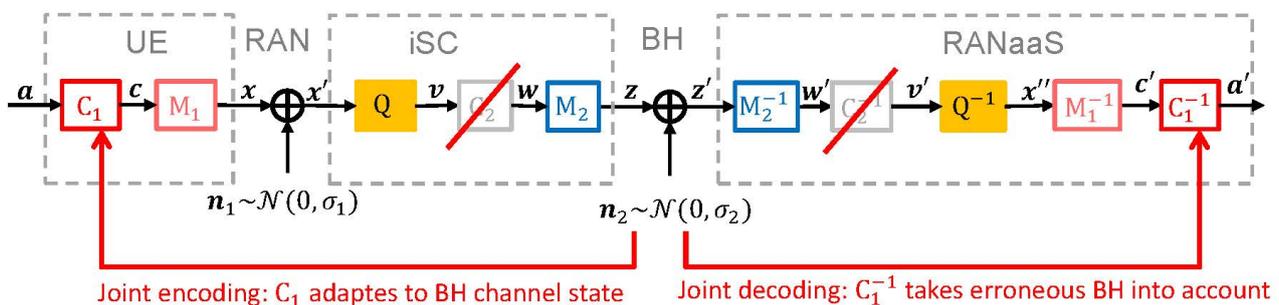


Figure 4-46: Overview of the approach for joint design of access and backhaul

Joint Encoding

In the conventional system, the UE sets its code rate R_c^{RAN} depending on the channel measurements on the RAN channel. The procedure itself is not standardized but in general the eNB tells the UE which MCS to use by sending downlink control information (DCI) data during the uplink grant [16]. Usually, the code rate is chosen to achieve a certain target FER. In the same way the code rate for the BH channel would be set,

depending on the BH channel's SNR. CT 2.7's approach for reducing the latency is now to omit the BH FEC and already consider the BH channel's quality when setting the RAN code rate, which is illustrated in Figure 4-47. For this, the end-to-end FER is monitored in the RANaaS, which already accounts for any degradation by an imperfect BH channel. The code rate R_c^{RAN} is then set to achieve the target FER. However, this involves more investigation towards RRM, which is not a focus of WP2. At this point we merely want to point out that an uncoded, yet imperfect BH integrates very easily into the LTE architecture.

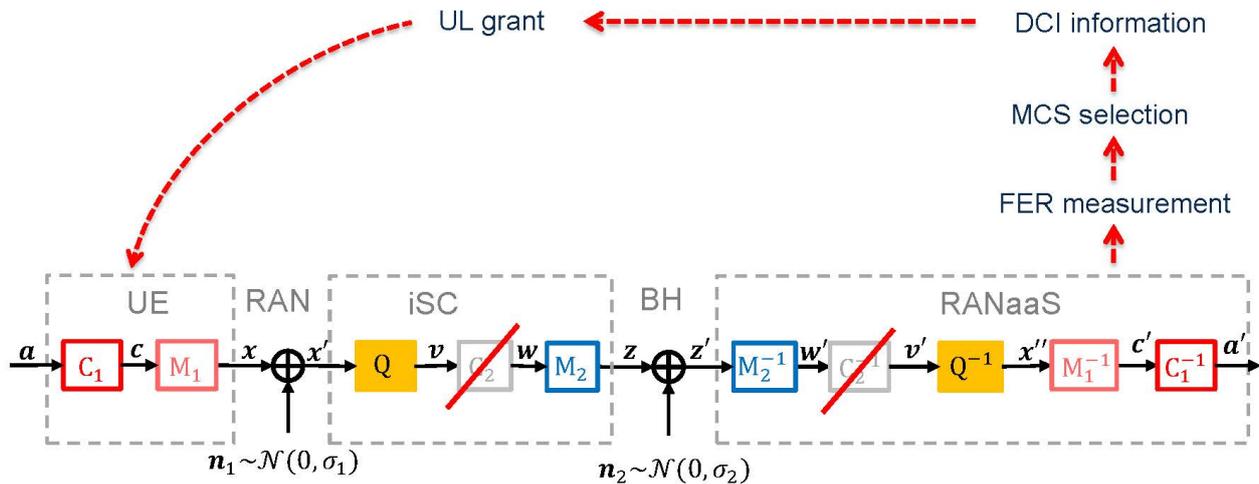


Figure 4-47: Joint encoding procedure

Joint Decoding

When omitting the BH FEC, the I/Q samples received at the RANaaS are potentially erroneous. These errors have to be corrected by the RAN FEC. However, modern codes like turbo-codes that are used for RAN coding draw their strength from exploiting the concept of *soft-information*, in which reliability information is extracted from received symbols. In a conventional setup, this soft information would include information only from the detection of the radio access channel, e.g. the conditional probabilities $p(\mathbf{c}'' | \mathbf{x}'')$. The approach taken here is to also utilize the soft information that can be provided from the BH channel's detection, e.g. $p(\mathbf{v}' | \mathbf{z}')$.

One approach for this is to employ a *soft-input/soft-output dequantizer* (SISODQ) [34]. The problem of the conventional setup without SISODQ is that the bits \mathbf{v}' received over the BH channel are dequantized to reconstruct the amplitudes \mathbf{x}'' that were originally received at the iSC. During dequantization the soft information from the detection of the BH channel $p(\mathbf{v}' | \mathbf{z}')$ is lost. The SISODQ, however, conserves the soft-information during dequantization and calculates $p(\mathbf{c}' | \mathbf{z}')$. As can be observed in [34], this improves the BER performance for uncoded BH, while decreasing the latency compared to an additional decoder by up to 72 %.

In a second approach, we follow the idea of *error resilient decoders* (ERD) proposed in [35]. The calculation of soft information is based on the probability density function (PDF), yet conventionally a normal distribution is assumed for \mathbf{x}' , which is close to the actual distribution in most cases. However, bit errors on the BH channel, i.e. errors between \mathbf{v} and \mathbf{v}' , change the distribution of \mathbf{x}'' . This modifies the PDF of the received signal in a deterministic way.

An example of this modified distribution is shown in Figure 4-48, the detailed description can be found in [81].

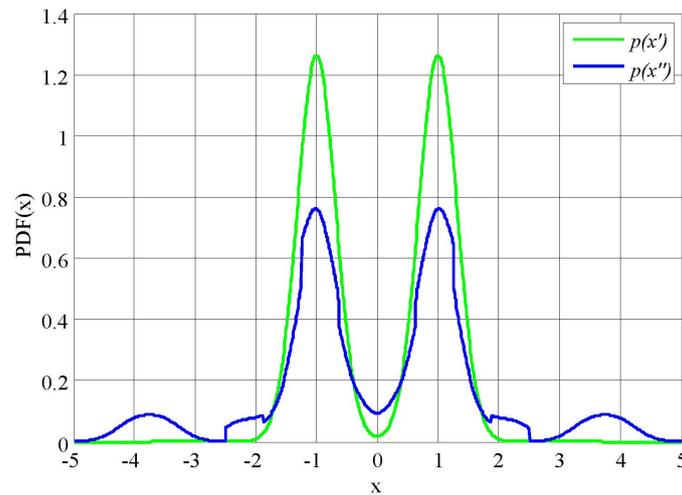


Figure 4-48: Distribution of received symbols after erroneous backhaul

The calculation of the soft information can now be based on this modified PDF to yield more accurate soft information as input to the decoder. This modified calculation merely requires the knowledge of the format in which the samples are stored and the average BER. The BH BER can be either directly measured or estimated from the channels SNR.

To summarize, in the first approach soft information available on the received symbols of the BH is used to improve the decoder input, while in the second approach more general statistical information about the BH channel in form of an BER is used for an improved decoder input.

4.7.2 Implementation in the iJOIN Architecture

As described in Section 4.7.1, mmWave backhaul is a basic transport technology that can be used to enable various different functional splits, as any type of data can be transported. The functional split then requires certain parameters like data rate and delay to be fulfilled to enable each split. The investigations on data compression of CT2.6 apply for mmWave as well.

Apart from that, this CT investigates joint access and BH channel coding, which requires a shift of detection and decoding of the access link into the RANaaS. In previous reports we discussed three possible functional split variants.

Functional split A.1: all digital processing in RANaaS

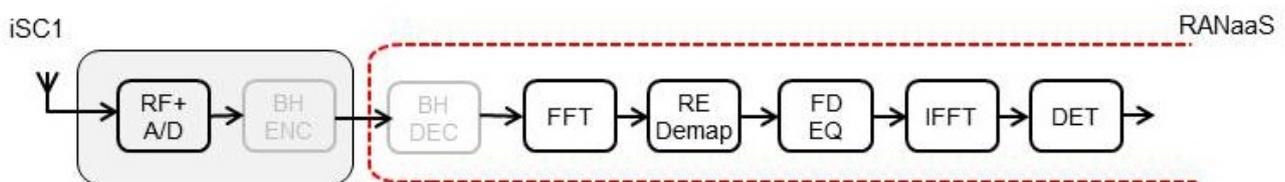


Figure 4-49: Functional split A.1: All digital processing in RANaaS

In this variant, only analogue and A/D conversion is performed at the iSC, meaning that unprocessed I/Q-data is backhauled. For this variant, uncoded BH is applicable; however, the approaches of joint decoding are not. The main benefit of this variant is that the iSCs do not need to be equipped with any baseband processing hardware and can thus be made smaller, lighter and cheaper. The main disadvantage is the high data rate requirement on the BH. It is also the variant that corresponds to the CPRI standard [14].

Functional split B.2: Detection and decoding in RANaaS

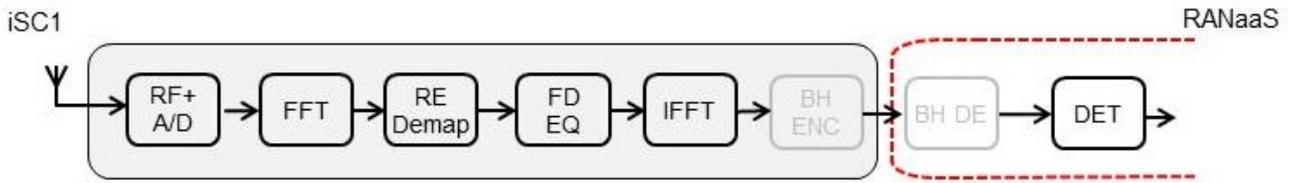


Figure 4-50: Functional split A.4: Detection and decoding in RANaaS

In this variant, detection and decoding are shifted to the RANaaS, which means that an I/Q-data stream of received signals is being backhauled. However, the received signals are already pre-processed in terms of RE demapping and channel equalization. This variant assumed for the approaches of joint decoding discussed above as these assume equalized I/Q-samples.

Backhaul Load per Functional split

For joint decoding, CT 2.7 requires the received signals to be equalized at the iSC, so RxCSI needs to be measured there similar to the standard LTE. For joint decoding it furthermore requires information on the access and backhaul channel quality. The SNR_{RAN} is required for soft-demodulating the received symbols on the access channel. For the SISODQ the SNR_{BH} is additionally required for soft-demodulating the received signals on the BH channel, while the ERD requires the BER_{BH} to calculate the modified PDF. The BER_{BH} can also be derived from the SNR on the BH using standard formulas. Since the BH channel information can be directly calculated in the RANaaS, only the SNR_{RAN} needs to be forwarded over the BH.

Table 4-23: PHY Measurement placement and exchange links for CT2.7

Functional Split Approach	Measurement	Measurement Location	Exchange link
A.1 All digital processing in RANaaS	RxCSI, SNR_{RAN}	RANaaS	N/A
	SNR_{BH}/BER_{BH}	RANaaS	N/A
B.2: Detection and decoding in RANaaS	SNR_{RAN}	iSC	iSC-RANaaS
	RxCSI	iSC	N/A
	SNR_{BH}/BER_{BH}	RANaaS	N/A

To operate in accordance to the LTE standard, CT2.7 also requires additionally common input parameters and delivers corresponding output. This is summarized in Table 4-24

Table 4-24: Backhaul control information of CT 2.7

Link	Description
iSC → RANaaS	PHY Information (access link) <ul style="list-style-type: none"> - O2.7: ACK/NACK, HARQ process - O2.8: User data to upper layers, if ACK
RANaaS → iSC	Network parameters <ul style="list-style-type: none"> - I2.1: System Parameter - I2.2: Estimated distance or SNR for wireless BH link - I2.3: RAN connection table of iSCs and UEs - I2.4: BH connection table for iSC-iSC links and iSCs-RANaaS links
	PHY information (access and BH link) <ul style="list-style-type: none"> - I2.8: RxCSI, SNR, noise variance of UE-iSC links - I2.9: Received signal at iSC - I2.12: RxCSI, SNR, noise variance of iSC-RANaaS link - I2.13: Parameter of iSC-RANaaS
	CT specific control information <ul style="list-style-type: none"> - I2.5: Functional control for shifting/splitting functions - I2.6: RRM information per UE - I2.7: MCS for backhaul link

The main contribution to BH load in this CT is the uplink user data that needs to be forwarded, either in the form of directly sampled or pre-processed I/Q data. The required BH data rate can be calculated using the standard formulas derived in section 3.2.1.2. Additionally, the SNR_{RAN} needs to be forwarded to the RANaaS for joint decoding. Since this is only one float number per UE and symbol, its contribution to the overall traffic is negligible.

Table 4-25: Backhaul Load of CT 2.7

Link	Description	Msg Size	Msg Frequency	Latency
Functional Split A.1: A digital processing in RANaaS				
iSC → RANaaS	Sampled received signal	$D_P^{A.1} = 2 \cdot N_R^{\text{iSC}} \cdot N_Q \cdot OF \cdot (N_{\text{FFT}} + N_{\text{CP}}) \cdot N_{\text{SYMB}}^{\text{SUB}}$	T_{SUB}^{-1}	1 ms
Functional Split B.2: Detection and decoding in RANaaS				
iSC → RANaaS	Quantized user signal	$D_P^{B.2} = \sum_{u=1}^{N_{\text{UE}}} N_{\text{sc},u}^j \cdot Q_m \cdot N_Q \cdot N_L \cdot N_{\text{SYMB}}^{\text{SUB}}$	T_{SUB}^{-1}	1 ms
	SNR_{RAN}	<16 bit per UE	T_{SUB}^{-1}	1 ms

4.7.3 Evaluation of the CT

Compliance with iJOIN objectives

By the aforementioned approaches we aim to provide a BH technology that is able to meet the other CTs requirements in terms of capacity, range, latency and reliability. In that regard it is an enabler for other CTs, especially those aiming to increase *utilization* of hardware by centralized processing.

The *cost* of a network can be decreased, since mmWave links do not require earthworks and thus, the deployment cost will be lower as compared to fibre. Compared to traditional microwave links they are also cheaper in terms of licensing as only “light licensing” (70-90 GHz) or no licensing (60 GHz) is required. By removing an additional encoder, the hardware of the mmWave links will also be simpler resulting in lower hardware cost and less energy consumption.

By increasing the reliability of the BH link, the range can be extended, further lowering the *costs* as less links per area are required, or enabling otherwise impossible topologies. An increased reliability can also be traded off to reduce transmit power, increasing the overall *energy efficiency* of the network, or to achieve a higher *throughput* when facing unfavourable BH channel conditions.

Description of the baseline used for the evaluation

The baseline of CT2.7 is a mmWave BH architecture featuring a backhaul en/decoder and a standard RAN decoder. This will be replaced by an architecture without BH en/decoder and an optimized RAN decoder.

To compare performance, we evaluate the end-to-end BER between the UE and the RANaaS after decoding the RAN FEC. We use a link-level simulator implemented in Matlab and simulate a single mmWave BH link. The mmWave link is characterized by an AWGN channel with a certain SNR. From the BER and the SNR we can evaluate the following parameters

- The SNR can be combined with a link budget to calculate a range and transmit power. These in turn can be mapped to network costs and energy consumption
- From the BER we can calculate the end-to-end throughput as $\eta = (1 - \text{BER})^{N_{\text{block}}} \cdot S$

Discussion of results

Figure 4-51 shows the BERs of an architecture employing coded BH (green lines) as compared to one without BH FEC that employs the ERD (blue lines) and the SISODQ (red lines). The RAN has an SNR of 10 dB, which will result in a BER of 10^{-4} when 4-QAM, a RAN code rate of 0.8 and a *perfect* BH link is used. This can be considered as the “operating point” of the RAN. The plots now show how an imperfect RAN (lower SNR) can be mitigated by using either an additional BH coder, or the joint encoding and decoding schemes. In the latter case, the RAN code rate is lowered instead. Since an additional BH code introduces

additional overhead in the transmission, so we introduce the notion of a product code rate $R_c^{tot} = R_c^{RAN} \cdot R_c^{BH}$ for fair comparison. An equal R_c^{tot} corresponds to the same bandwidth requirement on the BH.

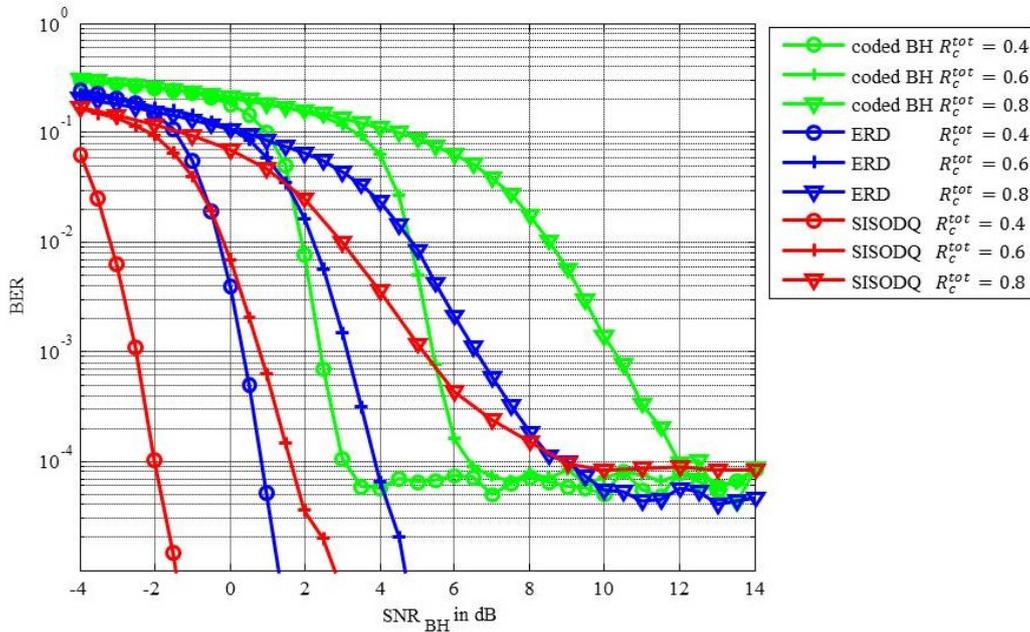


Figure 4-51: BER when using encoded BH (green lines) as compared to an uncoded BH with the ERD (blue lines) and with a SISODQ (red lines)

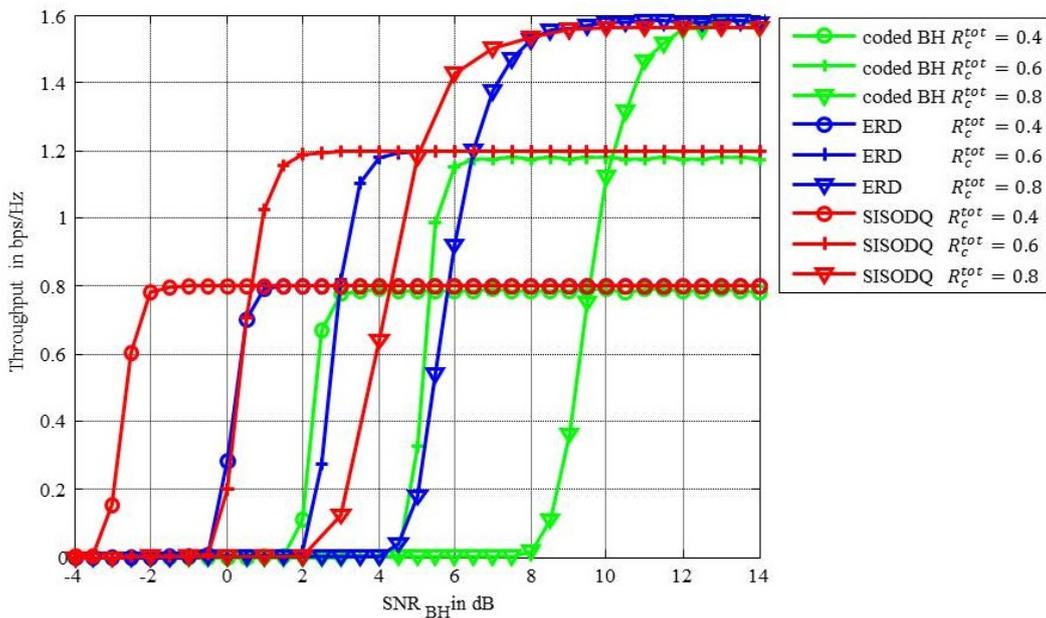


Figure 4-52: Throughput when using encoded BH (green lines) as compared to an uncoded BH with the ERD (blue lines) and with a SISODQ (red lines)

In D2.1 [36], we observed that the omission of the BH encoder without any of the joint decoding schemes reduces the performance by approximately 1-2 dB. It follows that without an optimized RAN decoder, a reduced latency is traded off against a higher BER. However, when employing the ERD or the SISODQ, the performance is increased even beyond the coded BH case, as can be seen from Figure 4-51. If we assume that latency constraints do not allow for a coded BH, the SISODQ increases the throughput by up to 30%, as can be seen from Figure 4-52. To summarize, by employing both an uncoded BH and a SISODQ we are able to meet tight latency constraints as well as deliver a high throughput.

5 Overall Evaluation

After defining the set of WP2 candidate technologies and their implementation in the iJOIN architecture together with the discussion of performance results in Section 4, we discuss in this section the interaction of WP2 CTs to address the global iJOIN targets. By this structured approach the final evaluation phase leading to deliverable D2.3 is prepared.

For PHY layer CTs Table 5-1 defines the primary function, the affected area as well as the processing nature. CT2.1, CT2.2, and CT2.3 are alternative approaches for the estimation of UE signals in the uplink following different philosophies of distributed or centralized processing. As demonstrated by performance evaluations, the different functional splits per CT offer gains in different set ups. Thus, based on the actual communication scenario, the iverC selects the appropriate approach for the uplink phase. In contrast, CT2.4 and CT2.5 are both addressing DL transmission in the iJOIN architecture and can in general be implanted together. Such a combined evaluation is foreseen for the final project phase. In addition, CT2.6 and CT2.7 address principle questions of BH operation being applicable for uplink and downlink communications. Thus, both CTs may support the other WP2 CTs and allow for a joint operation.

Table 5-1: Categorization and compatibility of WP2 CTs

CT	Function	Affected area	CT scope	Interoperation
2.1	UL Detection	RAN	Distributed	Different alternatives for UL Detection
2.2	UL Detection	RAN	Distributed / Centralized	
2.3	UL Detection	Joint RAN/BH	Centralized	
2.4	DL CoMP	RAN	Centralized	Interoperation is potentially possible
2.5	DL CoMP	Joint RAN/BH	Distributed	
2.6	Functional Split/BH for UL and DL	BH	P2P	Support for other UL & DL approaches
2.7	mmWave / coding for UL	Joint RAN/BH	P2P	

In principle, all UL and the DL oriented CTs can interoperate due to the independence of both communication flows. Thus, it is also possible to apply different functional splits for UL and DL at the same time due to the actual scenario. For example the latency introduced by the J2 link may prevent the distributed calculation of precoders for DL CoMP in case of low channel coherence time. Thus, a lower PHY layer functional split (A according to Figure 3-3) with centralized calculation of precodes may be favourable. Nevertheless, as the main limiting factor for UL detection is caused by LTE HARQ timing, a higher PHY layer functional split like B or C with distributed processing of local receive signals is still possible.

The iJOIN project has defined in deliverable D5.1 [17] the following four metrics to be improved by using a joint design of RAN and BH for the envisioned dense deployment of small cells with cloud processing.

- **Area Throughput (AT)** is the main metric targeted by CTs in WP2 by applying sophisticated PHY layer approaches. The area throughput is given by the data rate successfully delivered to or received from the users in a given geometrical area over time. Thus, it is directly measurable using PHY layer related performance evaluations.
- **Energy Efficiency (EE)** considers the power consumption of the different entities in the iJOIN network and depends for example on the complexity of algorithms, the placement of functions due to the applied split, the RF transmit power, the BH technology, etc. As an approach which improves the AT at a given transmit power, will usually also reduce the required energy to achieve a fixed data, the impact of the other components is less obvious.
- **Utilization Efficiency (UE)** measures how well the utilized resources are used to realize a given system performance. CTs which allow for flexible functional split depending on the current needs (e.g. BH load, required user data rate, etc.) are appropriate to utilize the different resources well-balanced and prevent over-provisioned system design.

- **Cost Efficiency (CE)** depends on the number and kind of nodes involved in the network and recognizes in addition the cost of the backhaul infrastructure determined by factors like distances, BH technology, etc.

Table 5-2 summarizes the qualitative impact of the CTs for these metrics. A “++” symbol indicates a large positive impact of a CT on a specific objective, the “+” symbol indicates a positive impact of a CT , and the “0” symbol implies a negligible impact. The main target of all CTs is area throughput; however the other objectives are addressed as well.

Table 5-2: Qualitative impact of WP2 CTs on iJOIN objectives

CT	Area Throughput	Energy Efficiency	Utilization Efficiency	Cost Efficiency
2.1	++	+	+	0
2.2	++	0	+	0
2.3	++	0	+	0
2.4	++	+	0	0
2.5	++	+	0	0
2.6	+	0	+	+
2.7	++	+	0	+

In the next deliverable D2.3 quantitative results for the different objectives will be provided. In particular, the scalability of area throughput by means of network entities like number of iSCs, number of users, and availability of BH resources will be investigated. In order to show the relative gains with respect to 3GPP LTE related approaches, two evaluation scenarios are defined in Appendix II. Scenario I defines an outdoor and Scenario II an indoor small cell deployment.

6 Summary and Conclusions

This deliverable defines the set of candidate technologies applicable at the physical layer for a dense deployment of small cells and a joint optimization of access and backhaul networks. Each CT is described in detail, providing preliminary evaluation results, their integration into the iJOIN architecture and functional split options, as well as the resulting impact on the backhaul. These aspects are of particular relevance considering that one of the main benefits of the network architecture proposed by iJOIN is the ability to flexibly assign functionality to either the iSC or the RANaaS, and that this functional split can differ in both space and time throughout a network, according to the traffic demand or the deployment scenario.

Because of their importance, the functional split options are investigated in more detail, highlighting the different required backhaul rates and their general advantages and disadvantages. Generally, the lower the split is performed, the higher the required backhaul rate and the higher the degree of joint processing. This investigation is supplemented with an analysis of the requirements imposed by LTE on the PHY layer, especially latency and the quantization on the backhaul, which are fundamental for centralized processing.

In the context of a flexible cloud-RAN, more detailed investigations concerning the precoder and decoder implementation in a cloud platform based on General Purpose Processors (GPP) are also provided. The precoding operation can be implemented locally in either iSC or centrally in the RANaaS, depending on the downlink transmission chain functional split. Even in the case that the precoding operation is executed centrally, the RANaaS high computational capability will be challenged when the CoMP set comprises a large amount of iSCs and UEs as the processing computational complexity is extremely high, growing cubically with the overall number of antennas in the CoMP set. One feasible solution is to divide the total number of iSCs into a certain number of cooperation clusters. This leads to an optimisation problem where the objective will be to optimise system sum rate by being able to choose the optimal cooperation cluster size.

Concerning the uplink, the cloud-based implementation of modern FEC decoders such as LDPC and Turbo codes is a challenging task and will require the simultaneous optimization at both the algorithmic and the architectural abstraction levels, to reach the increasing throughput requirements of 5G cellular systems.

Finally, for the purpose of a comparable performance evaluation among the different CTs, a common simulation scenario and a set of harmonized parameters have been defined. One of the iJOIN objectives addressed by all CTs is the throughput per geometrical area, which will be evaluated and compared in the defined common scenario by indicating the gains and the scaling with respect to iJOIN parameters.

Acknowledgements and Disclaimer

This work was partially funded by the European Commission within the 7th Framework Program in the context of the ICT project iJOIN (Grant Agreement No. 317941).

Appendix I Input/Output Interfaces

Based on the functional architecture defined in deliverable D5.1 [17], the input and output interfaces for WP2 CTs were defined in deliverable D2.1 [36]. In Table 6-1 we summarize the information required by each CT and specify the source of information by means of other CTs or system functions as well as by logical network entity in the IJOIN architecture. Thus, connections to the MAC functionality described D3.1 [31] and mobile network functions discussed in D4.1 [32] are established. Furthermore, the principle interfaces are defined by means of parameter lists. Similarly, Table 6-2 lists the provided output information per CT with sink of information and interface definition. Furthermore, Table 6-3 describes the used acronyms and indicates whether the related parameter is already defined by 3GPP LTE or has to be introduced by iJOIN.

Table 6-1: Requested Input of WP2 CTs

In	CT	Requested Input	Source of Information		Parameter
			CT or system function	Logical network entity	
Network Information					
I2.1	2.1 2.2 2.4 2.5 2.6 2.7	System Parameter		NMS, EMS	<RAN_BW, N_PRB, CP_length, N_sub, N_iSC, {N_Tx_iSC}, {N_Rx_UE}>
I2.2	2.4 2.7	Estimated distance or SNR for wireless BH link	BH Measurement	iNC	<BH_ID, BH_SNR, BH_PathLoss>
I2.3	2.1 2.2 2.3 2.4 2.5 2.6 2.7	RAN connection table of iSCs and UEs	RRM/RRC - 2.1: RRM for INP (CT3.8) may override	veNB	<Conn_UE_iSC{UE_ID, iSC_ID}>
I2.4	2.1 2.3 2.4 2.5 2.7	BH connection table for iSC-iSC links and iSCs-RANaaS links	Network layer	iNC	<Conn_iSC_iSC {iSC_ID1, iSC_ID2}, Conn_iSC_RANaaS {iSC_ID}>
Functional Control					
I2.5	2.1 2.2 2.3 2.4 2.5 2.6 2.7	Functional control for shifting/splitting functions	iveC per CT	veNB	<FuncCtrl>

RRM Information					
I2.6	2.1 2.2 2.3 2.4 2.5 2.6 2.7	RRM information per UE	RRM (MAC Layer) - 2.1: RRM for INP (CT3.8) - 2.2: RRM for MPTD (CT3.7) - 2.3: RRM for JNCC (CT3.2) - 2.5: RRM for ICIC (CT3.4, CT3.9)	veNB	<UE_ID, N_Tx_UE, {RI, PMI}, I_MCS, RV, HARQ_ID, NDI, TBS, RB allocation type (RBstart, L_CRB) >
I2.7	2.5 2.7	MCS for backhaul link	BH RRM	iSC or RANaaS	<BH_ID, I_MCS_BH>
PHY Information (Access Link)					
I2.8	2.1 2.2 2.3 2.6 2.7	Instantaneous RxCSI and SNR and/or noise variance of each UE-iSC link	UL Channel Estimation (PHY functionality)	iSC or RANaaS	<UE_ID, RxCSI, RxSNR, RxNVar>
I2.9	2.1 2.2 2.3 2.6 2.7	Received signal at iSC in frequency domain	UL Baseband Processing	iSC	<yRx>
I2.10	2.4 2.5 2.6	TxCSI, noise variance, and path-loss of each iSC-UE link	DL Channel Estimation (TDD UL CE, FDD Feedback) (PHY functionality)	veNB	<UE_ID, TxCSI, iSC_UE_PathLoss, TxNVar>
I2.11	2.4 2.5 2.6	User data to be transmitted in DL	MAC layer	veNB	<UE_ID, data_UE>
PHY Information (Backhaul Link)					
I2.12	2.1 2.2 2.3 2.6 2.7	Instantaneous RxCSI, SNR and/or noise variance of iSC-RANaaS link	BH Channel Estimation for iSC-RANaaS link	RANaaS	<iSC_ID, RxCSI_iSC_R, RxSNR_iSC_R, RxNVar_iSC_R>
I2.13	2.1 2.2 2.3 2.6 2.7	Parameter of iSC-RANaaS link	MM	iNC	<BH_ID, N_quant_iSC_R, BH_BW, BH_lat>
I2.14	2.4 2.5 2.6	Instantaneous RxCSI, SNR and/or noise variance of RANaaS-iSC link	BH Channel Estimation for iSC-RANaaS link	iSC	<RxCSI_R_iSC, RxSNR_R_iSC, RxNVar_R_iSC >

I2.15	2.4 2.6	Parameter of RANaaS-iSC link	MM	iNC	<BH_ID, N_quant_R_iSC, BH_BW, BH_TxPow>
I2.16	2.3 2.5	Instantaneous RxCSI, SNR and/or noise variance of iSC-iSC link	Channel Estimation for iSC-iSC link (PHY functionality)	iSC	<iSC_ID, RxCSI_iSC_iSC, RxSNR_iSC_iSC, RxNVar_iSC_iSC>
I2.17	2.1 2.2 2.3 2.4 2.5	Parameter of iSC-iSC link	MM	veNB	<BH_ID, N_quant_iSC_iSC, BH_BW, BH_TxPow>

Table 6-2: Provided Output of WP2 CTs

Out	CT	Provided Output	Sink of Information		Parameter
			CT or system function	Logical network entity	
Functional Control					
O2.1	2.4 2.5	Estimated backhaul load (including overhead) for CoMP schemes	- 2.4: Coop. RRM for ICIC in RANaaS (CT3.5) - 2.5: SD Scheduler (CT3.4) and HL Scheduler (CT3.9)	veNB	<BH_cap, BH_energ>
O2.2	2.6	Estimated BH parameter (latency, capacity/load, energy, BER)	MM	iNC	<BH_lat, BH_cap, BH_load, BH_energ, BH_BER>
RRM Information					
O2.3	2.1	Effective SINR per UE	RRM for INP (CT3.8)	veNB	<UE_ID, effRxSINR_INP>
O2.4	2.1	Effective SINR per UE-iSC link	RRM for INP (CT3.8)	veNB	<UE_ID, iSC_ID, effRxSINR_iSC>
O2.5	2.1 2.2 2.3	Estimated PER	- 2.1: RRM for INP (CT3.8) - 2.2: RRM for MPTD (CT3.7) - 2.3: RRM for JNCC (CT3.2)	veNB	<UE_ID, PER>
O2.6	2.4 2.5	Effective SNR/SINR of each UE (with precoding)	RRM (MAC layer)	veNB	<UE_ID, effTxSINR>

PHY Information (Access Link)					
O2.7	2.1 2.2 2.3 2.7	ACK/NACK (based on CRC), HARQ process	HARQ (MAC layer)	veNB	<HARQ_ID, ACK/NACK>
O2.8	2.1 2.2 2.3 2.6 2.7	If ACK, user data (decoded Layer 2 PUSCH) to upper layers	MAC layer (MAC PDU)	veNB	<UE_ID, HARQ_ID, data_UE_est>
O2.9	2.4 2.5 2.6	Transmit signal at iSC in frequency domain on RAN	DL Baseband Processing	iSC	<xTx>
O2.10	2.4	Access link capacity and energy efficiency	Cooperative RRM for ICIC in RANaaS (CT3.5)	veNB	<Access_Cap, Access_Energy>

In the following table the parameters exchanged on the interfaces are described. Either definitions from 3GPP LTE are used or specific definitions will be introduced by the CTs.

Table 6-3: Explanation of Parameters

Parameter	Full Name (including explanation if necessary)	LTE or CT specific
Identifier		
BH_ID	Backhaul link identifier (for logical network) <ul style="list-style-type: none"> iSC_ID_tx – iSC_ID_rx iSC_ID_tx – RANaaS_rx RANaaS_tx – iSC_ID_rx 	iJOIN
veNB_ID	ID of virtual eNodeB (for logical network)	iJOIN
CELL_ID	ID of macro cell (specifies location of pilots)	LTE
iSC_ID	ID of iSC (for logical network) <ul style="list-style-type: none"> iSC_ID_tx: ID of transmitting iSC iSC_ID_rx: ID of receiving iSC iSC_ID_i: ith- iSC to differentiate iSCs 	iJOIN
UE_ID	UE ID	LTE
Conn_iSC_iSC	Table listing available links between iSCs	iJOIN
Conn_iSC_RANaaS	Table listing available links between iSCs and RANaaS	iJOIN
Conn_UE_iSC	Table listing available links between UEs and iSCs	iJOIN
RAN transmission parameter		
N_PRB	Number of available RBs within the system bandwidth	LTE
N_Tx_iSC	Number of transmit antennas per iSC	LTE
N_Rx_iSC	Number of receive antennas per iSC	LTE
N_sub	Number of active (i.e. modulated) subcarriers	LTE
N_Tx_UE	Number of transmit antennas of UE	LTE
N_Rx_UE	Number of receive antennas of UE	LTE
RAN_BW	System bandwidth [MHz] (e.g. BW=20 MHz for a 20 MHz LTE carrier)	LTE
CP_length	Cyclic prefix length [μ s]	LTE
N_iSC	Number of iSCs serving a defined group of UEs	LTE
N_UE_per_iSC	Number of users per iSC	LTE
BH parameter		
BH_SNR	Signal to Noise Ratio in backhaul	BH Measurement

BH_load	Estimated BH load [%]	CT specific
BH_cap	Estimated max. BH capacity [Mbps]	BH Measurement
BH_lat	Estimated BH latency [μ s]	BH Measurement
BH_energ	Estimated BH energy consumption [mW]	BH Measurement or CT2.4
BH_BER	Estimated BH BER [float]	BH Measurement
BH_PathLoss	Path loss of BH link	CT2.4
BH_BW	Bandwidth of BH link	CT2.4
BH_TxPow	Tx Power of BH link	CT2.4
PHY parameter		
PMI	Pre-Coding Matrix Index	LTE
RI	Rank Indicator	LTE
CDI	Channel Direction Information	LTE
I_MCS	Modulation and Coding Scheme (0... 31)	LTE
I_MCS_BH	Modulation and Coding Scheme for BH	CT2.7
HARQ_ID	ID of HARQ process	LTE
NDI	New Data Indicator	
TBS	Transport Block Size (0 ... 24)	LTE
RB allocation type	DL {0,1,2} UL {0,1} (bold = default)	LTE
RBstart	UL: First allocated RB	LTE
L_CRB	UL: Number of continuous RBs allocated (starting from RBstart)	LTE
RV	Redundancy Version (0 ... 3)	LTE
ACK/NACK	ACK/NACK per packet	LTE
PER	Estimated packet error rate	LTE
RE	Resource Elements	LTE
RI	Rank Indicator	LTE
RRM	Radio Resources Management	LTE
RRC	Radio Resources Control	LTE
NMS	Network Management Subsystem	LTE
EMS	Element Management System	LTE
Access link: Signals and transmission variables (CSI, Noise Variance, SINR, etc.)		
RxCsi	Receive side Channel State Information for UE-iSC link	LTE
RxSNR	Signal to Noise Ratio on UL for UE-iSC link	LTE
RxNVar	Noise variance on UL for UE-iSC link	LTE
TxCsi	Transmit side Channel State Information for iSC-UE link	LTE
TxNVar	Noise variance on DL for iSC-UE link	LTE
xTX	Transmit signal on DL for iSC-UE link in frequency domain	LTE
yRx	Receive signal on UL for at iSC in frequency domain	LTE
data_UE	Transmit data of user delivered by upper layer	LTE
data_UE_est	Estimated data of User to pass to the upper layer	LTE
effRxSINR_iNP	Effective SINR per UE after INP processing	CT2.1
effRxSINR_iSC	Effective SINR per UE-iSC link after INP processing	CT2.1
effTxSINR	Effective SNR/SINR of UE after precoding	CT2.4, CT2.5
Access_Cap	Access link achievable Capacity	CT2.4
Access_Energy	Access link achievable Spent Energy (or efficiency)	CT2.4
iSC_UE_Pathloss	Pathloss from iSC to UE	CT2.4
BH links: Signals and transmission variables (CSI, Noise Variance, SINR, etc.)		
RxCsi_iSC_R	Receive side Channel State Information for iSC-RANaaS link	LTE
RxCsi_R_iSC	Receive side Channel State Information for RANaaS-iSC link	LTE
RxCsi_iSC_iSC	Receive side Channel State Information for iSC-iSC link	LTE
RxSNR_iSC_R	Signal to Noise Ratio on iSC-RANaaS link	LTE
RxSNR_R_iSC	Signal to Noise Ratio on RANaaS-iSC link	LTE
RxSNR_iSC_iSC	Signal to Noise Ratio on iSC-iSC link	LTE
RxNVar_iSC_R	Noise variance on iSC-RANaaS link	LTE

RxNVar_R_iSC	Noise variance on RANaaS-iSC link	LTE
RxNVar_iSC_iSC	Noise variance on iSC-iSC link	LTE
N_quant_iSC_R	Number of quantization levels on iSC-RANaaS link	iJOIN
N_quant_R_iSC	Number of quantization levels on RaNaaS-iSC link	iJOIN
N_quant_iSC_iSC	Number of quantization levels on iSC-iSC link	iJOIN

Appendix II Evaluation Methodology

II.1 Evaluation methodology

In deliverable D5.1 [17] of WP5 the following four common scenarios (CS) are defined for the iJOIN project in order to evaluate the performance of the iJOIN approach by means of different application areas

- CS 1: Stadium
- CS 2: Square
- CS 3: Wide-area continuous coverage
- CS 4: Shopping Mall / Airport

In order to derive a first set of evaluation parameters, WP2 and WP3 derived jointly a list of reference system parameters for an outdoor and for an indoor scenario. Appendix I of deliverable D3.2 [87] describes the WP3 definition of evaluation scenarios which aims to give a global vision about the evaluation methodology and provides a common reference of system parameters. The evaluation methodology is based on 3GPP parameters and settings (3GPP TR 36.814 [61], TR 36.932 [62] and TR 36.872 [63]).

In Appendix II.2 we list the PHY specific parameters for the evaluation scenario 1 “Outdoor Deployment” which mainly follows CS 2 (Square) and CS 3 (Wide-area continuous coverage). The parameters for the second evaluation scenario “Indoor deployment” defined in Appendix II.3 correspond to CS 4 (Airport / Shopping Mall).

One of the main objectives of all WP2 CTs is to evaluate the throughput per geometrical area, which will be specified by its dimension, the number of UEs and iSCs. The throughput will be compared to the one corresponding to the baseline approach by indicating the gains and the scaling with iJOIN parameters. In order to evaluate the objectives and how they are reached, we can select successively some CTs with the same (or very similar) baseline assumptions and draw out a common analysis.

II.2 Scenario 1: Outdoor deployment

In the following we discuss of the scenario model deployment of outdoor small cells for serving outdoor UEs. The small cells may operate either on co-channel on the 2 GHz band or on a dedicated carrier at 3.5 GHz. The density of small cells can be varied to represent different use case scenarios; however, this scenario mainly fits with the iJOIN CS2 (the square).

Table 6-4 presents common parameters that summarise the assumptions for the evaluation of mechanisms devoted to outdoor small cells. While in Table 6-5 we introduce different parameters for each CT that are involved in the outdoor scenario.

Table 6-4: 3GPP outdoor deployment assumptions

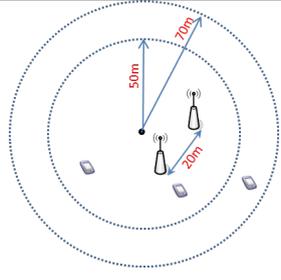
Small cell	
Layout	<p>Small cells uniformly random dropping and regularly placed within the cluster area. A cluster also denoted as <i>veNB domain</i> contains several iSCs (which communicate through the J2 interface) linked with the RANaaS by the J1 interface</p> 
System bandwidth per carrier	10MHz
RAN channel model (UE-iSCs)	Block Rayleigh Fading with distance-dependent path loss
Number of iSCs per veNB domain	Typically 4 ... 10, in extreme also up to 256
Number of UEs per veNB domain	Maximum 60 UEs
UE dropping	Baseline: 2/3 UEs randomly and uniformly dropped within the clusters, 1/3 UEs randomly and uniformly dropped throughout the geographical area
Radius for iSC dropping in a veNB domain	50m
Radius for UE dropping in a veNB domain	70m
Minimum distance (2D distance)	iSC-iSC: 20m iSC-UE: 5m
UE speed	Static UEs (0km/h) or pedestrian (3km/h) or 30Km/h
UE-iSC attachment	The ivEC assigns UEs to iSCs according to the CT properties
HARQ	On

Table 6-5: Mapping of the iJOIN WP2 assumptions to the 3GPP outdoor model

Parameters	UoB CT2.1	CEA CT2.3	UniS CT2.4	IMC CT2.5	TI CT2.6	TUD CT2.7
Mapping with iJOIN CS	Square	Wide-area continuous coverage (stadium)	Square	Square	Wide-area continuous coverage	Square
System bandwidth per carrier	☑	☑	☑	☑	☑	☒
Distance-dependent path loss (RAN)	☑	☒	☑	☑	N/A	☑ for BH
Fast fading	Rayleigh Block fading	Rayleigh/Rice Block fading	Temporal-correlated channel	Rayleigh Block fading	☒	Rayleigh Block fading
Antenna configuration	2x2	1x1	1x1	1x1	2x2	1x1
Number of small cells per cluster (e.g. 10)	3...10	2...10	1...10, extreme 256	1...10	1 (scaling possible)	1
Number of UEs	Varying	2 UEs per MARC-entity	Varying	Varying	Varying (1)	Varying (1)
UE dropping	Random	Predefined	Random	Random	N/A	N/A
Minimum distance	☑	☒	☑	☑	☒	☒
Traffic model	Full Buffer	Full Buffer	Full Buffer	Full Buffer	Full Buffer	Full Buffer
Backhaul Rate	to be investigated	to be investigated	to be investigated	to be investigated	to be investigated	to be investigated
Backhaul Latency	☒	☒	☑ (to be investigated)	☑ (to be investigated)	☒	☒
DL Modulation	N/A	N/A	☒	☒	☑	N/A
UL Modulation	QPSK, 16QAM, 64QAM	QPSK, 16QAM, 64QAM	N/A	N/A	QPSK, 16QAM, 64QAM	QPSK, 16QAM, 64QAM
Link adaptation	Discussed in WP3	Discussed in WP3	N/A	☒	☒	☑ for RAN and BH
Channel equalization	MMSE	ZF	MMSE	☒	N/A	MMSE
Synchronization	Perfect	Perfect	Perfect	Perfect	Perfect	Perfect
Coding for data channel	LTE Turbo Code	LDPC	N/A	N/A	N/A	LTE Turbo Code
Channel estimation	Perfect	Perfect	Perfect	Perfect	N/A	Perfect

II.3 Scenario 2: Indoor deployment

In the following we discuss the deployment scenario of indoor small cells for serving indoor UEs. The small cells operate on a dedicated carrier at 3.5 GHz. In any case, cross-tier interference is not considered due to the dedicated carrier deployment. Both the ITU indoor Hotspot and the 3GPP dual stripe models can be used in the simulations. The density of small cells can be varied to represent different use case scenarios; however, this scenario mainly fits with the iJOIN CS4 (Airport / Shopping Mall).

Table 6-6 presents relevant parameters that summarise the different assumptions for the evaluation of mechanisms devoted to indoor small cells. While in Table 6-7 we introduce different parameters for each CT that are involved in the indoor scenarios.

Table 6-6: ITU indoor deployment assumptions

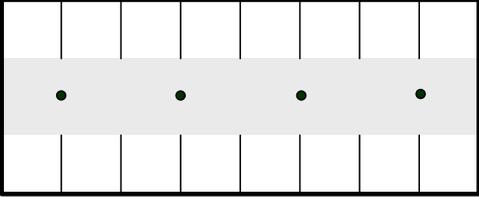
ITU indoor Hotspot	
Layout	2/4 small cells per floor, 1/2 floors 
System bandwidth per carrier	10MHz
RAN channel model (UE-iSCs)	<ul style="list-style-type: none"> Block Rayleigh with distance-dependent path loss ITU InH [referring to Table B.1.2.1-1 in TR36.814]
Number of UEs per small cell	5/10
UE dropping	Randomly and uniformly distributed over area per floor
Minimum distance (2D distance)	Small cell-UE: 3m
UE speed	Static UEs (0km/h) or pedestrian (3km/h)
UE-iSC attachment	The iverC assigns UEs to iSCs according to the CT properties
Link adaptation	On
HARQ	On / Off

Table 6-7: Mapping of the iJOIN WP2 assumptions to the 3GPP indoor model

Parameters	UoB CT2.1	SCBB CT2.2	UNIS CT2.4	IMC CT2.5	TI CT2.6
Mapping with iJOIN CS	Airport Shopping Mall				
System bandwidth per carrier	<input checked="" type="checkbox"/>	10MHz (few RBs will be used)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Distance-dependent path loss	<input checked="" type="checkbox"/>	N/A	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Fast fading	Rayleigh Block fading	ITU-R InH	Temporal-correlated channel	Rayleigh Block fading	<input checked="" type="checkbox"/>
Antenna configuration	2x2	1x2 (UL, ULA)	1x1	1x1	2x2
Number of small cells per veNB (e.g. 10)	2...10	2	1...10	1...10	1
Number of UEs	3-5	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Varying (1)
UE dropping	Random	N/A	N/A	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Minimum distance	<input checked="" type="checkbox"/>	N/A	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Traffic model	Full buffer				
UE speed	0Km/h	3km/h	0Km/h	0Km/h	0Km/h
Backhaul Rate	to be investigated				
Backhaul Latency	<input checked="" type="checkbox"/>				
DL Modulation	N/A	N/A	Gaussian	Gaussian	<input checked="" type="checkbox"/>
UL Modulation	QPSK, 16QAM, 64QAM	QPSK, 16QAM, 64QAM	N/A	N/A	QPSK, 16QAM, 64QAM
Channel equalization	MMSE	MMSE-IRC Turbo Detection	MMSE	N/A	<input checked="" type="checkbox"/>
Synchronization	Perfect	Perfect	Perfect	Perfect	Perfect
HARQ	N/A	<input checked="" type="checkbox"/>	N/A	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Coding for data channel	LTE Turbo Code	LTE Turbo Code	N/A	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Channel estimation	Perfect	Perfect	Perfect	Perfect	Perfect

References

- [1] D. Gesbert, S. Hanly, H. Huang, S. Shitz Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE Trans. on Inf. Theory*, vol. 28, no. 9, pp. 1380–1408, 2010.
- [2] H. Paul, J. Fliege, and A. Dekorsy, "In-network-processing: Distributed consensus-based linear estimation," *IEEE Commun. Lett.*, vol. 17, no.1, pp. 59-62, Jan.2013.
- [3] H. Paul, B.-S. Shin, and A. Dekorsy, "Distributed Consensus-Based Linear Estimation with Erroneous Links," in *Proc. VDE/ITG International Workshop in Smart Antennas (WSA)*, 2013.
- [4] H. Paul, B.-S. Shin, D. Wübben, and A. Dekorsy, "In-Network-Processing for Small Cell Cooperation in Dense Networks," in *Proc. IEEE Vehicular Technology Conference (VTC) workshops, First International Workshop on Cloud Technologies and Energy Efficient in Mobile Communications Networks (CLEEN)*, 2013.
- [5] B. S. Shin, H. Paul, D. Wübben, and A. Dekorsy, "Reduced Overhead Distributed Consensus-Based Estimation Algorithm," *Proc. IEEE Global Communications Conference (GLOBECOM) workshops, The first International Workshop on Cloud-Processing in Heterogeneous Mobile Communication Networks (IWCPM)*, 2013.
- [6] X.Y. Hu, E. Eleftheriou, and D.M. Arnold, "Regular and irregular progressive edge-growth tanner graphs" *Information Theory, IEEE Transactions on*, 51(1), 386-398, 2005.
- [7] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat-accumulate codes" In *Proc. 2nd Int. Symp. Turbo codes and related topics* (pp. 1-8), September 2000.
- [8] N. Wiberg, "Codes and decoding on general graphs," *Dissertation, Department of Electrical Engineering, Linköping University*, 1996.
- [9] C. Di, D. Proietti, I.E. Telatar, T.J. Richardson, and R.L. Urbanke, "Finite-length analysis of low-density parity-check codes on the binary erasure channel. *IEEE Transactions on Information Theory*, 48(6):1570-1579, 2002.
- [10] IEEE-802.16e, "Physical and medium access control layers for combined fixed and mobile operation in licensed bands," 2005, amendment to Air Interface for Fixed Broadband Wireless Access Systems.
- [11] T. Richardson, "Error floors of LDPC codes," In *Proc. Annual Allerton Conference on Communication Control and Computing*, volume 41, pages 1426-1435, 2003.
- [12] J. Bartelt and G. Fettweis, "Radio-over-radio: I/Q-stream backhauling for cloud-based networks via millimetre wave links," to appear in *Proc. IEEE Global Communications Conference (GLOBECOM) workshops, The first International Workshop on Cloud-Processing in Heterogeneous Mobile Communication Networks (IWCPM)*, 2013.
- [13] J. Bartelt, G. Fettweis, D. Wübben, M. Boldi, and B. Mellis, "Heterogeneous Backhaul for Cloud-Based Mobile Networks," in *Proc. IEEE Vehicular Technology Conference (VTC) workshops, First International Workshop on Cloud Technologies and Energy Efficient in Mobile Communications Networks (CLEEN)*, 2013.
- [14] Common Public Radio Interface, [online: <http://www.cpri.info/>]
- [15] 3GPP TS 36.211, "Radio Access Network; Evolved Universal Terrestrial Radio Access; (E-UTRA); Physical channels and modulation," Release 10, (2012 - 12).
- [16] 3GPP TS 36.213, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 10)," Release 10, (2012 - 12)
- [17] iJOIN Project, "D5.1- Revised definition of requirements and preliminary definition of the iJOIN architecture," November 2013. Available: <http://www.ict-ijoin.eu/wp-content/uploads/2014/01/D5.1.pdf>

- [18] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback", *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 2845-2866, June 2010.
- [19] S. Park, O. Simeone, O. Sahin, and S. Shamai, "Robust and efficient distributed compression for cloud radio access networks", *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 692-703, Feb. 2013.
- [20] Y. Zhou and W. Yu, "Optimized backhaul compression for uplink cloud radio access network", *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1295-1307, June 2014.
- [21] S. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks", *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5646-5658, Nov. 2013.
- [22] P. de Kerret and D. Gesbert, "Degrees of Freedom of Network MIMO with Distributed CSI," *IEEE Trans. Inf. Theo.*, vol. 58, no. 11, pp. 6806-6824, Nov. 2012.
- [23] P. de Kerret, J. Hoydis, and D. Gesbert, "Rate loss analysis of transmitter cooperation with distributed CSIT", *IEEE Signal Processing Advances in Wireless Communications (SPAWC)*, 2013.
- [24] P. de Kerret, R. Fritzsche, D. Gesbert, and U. Salim, "Robust precoding for network MIMO with Hierarchical CSIT", *11th International Symposium on Wireless Communication Systems (ISWCS 2014)*, Barcelona, Spain, Aug. 2014.
- [25] R. Radner, "Team decision problems", *the Annals of Mathematical Statistics*, 1962.
- [26] J. Nash, "Non-cooperative games", *Annals of Mathematics*, 1951.
- [27] Y. Ho, and K. Chu, "Team decision theory and information structure in optimal control problems-Part I", *IEEE Trans. Automat. Contr.*, vol. 17, No. 1, pp. 15-22, Feb. 1972.
- [28] S. S. Christensen, R. Agarwal, E. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Communications*, vol. 7, No. 12, pp. 4792-4799, Dec. 2008.
- [29] R. Fritzsche and G. P. Fettweis, "Robust precoding with general power constraint considering unbounded channel uncertainty," *9th International Symposium on Wireless Communication Systems (ISWCS 2012)*, Paris, France, Aug. 2012.
- [30] N. K. Babu and P. S. S. Babu, "Design of Physical Coding sublayer using 8B/10B algorithm," *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN: 2277-3878, Volume 2, Issue 2, May 2013
- [31] iJOIN Project, "D3.1- Final report on MAC/RRM state-of-the-art, Requirements, scenarios and interfaces in the iJOIN architecture," November 2013. Available: <http://www.ict-ijoin.eu/wp-content/uploads/2014/01/D3.1.pdf>
- [32] iJOIN Project. "D4.1- Report on SotA and requirements for network-layer algorithms and network operation and management," November 2013. Available: <http://www.ict-ijoin.eu/wp-content/uploads/2014/01/D4.1.pdf>
- [33] U. Dötsch, M. Doll, H. P. Mayer, F. Schaich, J. Segel, and P. Sehier, "Quantitative Analysis of Split Base Station Processing and Determination of Advantageous Architectures for LTE," *Bell Labs Technical Journal*, vol. 18, no. 1, pp. 105-128, May 2013.
- [34] J. Bartelt and G. Fettweis, "A Soft-Input/Soft-Output Dequantizer for Cloud-Based Mobile Networks," *15th IEEE International Workshop on Signal Processing Advances in Wireless Communications (IEEE SPAWC 2014)*, Toronto, Canada, 2014.
- [35] A.M.A. Hussien, M.S. Khairy, A. Khajeh, K. Amiri, A.M. Eltawil, F.J. Kurdahi, "A combined channel and hardware noise resilient Viterbi decoder," *Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pp. 395-399, Nov. 2010
- [36] iJOIN Project, "D2.1- State-of-the-art of and promising candidates for PHY layer approaches on access and backhaul network," November 2013. Available: <http://www.ict-ijoin.eu/wp-content/uploads/2014/01/D2.1.pdf>

- [37] D. Wübben, P. Rost, J. Bartelt, M. Lalam, V. Savin, M. Gorgoglione, A. Dekorsy, and G. Fettweis, "Benefits and Impact of Cloud Computing on 5G Signal Processing," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 35-44, Nov. 2014.
- [38] G. Xu, H. Paul, D. Wübben, and A. Dekorsy, "Fast Distributed Consensus-based Estimation for Cooperative Wireless Sensor Networks," 18th International ITG Workshop on Smart Antennas (WSA 2014), Erlangen, Germany, Mar. 2014.
- [39] D. Wübben, H. Paul, B.-S. Shin, G. Xu, and A. Dekorsy, "Comparative Study of Distributed Consensus-based Estimation Schemes for Small-Cell Networks," *European Conference on Networks and Communications (EuCNC 2014)*, Bologna, Italy, June 2014.
- [40] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," in *Soviet. Math. Dokl.*, Vol. 27, pp. 372-376, 1983.
- [41] T. Goldstein, B. O'Donoghue, and S. Setzer, "Fast Alternating Direction Optimization Methods," CAM report 12-35, UCLA, May 2012.
- [42] B. Sklar, "Rayleigh Fading Channels in Mobile Digital Communication Systems Part I: Characterization," *IEEE Communications Magazine*, vol. 35, no. 9, pp. 90-110, July 1997.
- [43] 3GPP, "Release 12 Work Item on Network-Assisted Interference Cancellation (NAIC)," Available: <http://www.3gpp.org/DynaReport/WiVsSpec--590018.htm>
- [44] REPORT ITU-R M2135, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," 2008.
- [45] R.G. Gallager, "Low density parity check codes," MIT Press, Cambridge, 1963, Research Monograph series.
- [46] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. on Communications*, vol. 44, no. 10, pp. 1261-1271, 1996.
- [47] M. M. Mansour and N. R. Shanbhag, "High-throughput LDPC decoders," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 11, no. 6, pp. 976-996, 2003.
- [48] A. Darabiha, A.C. Carusone, and F.R. Kschischang, "Multi-Gbit/sec low density parity check decoders with reduced interconnect complexity," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2005, pp. 5194-5197.
- [49] Z. Cui, Z. Wang, and Y. Liu, "High-throughput layered LDPC decoding architecture," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, no. 4, pp. 582-587, 2009.
- [50] C. Zhang, Z. Wang, J. Sha, L. Li, and J. Lin, "Flexible LDPC decoder design for multigigabit-per-second applications," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 1, pp. 116-124, 2010.
- [51] Z. Zhang, V. Anantharam, M. J. Wainwright, and B. Nikolic, "An efficient 10GBASE-T ethernet LDPC decoder design with low error floors," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, pp. 843-855, 2010.
- [52] S. Seo, T. Mudge, Y. Zhu, and C. Chakrabarti, "Design and analysis of LDPC decoders for software defined radio," *IEEE Workshop on Signal Processing Systems*, 2007.
- [53] G. Falcao, V. Silva, L. Sousa, and J. Marinho, "High coded data rate and multi-codeword WiMAX LDPC decoding on Cell/BE," *IET Electronics Letters*, vol. 44, no. 24, pp. 1415-1417, 2008.
- [54] Y. Zhu and C. Chakrabarti, "Architecture-aware LDPC code design for multiprocessor software defined radio systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3679-3692, 2009.
- [55] H. Ji, J. Cho, and W. Sung, "Massively parallel implementation of cyclic LDPC codes on a general purpose graphics processing unit," *IEEE workshop on Signal Processing Systems (SiPS)*, 2009, pp. 285-290.

- [56] J. Zhao, M. Zhao, H. Yang, J. Chen, X. Chen, and J. Wang, "High performance LDPC decoder on CELL BE for WiMAX system," IEEE International Conference on Communications and Mobile Computing (CMC), 2011, pp. 278-281.
- [57] G. Falcao, L. Sousa, and V. Silva, "Massively LDPC decoding on multicore architectures," IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 2, pp. 309-322, 2011.
- [58] G. Wang, M. Wu, Y. Sun, and J.R. Cavallaro, "A massively parallel implementation of QC-LDPC decoder on GPU," IEEE Symposium on Application Specific Processors (SASP), 2011, pp. 82-85.
- [59] A. Diavastos, P. Petrides, G. Falcao, and P. Trancoso, "LDPC decoding on the Intel SCC," IEEE 20th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), 2012, pp. 57-65.
- [60] M. Wu, G. Wang, B. Yin, C. Studer, and J.R. Cavallaro, "HSPA+/LTE-A Turbo Decoder on GPU and Multicore CPU," 47th IEEE Asilomar Conference on Signals, Systems, and Computers (ASILOMAR), Nov. 2013.
- [61] 3GPP TR 36.814, "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release9)," V9.0.0, Mar. 2010.
- [62] 3GPP, "TR 36.932 V12.0.0; Scenarios and Requirements for Small Cell Enhancements for E-UTRA and E-UTRAN," Dec. 2012
- [63] 3GPP TR 36.872, "Small cell enhancements for E-UTRA and E-UTRAN - Physical layer aspects (Release12)," V12.1.0, Dec. 2012.
- [64] M.P.C. Fossorier, M. Mihaljevic, and H. Imai, "Reduced complexity iterative decoding of low-density parity check codes based on belief propagation," IEEE Trans. on Communications, vol. 47, no. 5, pp. 673-680, 1999.
- [65] F. R. Kschischang and B. J. Frey, "Iterative decoding of compound codes by probability propagation in graphical models," IEEE Journal on Selected Areas in Communications, vol. 16, no. 2, pp. 219-230, 1998.
- [66] J. Zhang and M. P. C. Fossorier, "Shuffled iterative decoding," IEEE Trans. on Communications, vol. 53, no. 2, pp. 209-213, 2005.
- [67] E. Sharon, S. Litsyn, and J. Goldberger, "Efficient serial message-passing schedules for LDPC decoding," IEEE Trans. on Inf. Theory, vol. 53, no. 11, pp. 4076-4091, 2007.
- [68] D. E. Hocevar, "A reduced complexity decoder architecture via layered decoding of LDPC codes," in Proc. of IEEE Workshop on Signal Processing Systems (SIPS), 2004, pp. 107-112.
- [69] J. Zhang, Y. Wang, M. P. C. Fossorier, and J. S. Yedidia, "Iterative decoding with replicas," IEEE Transactions on Information Theory, vol. 53, no. 5, pp. 1644-1663, 2007.
- [70] M.P.C. Fossorier, "Quasi cyclic Low-Density Parity-Check Codes from Circulant Permutation Matrices," IEEE Trans. on Information Theory, vol. 50, no. 8, pp. 1788-1793, 2004.
- [71] M J. Quirm, "Parallel Programming in C with MPI and OpenMP," McGraw-Hill Press MIT Press, New York, 2004.
- [72] M. Kaiser, W. Fong, M. Sikora, and D.J. Costello Jr., "A Comparison of Decoding Latency for Block and Convolutional Codes," Tenth International Symposium on Communication. Theory and Application (ISCTA'09), Ambleside, UK, July 2009.
- [73] K. Hueske, J. Geldmacher, and J. Götze, "Multi Core Implementation of a Trellis Based Syndrome Decoder with Adaptive Complexity," 7th International Symposium on Wireless Communication Systems (ISWCS 2010), York, UK, Sept. 2010.
- [74] L. Li, R.G. Maunder, B.M. Al-Hashimi, and L. Hanzo, "A Low-Complexity Turbo Decoder Architecture for Energy-Efficient Wireless Sensor Networks," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol.21, no.1, pp.14-22, Jan. 2013.

- [75] G. Bosco, G. Montorsi, and S. Benedetto, "Soft Decoding in Optical Systems," *IEEE Transactions on Communications*, Vol. 51, NO. 8, August 2003.
- [76] RAN Evolution Project, "Further Study on Critical C-RAN Technologies," Version 1.0 (approved), 20 May 2014.
- [77] K. F. Nieman and B. L. Evans, "Time-Domain Compression of Complex-Baseband LTE Signals for Cloud Radio Access Networks," *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Austin, TX, USA, Dec. 2013.
- [78] IDT, Integrated Device Technology, "Compression IP for Wireless Infrastructure Applications," <http://www.idt.com/>
- [79] EU project ACCORDANCE, Deliverable D5.3, "Combined optical and wireless/wireline access based on existing requirements," Framework Programme (FP7/2007-2013), grant 248654
- [80] D. Wübben, H. Paul, P. Balleydier, V. Savin, and P. Rost, "Decoder Implementation for Cloud Based Architectures," *European Conference on Networks and Communications (EuCNC)*, Bologna, Italy, June 2014.
- [81] J. Bartelt and G. Fettweis, "An Improved Decoder for Cloud-Based Mobile Networks under Imperfect Fronthaul," *Globecom 2014 Workshop – Wireless optical network conversion in support of cloud architectures (GC14 WS- WONC)*, Austin, USA, Dec. 2014.
- [82] D. Wübben, H. Paul, B.-S. Shin, G. Xu, and A. Dekorsy, "Distributed Consensus-Based Estimation for Small Cell Cooperative Networks," *Globecom 2014 Workshop – Broadband Wireless Access (GC14 WS-BWA)*, Austin, USA, Dec. 2014.
- [83] V. Suryaprakash, A. Fehske, A. F. dos Santos, and G. P. Fettweis, "On the impact of sleep modes and BW variation on the energy consumption of radio access networks," *IEEE 75th. Vehicular Technology Conference (VTC Spring)*, Yokohama, Japan, May 2012.
- [84] iJOIN Project. "D4.2- Network-layer algorithms and network operation and management: candidate technologies specification," November 2014.
- [85] China Mobile Research Institute, "C-RAN The Road Towards Green RAN," White Paper, version 2.5, October 2011.
- [86] G. Xu, H. Paul, D. Wübben, and A. Dekorsy, "Distributed Augmented Lagrangian Method for Cooperative Estimation in Small Cell Networks," *10th International ITG Conference on Systems, Communications and Coding (SCC 2015)*, Hamburg, Germany, Feb. 2015.
- [87] iJOIN Project. "D3.2- Definition of MAC and RRM approaches for RANaaS and a joint backhaul/access design," November 2014.
- [88] A. Maeder, M. Lalam, A. De Domenico, E. Pateromichelakis, D. Wübben, J. Bartelt, R. Fritzsche, and P. Rost, "Towards a Flexible Functional Split for Cloud-RAN Networks," *European Conference on Networks and Communications (EuCNC 2014)*, Bologna, Italy, June 2014.
- [89] iJOIN Project. "D6.1- Preliminary proof-of-concept results for selected candidate algorithms," November 2014.
- [90] M. Joham, W. Utschick, and J. Nossék, "Linear transmit processing in MIMO communications systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2700–2712, Aug 2005.
- [91] S. Colieri, M. Ergen, A. Puri, and A. Bahai, "A study of channel estimation in OFDM systems," in *IEEE Vehicular Technology Conference*, vol. 2, 2002, pp. 894–898 vol.2.
- [92] J. Lee, N. Jindal, "Dirty Paper Coding vs. Linear Precoding for MIMO Broadcast Channels," *Fortieth Asilomar Conference on Signals, Systems and Computers*, Oct. 2006, pp.779-783.
- [93] R. D. Wesel, J. M. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Transactions on Information Theory*, vol.44, no.2, pp.824-831, Mar 1998.
- [94] J. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122–3134, November 2011.

- [95] M. K. Simon and M.-S. Alouini, "Digital Communication over Fading Channels," 2nd ed. Wiley, 2005.
- [96] D. Kaltakis, M. Imran, and C. Tzaras, "Information theoretic capacity of cellular multiple access channel with shadow fading," *IEEE Transactions on Communications*, vol. 58, no. 5, pp. 1468–1476, May 2010.
- [97] iJOIN Project, "D5.2- Final definition of Requirements and Scenarios," November 2014.
- [98] V. S. Ryaben'kii and S. V. Tsynkov, "A Theoretical Introduction to Numerical Analysis," Florida, USA: CRC Press, 2006.
- [99] Y. Qi, M. Ali Imran, A. Quddus, and R. Tafazolli, "Achievable Rate Optimization for Coordinated Multi-Point Transmission (CoMP) in Cloud-Based RAN Architecture," *IEEE International Conference on Communications (ICC)*, London, UK, June 2014.
- [100] M. R. Rurter, "An Introduction to Computing", Lecture notes University of Cambridge, Available: <http://www.tcm.phy.cam.ac.uk/~mjr/lect1.pdf>, 2001.
- [101] R. Fritzsche, G. Schulzig, and G. Fettweis, "Forward signaling strategies for MU-MIMO downlink transmission," *20th International Conference on Telecommunications (ICT)*, 6-8 May 2013.
- [102] R. Zakhour and D. Gesbert, "Optimized Data Sharing in Multicell MIMO With Finite Backhaul Capacity," *IEEE Transactions on Signal Processing*, vol.59, no.12, pp.6102-6111, Dec. 2011.
- [103] M. Grieger, P. Marsch, G. Fettweis, and J.M. Cioffi, "On the Performance of Compressed Interference Forwarding for Uplink Base Station Cooperation," *IEEE Global Telecommunications Conference (GLOBECOM 2009)*, Nov. 30 2009-Dec. 4, 2009.
- [104] M. Grieger, P. Helbing, G. Fettweis, and P. Marsch, "Field trial evaluation of compression algorithms for distributed antenna systems," *34th IEEE Sarnoff Symposium*, 3-4 May 2011.
- [105] B. Guo, W. Cao, A. Tao, and D. Samardzija, "CPRI compression transport for LTE and LTE-A signal in C-RAN," *7th International ICST Conference on Communications and Networking in China (CHINACOM)*, Aug. 2012.
- [106] S. Nanba and A. Agata, "A new IQ data compression scheme for front-haul link in Centralized RAN," *IEEE 24th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC Workshops)*, 2013.
- [107] J. Max, "Quantizing for minimum distortion." *IRE Transactions on Information Theory*, pp. 7-12, 1960.
- [108] S. Lloyd. "Least squares quantization in PCM." *IEEE Transactions on Information Theory*, pp. 129-137, 1982.
- [109] P. Marsch and G. Fettweis, "Coordinated Multi-Point in Mobile Communications - From Theory to Practice", Cambridge University Press, Aug. 2011.
- [110] N. Cassiau, D. Ktésas, and J-B. Doré, "Time and frequency synchronization for CoMP with FBMC", *10th International Symposium on Wireless Communication Systems (ISWCS 2013)*, 27-30 Aug. 2013, Ilmenau, Germany.
- [111] D. Bladsjo, M. Hogan, and S. Ruffini, "Synchronization aspects in LTE small cells," *IEEE Communications Magazine*, vol.51, no.9, pp.70-77, Sept. 2013.
- [112] 3GPP TS 36.104, "Base station (BS) radio transmission and reception," V12.0.0, Jul. 2013.
- [113] IEEE 1588-2008, "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," 2008
- [114] Semtech, "ACS9550 product brief," Feb. 2010, available <http://www.semtech.com/images/datasheet/acs9550pb.pdf>
- [115] R. Preuss and D.R. Brown III, "Two-way synchronization for coordinated multi-cell retrodirective downlink beamforming," *IEEE Transactions on Signal Processing*, vol.59, no.11, pp.5415–5427, Nov. 2011.

-
- [116] V. Jungnickel, T. Wirth, M. Schellmann, T. Haustein, and W. Zirwas, "Synchronization of cooperative base stations," IEEE International Symposium on Wireless Communication Systems (ISWCS 2008), pp 329-334, Oct. 2008, Reykjavik, Island.
- [117] V. Kotsch and G. Fettweis, "On synchronization requirements and performance limitations for CoMP systems in large cells," 8th International Workshop on Multi-Carrier Systems & Solutions (MC-SS), Herrsching, Germany, May 2011
- [118] K. Takeda, Y. Yuda, M. Hoshino, T. Iwai, A. Matsumoto, and A. Nishio, "Impact of Channel Estimation Error on LTE-Advanced Uplink Using CoMP Joint Reception in a Heterogeneous Network," IEEE 77th Vehicular Technology Conference (VTC Spring 2013), June 2013, Dresden, Germany.
- [119] K. Manolakis, V. Jungnickel, C. Oberli, T. Wild, and V. Braun, "Impairments in Cooperative Mobile Networks: Models, Impact on Performance and Mitigation" European Wireless 2014, Barcelona, Spain, May 2014.
- [120] Enablers for Ambient Services and Systems Part C (EASY-C), BMBF project, <http://www.easy-c.org/>.
- [121] Advanced Radio Interface Technologies for 4G Systems (ARTIST-4G), FP7 project, <https://ict-artist4g.eu/>.
- [122] T. Wild, N. Le-Hang, and S. ten Brink, "Joint channel estimation across multiple cells in coordinated multi-point," International Symposium on Wireless Communication Systems (ISWCS), Paris, France, Aug. 2012.
- [123] P. Marsch and G. Fettweis, "Uplink CoMP under a Constrained Backhaul and Imperfect Channel Knowledge," IEEE Transactions on Wireless Communications, vol. 10 , no. 6, pp. 1730-1742, June 2011.
- [124] P. Marsch, P. Rost, and G. Fettweis, "Application Driven Joint Uplink-Downlink Optimization in Wireless Communications," International ITG Workshop on Smart Antennas, Bremen, Germany, Feb. 2010.
- [125] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzaresse, S. Nagata, K. Sayana, "Coordinated Multipoint Transmission and Reception in LTE-Advanced: Deployment Scenarios and Operational Challenges" IEEE Communication Magazine, vol. 50, no. 2, pp. 148-155, Feb. 2012.
- [126] S. Sesia, I. Toufik, M. Baker "LTE - The UMTS Long Term Evolution: From Theory to Practice", Wiley, July 2011.
- [127] D. J. Love, R. W. Heath, V.K.N. Lau, D. Gesbert, B.D. Rau, and M. Andrews, "An overview of limited feedback in wireless communication systems", IEEE Journal on Selected Areas in Communications, vol. 26, no. 8, pp. 1341-1365, Oct. 2008.
- [128] Y. Cheng, V. K. N. Lau, and Y. Long, "A Scalable Limited Feedback Design for Network MIMO Using Per-Cell Product Codebook," IEEE Trans. on Wireless Communications, vol. 9, no. 10, pp. 3093-3099, Oct. 2010.