Helsinki University of Technology Department of Signal Processing and Acoustics
Teknillinen korkeakoulu Signaalinkäsittelyn ja akustiikan laitos
Espoo 2010 Report 15

CONVERGENCE OF PACKET COMMUNICATIONS OVER THE EVOLVED MOBILE NETWORKS; SIGNAL PROCESSING AND PROTOCOL PERFORMANCE

Mika Rinne

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Aalto University School of Science and Technology
Department of Signal Processing and Acoustics
Espoo 2010 Report 15

CONVERGENCE OF PACKET COMMUNICATIONS OVER THE EVOLVED MOBILE NETWORKS; SIGNAL PROCESSING AND PROTOCOL PERFORMANCE

Mika Rinne

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Faculty of Electronics, Communications and Automation for public examination and debate in Auditorium S4 at the Aalto University School of Science and Technology (Espoo, Finland) on the 26th of March, 2010, at 12 noon.

Aalto University School of Science and Technology
Faculty of Electronics, Communications and Automation
Department of Signal Processing and Acoustics

Aalto-yliopiston teknillinen korkeakoulu
Elektroniikan, tietoliikenteen ja automaation tiedekunta
Signaalinkäsittelyn ja akustiikan laitos

Distribution:
Aalto University School of Science and Technology
Faculty of Electronics, Communications and Automation
Department of Signal Processing and Acoustics
P.O. Box 13000
FIN-00076 Aalto
Tel. +358-9-47023211
Fax. +358-9-4523614
E-mail: Mirja.Lemetyinen@tkk.fi

ISBN 978-952-60-3067-8 (Printed)
ISBN 978-952-60-3068-5 (Electronic)
ISSN 1797-4267

Multiprint Oy
Espoo 2010

Abstract

In this thesis, the convergence of packet communications over the evolved mobile networks is studied. The Long Term Evolution (LTE) process is dominating the Third Generation Partnership Project (3GPP) in order to bring technologies to the markets in the spirit of continuous innovation. The global markets of mobile information services are growing towards the Mobile Information Society.

The thesis begins with the principles and theories of the multiple-access transmission schemes, transmitter receiver techniques and signal processing algorithms. Next, packet communications and Internet protocols are referred from the IETF standards with the characteristics of mobile communications in the focus. The mobile network architecture and protocols bind together the evolved packet system of Internet communications to the radio access network technologies. Specifics of the traffic models are shortly visited for their statistical meaning in the radio performance analysis. Radio resource management algorithms and protocols, also procedures, are covered addressing their relevance for the system performance. Throughout these Chapters, the commonalities and differentiators of the WCDMA, WCDMA/HSPA and LTE are covered. The main outcome of the thesis is the performance analysis of the LTE technology beginning from the early discoveries to the analysis of various system features and finally converging to an extensive system analysis campaign. The system performance is analysed with the characteristics of voice over the Internet and best effort traffic of the Internet. These traffic classes represent the majority of the mobile traffic in the converged packet networks, and yet they are simple enough for a fair and generic analysis of technologies. The thesis consists of publications and inventions created by the author that proposed several improvements to the 3G technologies towards the LTE. In the system analysis, the LTE showed by the factor of at least 2.5 to 3 times higher system measures compared to the WCDMA/HSPA reference. The WCDMA/HSPA networks are currently available with over 400 million subscribers and showing increasing growth, in the meanwhile the first LTE roll-outs are scheduled to begin in 2010. Sophisticated 3G LTE mobile devices are expected to appear fluently for all consumer segments in the following years.

Tiivistelmä

Väitöskirja käsittelee pakettimuotoisen viestinnän yhdentymistä päätelaitteisiin kattaen uuden sukupolven kehittyvät mobiiliverkkojärjestelmät. Long Term Evolution (LTE) on kehityssarja kolmannen sukupolven standardointiprojektissa (3GPP), joka pyrkii tuottamaan teknologioita markkinoille jatkuvassa, uutta luovassa hengessä. Langattomat päätelaitteet ja niiden verkkoyhteydet maailmanlaajuisiin palveluihin ovat kehittymässä osaksi liikkumisen vapauden mahdollistamaa tietoyhteiskuntaa.

Aluksi väitöskirja esittelee monikäyttöjärjestelmien, lähetin-vastaanotin-tekniikoiden ja signaalinkäsittelyalgoritmien keskeisiä periaatteita ja teorioita. Seuraavaksi väitöskirja kuvaa IETF-standardeja käyttävät pakettimuotoisen viestinnän periaatteet, jotka toimivat Internetissä. Näiden osalta väitöskirja kohdentuu ominaisuuksiin, jotka ovat tärkeitä langattoman viestinnän ja päätelaitteiden kannalta. Tietoliikenteen pakettimallit ja niiden tilastolliset ominaisuudet esitellään siltä osin kun ne vaikuttavat radiojärjestelmän ominaisuuksiin ja suorituskykyyn. Työssä on kehitetty radioresurssien hallinta-algoritmeja, protokollia ja menetelmiä, joista järjestelmän suorituskyky olennaisesti riippuu. Väitöskirja kattaa WCDMA, WCDMA/HSPA ja LTE teknologioiden ja niiden teorioiden yhtäläisyydet ja eroavaisuudet. Väitöskirjan päätulos on LTE teknologian järjestelmäsuorituskyky, jonka selvittäminen alkaa järjestelmän perusominaisuuksien yksittäisillä analyyseilla ja etenee lopulta laajaan järjestelmän analyysiin. Järjestelmän suorituskyvyn analyysiin on valittu sekä pakettimuotoinen puheliikenne että pakettipurskeinen tietoliikenne, jotka ovat yleisiä Internetissä. Nämä tietoliikenteen tyypit edustavat suurinta ja tärkeintä osaa liikkuvan tietoliikenteen palveluista yhdentyvissä liikkuvan tietoliikenteen pakettiverkoissa, ja kuitenkin nämä valitut liikennetyypit ovat riittävän selkeitä, jotta ne soveltuvat malleiksi objektiiviseen ja yleiseen teknologiavertailuun.

Järjestelmäanalyysissä havaitaan LTE teknologian parantavan liikkuvan tietoliikenteen suorituskykyä ainakin 2.5 tai kolminkertaiseksi verrattuna aiempaan WCDMA/HSPA teknologiaan. WCDMA/HSPA verkoissa on yli 400 miljoonaa asiakasta, ja niiden käyttäjämäärät ovat jatkuvasti kasvussa, samaan aikaan kun LTE teknologian ja verkkojen käyttöönottoa päätelaitteissa suunnitellaan alkavaksi vuonna 2010. Tulevina vuosina huippukyvykkäät 3G LTE mobiililaitteiden sukupolvet ilmestyvät vaiheittain markkinoille kaikkia kuluttajaryhmiä palvelevina tuotemalleina.

Preface

The research work for this doctoral thesis was carried out at Nokia Research Center for the Department of Signal Processing and Acoustics, in Aalto University.

I would like to express my gratitude to Prof. Risto Wichman, the supervisor of the thesis, for his encouragement, guidance and support during the summary of this work. As well, I would like to give warm thanks to Prof. Olav Tirkkonen for sharing his innovative thinking and for mentoring my research work.

Then, I would like to thank the pre-examiners Prof. Mamoru Sawahashi at Tokyo City University and Dr. Janne Peisa at LM Ericsson. Next, I thank managers Klaus Hugl, Yrjö Kaipainen, Juha Laurila, Kari Rissanen at Nokia Research Center, Kari Pehkonen, Mikko Rinne at Nokia Devices, Antti Toskala and Matti Kiiski at Nokia Siemens Networks. I also like to thank Jussi Numminen, Sari Nielsen, Asbjörn Grövlen as well as their standardisation teams. I acknowledge the head of the laboratory Petteri Alinikula, his predecessors Jukka Soikkeli and Pekka Soininen, as well as Seppo Granlund and Ari Ahtiainen.

I would especially like to thank the co-authors of my conference and journal articles, and co-inventors of my patents and patent applications; Jean-Philippe Kermoal, Markku Kuusela, Gregory Manuel, Jussi Kähtävä, Päivi Purovesi, Kari Pajukoski, Frank Frederiksen, István Kovács, Tsuyoshi Kashima and others, with respectful thanks for cooperation to Prof. Markku Renfors, Prof. Preben Mogensen, Prof. Tapani Ristaniemi and their research teams. Great thanks belong naturally to my colleagues and friends at Nokia.

Finally, I thank my parents Airi and Esko, brother Jyrki, and express the warmest thanks to my wife Sirpa and children Karolina, Emilia and Markus for inspiration.

Espoo, March 2010,

Mika Rinne

Contents

List of abbreviations and symbols

Abbreviations

3G	third generation technology
3GPP	third generation partnership project
ACI	adjacent channel interference
ACK	acknowledgement (used e.g. in HARQ, RLC, TCP, ROHC protocols)
AMR	adaptive multirate voice codec
AMR-WB	adaptive multirate codec for wideband voice and audio
AP	application protocol over a defined RAN interface
APN	access point name
ATB	adaptive transmission bandwidth
AVI	Actual value interface for SINR to BLER mapping in a packet transport system
BCCH	broadcast control channel (a logical channel type)
BCH	broadcast channel (a transport channel type)
BEP	bit error probability
BLER	block error rate
BSR	buffer status report
C-RNTI	cell radio network temporary identity
CCCH	common control channel (a logical channel type)
cdf	cumulative distribution function
CDI	channel distribution information
CDMA	code division multiple access
cdma2000	evolution of data optimized CDMA standards (EV-DO)
CID	context identity
CoIP	communications over the internet protocol

CP	cyclic prefix
CQI	channel quality indication
CRC	cyclic redundancy check
CSCF	call state control function
CSI	channel state information
DCCH	dedicated control channel (a logical channel type)
DiffServ	differentiated services in the internet
DIR	dominant interference power ratio
DRX	discontinuous reception
DTCH	dedicated traffic channel (a logical channel type)
E-DCH	enhanced dedicated channel for HSPA (a transport channel type)
E-DPCCH	enhanced (HSPA) dedicated physical control channel (a physical channel type)
E-DPDCH	enhanced (HSPA) dedicated physical data channel (a physical channel type)
EDGE	enhanced data rates for global evolution
EESM	exponential effective SNR mapping
EFR	enhanced full-rate voice codec
eNodeB (eNB)	evolved NodeB i.e. EUTRAN base station
EPC	evolved packet core
EPS	evolved packet system
EUTRA	evolved UMTS terrestrial radio access
EUTRAN	evolved UMTS terrestrial radio access network
EVM	error vector magnitude
EVRC	enhanced variable rate codec
FACH	forward access channel (a transport channel type)
FBI	feedback indication
FDMA	frequency division multiple access
FDE	frequency domain equalizer
FO	first order state (header) of the robust header compression
FPLMTS	future public land mobile telecommunications system
FRAMES	future radio wideband multiple access system
FT	frame type, a protocol field in the AMR frame
FTP	file transfer protocol
GGSN	gateway GPRS support node (corresponds to the PDN gateway of LTE)

GPRS	general packet radio service
GSM	global system for mobile communications, orig. Groupe Spécial Mobile
GTP	GPRS tunneling protocol
HSCSD	high-speed circuit-switched data
HS-DPCCH	HSPA dedicated physical control channel (a physical channel type)
HS-DSCH	HSPA downlink shared channel (a transport channel type)
HS-PDSCH	HSPA physical downlink shared channel (a physical channel type)
HS-SCCH	HSPA physical shared control channel (a physical channel type)
HSPA	high-speed packet access
HSDPA	high-speed downlink packet access
HSS	home subscriber server
HSUPA	high-speed uplink packet access
HTTP	hypertext transfer protocol
I-branch	in-phase signal of a complex modulator
IBI	interblock-interference
IBL	information block length (number of information bits in a code block)
ICI	intercarrier-interference
ICMP	internet control message protocol
IEEE	Eye-triple-E
IETF	internet engineering task force
IFDMA	interleaved frequency division multiple access
i.i.d.	independent and identically distributed
IMAP	internet message access protocol (for email)
IMS	IP multimedia subsystem
IoT	interference over thermal noise
IOTA	isotropic orthogonal transform algorithm
IP	internet protocol
IR	initialization and refresh state of robust header compression
IRC	interference rejection combining receiver
IS-95	interim standard 95 a.k.a cdmaOne spread spectrum system
ISI	intersymbol-interference
IMT	international mobile telecommunications
ITU-R	international telecommunication union - radio (ITU-R)

LDPC	low density parity code
LLR	log-likelihood ratio, a logaritmic decision metric in a channel coder
LMMSE	linear minimum mean-squared error (estimator)
LTE	long term evolution, process and technology
MAC	medium access control protocol
MAI	multiple-access interference
MAP	GSM mobile application part
MCS	modulation and coding scheme
MIMO	multiple input multiple output
MME	mobility management entity
MMSE	minimum mean-squared error
MRC	maximum ratio combining
MTU	maximum transmission unit of a transport network e.g. Ethernet
MUD	multi-user detection
NACK	negative acknowledgement (used e.g. in HARQ, RLC, TCP, ROHC protocols)
NGMN	next generation mobile networks alliance
NodeB (NB)	WCDMA or WCDMA/HSPA base station
OFDMA	orthogonal frequency division multiple access
OSI	open systems interconnection, a reference model of layered protocols
OVSF	orthogonal variable spreading factor
PAM	pulse-amplitude modulation
PCCH	paging control channel (a logical channel type)
PCH	paging channel (a transport channel type)
PCI	precoding control information
PCS	personal communications service
PDCCH	physical downlink control channel (a physical channel type of EUTRA)
PDCP	packet data convergence protocol
pdf	probability density function a.k.a probability function or probability distribution
PDN gateway	packet data network gateway
PDP context	packet data protocol context
PDSCH	physical downlink shared channel (a physical channel type of EUTRA)
PDU	protocol data unit
PEP	packet error probability

PHR	power headroom report
PIC	parallel interference cancellation receiver
PING	software utility to test whether a host is reachable across IP network
	PING measures RTT, optionally with payload (use ICMP echo request/echo reply)
PMI	precoding matrix information
PN	pseudo-noise (sequence)
POP	post office protocol (for email)
PRB	physical resource block
PT	payload type, a protocol field in the RTP frame
PUCCH	physical uplink control channel (a physical channel type of EUTRA)
PUSCH	physical uplink shared channel (a physical channel type of EUTRA)
Q-branch	quadrature-phase signal of a complex modulator
QAM	quadrature amplitude modulation
QoS	quality of service
QPSK	quadrature phase shift keying
RACH	random access channel (a transport channel type)
RAN	radio access network
RIT	radio interface technology
RLC	radio link control protocol
ROHC	robust header compression
RRC	radio resource control protocol
rrc	root-raised-cosine pulse shaping
RRM	radio resource management
RS	reference symbol (a.k.a pilot or training symbol)
RSVP	resource reservation protocol
RTP	real-time transport protocol
RTT	radio transmission technology
RTT	round-trip time,
	latency of a packet from a node to a remote node and back (see PING)
SC-FDMA	single carrier frequency division multiple access (a.k.a DFT-spread-OFDMA)
S-TMSI	SAE temporary mobile subscriber identity
SAE	system architecture evolution of the evolved packet system
SCM	spatial channel model

SDU	service data unit
SF	spreading factor
SGSN	serving GPRS support node (corresponds to the serving gateway of LTE)
SIC	successive interference cancellation receiver
SID	silence descriptor, a special RTP frame for silence period
SINR	signal-to-interference-plus-noise ratio
SIP	session initiation protocol
SNR	signal-to-noise ratio
SO	second order state (header) of robust header compression
SR	scheduling request (a UE-specific signal in PUCCH of EUTRA)
SRS	sounding reference symbol
STTD	space-time transmit diversity
SUC	satisfied user criterion (for voice over IP)
TA	timing advance (uplink transmission phase relative to the downlink reference)
TCP	transmission control protocol
TDMA	time division multiple access
TEID	tunnel endpoint identifier
TF	transport format (selected modulation alphabet and effective code rate)
TFCI	transport format combination indicator
TFRI	transport format and resource indicator (in HSPA)
TFT	traffic flow template
TPC	transmit power control
TTI	transmission time interval (also triggering interval of a transport channel)
TU	typical urban channel model
Turbo code	code resulting from parallel (or serial) concatenation of convolutional codes
TxAA	adaptive antenna transmission, see closed-loop mode 1 diversity
UDP	user datagram protocol
UE	user equipment, the mobile terminal for WCDMA, WCDMA/HSPA, LTE
UMB	ultra mobile broadband
UMTS	universal mobile telecommunications system
URI	uniform resource indicator
URL	uniform resource locator
URN	uniform resource name

VAF	voice activity factor
VoIP	voice over the internet protocol
VSCRF	variable spreading and chip repetition factors, a CDMA scheme
WCDMA	wideband code division multiple access
WCDMA/HSPA	wideband code division multiple access/high-speed packet access
WiFi	wireless fidelity a.k.a WLAN
WiMax	worldwide interoperability for microwave access a.k.a IEEE 802.16e standard
WLAN	wireless local area network a.k.a WiFi a.k.a set of IEEE 802.11 standards
WSSUS	wide-sense stationary uncorrelated scattering (used for a channel model)
ZF	zero-forcing equalizer

Symbols

j	imaginary unit
$*$	linear convolution operator
$\circledast$	circular convolution operator
$\otimes$	Kronecker (tensor) product (element-by-element matrix multiplication)
$\|\cdot\|$	Euclidean norm
$\overline{(\cdot)}$	the mean operator
arg	argument of a complex number
$\arg\max(\cdot)$	maximizing argument of $(\cdot)$
$\arg\min(\cdot)$	minimizing argument of $(\cdot)$
$E[\cdot]$	the expectation operator
$\exp(\cdot)$	exponential function
$\max\{\cdot\}$	the maximum
$\min\{\cdot\}$	the minimum
$(\cdot)\ mod\ (\cdot)$	modulo operation
$P(\cdot)$	probability of an event
$p(\cdot\|\cdot\cdot)$	joint probability density function of $(\cdot)$ with the condition of $(\cdot\cdot)$
$\_th$	threshold, fractile of a pdf set to trigger an action
$Re\{\cdot\}$	real part of a complex scalar
$Im\{\cdot\}$	imaginary part of a complex scalar
$\propto$	proportional to
$(\cdot)^*$	complex conjugate
$[\cdot]^T$	transpose of a matrix
$[\cdot]^H$	Hermitian transpose (i.e. complex conjugate transpose) of a matrix
$[\cdot]^{-1}$	inverse of a matrix
$diag(\cdot)$	diagonal matrix of the argument vector
$Tr\{\cdot\}$	trace of a matrix
$vec(\cdot)$	operator to stack columns of a matrix on top of each other to a vector
$z(\cdot)$	transform precoding function
a_k	weight of the k^{th} correlator in a Rake receiver
$a_{k,l}^{(n_r)}$	complex-valued symbol a modulated to a resource element (k,l)
	at antenna port (n_r)

$a_{k,l}^{(n_t)}$	complex-valued symbol a modulated to a resource element (k,l)
	transmitted from antenna port (n_t)
a_s	QoS term for a service flow (s)
$B_{\mathbf{m}}(n)$	packet buffer of user $\mathbf{m}$ at scheduling interval (n)
c_{ch}	channelization sequence of a spread-spectrum signal
D	delay of the equalized signal
$d_{\mathbf{m}}$	delay of a packet in the transmit buffer for user $\mathbf{m}$
d_{req}	delay requirement (a QoS criterion)
	(the maximum delay of a packet not to be exceeded without discard operation)
E_s	symbol energy
f	frequency
f_{3dB}	3 dB (of the maximum signal power) cutoff frequency of the signal band
g	equalizer filter
g_{MMSE}	equalizer filter of a MMSE equalizer
$G_{p,j}$	processing gain
$g^{(n_r)}$	the equalizer filter for a signal received at antenna port (n_r)
g_{ZF}	equalizer filter of a zero forcing equalizer
$\mathbf{h}$	channel vector
$\mathbf{h}^{(n_r)}$	the channel experienced at receive antenna port (n_r)
$\widehat{\mathbf{h}}$	estimated channel
i	index
i	a bit in a channel coded block mapped to the I-branch of a modulator
i	interference
I_0	received interference power
I_{ACI}	adjacent channel interference power
I_D	interference power of the dominant interferer at the receiver
I_{DIR}	the ratio of the dominant interference power at the receiver
	to the sum of interference power of all other interferers
$\mathbf{I}$	identity matrix
k	Pareto index of Pareto distribution
L	length of the basis function of a lapped transform
L	path loss of the signal propagation path
L_f	the length of the equalizer

m	a modulated symbol
M	number of correlators (fingers) of a Rake receiver for a single channelization code
M_{sc}^{PUSCH}	number of frequency bins (or subcarriers sc) in the system band allocated to a user on PUSCH
M_{symb}	a symbol block for the allocated bandwidth of a user for the duration of a subframe
N_0	received thermal noise power
N_{BW}^{DL}	number of subcarriers in the effective system bandwidth in downlink
N_{BW}^{UL}	number of frequency bins (subcarriers) in the effective bandwidth in uplink
N_{CP}	number of samples in the Cyclic Prefix
n_k	a bit of index k in a channel coded symbol block
N_{PRB}^{DL}	number of PRBs in the effective system bandwidth in downlink
N_{PRB}^{UL}	number of PRBs in the effective system bandwidth in uplink
N_R	number of antenna ports in the receiver
N_{SC}^{PRB}	number of subcarriers (or frequency bins) in a PRB
N_T	number of antenna ports in the transmitter
$P_{\mathbf{m}}$	the scheduling function a.k.a the scheduler utility function a.k.a the cost function
P_0	reference power
$P_{0\_nominal}$	the nominal reference power (for a channel type)
P_{0_PUSCH}	PUSCH channel reference power
P_c	coverage reliability area
p_{max}	the maximum transmit power
P_r	coverage probability as a cdf exceeding an outage threshold
p_{rrc}	root-raised-cosine pulse shaping function
p_{rx}	received power
p_{sc}	transmit power of subcarrier (sc)
p_{SP}	probability function of a silence period
p_{tx}	transmit power
p_{VP}	probability function of a voice active period (talk-spurt)
q	a bit in a channel coded block mapped to the Q-branch of a modulator
$\mathbf{r}$	received signal vector of complex symbols
$r[n]$	received symbol block of [n] complex symbols
$\hat{r}_{\mathbf{m},b}(n)$	estimated instantaneous throughput for user $\mathbf{m}$ receiver in the frequency resources (b) at the scheduling interval (n)

$\widehat{\mathbf{r}}$	estimated receive signal vector
$\widetilde{\mathbf{r}}$	soft estimate of the receive signal vector $\mathbf{r}$ at a given iteration in the iterative receiver
$\widehat{\widehat{\mathbf{r}}}$	received signal vector combined from multiple receive antenna ports
$\widetilde{\widehat{\mathbf{r}}}$	soft estimate of the received signal vector $\mathbf{r}$ combined from multiple receive antenna ports at a given iteration in the iterative receiver
$\widetilde{\widehat{\mathbf{r}}}^{(n_t)}$	soft estimate of the received signal $\mathbf{r}$, transmitted from antenna port (n_t), and combined from multiple receive antenna ports at a given iteration in the iterative receiver
R_{codec}	voice codec frame rate
R_j(eff)	effective code rate
R_{GBR}	the guaranteed bit rate (appears only as the long term mean value)
$\boldsymbol{\mathcal{R}}$	channel covariance matrix
$\boldsymbol{\mathcal{R}}_{rr}$	received signal covariance matrix
$\boldsymbol{\mathcal{R}}_{vv}$	interference plus noise covariance matrix
s	a service flow of packets
SF	spreading factor (a.k.a the processing gain) of a spread-spectrum signal
T_c	chip rate
t	distribution factor (or cluster distribution factor) in subcarrier mapping
T_{codec}	voice codec frame duration
T_s	a symbol period
$w_{\mathbf{m}}(s)$	QoS weight factor of a service flow s for user $\mathbf{m}$
$\underline{W}$	estimation filter coefficients
$\mathbf{W}$	equivalent channel matrix
(x, y)	a small area element i.e. a differential area element (dx, dy)
$y(t)$	Rake receiver output
z_k	complex output signal of the k^{th} correlator in a Rake receiver
$\mathcal{A}$	network area (simulated network area)
γ	a coefficient of Cauchy distribution
$\mathbb{C}$	channel coding gain
$\mathbf{c}$	channel capacity
$\mathcal{C}$	ergodic capacity
$\mathcal{C}^{\mathcal{A}}$	wide area capacity

$\mathbf{E}$	entropy rate of a symbol source
$\mathbf{F}, \widetilde{\mathbf{F}}$	precoding vectors
$\mathbf{G}$	geometry factor (scalar)
$\mathbf{H}$	channel matrix of a MIMO system
$\mathbf{H}_w$	random channel matrix
$\boldsymbol{\mathcal{H}}$	MIMO channel matrix with correlation
$\boldsymbol{\mathcal{I}}$	frequency selective interference vector
$\mathcal{L}$	load factor
$\mathbf{m}$	a user index (scalar) in the scheduler funtion
$\mathbf{m}'$	the index of the scheduled user
s	signal in time domain
S	signal in frequency domain
S	analog signal or signal power
$\mathbf{s}$	transmit signal vector of complex symbols
s[n]	transmit symbol sequence of [n] complex symbols
$\widehat{\mathbf{s}}$	estimated transmitted signal (from the received noisy and interfered signal)
$\mathbf{S}_{QoS}$	QoS weight set of service flows
z	a precoded, DFT transformed, modulated symbol block
z'	a DFT transformed output $\mathbf{z}$ mapped to the IFFT input
$\boldsymbol{\mathcal{W}}$	the combining vector for a multiantenna receiver
$\mathbf{W}$	chip rate (scalar)
α	orthogonality factor
α	the combining loss factor
$\underline{\alpha}$	a scheduler coefficient for aggressiveness of operation
α_{PL}	fractional pathloss compensation factor
β	power weight factor of an uplink code channel
$\underline{\beta}$	a scheduler coefficient for aggressiveness of operation
β_{PUSCH}	amplitude scaling factor (i.e. power weight) of a user signal on PUSCH
$\boldsymbol{\delta}_D$	zero vector with a single element of value 1 at position D unit delay
Δf_{sc}	subcarrier spacing in OFDM
$\delta_{\mathbf{m}}$	aggressiveness factor of a delay criterion (for user $\mathbf{m}$)
ƒ	fast fading process
γ_0	protection ratio as a defined probability tail of a pdf

ξ	user specific frequency shift in the uplink symbol mapping to the allocated bandwidth
κ	the code word of a multistream MIMO transmission
κ_0, κ_1	codeword of index 0, codeword of index 1 of a multistream MIMO
$\mathbf{\Lambda}$	diagonal Fourier matrix
λ	rate parameter of a process or a distribution (intensity of session activation or voice activity factor or traffic source activity)
μ	the mean of a statistical distribution
η	thermal noise power
η_{ph}	phase noise
Ω	shadow fading process
ω	phase of the transmitted signal
ω_D	Doppler shift
ϕ	phase shift of a complex signal
$\mathbf{\Phi}$	covariance matrix of MIMO signal received from multiple transmit antennas
ρ	the mean SNR
ρ_{cmplx}	complex correlation of two complex variables
ρ_{env}	envelope correlation of two complex variables
ρ_{pow}	power correlation of two complex variables
σ	Gaussian distributed random variable
σ	the standard deviation of a statistical distribution
σ_v^2	variance of the (thermal) noise power
$\mathbf{\Sigma}_R$	receive antenna correlation matrix
$\mathbf{\Sigma}_T$	transmit antenna correlation matrix
τ_{AE}	probability distribution of the time between successive talk-spurts
τ_{SP}	silence period
τ_{VP}	voice activity period
θ_l	angular shift of the complex channel on propagation path (l)
ϑ	cluster repetition index (for clustered symbol mapping)
ϑ_{CL}	number of clusters (for clustered symbol mapping)
χ	Chi-squared distributed random variable
ζ_{dl}	cell-specific scrambling code
ζ_{UE}	UE-specific scrambling code

Chapter 1

Introduction and scope

1.1 Motivation of the thesis

Technical development of mobile communications for the Mobile Information Society is progressing rapidly to the third generation (3G) technologies and beyond, along the technical requirements and targets set for International Mobile Telecommunications (IMT). Untill today, majority of global mobile communications happened by the second generation (2G) technologies. Recently, 3G technology and subscriptions are rapidly increasing, and the commercial success of any high end product evidently requires 3G implementation. 2G refers to the technologies for the Global System for Mobile Communications (GSM) and 3G to the technologies for the Universal Mobile Telecommunications System (UMTS). Currently, 3G technologies are standardized by the Third Generation Partnership Project (3GPP [1] and also 3GPP2 [2]), and they create the only high performing, global, mobile wide area solution that is adopted largely among the operators and vendors. An example of the vitality of 3G is Apple iPhone, which was launched with large publicity (originally including 2G radio and wireless access). Soon after the launch, Apple had to provide a revised product, Apple iPhone 3G, where 3G technology is a real enabler of success. This happened, even when Apple is well known in usability, computing and operating systems. All this just was not sufficient without competitive radio and mobility technolgies in the product.

In the near future, 3G is challenged by outperforming new technologies like Evolved UMTS Terrestrial Radio Access (EUTRA)[1] also known as the Long Term evolution (LTE) and Worldwide Interoperability for Microwave Access (WiMax)[3], which have a benefit compared to the legacy technology, beccause of their scalability to different band-

widths. The longer term development of IMT-Advanced seems to appear based on these two technology tracks complemented by the Wireless Local Area Networks (WLAN a.k.a Wireless Fidelity, WiFi [4]). In 3GPP2, there existed a candidate called Ultra Mobile Broadband (UMB) technology, which was soon withdrawn and integrated to the other two tracks.

Starting from multiple-access studies in late 90's, technology convergence in the early 21st century has taken place and it is now possible to position spread spectrum and multicarrier technologies in their most competitive way to the products and markets. The future of radical innovations is expected to happen among services and software in devices, whereas radio technology will continue on evolutionary tracks. This however does not reduce the value of radio technologies as essential for the seamless (broadband) services.

The era of this thesis covers the periods from the prestudies of 3G to the first standard specifications of the Wideband Code Division Multiple Access (WCDMA), WCDMA/High-Speed Packet Access (HSPA) and further to the first release of LTE, for which product launches will begin within the next couple of years. The scientific contribution of the thesis is included in publications and in Intellectual Property Rights[1] created for Nokia by the author together with the co-authors and co-inventors from Nokia, Nokia Siemens Networks, Helsinki University of Technology, Tampere University of Technology and University of Jyväskylä. The monograph summarizes fundamentals of these technologies and sets the publications and inventions by the author to a wider context. Due to the wide scope of the thesis, it is not possible to cover all the details in depth, but literature references are given for further information. The scientific experience gained in studying, specifying, modelling, practising and analyzing components for these technologies is evident from the referenced publications. Finally, some of these solutions may appear in commercial products.

The approach for this thesis is to take a longer time frame into focus and point the areas, where the author has scientific contribution in publications and in creating new solutions in the form of patents and patent applications. This thesis points out the theories behind the third generation WCDMA, WCDMA/HSPA and LTE that changed the UMTS system specification along these years. An alternative approach could have been, say, to take a feature or a building block as a candidate to the standard and compare all its alternatives and parametrisations and their impact to the system performance. This approach would,

[1]The importance and share of 3G cellular patents is studied in [5].

however, have lead to a documentation of many discarded solutions and those already presented in the literature. The difference of this thesis to a literature study is clearly in the own view of the author, long term in depth scientific project work in the field, published performance results on selected topics and inventions in the related art.

The thesis is particularly timely, because the standard specifications of WCDMA and WCDMA/HSPA are mature and deployed for use in commercial products, the standard specification of LTE is in its completion phase to be commercialized in a couple of forthcoming years [6][7][8][9][10] and simultaneously a call for proposals of inclusive advanced components is launched in [11].

1.2 Scope and structure of the thesis

The thesis about the convergence of packet communications over the evolved mobile networks covers signal processing and protocol design and performance analysis of the WCDMA, WCDMA/HSPA and LTE technologies including the contribution of the author. The thesis is a monograph, hence it addresses the concept research and system analysis results, where the several publications and inventions of the author are included on selected research results to supplement the scientific and novel contribution.

The structure of the thesis is as follows. After the introduction in Chapter 1, Chapter 2 covers a briefing on the mobile evolution and standards. Chapter 3 describes the multiple-access technology, the transmission schemes and their importance to the system performance. Chapter 4 visits the receiver algorithms. Chapter 5 starts from packet communications and those networking protocols, which have an impact on the radio performance. Chapter 6 describes the mobile networking architecture and its evolution, where the architecture impacts the split of protocol functions and where the radio protocols concretely determine the access and system performance. Chapter 7 visits some of the traffic models in a rich media protocol environment. In Chapter 8, Radio Resource Management algorithms, protocols and their operational principles are depicted. Finally, Chapter 9 analyses selected results on the system performance, where the issues addressed in Chapters 3 to 8 have a dominant role. Chapter 10 summarizes briefly the thesis. The structure of the thesis is illustrated in Fig. 1.1.

The session organization covers the WCDMA, WCDMA/HSPA and LTE access systems throughout Chapter 3 to Chapter 8. The solutions of these Chapters appear in

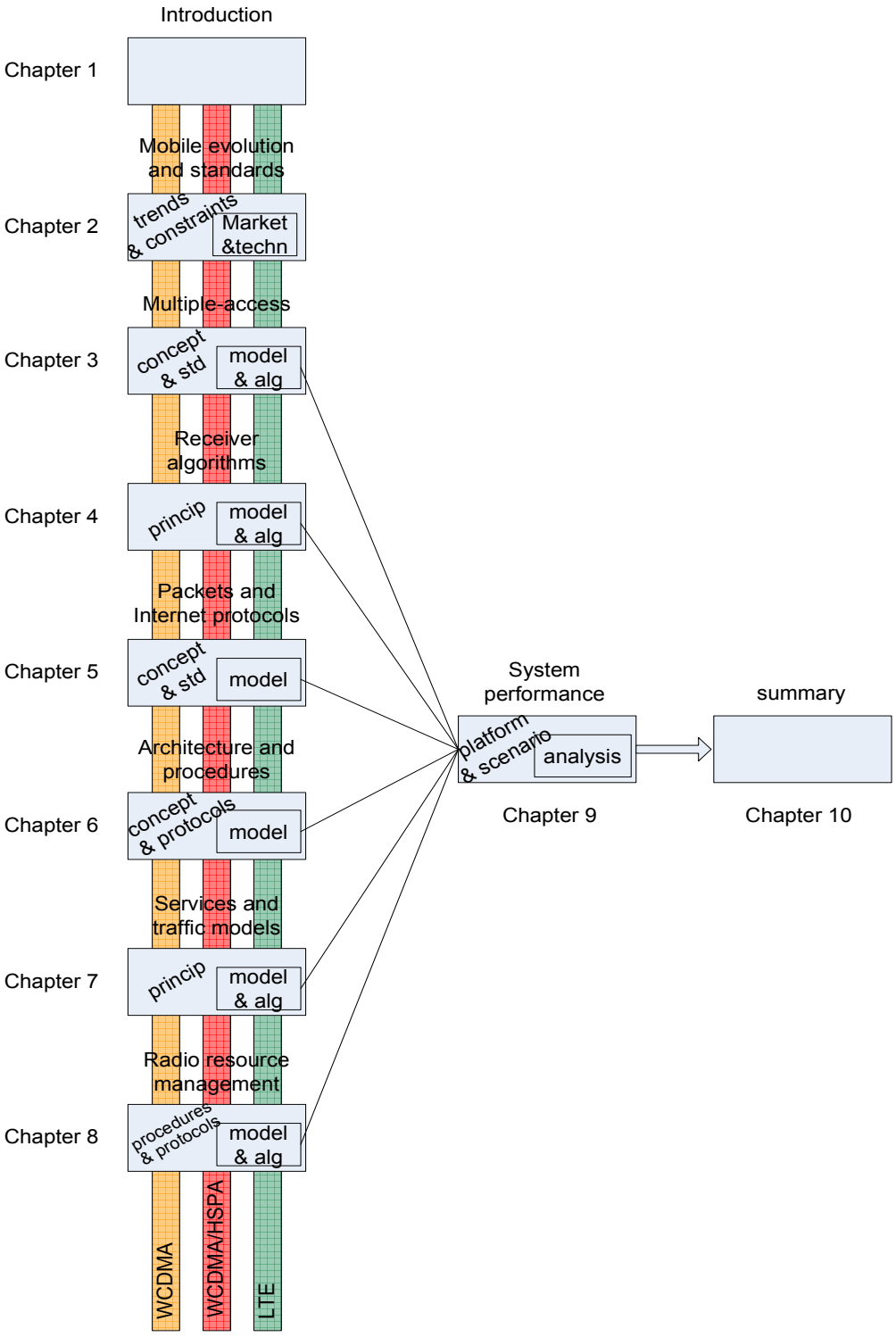


Figure 1.1: Structure of the thesis.

details in the 3GPP standard specifications and lead to mathematical models that have been implemented to the system simulator platforms. These simulators and their results are known by the 3GPP community via the contributions of Nokia. The quality and reli-

ability of the results were proven e.g. as approved 3GPP system evaluation contributions in the technical specification group Radio Access Networks. The models of each Chapter, especially multiple-access schemes from Chapter 3, receiver algorithms from Chapter 4, Internet protocol fields of Chapter 5, mobile network architecture, procedures, protocols and channels in Chapter 6, traffic models in Chapter 7, radio resource management in Chapter 8 are selectively imported to the system simulator platforms in order to create the simulation results in Chapter 9. For each simulation result, the given working assumption, model and parametrisation is stated and referenced in Chapter 9. The use of a given model in a simulation depends on the analysis and scenario, which were defined according to the agreed references by the 3GPP, IEEE or the Next Generation Mobile Networks (NGMN) alliance.

1.3 Contributions of the author

The contributions of the author in the technology areas of early 3G, WCDMA, WCDMA/HSPA and LTE cover the long term research work from 1996 till 2009 as a researcher, project manager, team-leader, principal scientist and chief architect in Nokia Research Center, in the field of systems research and standardization. Along these years, Nokia and since 2007 also Nokia-Siemens-Networks created several technology proposals that drive the 3GPP standards, together with the partnership companies. The work by the author has created many in-depth solutions and scientific experiments, which have been modelled into the statistical simulator platforms, where the impact of features to the system performance of a technology was analysed, problems identified and improvements created. The mathematical models of the proposals in a simulator let us provide numerable measures, so that appropriate knowledge of the system behaviour and performance becomes available. Many of the proposals studied in the scope of this thesis have actually not entered into the standards, but have instead resulted reasoned exclusion of non-favourable candidates. Some of the proposed solutions have successfully been adopted to the standardization process, where after a synthesis they have transformed to a standardizable solution accepted by the community. In this way, the scientific work has contributed to the overall success of the standard specifications.

During the long term evolution process, the author has created over 40 patent application families in the scope of the thesis, some 21 of which have been selected for a closer

analysis and included in the thesis. The publications have been created together with other researchers, because the modeling, software development and simulations require an effort that typically requires team work. In the simulations, the partition into platforms and the platform approach itself benefit from additive features, so that the analysis of new candidates is feasible in collaboration with tested baseline solutions. The role of the author in these research articles is significant, for setting the framework, assumptions, models, algorithms and guidance for the simulation work to produce data, and particularly for analyzing and reporting the results. Having system concept point of view, the focus of the author was in the radio resource management, signal processing and protocols. The presented arguments and results of the thesis are new, they either introduce a new feature or a new aspect of the WCDMA, WCDMA/HSPA and LTE, or they fully analyze the system technology and its performance.

In the spirit of applied sciences, the analyses published in the articles have created knowledge attained through studies and practices, which converge to measurable quantities of a technology. The methodology is mainly based on statistical models of processes that act in reality, the models of which let us gain systematic knowledge from complex systems that were otherwise hard to analyse in fair and even conditions. The merit of scientific work in the thesis concretizes to the improved system performance along the evolution of WCDMA, WCDMA/HSPA and LTE technologies.

1.4 Summary of publications

The undersigned has authored or co-authored 23 articles for scientific conferences and four (4) journal articles. These research articles have been subject to peer-review prior to publication, and they appear in the conference Proceedings, in a transaction and journals, published by IEEE, Eurasip and comparable organizations. The author is the inventor or a co-inventor of 7 patents and over 14 pending patent application families in the scope of this thesis. The published material covers mainly the Radio Resource Control (RRC), Radio Link Control (RLC), Medium Access Control (MAC) and Radio Resource Management (RRM) solutions and their system analyses based on simulated results.

Publications[12][13][14][15] analyse WCDMA, [16][17][18][19] WCDMA/HSPA and [20][21][22][23][24] LTE. System simulator resolution was verified in [25]. RRM algorithms were in the focus of [12][13] [14][15][16][17][18][19] [22][26][27][28][29]. Protocol design and

analyses were in the focus of [20][26][27][30]. System analysis of LTE are represented in [21][22][24]. Co-authoring for other than 3G systems is present in [29][31][32]. Co-authoring specific signal processing techniques are present in [28][33][34] [35][36][37][38].

Patents and applications [39][40][41][42] are pending for WCDMA, [43][44] for WCDMA/HSPA and [41] [45][46][47][48][49] [50][51][52][53][54] [55][56] for LTE. Patents and applications [43] [48][49][52][53][55] [56][57] are pending for protocols and [40][41] [45][57][58] for the system architecture.

Chapter 2

Briefing on mobile evolution and standards

2.1 Mobile evolution and its main characteristics

Mobile network operators have set their priorities for the key properties of the mobile communication systems in [59], as summarized in the spidernet of Fig. 2.1. Seamless mobility, reliability, cost and simplicity are priorised as items of "no compromise" as well as the spectral efficiency. The next most priorized items as "strong requirements"are throughput, latency and quality of service. The spidernet depicts that achieving improvements in one domain shall not imply compromises in the other domains, instead an overall balance need to be maintained.

Throughout this thesis spectral efficiency, throughput and latency are studied, because they are the most valued measurable properties of a mobile communication system. These are possible to be analysed and compared by statistical simulations in comparable reference cases [60][61][62][63][64]. Analysis in defined reference scenario is considered fair and objective. Quality of Service (QoS) is addressed in the context of architectural choices, protocol designs and algorithms. The other mentioned properties are more subjective and may not be directly measurable, hence they are not in the core of this thesis.

2.2 Trends in mobile communications

A vision of the mobile network evolution according to the contents of the thesis is presented in Fig. 2.2. The Figure shows coexistence of various technologies that motivates to reduce

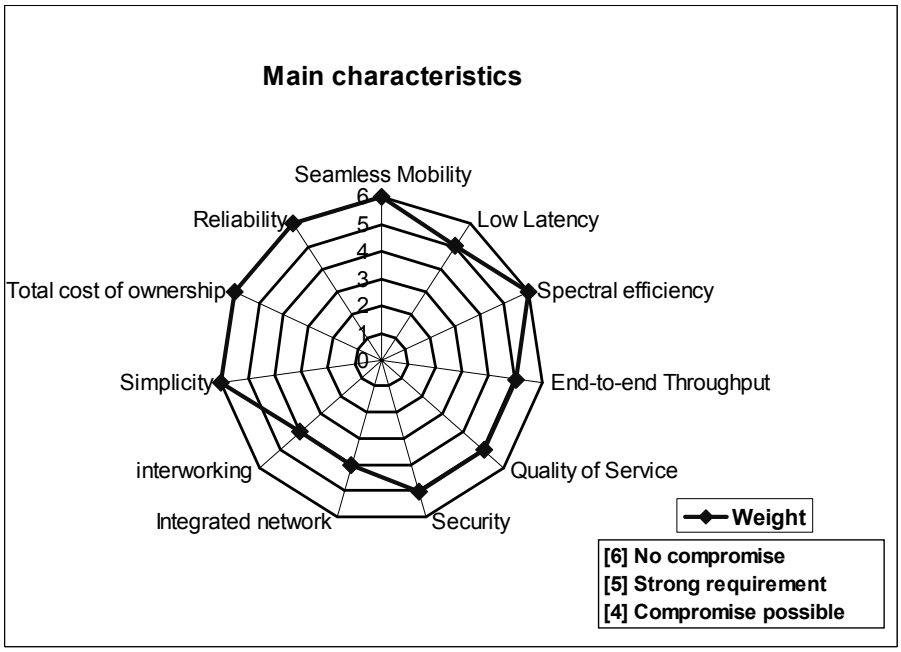


Figure 2.1: Relative priorities of the main characteristics of the mobile communications system. *Source: Next Generation Mobile Networks.*

their divergence. Preferred evolution is convergence towards one integrated network [59]. The evolution of user throughput (shown in log domain on the right) in a wide area network is the booster of mobile services. The Figure emphasizes 3GPP technologies.

The evolution of mobile communication systems is the great outcome of the co-operation of partners in 3GPP, 3GPP2, IEEE, Internet Engineering Task Force (IETF), which include international standardization organizations (e.g. Association of Radio Industries and Businesses in Japan, China Communications Standards Association, European Telecommunications Standards Institute, Telecommunications Technology Association in Korea, Telecommunications Technology Committee in Japan and Alliance for Telecommunications Industry Solutions in USA). Finally, the competition among technologies and vendors in the market place has realized the mobile communications of today. The most dominant paths, globally, have been developed either by the 3GPP or IEEE. The landscape of mobile communication systems and standards is shown in Fig. 2.3, where the dominant technology shifts are identified as well.

The roots of the Global System for Mobile Communications (GSM) are in Time Division Multiple Access (TDMA) and in circuit-switched voice connection core network that have later elaborated to include high-speed circuit-switched data (HSCSD) connections.

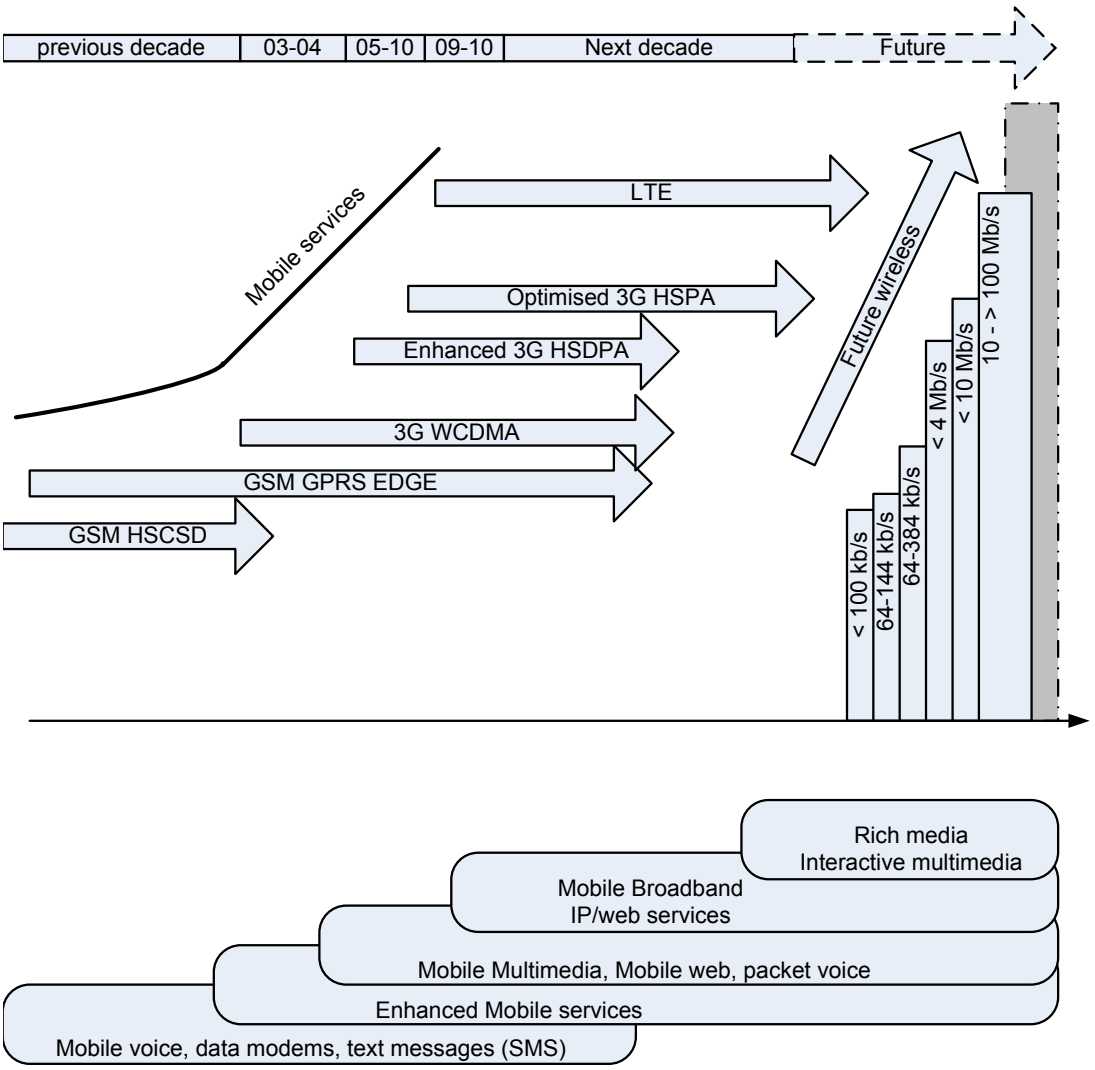


Figure 2.2: A vision of the mobile network evolution and services. The Figure shows coexistence of various technologies that motivates to reduce their divergence. The preferred evolution is convergence towards one integrated network. The evolution of user throughput (shown in log domain on the right) in a wide area network is the booster of mobile services.

Later, with the introduction of packet core networks, the radio access was completed to the General Packet Radio Service (GPRS) architecture. The GSM air interface soon turned out to be too limited for practical network data and its definitions were upgraded to Enhanced Data rates for Global Evolution (EDGE) and Enhanced GPRS respectively [65]. The difficulty of multiplexing variable rate data services and voice, triggered the 3G evolution, where spread spectum communications with variable spreading was the prospected solution (see Chapter 3). This clearly meets the targets of UMTS service requirements up to say 384 kb/s in practice and up to 2 Mb/s in theory, but has limits in a multiuser

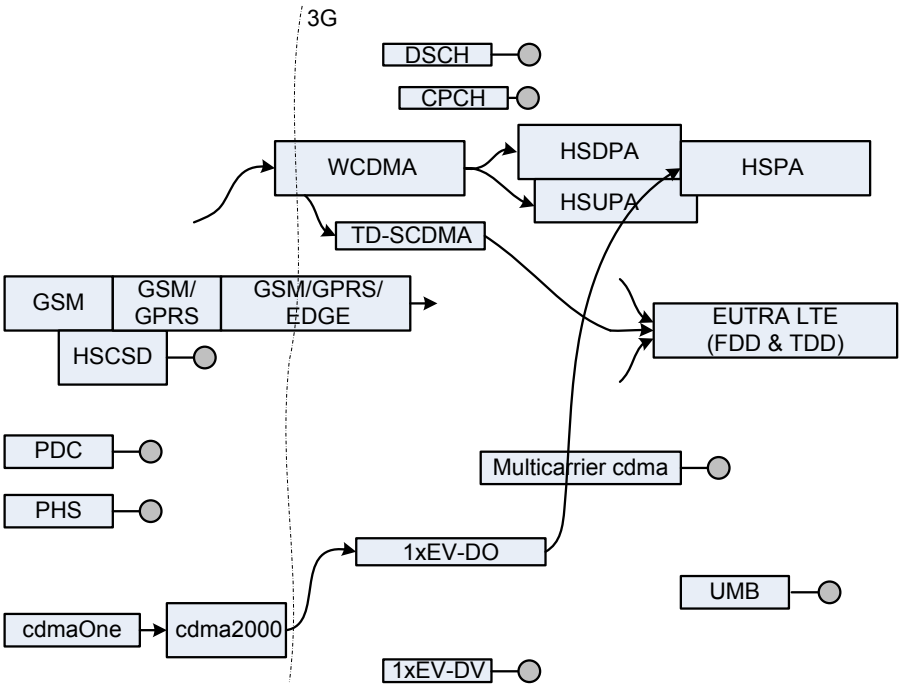


Figure 2.3: The landscape of mobile communication standards and systems. The dominant technology shifts are marked by arrows, and the terminated or completed system work is marked by circles.

network for higher, bursty data rates. Also, power consumption of the air interface due to the direct sequence spreading is a notable concern. The HSPA technique introduced power efficient, short term, time multiplexed allocations for bursty data, first in the downlink (HSDPA) and soon after in the uplink (HSUPA). These techniques are in the commercial deployment today.

The next upgrade requires a technology change, because going for higher bandwidths by spread spectrum techniques would lead to very short symbols, which suffer from inter-symbol interference and would be complex to equalize. The high bursty data rates would further imply low spreading factor of the code channels, which mitigates spreading gains and leads to multiuser detection problems. The scalability of bandwidth to versatile frequency bands available in different regions is yet another challenge. All these challenges are faced with less complexity compared to spread spectrum processing by transform domain techniques, which apply parallel processing of symbols, have easy bandwidth scalablility and keep the symbols long and resistant to the intersymbol interference.

Another branch of circuit-switched communications originated from code division multiple access with short Walsh-codes, known as Interim Standard (IS-95) or cdmaOne [66]. Here, the spreading band is narrow, tailored to the Personal Communications Service (PCS) bands in the US and Korea, and hence its benefits of spread spectrum communications remain limited [67]. An attempt to enhace the air interface by multicarrier transmissions was included in the cdma2000 family, but it does not appear technically competitive due to the expenses of RF filtering.

Another compelling approach has started from local area packet communications, which arised from the need to extend fixed communications like Ethernet protocol towards wireless communications. The technical basis of WLAN a.k.a IEEE 802.11 standard or WiFi system lies in the carrier sensing protocols, which perform well in small scale cells with bursty traffic arrivals, but whose offered load decreases, when the number of users or traffic activity increase. This technology was proven viable especially in small cell areas (or hotspots), as originally intended [68]. However, its extensions to wide are networks has not become a great success due to the problems of mobility, coverage holes and cell edge interference (observed e.g. in a large scale experiment in San Francisco). Especially, seamless mobility (see Fig. 2.1) that was priorised high by the operators has not been realized in practise.

Other technologies, like WiMax (IEEE 802.16e standard [3][69][70]), were targeted to overcome shortages of WLAN in the wide area networks. WiMax however resulted in a complex specification and solutions, which suit more for fixed wireless links than mobile communications.

Along the evolution of standards and systems, there has been an expansion of the coverage and mobility of practical deployments as shown in Fig. 2.4. Further, the classical measures of coverage and voice capacity have converted to the new measures of coverage-throughput with user perception of low latency and service quality. These are currently the true drivers of mobile communication technologies to provide seamless rich services for mobile users. The mobility vs data rate as a function of propagation environment is shown in Fig. 2.5.

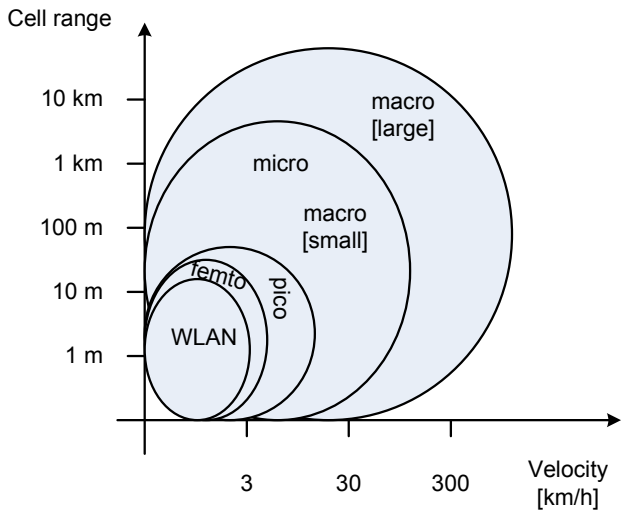


Figure 2.4: Deployment of cell sizes ranging from large macro cells for high mobility, large coverage but low capacity to the micro cells for lower mobility, larger coverage probability and high capacity. Extreme traffic densities can be covered by even smaller cells or WLAN hotspots.

2.3 Political goal setting and technical requirements

Requirements for the Radio Transmission technology (RTT) and Radio Interface Technology (RIT) have originated from high level political goal settings for the Mobile Information Society, which were mapped to concrete technical requirements. These requirements are the prerequisite for frequency allocations and regulations to take place.

UMTS is the third generation mobile communication system, which targets to extend the services of fixed network to mobile users. Derived from this high level goal, a set of technical requirements were given in [71] taking into consideration different operational environments, mobility and architectures. Today, UMTS covers the International Telecommunication Union - Radio (ITU-R) conceptions for the IMT-2000 and the Future Public Land Mobile Telecommunications System. The trend for convergence has structured these technologies under the common IMT-family.

The third generation WCDMA technology was defined and standardized by 3GPP [1] to satisfy the requirements set in [71]. However, due to the very high targets, and for the competitiveness among IMT-2000 technologies [72], WCDMA system had to be complemented by HSPA technology [6][73]. HSPA was integrated to the WCDMA specification as a new set of transport channels, but actually it had a much wider impact over the

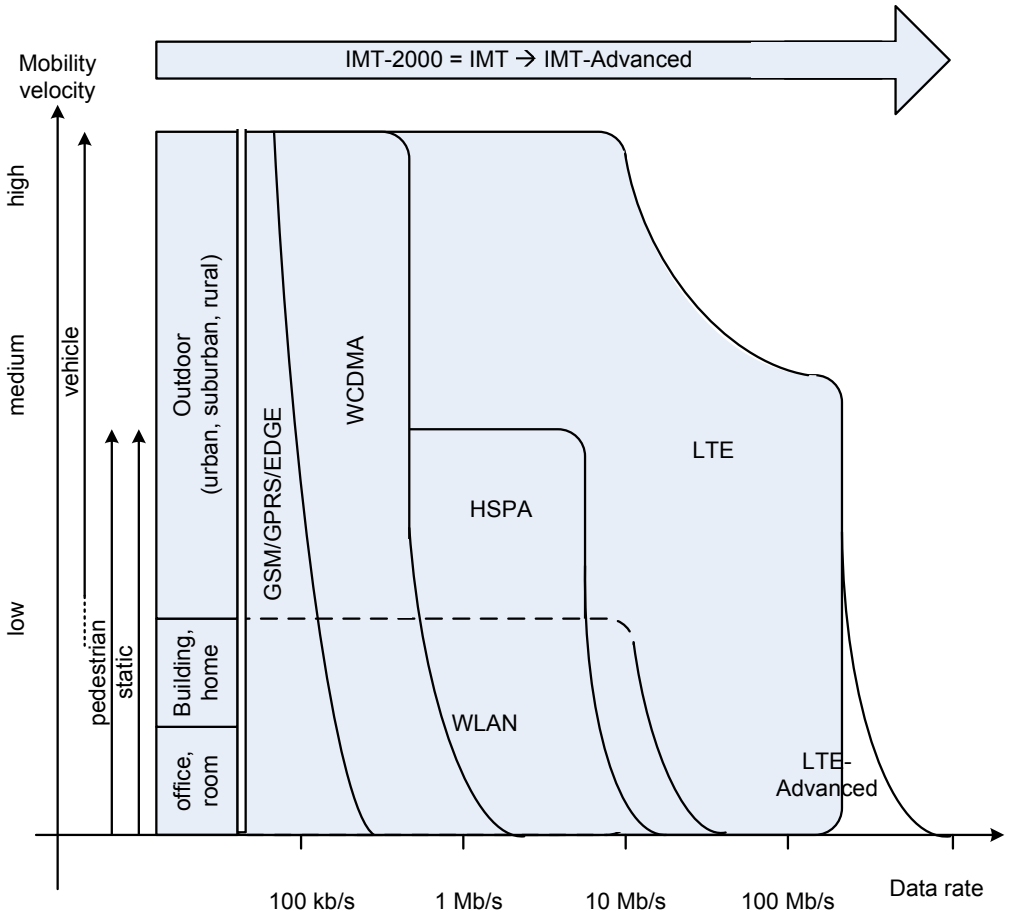


Figure 2.5: Mobility vs data rate as a function of the propagation environment.

protocol architecture, physical channels and RRM algorithms.

After the WCDMA/HSPA standard, further objectives were set by 3GPP to ensure the Long Term Evolution (LTE) process. EUTRA technology, which hence also became known as the LTE technology, targets at competitiveness in a longer time frame towards a high data rate (broadband), low latency and packet optimized system, which improves 3G system capacity and coverage by the average spectral efficiency factor of three to four and the cell edge spectral efficiency by the factor of two to three, and further reduces the cost of operation [6][8][7]. Considering the desire for higher performance [74][75], and also taking into account the future additional spectrum allocations, the evolution of UTRA and UTRAN included a request for transmission on bandwidths wider than 5 MHz at the same time, when also bandwidths below 5 MHz were requested for more flexible deployment in those frequency bands that will be available e.g. from refarming spectrum allocations of

outdated systems. The set of technical requirements for EUTRA and EUTRA Network (EUTRAN) were set in [76].

Recently, new spectrum allocations are identified again and all previously allocated IMT-2000 spectrum was generalized to IMT-bands. Further system proposals are called for the IMT-bands in a Circular Letter [11]. EUTRA and EUTRAN technology targets at meeting the requirements given in [77] and [78] creating the heart of the IMT-Advanced system [79].

2.4 A market view

This section represents a market view of the third generation technology. The view is formed from sources in the Internet, newspapers and market research reports by Informa Telecoms and Media [80][81]. Informa Telecoms and Media provided statistics from 2003 to 2008 and forecasts from 2009 to 2012. In public, years 2009 and 2010 are assumed uncommon within a trend, because of the uncertainty of economics, and therefore the accuracy of technology market predictions may decrease.

Statistics is scattered by the mobile device type, by the service classification (voice, mobile data, fixed portable or mobile broadband) or by the application (web browsing, short messaging, entertainment, enterprise). The information is provided by geographical regions, where the markets may behave diversely and where the deployment of technologies may be in different phase. Worldwide statistics is not always available, or it may not be as comprehensive as the regional or countrywide statistics.

Categorization of devices by technology is not straightforward, because devices often include several technologies. Hence, if the total number of devices equals the estimated total market size (1,220 million units in 2008), calculating the technology shares may be numerically folded inside the range. Hence, 2G/GSM/GPRS and WCDMA and WCDMA/HSPA and LTE technologies in a device are not exclusionary additive but complementary additive instead.

Regarding use of a technology in a given service category, availability of statistics is not that obvious. Network deployment of a new air interface technology proceeds gradually. The actual use of a technology requires that it is available in the cell area for a user, who has a device operating on that technology, the user has paid for a subscription that allows use of that technology and yet the service selection triggers device to operate on

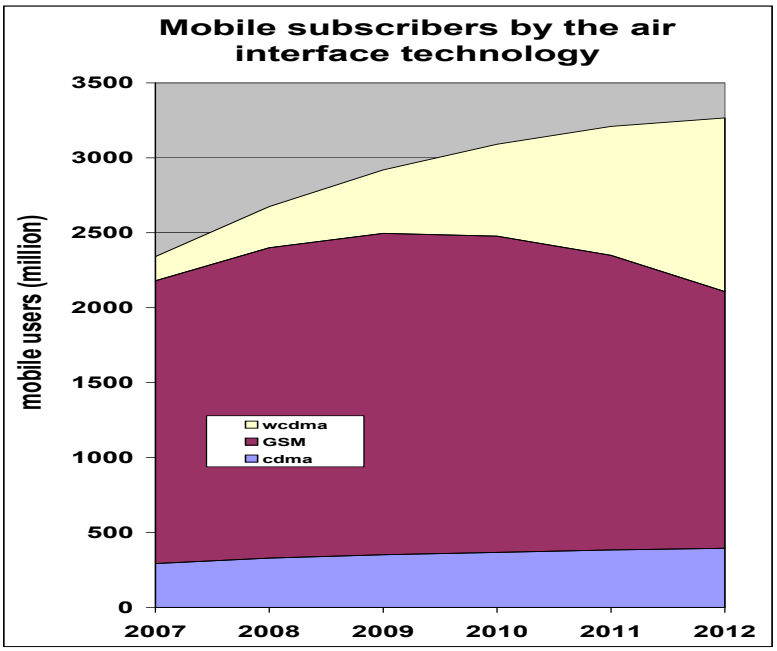


Figure 2.6: Mobile subscribers (as the number of mobile users) worldwide by the air interface technology. *Source: Informa Telecoms & Media.*

that technology.

Market indications are outlined by Informa Telecoms and Media e.g. in [80][81]. The deployment of WCDMA and WCDMA/HSPA technologies is expected to happen fast and they gradually will replace the use of 2G/GSM, as shown in Fig. 2.6. GSM extensions will remain implemented in the devices and they will continue operation in network areas, where 3G coverage is not yet deployed. Anyhow, 2G technology is expected to face negative growth in the number of users, even if its device base would continue increase. WCDMA is currently reported as the second-most deployed air interface, with 195 live networks worldwide as of March 2008, significantly behind GSM but just ahead of CDMA. Over the year 2008, there were 128 HSPA launches worldwide and 45 cdma2000 (1xEV-DO) launches respectively.

Mobile data and broadband data volmes are expected to increase nearly exponentially mainly due to the growth of information services, entertainment and enterprise applications that are available in the World Wide Web of the Internet. In the near future, this demand will almost entirely be served by 3G technologies and dominantly so by WCDMA and WCDMA/HSPA. LTE is expected to capture a significant share of data growth on longer run.

According to Informa Telecoms and Media, the worldwide network traffic is expected

to increase by over 8% for voice and nearly 65% for data in period from 2007 to 2012. This means that the 70% share for voice traffic over 30% share for data traffic in 2008 would turn to negative relative growth down to 30% for voice and rapid relative growth up to 70% for data. From 2007 to 2012, there would be over tenfold (1000%) increase of volume in the mobile data traffic.

According to Informa Telecoms and Media, the use of communication services (voice) is expected to be fairly flat in regions (e.g. Europe), where penetration is already very high and is expected to increase rather linearly in regions (e.g. Asia Pacific), where subscription rates will still increase steadily. This structural growth will cause a dramatic impact to the relative (percentage) numbers, which are expected to show negative growth for 2G and exponential-like growth for 3G. More realistically, the growth of 2G implementations in devices could still be steady, but 3G would dominantly capture the use cases from 2G. Japan is in this respect exceptional, because there 2G was ramped down rapidly and was entirely replaced by 3G. In Japan, this was due to very scattered shares of 2G technologies, due to spectral issues and due to immediate needs to satisfy voice capacity demand by WCDMA in dense urban areas. In China, the State Council agreed on start of issuing 3G technology licenses in the beginning of 2009 i.e. years later than in the western developed countries, but the progress is expected to be fast due to high demand of coverage and capacity for mobile services. The expected growth of WCDMA users by region is shown in Fig. 2.7.

Based on indications by Informa Telecoms and Media, the revenues of network operators would shift more from voice to data services but could not increase even nearly by the same ratio as the traffic amount (measured in petabytes [PB] per year). Informa Telecoms and Media report about 160 PB global data traffic in 2007 (238 PB in 2008), the compound annual growth rate from 2007 to 2012 is expected to be of order 65% with increasing trend. The compound annual growth rate of revenues is expected to be above 10%, however without a clear increasing trend.

The transition of use cases from 2G to 3G technologies is happening fast and the crossover of subscribers (as the number of mobile users) for 3G has already happened in many developed countries. The evolution of 3G as an upgrade from WCDMA to WCDMA/HSPA happens fast in deployments, where the use of data services is already intensive, but may not extend that soon over the whole footprint of a network, where GSM/EDGE/GPRS enhancements may still offer sufficient bit rate coverage.

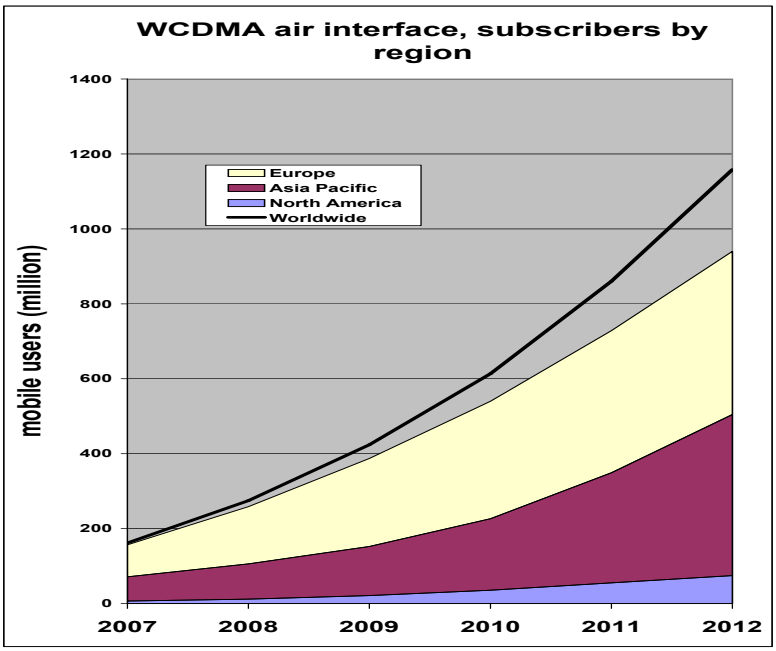


Figure 2.7: WCDMA air interface subscribers (as the number of mobile users) by region. *Source: Informa Telecoms & Media.*

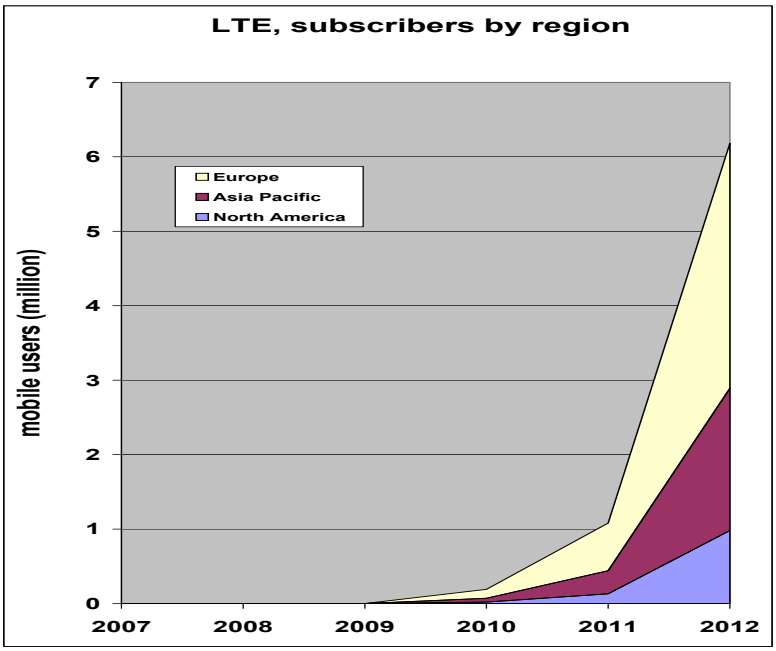


Figure 2.8: Forecast of LTE subscribers (as the number of mobile users) by region. *Source: Informa Telecoms & Media.*

LTE is in the horizon to enter commercial use after the ongoing field trials by several operators in a few years, see Fig. 2.8. LTE is gradually expected to serve a significant share of the growth of broadband data traffic [82]. WCDMA/HSPA and LTE are designed

for evolutionary system architecture and even if their air interfaces differ, their integration into UE devices and network elements is fairly smooth. After extensive discussions with the leading operators, Informa Telecoms and Media expects initial LTE network deployments in 2010 in the advanced mobile broadband markets of Western Europe, Japan, South Korea and the US.

The market views by Informa Telecoms and Media emphasize the role of 3G technology serving both voice and data communications and particularly serving the rapidly increasing traffic volumes including broadband communications. The revenues are increasingly expected from data services, and hence the percentage revenues of voice services could decrease, even if the absolute revenues would not. The expected increase of traffic volume is relatively much larger than the expected revenue, particularly due to shifting towards flat rate data charging per bit rate (limited subscription) instead of charging per transferred amount of kilobytes. Potential revenues of the service content are not included in this consideration. The growth rate of services is expected to be large in the US, also in Japan, and selectively in Western Europe. In terms of mobile communication services, China is still modest, but its annual growth rate is notably higher reaching two digit numbers.

Chapter 3

Multiple-access and transmission schemes

3.1 Multiple-access

Multiple-access is a key nominator of the cellular/wireless communication system and a fundamental choice of the system design. Its properties are largely inherited from the signal processing theory of the transmitters and receivers exchanging information at a service rate over a time variant selective channel with spatial characteristics and correlation in the presence of interference [83].

The frame structure serves for cellular multiple-access with radio resource management algorithms including power control, link adaptation, scheduling and handover. Also load management and admission control play a role in a loaded or hardware limited network. How to access, how to control and how to provide feedback for distributed communication links are set by the Medium Access Control protocols applying channel structures coping the frame design. Higher level management and connectivity are solved by the radio network dependent *access stratum* and radio network independent *non-access stratum* protocols that operate by peer-to-peer messaging and local state machines. The multiple-access technique dramatically determines the achievable system performance (see Chapter 9). However, equally important concerns are the management of complexity, detailed system design and algorithms that further differentiate the implementation constraints and simplicity of the system.

3.1.1 Classical multiple-access techniques

Frequency Division Multiple Access (FDMA) is used in all communication channels, partly due to the spectrum regulations and licensing, and partly due to practical filtering and carrier search requirements of the mobile device. Even in a licence-exempt band, FDMA is the major carrier access principle, even though the spectral placement of carriers may omit tight regulation.

Time Division Multiple Access (TDMA) burst transmission is the basis of 2G wireless communications. Every transmission burst includes a modulated sequence of symbols and a Training Sequence (as the burst midample) for channel equalization of the burst period. Burst transmission is orthogonal in the serving cell due to time alignment of bursts at the receiver. The bursts include sufficient number of gap samples at the end of each burst, so that burst ramp up and ramp down periods do not overlap in a dispersed propagation environment. This largely avoids interburst interference. The dominant interference mechanism is co-channel interference from all those neighboring cells that operate in the same frequency. Frequency reuse is a conventional solution to increase distance to the nearest co-channel interferers. For a sufficiently high reuse factor, the frequency band is divided to narrow carriers (200 kHz in GSM), which have a very limited transmission capacity. Even if multislot communications is defined and even if GPRS flexibly utilizes carrier and slot resources for packet communications, it cannot satisfy the wideband requirements in a spectrum efficient way.

Code Division Multiple Access (CDMA) spread spectrum communications has its origins in military communications to resist narrowband interference, jamming and make fraudulent carrier access complicated. Spread spectrum was found by defining frequency hopping of narrowband sequences over a wider transmission bandwidth. Apart from military reasons, frequency hopping also increases frequency diversity and was later applied to 2G/TDMA as well. Slow frequency hopping happens in periods longer than the symbol period, and frequency hopping in periods shorter than the symbol period led to spread spectrum communications. In spread spectrum, every coded information bit is spread to a wideband signal by multiplying it by a spreading sequence to multiple chips. The ratio of a bit period to the chip period is the spreading factor a.k.a the processing gain.

Ideally, the chip sequences for separate transmission links should be orthogonal to avoid interference in the serving cell. Orthogonal sequences can however only be found in known time shifts, and no general orthogonalization is feasible with arbitrary phase

shifts. In mobile communications, perferct orthogonality is available at the base station transmitter of downlink multiple-access signals. However, the orthogonality is partly lost at the mobile receiver due to multipath propagation, which causes interpath interference. The mobile transmitters in uplink are in random locations, thus their multiple access signals are inherently non-orthogonal. In both downlink and uplink, the receiver executes a tracking algorithm to receive each multipath propagated signal component independently by a sequence correlator, known as the *Rake receiver*. A tracker allows phasing of the correlation operation accurately in a time variant channel. A searcher correlator is continuously searching for signal components from noise so that the Rake correlators (Rake fingers) can be allocated to the strongest signal components at a time. The received signal to be decoded is formed by Maximum Ratio Combining (MRC) the signal components [66][84].

3.1.2 WCDMA

Wideband CDMA (WCDMA) was studied earlier in [85][86][87] and later developed to a standard by 3GPP for commercial systems to meet the IMT-2000 requirements in [71]. WCDMA contains largely different features and system design compared to the classical CDMA. WCDMA applies direct sequence spread spectrum technique in a frequency reuse one network without network synchronization requirements (unlike cdmaOne). WCDMA provides multiple-access channelization by the orthogonal long spreading sequences and isolation of these codes in non-orthogonal propagation conditions by long Pseudo-Noise (PN) scrambling sequences. Due to non-orthogonality, *multiple-access interference (MAI)* is inherently present as noise rise of the cross correlated scrambling sequences. The presence of MAI reduces the received signal-to-interference-plus-noise ratio (SINR) and decreases the coverage-throughput compared to an ideal orthogonal signal composition. The WCDMA transmitter and reference receivers are described in Section 3.3 with both the User Equipment (UE) and the base station a.k.a the NodeB (NB) functions.

3.1.3 WCDMA/HSPA

HSPA is a derivative of spread spectrum communications with multicode transmission and fast scheduling. HSPA applies together with the WCDMA system as a set of transport channels and physical channels having direct sequence properties. In HSPA, however, the code channels may be discontinued when needed. In this thesis, WCDMA/HSPA notation

is used when properties of HSPA are addressed as an add-on technology. It may also be feasible to deploy HSPA-only networks in certain propagation environments. This kind of network implementation is called Internet HSPA (iHSPA) or Evolved HSPA. The basics of WCDMA/HSPA transmitter and reference receivers are described in Section 3.3.

3.1.4 LTE

Reaching for significantly higher performance than WCDMA/HSPA, a new multiple-access technique based on Orthogonal Frequency Division Multiple Access (OFDMA) is motivated for the LTE. The WCDMA chip rate cannot be nicely increased for wide bandwidths (above 5 MHz), because that would increase the intersymbol-interference and would make equalization of the wideband carrier very complex. Also, the scalability of WCDMA to any other bandwidth options than the nominal 5 MHz is poor, because the spreading rate (chip rate) is a system specific constant and would require new designs per varied bandwidth. On the contrary, OFDMA is a scalable multicarrier technology that operates both in time and frequency domains with constant subcarrier spacing and constant symbol length [88]. This property makes OFDMA nicely scalable to many practical bandwidth options. Once the subcarrier spacing is defined constant, the symbol length remains constant despite of the Fourier transform size, and the number of active subcarriers can be adjusted properly to the transmission band. Further, OFDMA allows all baseband processing in frequency domain, where the computational complexity increase as a function of increased bandwidth, or as a function of increased channel length, is much lower than in the corresponding time domain operations. The basics of LTE transmitter and reference receivers are described in Section 3.4 with both the User Equipment (UE) and the base station a.k.a the Evolved NodeB or EUTRAN NodeB (eNB) functions.

3.2 WCDMA transmission

WCDMA transmission is based on the direct-sequence spreading technique. In downlink, each NodeB applies a set of orthogonal sequences to multiplex transmissions to the served UEs. The transmissions of the NodeB include cell specific pilot sequences and code multiplexed UEs, which are additionally protected by a long scrambling sequence relative to the signals of the other cells. Each channelization code of a UE carries control information time multiplexed with the data in each transmission frame. In uplink, each UE applies a

set of unique channelization sequences, which are code multiplexed and additionally protected by a long scrambling sequence relative to the transmissions of the other UEs served in the multiple access.

WCDMA transmission frame is rather short (10 ms) to allow frequent change of data rates and to change the signals that are actually multiplexed into the channel. This kind of transmission scheme improves multiplexing efficiency and channel utilization compared to more permanent channel reservation schemes. Detection and decoding of a data block in a time-variant channel, however, benefits of constant signal quality, and hence the WCDMA transmission frame is split to shorter slots (15 slots per frame), whose power can be controlled against the short term channel variation. Each slot in the WCDMA system contains 2560 spread symbols (chips).

Information from the higher layers is actually carried over the air interface in the form of a *transport block* that is coded and modulated for transport on the physical channels. The transport block is of defined size and format, known as the *transport format*. WCDMA may carry a *transport block set* during a *transmission time interval*, which is a defined multiple of the transmission frame.

HSPA transmission maintains the fundamantal multiplexing properties of the WCDMA downlink (HSDPA) and uplink (HSUPA). HSPA however adds the opportunities to use transmissions on many parallel code channels, and to adapt the modulation and coding format of the data block. HSPA further applies short term allocations for the duration of 2 ms (three slots), which allows even faster change of data rates and updates of the composition of the multiplexed signals. HSPA carries a single transport block in a transmission time interval of 2 ms.

3.3 WCDMA transmitter-receiver techniques and their imperfections

3.3.1 WCDMA/HSPA transmitter

In WCDMA downlink transmitter shown in Fig. 3.1, each pair of two consecutive real-valued symbols (bits) n_k, n_{k+1} is first serial to parallel converted to become the in-phase (I-branch) and quadrature-phase (Q-branch) symbols, and then spread to the constant chip rate (3.84 Mcps) with symbol period of about 260 ns. The channelization actually happens by multiplying every bit on the I-branch and Q-branch by the same, unique,

real-valued channelization sequence c_{ch} selected from the Orthogonal Variable Spreading Factor (OVSF) code tree, with spreading factor SF and code index m (hence the notation $c_{ch,SF,m}$). The spread sequences on the I-branch and Q-branch are actually treated as a single complex-valued sequence of chips that consists of Quadrature Phase Shift Keying (QPSK) modulated symbols ($I + j \cdot Q$). Additionally, each complex-valued spread channel may be separately weighted by a power weight factor. All downlink physical channels at the transmitter are summed using complex addition, and the spread sequences are scrambled by a complex-valued cell-specific scrambling code ζ_{dl}. The signal is finally pulse shape filtered for transmission. Orthogonality is kept among all physical channels, except the system specific synchronization sequences that are not allocated from the orthogonal space for the ease of cell search procedure.

For HSPA, higher order modulation mappers are available. In case of Quadrature Amplitude Modulation (QAM) 16QAM, a set of four consecutive binary symbols (bits) $n_k, n_{k+1}, n_{k+2}, n_{k+3}$ is serial-to-parallel converted to two consecutive binary symbols ($i_1 = n_k,\ i_2 = n_{k+2}$) on the I-branch and two consecutive binary symbols ($q_1 = n_{k+1},\ q_2 = n_{k+3}$) on the Q-branch, and then mapped to a 16QAM symbol. The I-branch and Q-branch are both spread to the chip rate by the same real-valued channelization code $c_{ch,16,m}$ of constant spreading factor ($SF = 16$). The sequence of chips from all HSPA code channels are summed and scrambled by the cell-specific scrambling code (ζ_{dl}).

In the case of 64QAM, a set of six consecutive binary symbols $n_k, n_{k+1}, n_{k+2}, n_{k+3}, n_{k+4}, n_{k+5}$ is serial-to-parallel converted to three consecutive binary symbols ($i_1 = n_k, i_2 = n_{k+2}, i_3 = n_{k+4}$) on the I-branch and three consecutive binary symbols ($q_1 = n_{k+1}, q_2 = n_{k+3}, q_3 = n_{k+5}$) on the Q-branch and then mapped to a 64QAM symbol by the modulation mapper. The I-branch and Q-branch are then both spread to the chip rate by the same real-valued channelisation code $c_{ch,16,m}$. The sequence of chips from all HSPA code channels are summed and scrambled by the cell-specific scrambling code. The modulation constellations are visualized for QPSK, 16QAM and 64QAM in Fig. 3.2. The channels are combined for transmission according to Fig. 3.3. The root-raised-cosine (rrc) pulse shaping is applied to the complex-valued, modulated chip sequences according to Fig. 3.4.

In the WCDMA uplink transmitter shown in Fig. 3.5, each bit on a dedicated physical channel of a UE is spread by multiplying it by a real-valued channelization sequence selected from the set of orthogonal sequences. The spreading factor is variable to match

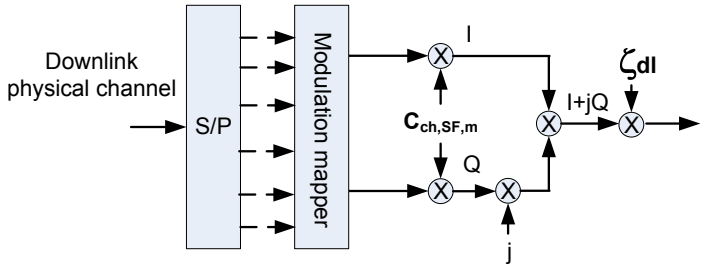


Figure 3.1: Downlink modulator with the spreading and scrambling functions.

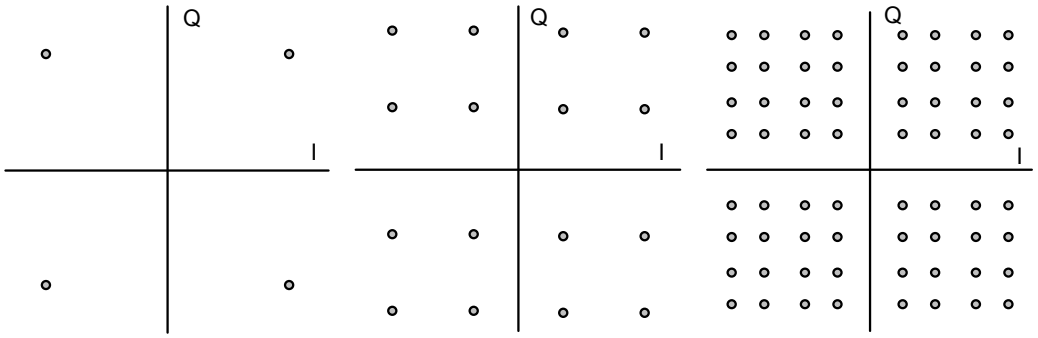


Figure 3.2: Modulation constellations of QPSK, 16QAM and 64QAM alphabets. Serial to parallel conversion of the coded bit stream is applied to 2 bit, 4 bit or 6 bit words per modulation symbol respectively.

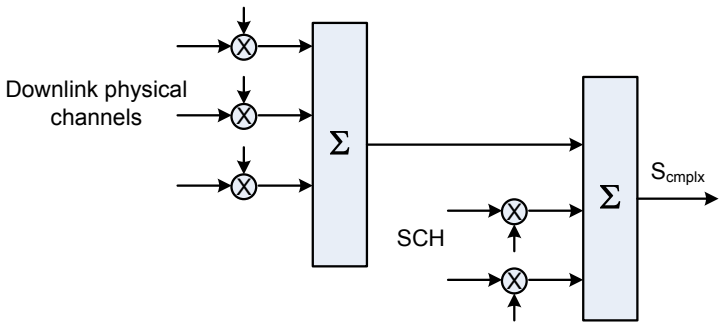


Figure 3.3: Combining of the downlink physical channels for transmission. Other channels except pilot and synchronization channels (SCH) satisfy the orthogonality condition at the transmitter.

the variable rate bit sequence to the constant chip rate. After the channelization, power weighting (β) is applied for each code channel, and the spread sequence is mapped either

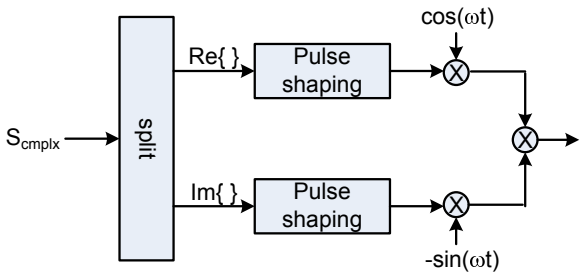


Figure 3.4: Modulation and pulse shaping of the complex downlink signals.

to the I-branch or Q-branch of the modulator. In the case of multicode transmission, the physical channels are mapped in alternating order to the I-branch and Q-branch to keep the peak-to-average power ratio of the transmission small. In WCDMA, when multicode transmission is not applied, there are only two physical channels (a Dedicated Physical Control Channel, DPCCH, and a Dedicated Physical Data Channel, DPDCH) present. In a multicode transmission with multiple dedicated data channels, all the channels on the I-branch ($\mathcal{S}_{dpdch,i}$) are summed, as all the channels on the Q-branch ($\mathcal{S}_{dpcch}$, $\mathcal{S}_{dpdch,q}$) are summed to form a complex-valued sequence of QPSK symbols. The resultant complex-valued symbol sequence is further multiplied by a UE-specific complex-valued scrambling sequence ζ_{UE}. The signal is finally pulse shape filtered for transmission. Orthogonality is kept among physical channels of a single UE. Orthogonality of the channelization sequences is completely lost between signals from different UEs anyway and the signals remain solely isolated by the scrambling code, therefore the same code tree may be reused without additional losses. Signal isolation between different UEs is kept sufficient by the long scrambling sequences that produce low cross-correlation noise.

For WCDMA/HSPA downlink transmission, one more physical channel ($\mathcal{S}_{hs-dpcch}$) is added in the uplink for the sake of downlink feedback. For WCDMA/HSPA uplink transmission, a number of enhanced dedicated physical channels (one enhanced dedicated physical control channel, e-dpcch, $\mathcal{S}_{e-dpcch}$ and a number of enhanced dedicated physical data channels e-dpdch, $\mathcal{S}_{e-dpdch}$) may be added and again multiplexed in alternating order to the I-branch or Q-branch respectively. In WCDMA/HSPA, physical channel specific power weighting ($\beta_{hs}, \beta_{ec}, \beta_{ed,k}$) may apply.

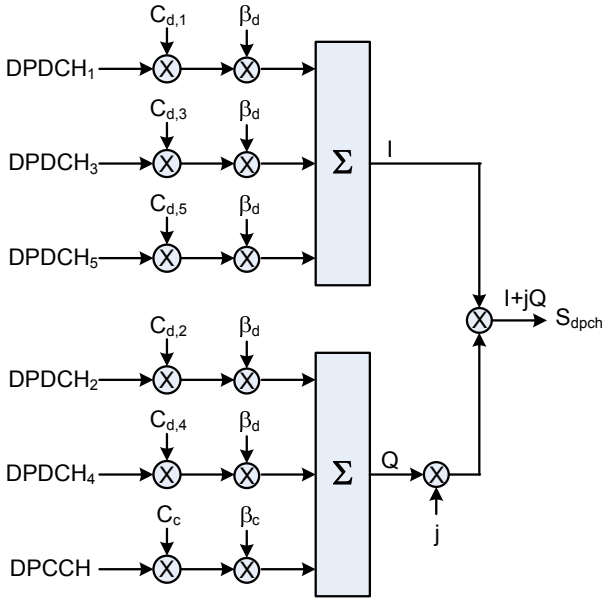


Figure 3.5: Spreading and code multiplexing of an uplink dedicated physical control channel (DPCCH) and dedicated physical data channels (DPDCH).

3.3.2 Transmitter with multiple antennas

WCDMA/HSPA transmission is possible from multiple antennas. The various transmission schemes include at least the open-loop space-time transmit diversity (STTD) and closed-loop mode 1 transmit diversity [89] also known as the adaptive antenna transmission (TxAA) [90][91][92]. In multiantenna diversity transmission, the coded and modulated symbol sequence is duplicated to the transmit antenna branches. In STTD, the space time Alamouti-transform is applied to the symbol block as precoding prior to transmission [93][94][91]. In TxAA, a phase shift ($e^{-j\phi}$) is multiplied to the symbols of a symbol block, transmitted simultaneously from the antenna branches. The phase shift is selected per symbol block, assisted by the UE receiver feedback, to obtain the largest instantaneous SINR at the receiver. In the diversity transmission mode of Fig. 3.6, the transport block fed to different antenna branches is duplicated, but the physical layer symbol sequences of the replica are different by a precoding operation (STTD) or by a phase shift (TxAA). In this case branch 3 and branch 4 in Fig. 3.6 may be omitted. Closed-loop mode 1 transmit diversity was analysed for WCDMA/HSPA transmission in [28].

Transmit diversity is the nominal multiantenna transmission scheme for WCDMA, but multistream transmission has been studied as well. For a multistream transmission, the symbol sequence is split to multiple antenna branches without duplication. In the multistream transmission of Fig. 3.6, the transport blocks fed to different antenna branches are independent primary transport blocks. In per antenna rate control scheme, each of

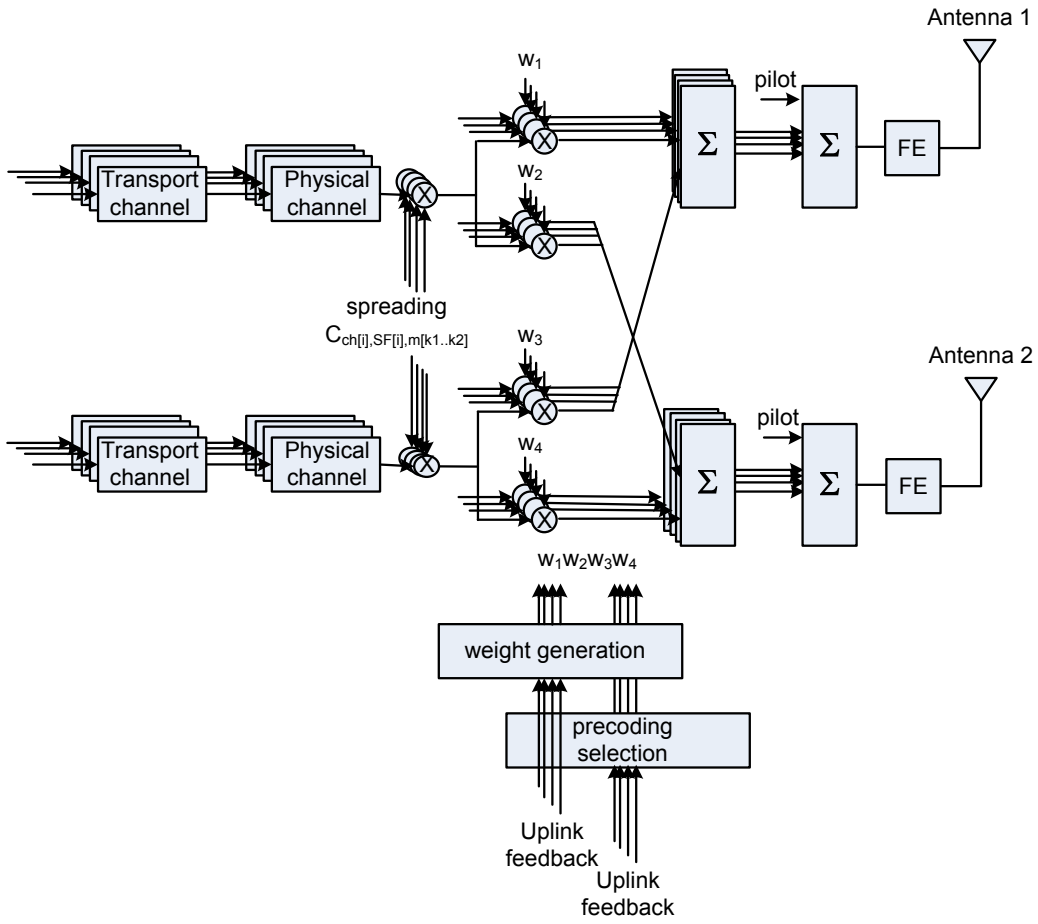


Figure 3.6: Block diagram of the WCDMA transmitter with multiple antennas.

the transport blocks are transmitted from a single antenna port so that e.g. branch 2 and branch 3 may be omitted in Fig. 3.6. In the multistream layered matrix modulator, each transport block is transmitted from any number of antenna ports simultaneously, after multiplied by the complex weights of the precoding matrix.

3.3.3 WCDMA/HSPA reference receivers

The conventional *Rake* reference receiver for the spread spectrum signals has correlators or matched filters, which are set to receive a known channelization sequence created in the transmitter. The signal phase and amplitude distort in the channel, and equalization is necessary in the receiver. The signal scatters to multipath components in the channel, and reception of the signal components in multiple phase shifts is necessary. The corre-

lator receiver executes this by setting a separate correlator process to receive each signal path and then sum all received multipath components to a resultant signal. The receiver tunes the correlator phase to a signal component and tracks its timing changes by a code tracker algorithm. The correlator receiver is called Rake receiver and it consists of a correlator bank. In a practical WCDMA receiver, separate correlators are needed for every channelization code to be received, and actually to each and every signal component of them, which sufficiently contribute to increase the received SINR of the resultant signal. In addition, searcher correlators are needed to track new signal components with unknown phases. Further, pilot sequences need to be continuously received in the serving cells for equalization and channel estimation. Also, the pilot sequences of the neighboring cells need to be searched for in the cell search algorithm (see Section 8.4) for the measurements. The complexity of Rake reception gets particularly large, when the number of code channels to receive increases, when the code channels additionally need to be received from multiple cells at the same time (in the macrodiversity handover) and yet when the multiantenna transmission applies. One remark is that operation in the macrodiversity handover requires processing of all the physical control channels separately from all cells that belong to the *active set* that form the macrodiversity connection, in addition to soft combining the signal paths of the dedicated data channels that are transmitted from multiple cells of the active set. For the reception of (closed loop) multiantenna transmission, the receiver needs to estimate the channel for each signal transmitted from one antenna port. The receiver may estimate the channel phase from the common pilot signal. The receiver also has to provide the channel phase reference as a feedback to the transmitter. To receive data channels, the receiver may set correlators directly to the strongest observed signal components of the (multiantenna) joint-channel that provide transmit diversity gain at the receiver (both in STTD and in TxAA).

In case of multistream transmission, multiple receiver antennas are required to form a true Multiple Input Multiple Output (MIMO) configuration. Multistream transmission roughly multiplies the receiver complexity, as separate symbol streams need to be simultaneously detected from multiple antennas, and the channelization sequences may additionally have experienced multipath dispersion and need reception in multiple phases. The mutual interference between symbol streams is often too high, and advanced interference cancellation schemes (see Chapter 4) are necessary to observe throughput gains. The advanced interference cancellation schemes add complexity apart of the complexity

of multiantenna reception as such.

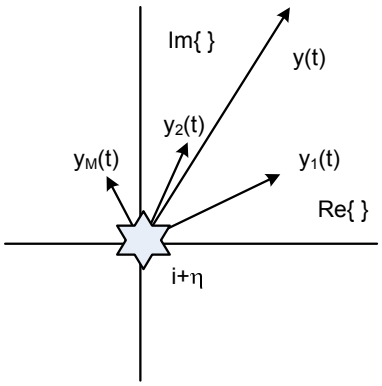
The principle of Rake receiver is shown in Fig. 3.7[1]. In Rake, the correlator outputs are constructively combined by phasing the correlator timing to each multipath omponent and by weighting the correlator outputs so that the stronger components dominate the weaker ones. The receiver combines the phase aligned signal components by the maximum ratio combining. For Rake, the optimum receiver property holds in the additive white Gaussian noise channel without correlation. It leads to the highest expected SNR by reducing channel fading of the sum signal [84]. In maximum ratio combining, a linear combination of the phase aligned signal components is formed by (3.1).

$$y(t) = \sum_{k=1}^{M} a_k \cdot z_k(t), \tag{3.1}$$

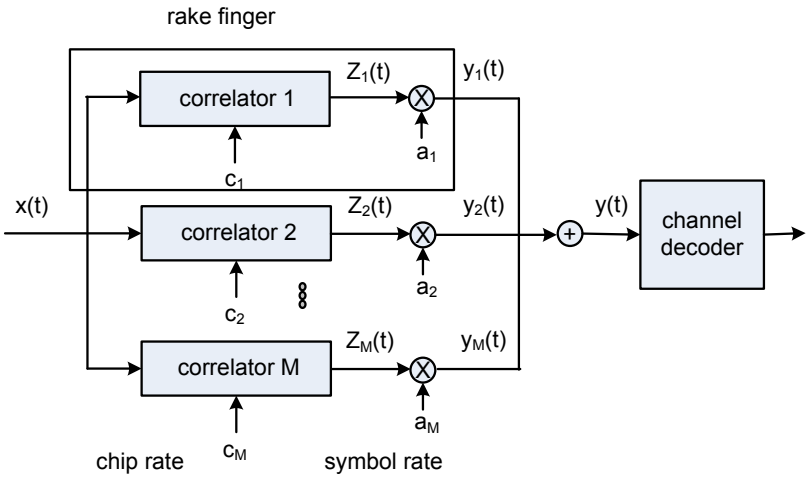
where y is the receiver output, a_k is the k^{th} weight, z_k is the k^{th} correlator output and M is the number of Rake fingers. The weights a_k are chosen to be the complex conjugates of the corresponding channel coefficients, estimated from the common pilot signal. That is, let the k^{th} channel coefficient be h_k, the weighting coefficient a_k is set to the estimated $\widehat{h_k}^*$.

The *Minimum Mean-Squared Error (MMSE)* receiver is an alternative reference receiver for WCDMA, especially for the multiantenna reception. MMSE does not need to resolve the multipath propagated signal components separately, which is an advantage that simplifies the receiver a lot. On the other hand, MMSE needs to equalize the complex joint propagation channel in a chip resolution (chip equalizer), which is a very demanding computation task due to short duration of the chip. For multiantenna reception, the chip sequences are equalized per antenna branch, and no separate reception per channelization code is necessary. Strong multipath propagation with long delays (larger than the chip period) makes the chip equalizer in time domain very challenging (see Chapter 4). Therefore, MMSE receiver is a promising technique particularly for HSPA, which by definition does not operate in macrodiversity and which is typically scheduled for a receiver, when the delay spread of the propagation environment is small.

[1] A static Rake receiver was originally introduced in Price, R. and Green, P.E. Jr., A communication Technique for Multipath channels, Proc. IRE 46, 1958, pp.555-570



(a)



(b)

Figure 3.7: (a) Composition of the received complex signal vector from the received multipath signal components. (b) The block diagram of the Rake receiver with correlators for an input to the sequence decoder.

3.4 LTE transmission

3.4.1 Design and parametrisation

LTE is designed for OFDMA transmission in a way having robustness agains multipath propagation and to keep orthogonality between transmitted signals. Multipath robustness is achieved by a transform, which takes a symbol block and creates a set of parallel tones that have a much longer symbol period than the original symbols [95][96]. The longer symbols suffer less from multipath propagation, as *intersymbol-interference (ISI)* can be avoided. Signal orthogonality is provided by isolating the symbol blocks by a guard interval

in time domain and by having a transform that overlaps the signal components in frequency domain satisfying the orthogonality condition [97]. The guard interval avoids ISI in time domain and the orthogonal transform avoids *intercarrier-interference (ICI)* in frequency domain.

OFDM is a type of multicarrier modulation, which uses frequency overlapped orthogonal sinusoids to divide frequency selective wideband channel into a number of narrowband flat fading channels. Ideally there is no need to equalize OFDM signal by other than a single complex subcarrier coefficient, and therefore the ideal OFDM waveform provides orthogonality without multiple-access interference. However, frequency offset errors and receiver phase noise impact so that the subcarriers are not perfectly orthogonal, which creates ICI and performance degradation [98][99][100]. The system design with proper dimensioning of the subcarrier bandwidth and the selected receiver technology largely mitigate these performance degradations [101].

One of the fundamental strengths of OFDM technique is the processing in frequency domain, which enables *Frequency Domain Equalizer (FDE)* and advanced receivers with low computational complexity [102]. Frequency transform is done orthogonally by the discrete DFT or FFT transforms, whose product in frequency domain equals the convolution in time. The property of the discrete Fourier transforms and their equivalent circular convolution in (3.2) are different from the continuous Fourier transforms and their linear convolution. (The notation of linear convolution is $*$ and circular convolution is $\circledast$.)

$$\begin{aligned} FT\{s(t) * h(t)\} &= FT\{s(t)\} \cdot FT\{h(t)\} \\ FFT\{s[n] \circledast h[n]\} &= FFT\{s[n]\} \cdot FFT\{h[n]\} \end{aligned} . \tag{3.2}$$

The continuous time signal s(t) and the channel h(t) are linearly convolved, hence a cyclic convolution property of the discrete transform for a signal block s[n] in a channel h[n] imposes an approximation problem in a multipath channel. Yet for a block transmission, a cyclic extension of the signal can be added between the symbol blocks (for the guard interval period), which causes the circular convolution become equivalent to the linear convolution, as long as the channel spread is shorter than the cyclic extension. In this way the ideal block transmission is preserved even in a linearly convolved channel, when the finite length block transmission is convolved with the finite length FIR filtering of the channel. The circular convolution provides a response of length N for the product of two DFTs of length N, and no *interblock-interference (IBI)* will arise.

Proper windowing of the finite symbol block in the transmitter is important, because otherwise the aliased signal appears infinitely periodic to the channel. Windowing is applied after the samples of the cyclic extension of the signal are inserted, so that the windowed symbol block will not become cyclically extended. In the OFDM receiver, windowing is done by adjusting the window position to include cyclic extensions on both sides of the received symbol block and hence to include channel delay spread of the symbol block without interblock interference. This is achieved by convenient sampling techniques and by discarding the samples of the cyclic extension before the transform. Windowing as such may be inherent in the FIR filtering process and it does not require a special concern. Because OFDM transmission technique is very simple and its theory well known, it is worth to describe its parametrisation and practical system design choices, which are a challenge instead.

3.4.2 Time-frequency design of the symbol block

The time-frequency representation of an OFDM signal is given by the sinusoids of the symbol block transformed to the frequency components as sinc-function according to Fig. 3.8. In LTE, the time-frequency allocation happens in the resolution of a set of subcarriers that is called a Physical Resource Block (PRB). The symbol resources on a subcarrier of an OFDM symbol are called resource elements. The time-frequency symbol blocks are shown in Fig. 3.9.

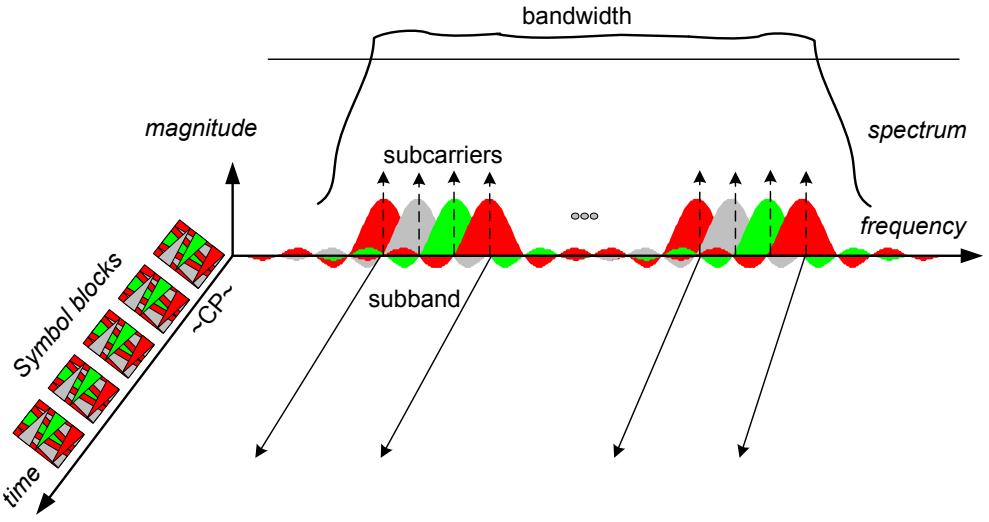


Figure 3.8: Time-frequency representation of the OFDM signal.

There exists a practical dual problem for solving the frequency properties of the subcar-

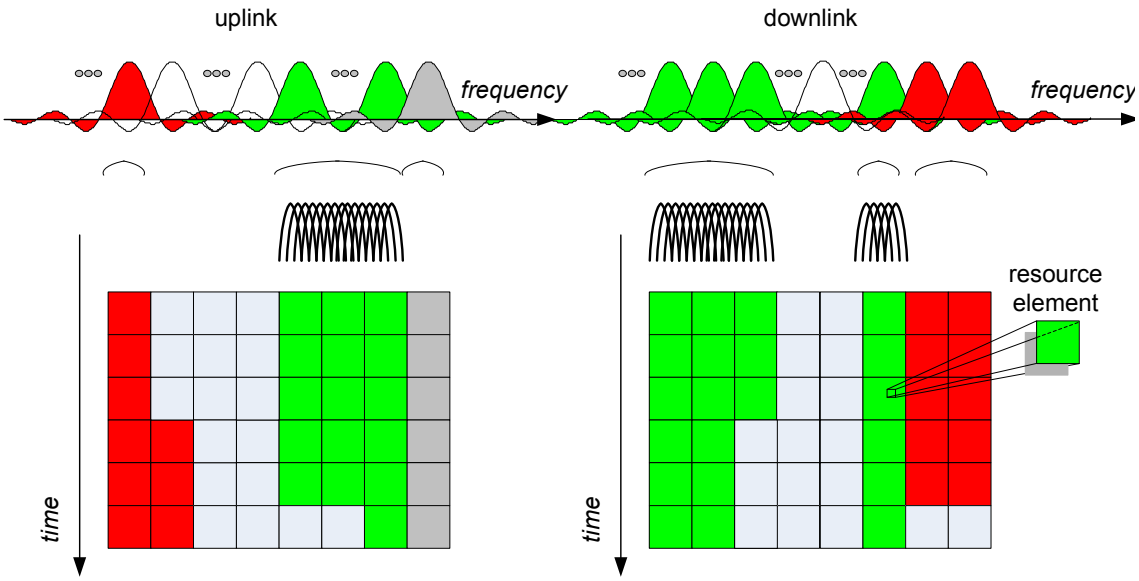


Figure 3.9: An example of the time-frequency allocations in uplink and downlink.

rier and the time properties of the symbol. Dimensioning of the cyclic extension, actually, is a separate problem, but is not trivial, when practical constraints of the frame length have to be preserved. In LTE, the symbol was defined for a large enough subcarrier bandwidth, say from 10 to 30 kHz, so that it well tolerates phase noise and fairly large Doppler spread experienced at high velocities (up to 500 km/h). Yet, the subcarrier bandwidth is sufficiently narrow to avoid complex subcarrier equalization and to allow fast synthesizer settling times. A wide subcarrier bandwidth creates short symbols in time domain, and vice versa a narrow subcarrier creates long symbols. A nearly optimal generic choice was selected for LTE with subcarrier spacing of 15 kHz that provies symbols of length about 67 μs. These parameters are excellent as they also nicely match the clock rate 38.4 MHz that is compatible to the clock rates of GSM/EDGE and WCDMA systems and to the frame length of WCDMA (10 ms). Further, this subcarrier spacing allows bandwidths at least up to 20 MHz be generated by a modest FFT size of 2k-points. For all the above mentioned reasons, 3GPP decided to make a universal symbol definition for LTE that has the basic time unit $1/\ (15kHz \cdot 2048)$ about 33 ns, subcarrier spacing 15 kHz and OFDM symbol duration of 67 μs.

For allocating time-frequency symbol resources fast and accurately, the frame was further divided to subframes of 1 ms each. The subframe consists of two halves, called slots,

each of which has an exact integer number of symbols for both the normal cyclic extension and the extended cyclic extension. This design has the property that the last symbol of every slot is equally spaced in time for both the cyclic extension lengths. Synchronization sequences are modulated to these symbols (once/twice per frame), which make cell search and symbol timing particularly easy.

Information from the higher layers is carried over the air interface in the form of a transport block that is coded and modulated to the time-frequency symbol resources. In LTE, one single transport block fits to an integer number of PRBs during a transmission time interval of a subframe (1 ms).

3.4.3 Frequency scaling

Classical TDMA and spread spectrum techniques are designed per system bandwidth, whereas OFDM has the property of bandwidth scaling. Once the subcarrier bandwidth and symbol length is designed, the IFFT/FFT transform can scale to any power of two. For modulating the effective system bandwidth, a wider Fourier transform can be computed and the subcarriers outside the effective system bandwidth may be transmitted at zero power and filtered away both in the transmitter and in the receiver. Increase of the FFT-size provides shorter sampling interval during a symbol period, which can hence present more frequency components, and the number of subcarriers may increase. The sampling interval does however not change the subcarrier spacing, which is proportional to the symbol period $T_s = 1/\Delta f_{sc}$. Larger FFT may fill a larger system bandwidth, which will increase the total symbol rate as a function of increasing number of subcarriers. For a narrowband transmission, FFT size may be decreased.

3.4.4 Cyclic signal extension

The cyclic signal extension is designed as the Cyclic Prefix (CP) extension of the sinusoids in the symbol block. The CP length should be relative to the channel delay spread, but that depends on the propagation environment i.e. on the presence of scatterers and reflectors. Plenty of channel measurements are available, and optimisation of the length of the CP is studied in the literature [103]. In addition to RMS delay spread of the channel, the channel filtering and windowing responses have to be included during the CP samples. The attenuation of about 20 dB both in the RMS delay spread and in channel filtering is typically required [104].

In LTE, adaptation to different propagation environments was done by parametrising the length of the CP. A normal CP of about 5 us is needed to keep the overhead small and still fit majority of all propagation environments. On the other hand, an extended CP of about $15\mu s$ is needed to cope with all possible environments (e.g. macro cells) experiencing long delay spreads. The same extended CP also serves Multimedia Broadcast Multicast Simulcast (MBMS) services, which by definition need to receive identical symbols from multiple cell sites with longer propagation time differences.

Any number of transform sizes, CP lengths and symbol lengths could be designed for a communication system to match different propagation environments. However, this complicates the system design and does not provide large gains. For LTE, actually two CP values were considered sufficient, and they were defined as the normal CP and the extended CP in [105]. The beauty of LTE design is in that the CP length does not change the symbol length itself. The symbol is invariant in time and frequency despite of the CP length, and the sequence of symbols exactly fills the universal subframe. (In other specifications based on OFDM e.g. WiMax and DVB, the choice of a CP may modify the transform size ($CP \propto 1/FFT\_size$) and hence the symbol length and subcarrier spacing will change.)

For environments with very small delay spread, even a shorter CP could have been chosen, which however does not satisy the two conditions of a universal symbol (length) and the fixed frame length. Further, squeezing the CP smaller is not favourable, because in situations of discontinuous reception, the automatic frequency and gain control circuitry require convergence time. If convergence is not good, the pilot symbols present in the first OFDM symbol may suffer from degraded quality, which can impact the equalization and channel estimation accuracy. In uplink, the burst power ramp-up and ramp-down to about $+/-20\ dB$ level may require about 1 to 2 μs with realistic channel filtering. If burst power ramp-up/down periods are shortened or they do not have sufficient dynamic range, the burst overlap will convert to IBI [98].

3.4.5 Symbol mapping

Once DFT-precoding was decided for LTE, there exist many alternative ways to multiplex UE allocations and to provide the multiple-access in frequency domain. The symbol mapping technique actually impacts the multiple-access scheme, addressable resolution of allocations, required word length of signalling, receiver complexity and receiver sensitivity.

3.4.5.1 Localized mapping

In the localized symbol mapping of Fig. 3.10, DFT-spread samples of one user are placed to a set of contiguous frequency bins of the IFFT, whereas zeros are mapped to the non-modulated subcarriers, according to (3.3). Multiple users at the same transmission time may use DFT output mapping to the orthogonal frequency bins of the IFFT, which are left non-modulated by the others. Generation of OFDM signal as parallel long symbol blocks is shown in Fig. 3.11, the frequency bins of localized mapping are shown in Fig. 3.12.

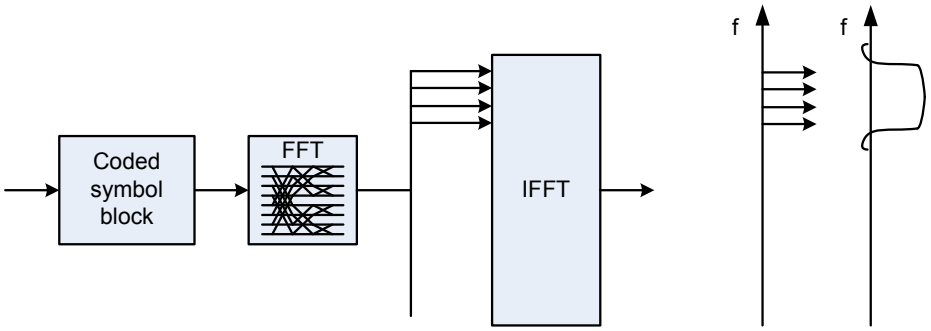


Figure 3.10: Transform domain processing and mapping to frequency localized signal components.

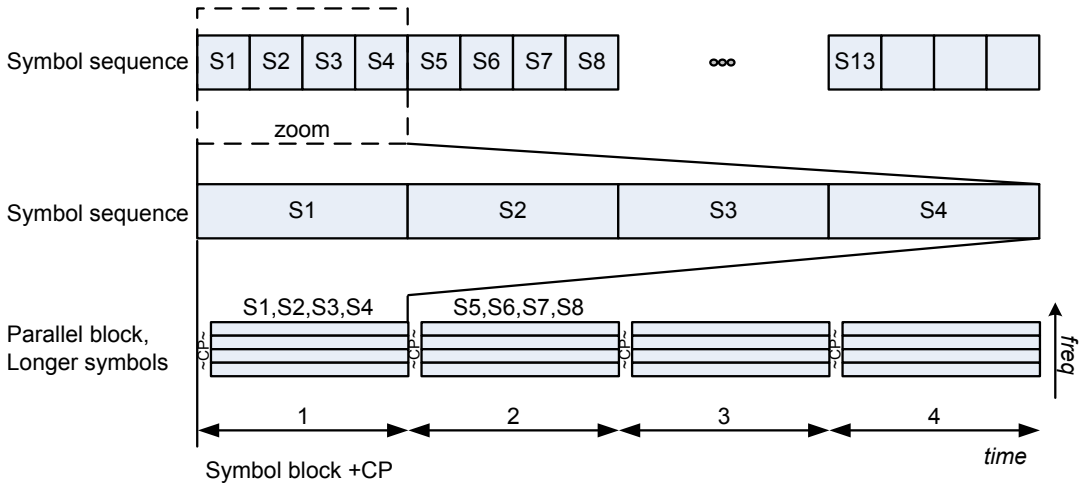
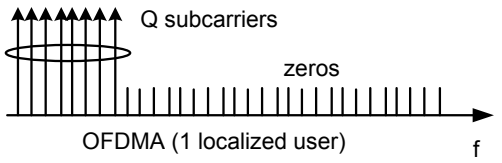


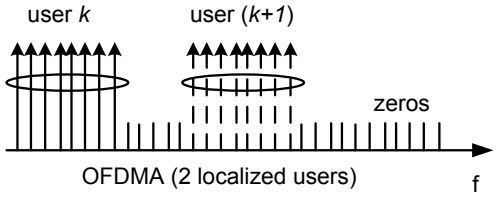
Figure 3.11: The OFDM, generation of long parallel symbol blocks and transform domain procssing for frequency multiplexing.

$$z'(i) = \begin{cases} z(l) & \text{, if } i = (l+\xi) \mod \text{IFFT}_{\text{size}} \\ 0 & \text{, otherwise} \end{cases}, \tag{3.3}$$

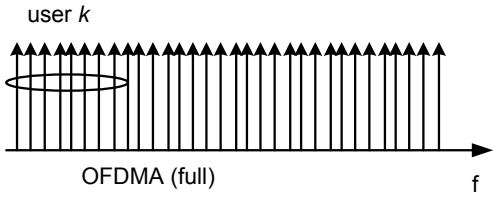
so that z is the modulated symbol vector at the output of the DFT precoder, and z' is the DFT output mapped to the input of the IFFT (of $\text{IFFT}_{\text{size}}$). The DFT with precoding is given by (3.7). $l = 0, \ldots, M_{sc}^{PUSCH} - 1$, ξ is the user specific subcarrier allocation offset (in PRB resolution) with $0 \leq \xi \leq \text{IFFT}_{\text{size}} - 1$ and $i = 0, \ldots, \text{IFFT}_{\text{size}} - 1$.



(a)



(b)



(c)

Figure 3.12: OFDMA representation of frequency localized mapping for (a) a single user, (b) two frequency multiplexed users and (c) fully allocated (multiuser) band.

3.4.5.2 Clustered mapping

The clustered mapping of Fig. 3.13 is created by taking a block of symbols for the contiguous frequency bins and by repetition coding that block of symbols according to (3.4). The block repetition process creates a comb-shaped mapping of symbols of one user to the frequency bins of the IFFT, which thus consists of a set of localized clusters distributed in frequency [106]. Multiple users at the same transmission time may use DFT output mapping to the orthogonal frequency bins of the IFFT, which are left non-modulated by

the others.

$$z'(i) = \begin{cases} z(l) & \text{, if } i = ((l+\xi) + \vartheta \cdot t) \mod t \\ 0 & \text{, otherwise} \end{cases}, \tag{3.4}$$

so that z is the modulated symbol vector at the output of the DFT precoder, and z' is the DFT output mapped to the input of the IFFT (of $\text{IFFT}_{\text{size}}$). The DFT with precoding is given by (3.7). $l = 0, \ldots, M_{sc}^{PUSCH}/\vartheta_{CL} - 1$, t is the cluster distribution factor, ξ is the user specific subcarrier allocation offset (in PRB resolution) with $0 \leq \xi \leq (t-1)$, ϑ is the cluster index, $\vartheta = 0, \ldots, \vartheta_{CL} - 1$ with the number of clusters ϑ_{CL} and $i = 0, \ldots, \text{IFFT}_{\text{size}} - 1$.

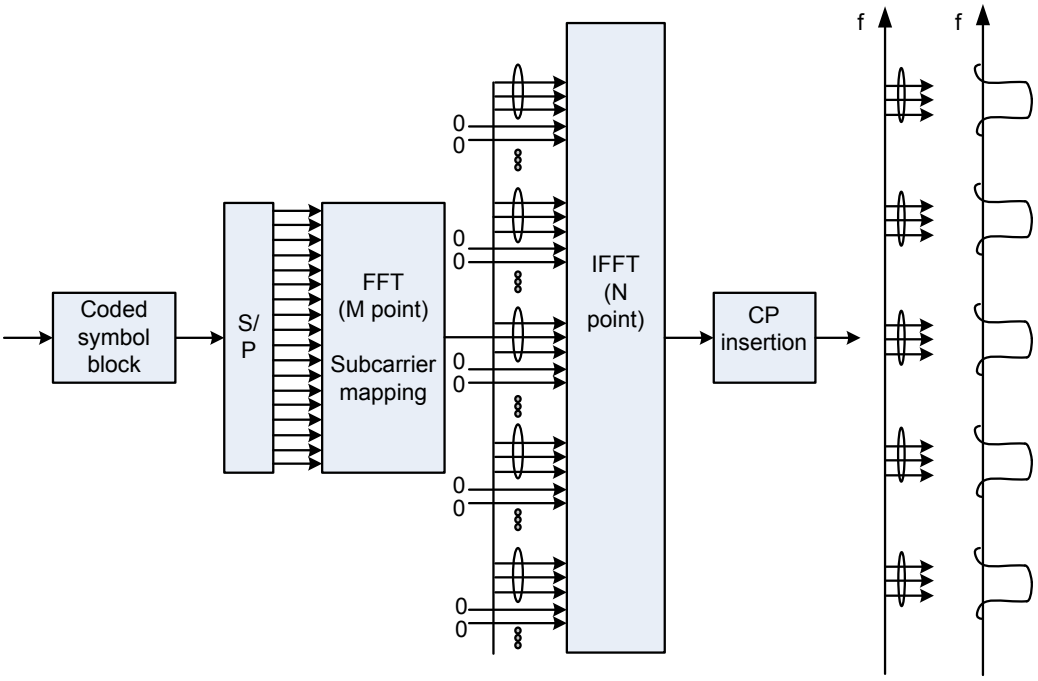


Figure 3.13: Transform domain processing and mapping to the distributed sets of frequency localized signal components.

3.4.5.3 Distributed mapping

In the distributed symbol mapping of Fig. 3.14, DFT-spread symbols of one user are evenly spaced to the frequency bins of the IFFT, whereas zeros are mapped to the non-modulated subcarriers, according to (3.5). Multiple users at the same transmission time may use DFT output mapping to the orthogonal frequency bins of the IFFT, which are left non-modulated by the others.

$$z'(i) = \begin{cases} z(l) & \text{, if } i = (tl + \xi) \\ 0 & \text{, otherwise} \end{cases}, \tag{3.5}$$

so that z is the modulated symbol vector at the output of the DFT precoder, and z' is the DFT output mapped to the input of the IFFT (of $\text{IFFT}_{\text{size}}$). DFT with precoding is given by (3.7). $l = 0, \ldots, M_{sc}^{PUSCH} - 1$, t is the distribution factor, ξ is the user specific subcarrier allocation offset (in subcarrier resolution) with $0 \leq \xi \leq (t-1)$ and $i = 0, \ldots, \text{IFFT}_{\text{size}} - 1$.

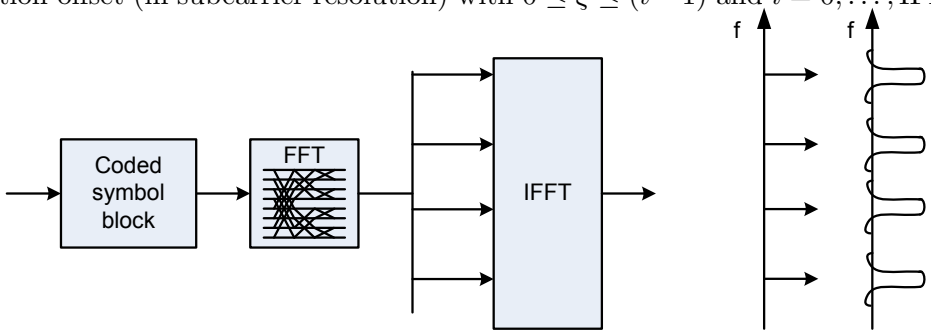


Figure 3.14: Transform domain processing and mapping to frequency distributed signal components.

3.4.5.4 Interleaved mapping

The interleaved symbol mapping can be created by repeating and rotating the symbols. This scheme is a.k.a Interleaved FDMA (IFDMA), proposed in [107] and shown in Fig. 3.15 and Fig. 3.16.

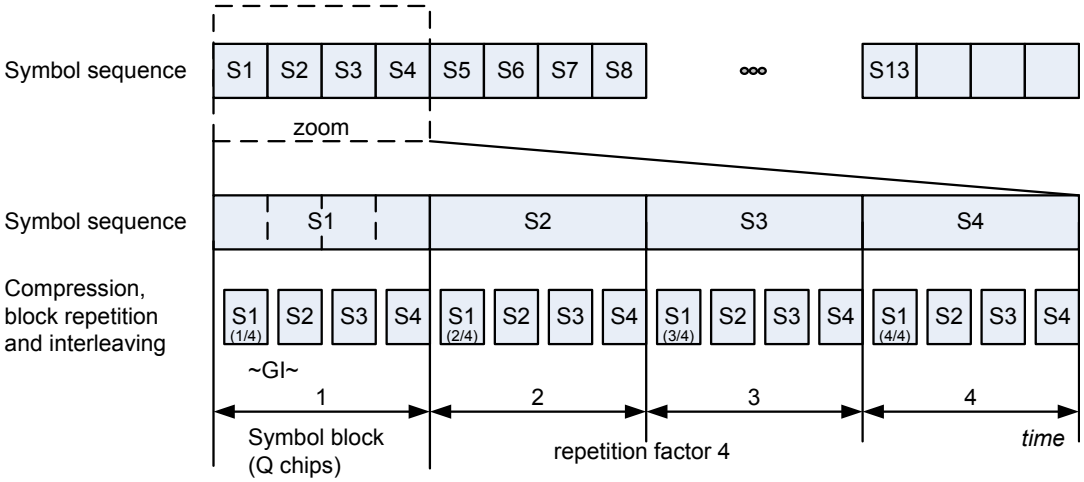


Figure 3.15: Symbol sequences with symbols compressed to the chips of short duration and guard interval added for block separation. Block repetition and interleaving generates a comb-shaped spectrum for frequency multiplexing.

Symbol repetition creates a distributed and interleaved comb-shaped spectrum over the symbol bandwidth. User allocations are separated by user-specific phase vectors, which map the user-specific allocations to mutually orthogonal frequency bins of the comb-shaped

spectrum. The symbol blocks need to have a guard interval between the symbol blocks in order to keep the orthogonality in a dispersed channel. The length of the symbol repetition block and the symbol repetition factor define the tightness of symbols in the frequency comb. The subcarriers of another user are avoided by assigning orthogonal phase vector patterns in the repetition. IFDMA is analysed e.g. in [108][109].

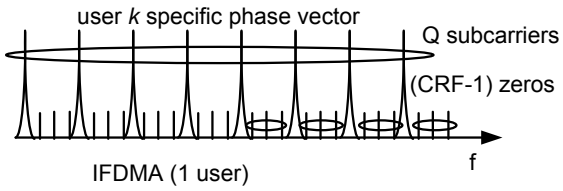


Figure 3.16: Representation of the block repetition interleaved signal components as frequency bins shifted by the user dependent phase vector.

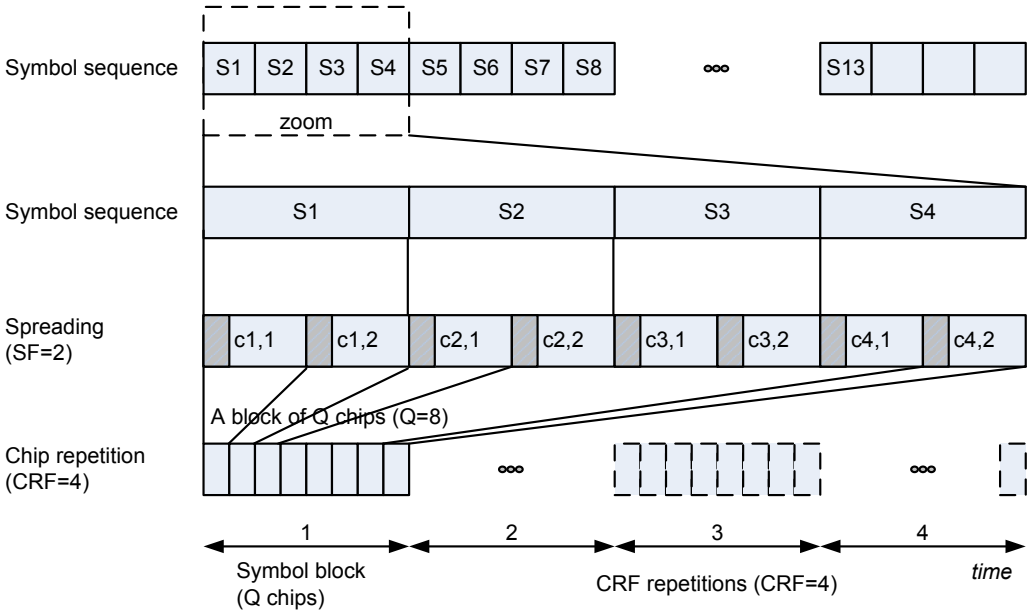


Figure 3.17: Spreading of the symbol sequences to the chips of short duration for frequency multiplexing.

IFDMA scheme is actually very close to a more recent Variable Spreading and Chip Repetition Factor (VSCRF) CDMA scheme, presented in [110] and shown in Fig. 3.17 and Fig. 3.18. In VSCRF-scheme, however, the symbols are both spread and block repetition coded, so that the symbols are first multiplied by a channelization sequence and then repetition is applied to the blocks of interleaved chip sequences. The variable spreading factor relative to the common chip rate is able to provide different data rates. Users are

separated by the user-specific phase vectors (or phase shifts) in the interleaved block of chips, which realizes the orthogonality condition in frequency domain. The frequency bins of distributed interleaved mapping are shown in Fig. 3.18 for one user, two frequency multiplexed users and for full band allocation.

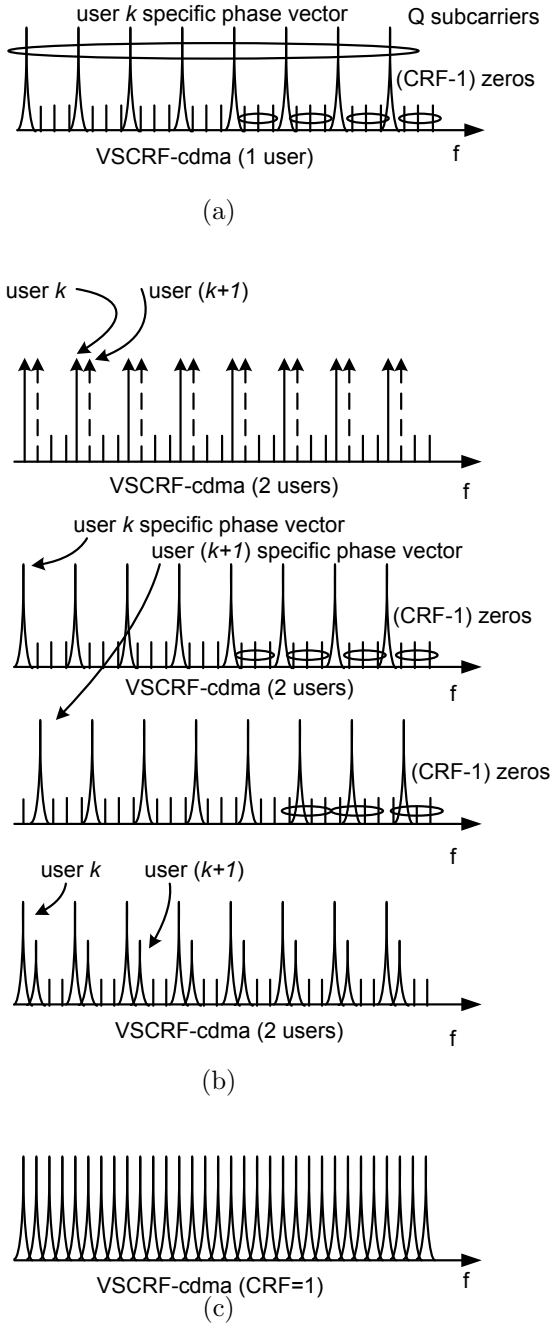


Figure 3.18: Representation of the block repetition interleaved signal components as frequency bins for (a) a single user, (b) two frequency multiplexed users and (c) fully allocated band.

Both IFDMA and VSCRF schemes perform comparably to the OFDMA with small

repetition factors ($CRF < 8$), but large repetition factors ($CRF > 8$) cause a performance loss of $0 - 2\ dB$ compared to OFDMA [111][112]. This implies a problem for low bitrate transmissions. Both IFDMA and VSCRF, similar to the subcarrier distributed mapping of OFDMA, are known to be sensitive to frequency error, phase noise and imperfect power control [113].

As a conclusion, frequency-localized symbol mapping was selected for LTE transmission both in downlink and in uplink. For such downlink transmissions that do not get frequency diversity by other means (e.g. by frequency adaptive retransmissions), it is further possible to create PRB-level distributed allocations. This is efficiently exploited by the frequency domain packet scheduler, and it does not suffer from imperfections as the subcarrier interleaved schemes do.

3.5 LTE transmitter-receiver techniques and their imperfections

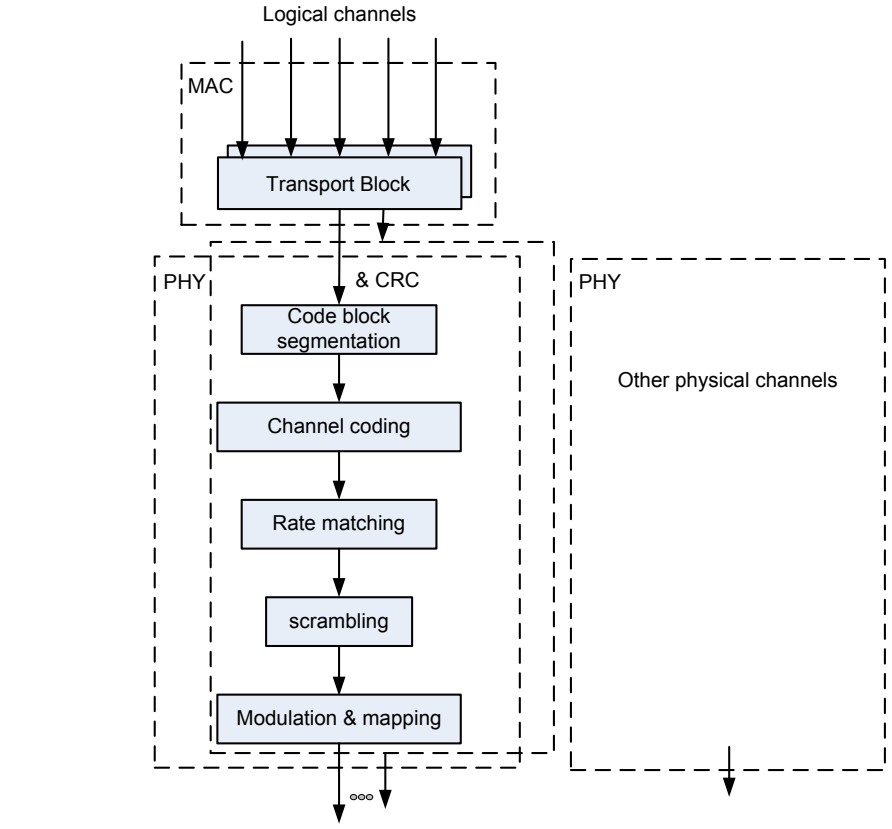
3.5.1 LTE transmitter

OFDM transmitter with multiple antennas consists of the functional blocks shown in Fig. 3.19. The transport channel input is first channel coded to a stream of coded bits. Each stream is first scrambled, modulated to a set of complex-valued modulation symbol alphabets and further mapped to one or several spatial layers. The modulation symbols may be precoded on each layer before transmission from the antenna ports. Multiantenna diversity transmission of the conventional STTD converts to the form of space-frequency block code for LTE. This is studied as bit interleaved coded modulation with space time block codes e.g. in [114] and [115].)

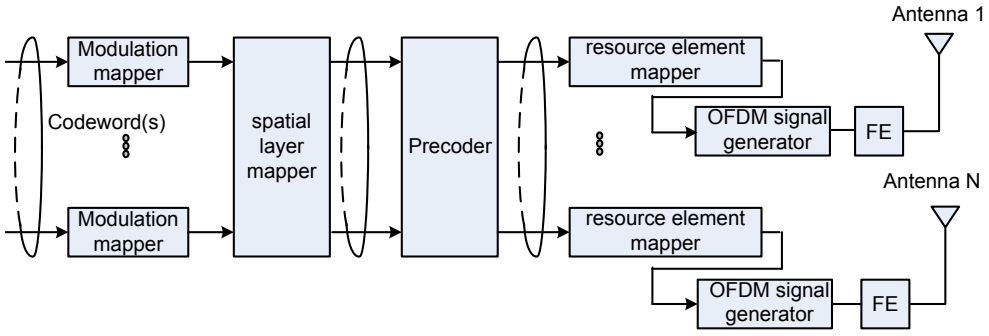
For each channel coded word κ, the block of bits in a code word is transmitted on the physical channel in one subframe. The block of bits are scrambled prior to modulation in order to provide bit pattern dispersion and on the other hand better isolation to the neighbour cell interference.

Up to two channel coded words can be transmitted in one subframe i.e. $\kappa \in \{0, 1\}$. In multiantenna transmission, each code word (κ_0 alone or both κ_0 and κ_1) may be transmitted from any number of n_t antenna ports, $n_t = 1, .., N_T$ with the number of transmit antennas N_T.

For each code word, the block of scrambled bits is modulated using one of QPSK,



(a)



(b)

Figure 3.19: Overview of the physical channel processing for the OFDMA. (a) Mapping of transport block(s) of the transport channel to a physical channel shown in (b) for multiple transmit antennas, which allows transmission of a single or multiple codewords.

16QAM, 64QAM alphabets resulting in a block of complex-valued modulation symbols. After splitting the block of symbols to spatial layers for multiantenna transmission and after precoding, the block of complex-valued symbols for each antenna port (n_t) are mapped in sequence to the subcarrier symbols i.e. frequency-time resource elements (k, l). This

mapping is done in increasing order of first frequency (k) then time (l), so that frequency diversity is achieved without a separate interleaving process.

Special resource element positions of the subframe are reserved for the Reference Symbols (RS) and synchronization sequences. The design of RS mapping patterns and the RS sequences have a large impact to the system performance due their overhead and due their large impact to channel estimation, equalization and handover measurements [51].

The continuous time signal $S_l^{(n_t)}$ on antenna port (n_t) in OFDM symbol l in a downlink slot is defined in [105] as (3.6),

$$S_l^{(n_t)}(t) = \sum_{k=-\lfloor N_{BW}^{DL}/2 \rfloor}^{-1} a_{k^{(-)},l}^{(n_t)} \cdot e^{j2\pi k \Delta f_{sc}(t - N_{CP,l}T_s)} + \sum_{k=1}^{\lceil N_{BW}^{DL}/2 \rceil} a_{k^{(+)},l}^{(n_t)} \cdot e^{j2\pi k \Delta f_{sc}(t - N_{CP,l}T_s)} \quad (3.6)$$

for $0 \leq t < (N_{CP,l} + N) \times T_s$, where $k^{(-)} = k + \lfloor N_{BW}^{DL}/2 \rfloor$, $k^{(+)} = k + \lfloor N_{BW}^{DL}/2 \rfloor - 1$, $N = 2048$, $N_{BW}^{DL} = N_{PRB}^{DL} N_{SC}^{PRB}$, the subcarrier spacing is $\Delta f_{sc} = 15kHz$ and $a_{k,l}^{(n_t)}$ is the complex value content of resource element (k, l) on antenna port n_t.

OFDMA signal is known to have high peak-to-average power ratio at high probability due to large number of subcarriers, where amplitude modulated symbols (QAM) are present at the same time. This phenomenon is not very critical in the eNodeB transmitter, where high linearity of the power amplifier extends over a large dynamic range. Anyway, clipping of the peaks will convert to wideband noise, part of which will remain as inband noise. The higher is the clipping threshold, the lower is the probability of exceeding the threshold, and the smaller is the effective noise power of the clipped signal.

In the UE transmitter, the constraints are very different, because the transmitted average power is strictly limited (by the emission law) and because the dynamic range of the power amplifier cannot be increased without a large impact to the *Cubic Metric*, which is proportional to the 6^{th} power of the signal amplitude [116][117]. For uplink transmission, a scheme with a low peak-to-average power ratio is always preferred [118][119], because it shows a lower Cubic Metric, which impacts on power efficiency (power consumption) and physical design (cubic dimension) of the device. The lower peak-to-average-power ratio allows operating at higher power inside the given average power constraints, which increases cellular coverage or increases throughput at the cell edge. Vice versa, offering coverage at desired BLock Error Rate (BLER) for a given amplitude modulated (QAM) symbol sequence, low peak-to-average power ratio consumes less transmit power, increases

coverage efficiency and contributes to a longer battery activity time.

The main reason for the preference of a single carrier transmission in uplink is the lower peak-to-average power ratio. In LTE, uplink transmission scheme is precoded OFDMA, where *DFT precoding* actually generates a single carrier waveform instead of the multicarrier one. Single Carrier FDMA (SC-FDMA) transmission of LTE is motivated by operating in the transform domain that enables freqeuncy domain equalization similar to that of the multicarrier OFDMA [120][121][118][10][102].

DFT precoding has further benefits, as DFT spreads each and every modulated symbol to the butterfly outputs over the full (frequency domain) subband, which is allocated to a user. DFT spreading is a powerful operation against narrowband (or coloured) interference that can harm OFDM transmission on a narrow subcarrier [100]. For its precoding choice, LTE uplink transmission scheme is also known as DFT-spread-OFDMA [122].

In LTE, uplink multi-user multiplexing is done by the IFFT that is executed over the full system bandwidth by all the UEs, see e.g. [122]. Each UE maps the modulated symbol block by DFT to the uniquely reserved (allocated) frequency bins, so that the transmissions of different UEs are orthogonal in frequency. In order the IFFT to keep the orthogonality, transmit timing of the UEs have to be sufficiently accurate, i.e. the transmissions by different UEs at the receiver need to have timing accuracy much smaller than the CP. This requirement is not too demanding, as the OFDM symbol period and its cyclic extensions are lengthy in time (e.g. compared to a spread chip period of WCDMA). OFDM transmitter and receiver are computationally very efficient especially because of their frequency domain operations.

In LTE standard, DFT precoding was selected because of its favourable properties [123][119]. Yet, the drawback of a generic DFT is its computational complexity for certain transform sizes. LTE channel allocation works in multiples of 12 subcarriers (one PRB), and in principle any allocation of $k \cdot 12$ would be needed. The analysis e.g. in [38] shows that the complexity increases, when radix set of the DFT is increased from two (FFT). By adding more radixes, a tighter space is created and more PRB options become feasible for allocation. Radix of 2, 3 and 5 at least are known to have efficient, well optimized, implementations [8]. On the other hand, they already provide a tight space of PRBs for small bandwidth allocations (a low number of PRBs). The granularity of PRBs for wider band allocations i.e. large values of k is not as tight, but it is well sufficient. Increasing the radix set e.g. by radix 7 would bring some new lattice points to the feasible set, especially

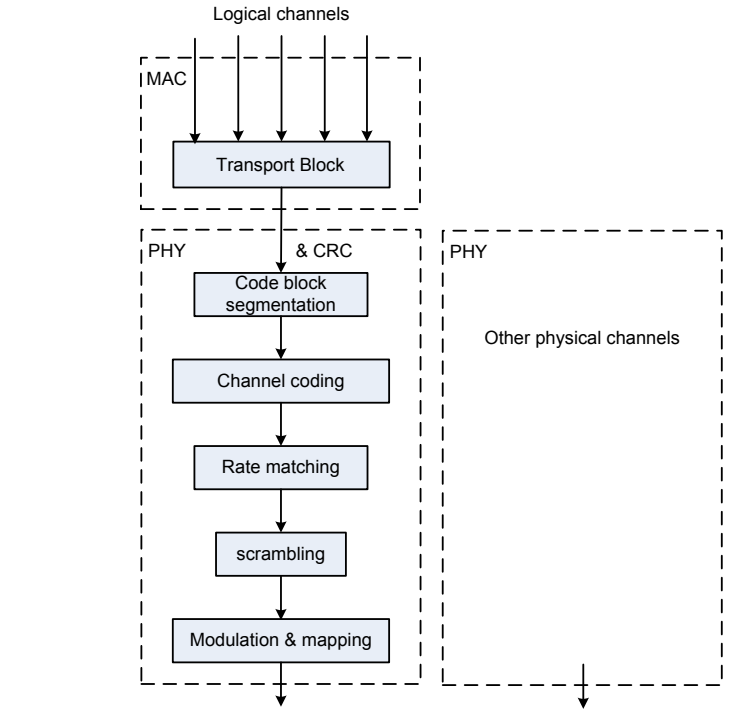
for the large values of k. This however adds DFT complexity significantly and does not pay much in added PRB granularity, because there exists a large degree of scheduling freedom anyway, when multiple users are present. Dropping radix 5 from the set is not feasible, because full band allocations of {72, 180, 300, 600, 900, 1200} subcarriers obviously require that.

3.5.2 Transform precoding

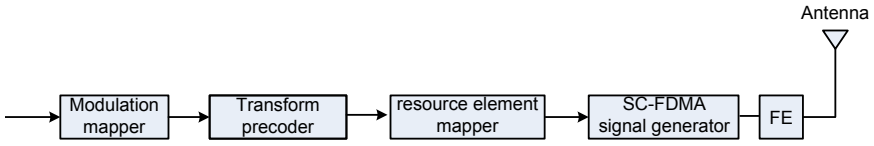
SC-FDMA transmitter with functional blocks is shown in Fig. 3.20. Each transport channel bit stream is first channel coded, scrambled and modulated to a set of complex-valued modulation symbol alphabets. The modulation symbols are DFT precoded and further mapped to the physical resource elements (subcarrier resources) to generate a complex-valued time domain signal. The block of bits is transmitted on the physical uplink shared channel in one subframe. The block of bits is scrambled with a UE-specific scrambling sequence prior to modulation. The modulation uses one of QPSK, 16QAM, 64QAM alphabets resulting in a block of complex-valued symbols. The block of complex-valued symbols $m(0), .., m(M_{symb-1})$ is divided into $(M_{symb})/M_{sc}^{PUSCH}$ sets, each corresponding to one SC-FDMA symbol. M_{sc}^{PUSCH} is the number of allocated subcarriers to a user. SC-FDMA symbols are isolated in time-domain from each other by the cyclic extension of signal waveforms to provide interblock orthogonality in a dispersed channel similar to OFDM symbols. Transform precoding is applied according to [105] as (3.7).

$$
\begin{aligned}
z(l \cdot M_{SC}^{PUSCH} + k) = \frac{1}{\sqrt{M_{SC}^{PUSCH}}} \sum_{i=0}^{M_{SC}^{PUSCH}-1} m\ (l \cdot M_{SC}^{PUSCH} + i) \cdot e^{-j\frac{2\pi ik}{M_{SC}^{PUSCH}}} \\
k = 0, .., M_{SC}^{PUSCH} - 1 \\
l = 0, .., M_{symb}/(M_{SC}^{PUSCH} - 1)
\end{aligned}
\tag{3.7}
$$

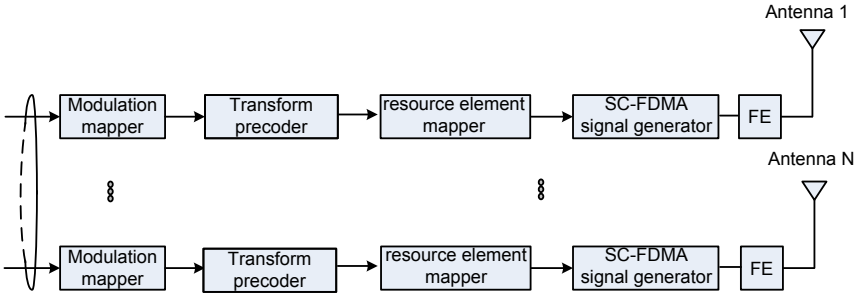
resulting in a block of complex-valued, precoded symbols $z(0), .., z(M_{symb-1})$. The block of complex-valued symbols on the Physical Uplink Shared Channel (PUSCH) are multiplied with the amplitude scaling factor β_{PUSCH}. The symbols are mapped to the frequency-time resource elements (k, l) corresponding to the subcarriers of the allocated physical resource blocks. The mapping is in increasing order of first frequency (k) then time (l), so that frequency diversity is achieved without a separate interleaving process. Special symbol positions in PUSCH are reserved for Demodulation Reference Symbols (DMRS).



(a)



(b)



(c)

Figure 3.20: Overview of the physical channel processing for the SC-FDMA or precoded OFDMA. (a) Mapping of a transport block of the transport channel to a physical channel shown in (b) for a single transmit antenna and in (c) for multiple transmit antennas. Precoded mapping and spatial layer multiplexing is omitted in the uplink transmission.

The continuous time signal $S_l(t)$ in SC-FDMA symbol l in an uplink slot is defined in [105] by (3.8),

$$S_l(t) = \sum_{k=-\lfloor N_{BW}^{UL}/2 \rfloor}^{\lceil N_{BW}^{UL}/2 \rceil - 1} a_{k^{(-)},l} \cdot e^{j2\pi(k+1/2)\Delta f_{sc}(t - N_{CP,l}T_s)} \tag{3.8}$$

for $0 \leq t < (N_{CP,l} + N) \times T_s$, where $N = 2048$, $N_{BW}^{UL} = N_{PRB}^{UL} N_{SC}^{PRB}$ determines the adaptive transmission bandwidth, the subcarrier spacing is $\Delta f_{sc} = 15kHz$ and $a_{k,l}$ is the complex value content of resource element (k, l). N_{BW}^{UL} takes values that are multiples of the PRB size of N_{SC}^{PRB} and can be expressed as radix $\{2,3,5\}$ so that $N_{BW}^{UL} = 2^{r1}3^{r2}5^{r3} \cdot N_{SC}$, where $r_1 \geq 0, r_2 \geq 0, r_3 > 0$.

3.5.3 LTE reference receivers

OFDMA and SC-FDMA signals, as block transmissions, can generically be received by the FFT algorithm. The receiver sample window is positioned to contain the dominant multipath propagated signal components. The filtering window may partly consist of the CP period, however, only the samples of the symbol (without CP) are transformed to the frequency domain for baseband processing of the channels. Oversampling is required in the receiver.

OFDMA receiver includes time and frequency synchronization, channel estimation, frequency domain equalization and channel decoding. Subcarrier demodulator and symbol demapper are fundamental blocks as well. The receiver algorithms are revisited in Chapter 4.

The signal model of OFDM block transmission receiver is,

$$r[n] = \sum_{k=0}^{\infty} h[n]\ s[n-k], \tag{3.9}$$

where $s[n]$ is the transmit symbol sequence or the signal vector $\mathbf{s} = [s_1\ s_2\ s_3 \cdots s_n]$ and $r[n]$ is the receive symbol sequence or the symbol vector $\mathbf{r} = [r_1\ r_2\ r_3\ \cdots\ r_n]$ over a time-variant, dispersive, frequency selective channel (doubly spread or doubly dispersive wide-sense stationary uncorrelated scattering (WSSUS) channel) h [84]. The expected normalized transmit power is $E\{\mathbf{s}\mathbf{s}^H\} = \sigma_s^2 \mathbf{I}_N$ with the identity matrix $\mathbf{I}_N$. For a multiantenna transmitter with N_T antennas, the input signal matrix becomes $\mathbf{s} = [\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \mathbf{s}^{(3)}, \ldots, \mathbf{s}^{(N_T)}]^T$. For a multiantenna receiver with N_R antennas, the received signal matrix becomes $\mathbf{r} = [\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, \mathbf{r}^{(3)}, \ldots, \mathbf{r}^{(N_R)}]^T$. LTE with multiantenna transmission is analysed recently in [124].

For a block transmission system (as OFDMA and SC-FDMA in LTE) the symbol blocks, with cyclic convolution and cyclic extension longer than the channel, will truncate and repeat. Therefore, once windowing and sampling are proper, the receiver may be implemented by the FFT.

The block transmission of input sequence $\mathbf{s}$ of symbol block size [n] can be written for the received signal as (3.10).

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{v}, \tag{3.10}$$

where the noise vector $\mathbf{v} = [v_1 \ v_2 \ v_3 \cdots \ v_n]$ received by multiple antennas is $\mathbf{v} = [\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \mathbf{v}^{(3)}, \ldots, \mathbf{v}^{(N_R)}]^T$ and its expected value is $E\{\mathbf{v}\mathbf{v}^H\} = \sigma_v^2\mathbf{I}$. With transform precoding $\widetilde{\mathbf{F}}$ and yet inserting the CP into the precoding vector we get,

$$\begin{pmatrix} \mathbf{I}_{CP} \\ \mathbf{I}_N \end{pmatrix} \widetilde{\mathbf{F}} = \mathbf{F}$$

we get,

$$\mathbf{r} = \mathbf{H}\ \mathbf{F}\ \mathbf{s} + \mathbf{v} \tag{3.11}$$

Here, $\mathbf{H}$ is $N \times N$ Toeplitz circulant matrix. The fundamental property of a circulant matrix is that it can be diagonalized as,

$$\mathbf{H} = \mathbf{W}\ \mathbf{\Lambda}\ \mathbf{W}^H \tag{3.12}$$

with $\Lambda(i,j) = \sum_{l=0}^{L} h(l) e^{\ -j \cdot \frac{2\pi i l}{N}}$ of the FFT transform. $\mathbf{\Lambda}$ is a diagonal matrix, whose diagonal elements are the values of the FFT calculated over N sinusoids. $\mathbf{W}$ includes entries $W(c,d) = \frac{e^{\ j \cdot 2cd/N}}{\sqrt{N}}$, and because $\mathbf{W}$ is unitary, $\mathbf{W}^{-1} = \mathbf{W}^H$ holds. Therefore, the only inversion needed in this reference receiver to calculate $\mathbf{H}^{-1}$ is the inversion of $\mathbf{\Lambda}$, which is diagonal. The multiplication by $\mathbf{W}$ and $\mathbf{W}^H$ can be implemented by the FFT.

Therefore, in the diagonalized form of (3.13),

$$\mathbf{r} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^H\ \mathbf{F}\mathbf{s} + \mathbf{v} \tag{3.13}$$

multiplying by $\mathbf{W}^H$ from the left side gives the MMSE equalized receive signal vector $\mathbf{z}$.

$$\mathbf{z} = \mathbf{\Lambda}\ \mathbf{W}^H\ \mathbf{F}\mathbf{s} + \mathbf{W}^H\ \mathbf{v}, \tag{3.14}$$

where $\mathbf{z} = \mathbf{W}^H\ \mathbf{r}$ gives the estimate of the transmit signal $\widehat{\mathbf{s}}$. Having precoding $\mathbf{F} = \mathbf{W}$, we get

$$\mathbf{z} = \mathbf{\Lambda}\mathbf{s} + \mathbf{w},^2 \tag{3.15}$$

and since $\mathbf{\Lambda}$ is diagonal we get $z_{sc} = \Lambda(sc : sc)\ s_{sc} + w_{sc} = H_{sc}\ s_{sc} + w_{sc}$ with subcarrier index $sc = 1, .., N_{sc}$. In this manner, OFDM converts the time dispersive channel into N_{sc} flat fading (narrowband) channels. The equalizer for OFDM is simple and it only requires a single complex channel coefficient per subcarrier [125].

$$MMSE:\ \widehat{s}_{sc} = \frac{\sigma_s^2 H_{sc}^*}{|H_{sc}|^2\sigma_s^2 + \sigma_v^2}\ z_{sc} \tag{3.16}$$

Ideally and without noise, the signal estimate is $\widehat{s}_{sc} = \frac{1}{H_{sc}} z_{sc}$. With noise, the division by H_{sc} may additionally result in *noise enhancement* while equalizing the channel for the wanted signal.

The imperfections of OFDM reference receiver are due to inaccuracy of frequency synchronization and phase noise. Accuracy of time synchronization and timing advance are not that critical because of the long symbol duration and CP. In LTE, subcarrier dimensioning is practical providing robustness for small frequency errors and Doppler tolerance, even in high velocity channels. Error due to classical Doppler spread can be kept below about +/-2% of subcarrier spacing. (For other OFDM systems like DVB and some WiMax profiles that use narrow subcarrier spacing of order 1 to 2 kHz, frequency accuracy and Doppler tolerance are critical issues indeed.) Accurate frequency operation is important to keep the *Error Vector Magnitude (EVM)* small, especially so for the high order modulations.

IFDMA and subcarrier distributed OFDMA mappings would seem to benefit of maximal frequency diversity for all the multiplexed payloads (as was the choice for WiMax). However, due to imperfect timing between users and due to synthesizer circuitry, the receiver experiences a noise floor between the distributed frequency bins, instead of the perfect sinc function of the OFDM transform. In presence of noise floor with non-perfect power control between different users, the SINR actually gets limited by the phase noise

[2] $\mathbf{w}$ is the equivalent noise vector with statistical properties equal to $\mathbf{v}$

floor of the other user received at higher signal power. The localized or clustered localized mapping of multiple adjacent frequency bins provide much lower phase noise floor between the disjoint frequency bins, and hence the other user does not become that limited compared to subcarrier distributed mappings. The frequency diversity of localized and clustered localized mapping can equally be achieved by frequency hopping, by reallocating retransmission resources or by PRB-level distribution.

The power spectral density of phase noise η_{ph} may be estimated by a Lorenzian distribution in (3.17).

$$\eta_{ph} = \frac{1}{\pi} \cdot \frac{\left(\int_{f_{3dB}} S(f)df + \int_{-f_{3dB}} S(f)df \right) \cdot f_{3dB}}{f^2 + f_{3dB}^2} \tag{3.17}$$

$$S(f) = \mathrm{FFT}\Big\{ \int_{-\infty}^{\infty} S^*(x)(t+x)\ dx \Big\}, \tag{3.18}$$

where signal power S is integrated over the signal representation in frequency domain, which is the Fourier transfom of the signal autocorrelation function in (3.18) instead of the signal itself, because a signal with non-zero mean power is not square integrable. The synthesizer bandwidth is f_{3dB} [98].

3.6 Alternative transmission techniques

3.6.1 Multicarrier waveforms

OFDM is the prevalent technique to generate multicarrier waveforms, however many other alternative techiques exist in the literature [126]. Some of the most attractive alternatives are based on filterbank transforms, also known as lapped transforms [127]. Lapped orthogonal transforms execute block transform coding, which preserve orthogonality both in time and in frequency over a block of symbols. In lapped transforms, time domain symbol block is localized by a finite pulse shape waveform other than a rectangular window. OFDM, on the contrary, uses a rectangular window for pulse shaping, which suffers from multipath propagation that turns to severe interblock interference, unless successive symbol blocks are isolated by a guard interval longer than the delay spread of the channel, see Section 3.4.1. A lapped transform, instead, can sustain situation, where the successive symbol blocks overlap, because their waveforms keep orthogonality also in the time domain sliding window. In a multipath delay dispersive channel, the orthogonality of overlapping

symbol blocks can still be restored in the receiver, whenever the channel is equalized perfectly and if the filterbank has the perfect reconstruction property [127][128][129]. Hence, lapped transform filterbanks provide a natural extension of a regular block transform.

OFDM is favoured in practical implementations for its computational simplicity and easy equalization. The guard interval adds overhead, but it can be kept percentually small by a proper system design of subcarrier spacing and symbol length. Filterbanks are studied as a challenger technique, whose computational complexity, especially for equalization, is critical. Benefits of filterbanks include lack of cyclic extension overhead and possibility to design wider bandwidth subcarriers compared to OFDM [130]. Wider subcarrier bandwidth reduces computational complexity, improves tolerance for narrowband interference and further increases tolerance to frequency inaccuracy and Doppler shift. Due to properties of lapped transforms, training sequence design and modulation technique will differ from OFDM. Briefly, traininig sequence symbols cannot be multiplexed with data symbols in a scattered manner and the modulation alphabet has to be pulse amplitude modulation rather than quadrature amplitude modulation because of the required phase properties.

The filterbank transforms, their time-frequency properties, their reconstruction properties and their implementation techniques impact largely on the obtained performance. The filterbank techniques include; Gaussian basis functions e.g. Isotropic Orthogonal Transform Algorithm (IOTA) based on FFT transform and Offset QAM modulation, better known DFT filterbanks [127][131], modulated lapped transforms, extended lapped transforms, cosine modulated filterbanks [132][133][98] (known from Digital Subscriber Lines, DSL, xDSL, vDSL) and cosine/sine-modulated (exponential) filterbanks. In Section 3.6.2, a filterbank solution based on cosine/sine modulated filterbank is introduced because of its favourable properties for a radio channel. Unlike single-sided DFT and IOTA transforms, which only allow modulation of the real-branch of the signal, the cosine/sine-modulated filterbank can provide complex modulated symbols that have information contents on both the I-branch and Q-branch. This obviously suits well for radio transmission, which happens over a complex coefficient radio channel. A transmission technique like this will increase spectral efficiency.

3.6.2 Modulated and extended lapped transforms

Modulated lapped transforms are based on cosine or sine modulation of a prototype filter $h(n)$, where the length of the basis functions $L = 2M$, where $2M$ is the number of sub-

carriers of the transform. Extended lapped transforms are similar to modulated lapped transforms, however with longer basis functions of arbitrary length $L > 2M$. Extended lapped transforms provide more flexibility to the practical system design, where subcarrier bandwidth and pulse shaping functions are critical.

An attractive filterbank design for radio communications is based on both cosine and sine modulation, hence also referred to as exponential or complex modulated filterbank ($e^{j\ \phi} = \cos(\phi) + \ j\ \sin(\phi)$) [134][135][136][36]. This structure preserves complex signal presentation, and it allows the use of both I-branch and Q-branch for information transmitted over the complex coefficient radio channel. Baseband modulation happens per subcarrier by the chosen alphabet (PAM or QAM) and it forms the modulated input vector for the filterbank structure. The structure is shown in [34], where transmultiplexer provides complex I/Q-symbol pair by cross-connecting a cosine-modulated synthesis filterbank and a sine-modulated synthesis filterbank. In the receiver, oversampling is necessary and the receiver has to consist of duplicated cross-connected cosine-modulated and sine-modulated analysis filterbanks so that both the I-branch and the Q-branch will have a complex symbol representation. The information of the symbol stream is fully present in the real-part of the I-branch and in the real-part of the Q-branch respectively. However, processing the whole complex received signal (both for the I-branch and for the Q-branch) allows equalization errors be cumulated solely to the imaginary part. As the imaginary part is finally discarded and purely the real-part is restored and demodulated, the cosine/sine-modulated filterbank enables perfect equalization in theory. In practice, equalization suffers, as always, from slight degradation of channel estimation inaccuracy. Channel equalization for a cosine/sine-modulated filterbank is described and analyzed e.g. in [34][36].

3.6.3 Single carrier waveforms

Similar to a frequency domain SC-FDMA receiver based on Fourier transform, it is possible to have a filterbank solution for frequency domain equalization of a single carrier transmission. This structure is based on cross-connected cosine and sine modulated analysis filterbanks, as shown in [134] which executes the transform of complex I-branch and Q-branch signals to the complex frequency domain subbands. After the subband signals are equalized, solely their real-part is restored and transformed back to time domain, again by cross-connected cosine and sine modulated synthesis filterbanks. Channel equalization and mitigation of narrowband interference in the cosine/sine-modulated filterbank structure is

described and analyzed e.g. in [33][35][37].

There exists yet an alternative way of creating the SC-FDMA waveform, instead of transform domain processing. We call this structure Single Carrier implementation using Fractional Interpolation (SCiFI), and it is prsented in [38]. SCiFI is based on equivalent matrix representation of the transform, matrix multiplication and fractional interpolation. This structure provides a generalization that does not need DFT and avoids DFT radix set selection. Further, SCiFI can be made computationally comparable to the overall DFT-IFFT signal generation, if channel filtering is included to the complexity comparison. On the other hand, SCiFI suffers slightly from an approximation error that is inherent to the polynomial interplolator. This approximation error is however low and it can be scaled well below the noise floor of the radio channel.

Chapter 4

Receiver algorithms

In Chapter 4, the essential receiver algorithms are represented and their signal processing formulation is derived in key contents. First, the fundamentals of channel estimation and signal detection by the maximum likelihood decoding in Section 4.2.1 and its computational simplifications in Sections 4.2.2 are introduced. Next, channel equalization with the algorithms for a Zero Forcing (ZF) equalizer and for a Linear Minimum Mean-Squared Error (LMMSE) equalizer are given for the spread spectrum chip sequences (WCDMA/HSPA) in Section 4.3. In Section 4.5, these algorithms are further extended to the multiantenna transmissions. However, before that, an important discussion on multiantenna reception is covered in Section 4.4. Multiantenna reception allows signal combining gains and interference rejection or interference cancellation opportunities, despite of the number of transmit antennas. The receiver algorithms, without loss of generality, can be tailored for different transmission techniques, and hence many of them are applicable to both WCDMA, WCDMA/HSPA and LTE. In this Chapter, the repetition of each algorithm for each and every transmission system is avoided, and the most relevant representation is given instead. Yet, the special characteristics of the transmission technique is that makes the use of a specific algorithm attractive for it, whereas another algorithm remains in a marginal role in practise. With this reasoning for LTE, the FFT equalizer is given first in Section 4.6.1, and next, the most powerful QRDM and turbo iterative receivers are derived in Section 4.6.2 and Section 4.6.3 respectively. The algorithms given in this Chapter are selectively in concrete use for the system performance simulations in Chapter 9.

4.1 Channel estimation and tap solver

Receiver algorithms require estimation of the channel, the signal and interference as a pre-requisite to equalize and decode the expected transmitted symbol sequence [137]. The WCDMA, WCDMA/HSPA and LTE provide reference symbol sequences (pilot signals) for pilot-aided channel estimation i.e. to solve the complex channel coefficients of the channel filtering taps. In WCDMA and WCDMA/HSPA, the pilot is a time-continuous direct-spread wideband sequence. In LTE, the pilot sequence is modulated to subcarrier symbol elements in dense grid of frequency and time domain positions. These signal sequences are specific to a cell and are known after the cell search procedure, which reliably correlates and detects these sequences from noise.

The received signal power can be estimated by integrating the transmitted sequence over the reference symbol periods. The interference is measured as the total received power over the reference symbol periods subtracted by the power of the desired signal. In case these measures additionally use integration over the data symbols, a data-aided estimation is in question. The channel estimates and power measurements are needed for channel equalization and to calculate the SINR metric.

4.2 Signal detection

4.2.1 ML receiver

In general, the maximum likelihood detector computes the squared Euclidean distance between the received vector and all possible transmitted vectors (of the training set i.e. the pilot sequence). The detected vector is the one with the minimum Euclidean distance. The ML detector is given e.g. in [138][139] by

$$\widehat{\mathbf{s}} = \arg\min_{\mathbf{s}} \left\| \mathbf{r} - \mathbf{\Lambda}\ \mathbf{s} \right\|^2, \tag{4.1}$$

where $\mathbf{\Lambda} = \mathbf{H}\ \mathbf{F}$ includes the complex channel matrix $\mathbf{H}$ and precoding $\mathbf{F}$, if present.

The receiver detects the symbol sequence $\mathbf{s}$ in a received signal with noise plus interference with the minimum probability of detection error, thus the detector maximises the *a posteriori* probability for sequence $\widehat{\mathbf{s}}$,

$$\widehat{\mathbf{s}} = \arg\max_{\mathbf{s}} \left\{ p(\mathbf{s}|\mathbf{r}) \right\} \tag{4.2}$$

$$\widehat{\mathbf{s}} = \arg\max_{\mathbf{s}} \left\{ \frac{p(\mathbf{r}|\mathbf{s}) \cdot P(\mathbf{s})}{P(\mathbf{r})} \right\}, \tag{4.3}$$

which is the maximum likelihood with $p(\cdot)$ as the joint probability density function and $P(\cdot)$ as the probability of an event. So, $p(\mathbf{r}|\mathbf{s})$ is the joint probability of the event that vector $\mathbf{r}$ was received in the condition that vector $\mathbf{s}$ was transmitted. $P(\mathbf{s})$ is the probability that vector $\mathbf{s}$ was actually transmitted. $P(\mathbf{r})$ is the probability that vector $\mathbf{r}$ was received.

As the computation of metrics for all the possible transmit vectors is heavy, simplifications of the ML algorithm are important, and some examples of them are presented in Sections 4.2.2 and 4.6.2. The Rake receiver (see Section 3.3.3) is also an ML receiver (optimum receiver) in the conditions of additive white Gaussian noise channel but not in a multipath fading nor correlating channels.

4.2.2 M receiver

M receiver is so called reduced maximum likelihood search detector, which instead of the comprehensive tree search algorithm of the ML detector, applies tree search pruning algorithms. In the M-algorithm, at each level of the tree all the possible symbol constellations are kept, but a reduced set of survivors of the detection path with the best metrics of the log-likelihood-ratio (LLR) are passed to the next search round. In order to get good decoding results, soft symbol output must be generated at each round. It is known in the literature [138][139] that max-log-map approximation can be exploited to compute the required LLR [140].

4.3 Channel equalization

Channel equalization can be done by conventional means such as the zero forcing or MMSE algorithms. Equalization of multi-antenna signals is an extension thereof. Here, equalization by LMMSE is shortly visited for the spread spectrum signals (WCDMA/HSPA) and for the block transmission OFDM signals (LTE).

The signal model for the impulse response of the channel between transmit antenna n_t and receive antenna n_r for a pulse shaped chip sequence,

$$h^{(n_r)}(t) = \sum_{k=0}^{L_h-1} h^{(n_r)}[k] \; p_{rrc}(t - kT_c), \tag{4.4}$$

where p_{rrc} is the pulse shaped root-raised cosine (rrc) spectrum and $L_h T_c$ is the length of the multipath delay spread of channel realization h. In (4.4) the channel coefficients are discretized to chip intervals kT_c including at most L_h taps. (The channel taps actually represent the discretized coefficients of the channel power delay profile. In a typical urban 5 MHz channel about six dominant taps are sufficent to describe the power delay profile, whereas a highly selective wideband channel may require at least twenty taps). The received signal is sampled at integer chip intervals $t = nT_c$ with an oversampled multiple of chip rate $1/T_c$.

The signal vector received at the $(n_r)^{th}$ antenna after convolution with the channel matched filter is,

$$r^{(n_r)}[n] = \sum_n s[n] \cdot \widehat{\mathbf{h}}^{(n_r)}(t - nT_c), \tag{4.5}$$

where the received signal $r^{(n_r)}[n]$ is sampled at instants $t = nT_c$ with the chip rate $1/T_c$. $s[n]$ is the transmit symbol sequence, or a symbol vector $\mathbf{s}$), to be equalized in the receiver. Each element s_i in the symbol vector $\mathbf{s}$ has the actual length of the chip sequence N_{ch}, and each chip is represented by a complex sample. The equalizer length is L_f for the channel from transmit antenna n_t to receive antenna n_r, where the estimated channel $\widehat{\mathbf{h}}^{(n_r)}$ is of length L_h. In the literature [141], it is shown that increasing the equalizer length increases the performance, and the equalizer length should preferably be much longer than the channel ($L_f >> L_h$).

Receiving the signal sequence from a single transmit antenna n_t by multiple receive antennas $n_r = 1, .., N_R$ operates over the estimated channels $\widehat{\mathbf{h}}^{(n_r)}$. Channels $\mathbf{h}^{(n_r)}$ are often assumed independent in analytical calculations, but in reality interpath correlation appears.

4.3.1 Zero forcing equalizer

The synthesis of the ZF equalizing filters $g^{(n_r)}$ by N_R antennas over a sequence of [n] symbols is based on minimizing the output noise power subject to the ZF constraint. That is, minimize according to (4.6).

$$\min_{g^{(n_r)}[n]} \sum_{n=0}^{L_f-1} \left\{ |g^{(1)}[n]|^2 + |g^{(2)}[n]|^2 + \cdots + |g^{(N_R)}[n]|^2 \right\} \tag{4.6}$$

subject to:

$$\left\{h^{(1)}[n] * g^{(1)}[n] + h^{(2)}[n] * g^{(2)}[n] + \cdots + h^{(N_R)}[n] * g^{(N_R)}[n]\right\} = \delta[n - D], \tag{4.7}$$

where D is an integer chip delay, L_f is the equalizer length and N_R is the number of receive antennas, $n_r = 1, .., N_R$. In matrix form, the ZF constraint may be expressed as

$$\mathbf{H}_{g\_ZF} = [\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \mathbf{H}^{(3)}, \ldots, \mathbf{H}^{(N_R)}][\mathbf{g}^{(1)}, \mathbf{g}^{(2)}, \mathbf{g}^{(3)}, \ldots, \mathbf{g}^{(N_R)}]^T = \boldsymbol{\delta}_D, \tag{4.8}$$

where $\boldsymbol{\delta}_D$ is a matrix of all zeros, but 1 in $(D-1)^{th}$ unit delay position. $\mathbf{H}^{(n_r)}$ is the $((L_h - 1) + L_f) \times L_f$ convolution matrix.

The zero forcing solution is

$$\mathbf{g}_{ZF} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1} \cdot \boldsymbol{\delta}_D \tag{4.9}$$

4.3.2 LMMSE chip equalizer

LMMSE chip equalizer is based on the minimum mean-squared error criterion in (4.10).

$$\arg\min_{\mathbf{g}} E\Big\{ \mid \mathbf{g}^T\big(\mathbf{H}^T\mathbf{s} + \mathbf{v}\big) - \boldsymbol{\delta}_D^T\ \mathbf{s} \mid^2 \Big\}, \tag{4.10}$$

where $\mathbf{s}$ is the transmitted signal vector of length $(L_h - 1) + L_f$. $\mathbf{v}$ is the noise vector of length L_f, zero mean and variance σ_v^2. For spread, scrambled sequences, the codes are random independent and identically distributed (i.i.d.) sequences. Thus, covariance of the signal is $E\{\mathbf{s}\mathbf{s}^H\} = \sigma_s^2\mathbf{I}$.

The LMMSE equalizer becomes,

$$\begin{aligned} \mathbf{g}_{MMSE} &= \sigma_s^2\big\{\sigma_s^2\ \mathbf{H}^H\mathbf{H} + \sigma_v^2\mathbf{I}\big\}^{-1}\mathbf{H}^H\boldsymbol{\delta}_D \\ &= \sigma_s^2\mathbf{H}^H\big\{\sigma_s^2\ \mathbf{H}\mathbf{H}^H + \sigma_v^2\mathbf{I}\big\}^{-1}\boldsymbol{\delta}_D, \\ &\propto \mathbf{H}^H\boldsymbol{\mathcal{R}}^{-1}\boldsymbol{\delta}_D \end{aligned} \tag{4.11}$$

where $\mathbf{H}\mathbf{H}^H = \boldsymbol{\mathcal{R}}^{-1}$ with the channel covariance matrix $\boldsymbol{\mathcal{R}}$, and the solution uses Cholesky decomposition. The optimal value is found at,

$$d = \arg\min diag(\mathbf{H}\mathbf{H}^H + \sigma_v^2\mathbf{I})^{-1}. \tag{4.12}$$

At low SNR, the LMMSE of (4.11) can be written,

$$\mathbf{g}_{MMSE} \propto \mathbf{H}^H \boldsymbol{\delta}_D, \tag{4.13}$$

which (4.13) is comparable to Rake. At high SNR, the LMMSE of (4.11) can be written,

$$\mathbf{g}_{MMSE} \propto \mathbf{H}^H \left\{\mathbf{H}\mathbf{H}^H\right\}^{-1} \boldsymbol{\delta}_D, \tag{4.14}$$

which (4.14) is comparable to ZF without the noise enhancement property. In the LMMSE equation, Cholesky decomposition acts as a tap solver for a well-conditioned system of linear equations. For ZF, the matrix inversion includes ill-conditioned channel matrix instead.

4.4 Multiantenna receivers, combining and interference rejection

4.4.1 Maximum Ratio Combining

Maximum Ratio Combining of the signal received by multiple antennas is well known for spread spectrum signals. Here, the equations are shown for a frequency selective OFDM transmission over a subcarrier index (sc). SINR of the combined signal, transmitted from antenna n_t and received by antennas $n_r = 1, .., N_R$ in a single tap channel per subcarrier is,

$$SINR_{sc} = \frac{P_{sc}^{(n_t)}}{N_T} \cdot \sum_{n_r=1}^{N_R} \frac{L\ \Omega\ |h_{sc}^{(n_r,n_t)} h_{sc}^{(n_r,n_t)*}|}{\sum I_{total} + \sigma_v^2}, \tag{4.15}$$

where $P_{sc}^{(n_t)}$ is the transmitted power to the receiver from antenna n_t on the subcarrier (sc), L is the pathloss, Ω is the shadow fading and $|h_{sc}^{(n_r,n_t)} h_{sc}^{(n_r,n_t)*}|$ is the frequency selective multipath fading component (a single coefficient) of the channel $\mathbf{h}$ between antennas (n_t, n_r) on the subcarrier (sc). MRC requires a single-stream transmission, which however could be transmitted from multiple antennas n_t sharing the total transmit power $P_T = \sum_{n_t=1}^{N_T} \sum_{sc=1}^{N_{sc}^{BW}} P_{sc}^{(n_t)}$.

The received subcarrier SINR including multipath propagation is,

$$SINR_{sc} = \alpha \, \frac{(\mathbf{h}_{sc}^H \, \mathbf{h}_{sc})^2}{\mathbf{h}_{sc}^H \, \boldsymbol{\mathcal{R}}_{sc,vv} \, \mathbf{h}_{sc}} \tag{4.16}$$

with combining loss α, $0 < \alpha \leq 1$ and the interference plus noise on the subcarrier (sc) is,

$$\boldsymbol{\mathcal{R}}_{sc,vv} = \sum_{j=0}^{N_I-1} \mathbf{i}_{sc,j} \, \mathbf{i}_{sc,j}^H + \sigma_v^2 \mathbf{I}, \tag{4.17}$$

where $\mathbf{i}_{sc,j}$ is the interferer vector ($\mathbf{i}$) on the subcarrier (sc) for a number of interferers $j = 0, .., N_I - 1$. For a block transmission with a block size equal to a PRB of LTE,

$$SINR_{PRB} = \alpha \, \frac{Tr\{\mathbf{H}_{PRB}^H \, \mathbf{H}_{PRB}\}^2}{Tr\{\mathbf{H}_{PRB}^H \, \boldsymbol{\mathcal{R}}_{PRB,vv} \, \mathbf{H}_{PRB}\}} \tag{4.18}$$

$Tr(\cdot)$ is the trace operator as the sum of the diagonal elements $\sum_i a_{ii}$. $\mathbf{H}_{PRB}$ is a matrix of elements $h_{i,sc}$ such that $h_{i,sc}^{(n_r,n_t)}$ is the i^{th} channel path on the subcarrier (sc) from antenna n_t to n_r. $i = 0, .., L_h - 1$, $sc = 1, .., N_{sc}^{PRB}$, $n_t = 1, .., N_T$ and $n_r = 1, .., N_R$.

4.4.2 Interference Rejection Combining

Interference Rejection Combining (IRC) applies for a single-stream transmission as MRC, but IRC provides notable interference rejection combining gains especially at low SINR in the presence of dominant interferers. IRC improves cell-edge performance compared to MRC by reducing structured interference.

The signal model for multiple receive antenna IRC is,

$$\mathbf{r}^{(n_r)} = \mathbf{H} \, \mathbf{F} \, \mathbf{s} + \mathbf{v}, \tag{4.19}$$

where $\mathbf{r}^{(n_r)}$ is the received signal by antenna n_r, for all antennas $n_r = 1, .., N_R$ and the received signal is $N_R \times 1$ vector. $\mathbf{H}$ is $N_R \times N_T$ channel matrix, when multiple transmit antennas also exist. $\mathbf{F}$ is $N_T \times 1$ precoding vector, $\mathbf{v}$ is $N_R \times 1$ noise plus interference vector and $\boldsymbol{\Lambda} = \mathbf{H} \, \mathbf{F}$ is the effective $N_R \times N_T$ channel. $P_T = \left\{|\mathbf{s}\mathbf{s}^H|\right\}$ is the average power for transmit vector $\mathbf{s}$.

After combining signals received from antennas $n_r = 1, .., N_R$, we get $\widehat{\mathbf{r}}$. For maximizing SINR while combining signals from N_R receive antennas, the optimum combining provides,

$$\boldsymbol{\mathcal{W}} = \arg\max_{\mathbf{w}} SINR_{sc}, \tag{4.20}$$

where $SINR_{sc}$ is the SINR per subcarrier and $\boldsymbol{\mathcal{W}}$ is the optimum combining vector of antenna weights w. Thus,

$$SINR = \frac{\boldsymbol{\mathcal{W}}^H \boldsymbol{\Lambda}\boldsymbol{\Lambda}^H \boldsymbol{\mathcal{W}}}{\boldsymbol{\mathcal{W}}^H \boldsymbol{\mathcal{R}}_{vv}\ \boldsymbol{\mathcal{W}}} \cdot P_T, \tag{4.21}$$

where $\boldsymbol{\mathcal{R}}_{vv}$ is the noise plus interference covariance matrix. IRC maximizes (4.21) by choosing the weight vector,

$$\boldsymbol{\mathcal{W}} = \boldsymbol{\mathcal{R}}_{vv}^{-1} \boldsymbol{\Lambda} \tag{4.22}$$

that is

$$\boldsymbol{\mathcal{W}}^H = \boldsymbol{\Lambda}^H \boldsymbol{\mathcal{R}}_{vv}^{-1} \tag{4.23}$$

or rather

$$\boldsymbol{\mathcal{W}}^H = \alpha\ \boldsymbol{\Lambda}^H \boldsymbol{\mathcal{R}}_{vv}^{-1} \tag{4.24}$$

with combining loss α. Now, for a subcarrier SINR,

$$SINR_{sc} = \alpha\ \mathbf{h}_{sc}^H\ \boldsymbol{\mathcal{R}}_{sc,vv}^{-1}\ \mathbf{h}_{sc}, \tag{4.25}$$

where $\mathbf{h}_{sc}$ is the $N_R \times 1$ channel matrix (a vector), $\boldsymbol{\mathcal{R}}_{sc,vv}$ is the noise plus interference covariance matrix and

$$\boldsymbol{\mathcal{R}}_{sc,vv} = \sum_{j=0}^{N_I - 1} \mathbf{i}_{sc,j}\ \mathbf{i}_{sc,j}^H + \sigma_v^2 \mathbf{I}, \tag{4.26}$$

includes the interferers on a subcarrier (sc). Here,

$$\begin{pmatrix} \ddots & & \\ & I_D & \\ & & \ddots \end{pmatrix} + \sigma_v^2$$

includes the interference generated at the receiver by the dominant interferer on the di-

agonal. The ratio of the dominant single interferer power I_D to the sum of the powers of other interferers, $DIR = 10 \cdot \log_{10}\left(I_D / \sum_{\substack{j=0 \\ j \neq D}}^{N_I - 1} I_j + \eta\right)$ and $I_D = \max_j\{I_j\}$. The dominant interference source per subcarrier may be different, particularly because of the frequency selective propagation of the interfering transmissions.

It is feasible, instead of solely maximizing the received $SINR$ according to (4.20), to joint optimize the $SINR$ with the LMMSE equalizer (4.12) i.e. to jointly maximize $SINR$ and minimize MMSE. Hence we get optimal weighting,

$$\boldsymbol{\mathcal{W}}^H = P_T \boldsymbol{\Lambda}^H \boldsymbol{\mathcal{R}}_{rr}^{-1}, \tag{4.27}$$

where

$$\boldsymbol{\mathcal{R}}_{rr} = E\{\mathbf{r}\mathbf{r}^H\} \tag{4.28}$$

is the covariance matrix of the received signal $\mathbf{r}$.

4.4.3 Parallel Interference Cancellation

Parallel Interference Cancellation (PIC) makes decisions based on parallel received signals of multiple antennas. The receiver cancels the strongest detected interferer from the parallel received streams. As many interferers can be cancelled as are reliably detected. Typically, the target is to cancel a single dominant interferer from the received symbol stream. If a single dominant interferer does not exist, PIC typically does not gain, because it would require all or at least most of the interferers cancelled. In presence of many interferers (DIR is small) and $SINR \sim 0\ dB$, reliable interference cancellation is typically not possible. In presence of a single dominant interferer (DIR is large), interference cancellation is typically possible and it gains especially when $SINR \ll 0\ dB$.

For properties of the dominant interferer WCDMA/HSPA and LTE are very different. In WCDMA/HSPA, a single dominant interferer is typically not found due to spreading and scrambling operations that generate Additive White Gaussian Noise -like interferers with close to equal powers (small DIR). In LTE, a receiver operates on a frequency selective block transmission. Hence, even in the presence of many interferer sources, it is typical to find a single dominant interferer per subcarrier (large DIR) due to independent frequency selectivity of the interferers.

4.4.4 Successive Interference Cancellation

Successive Interference Cancellation (SIC) makes sequential decisions in the iterative receiver. At each iteration stage, the receiver regenerates the signal and cancels the strongest detected interferer. The next iteration therefore includes signals with lower interference and higher SINR, provided the interference detection was reliable in the previous stage and the reliability labeling from the previous stage to the next stage is available without too large latencies. Thus, signal detection and decoding reliability may increase stage by stage with the expense of decoding latency.[142][143]

4.5 WCDMA/HSPA multiantenna receivers

The general principles of the chip equalizer presented in Section 4.3.2 are extended to MIMO chip equalizer in Section 4.5.1 and further to the MIMO chip equalizer solution with the LMMSE algorithm in Section 4.5.2. The chip equalizer performs well, but it has computational challenges in long multipath propagated channels, where the dimensions of the channel matrix will increase, which requires longer symbol vectors i.e. complex-valued chip sequences to be equalized.

4.5.1 MIMO chip equalizer

In MIMO system model [92], there are N_T transmit antennas and N_R receive antennas, hence the transmit signal vector $\mathbf{s}$ of (4.5) is indexed by the relevant transmit antenna n_t in (4.29) respectively. The bit stream is mapped to the modulation symbol alphabets after spreading every bit by multiplying it with the channelization sequence c_{ch}. The channelization sequence is of length N_c with chips $c_{ch}^{(1)}, c_{ch}^{(2)}, \ldots, c_{ch}^{(N_c)}$. The transmitted chip symbol vector at antenna n_t is,

$$\mathbf{s}^{(n_t)} = \left[s_n^{(n_t)} \;\; s_{n-1}^{(n_t)} \;\; \ldots \;\; s_{n-(((L_h-1)+L_f)-1)}^{(n_t)} \right], \tag{4.29}$$

where L_h is the length of the channel impulse response, L_f is the equalizer length and vector $\mathbf{s}^{(n_t)}$ contains $(L_h-1)+L_f$ recent transmit symbols to be equalized in the receiver. Now the transmitted symbol vectors $\mathbf{s}^{(n_t)}$ of all transmit antennas N_T, $n_t = 1, .., N_T$ can be stacked to the transmitted symbol matrix,

$$\mathbf{s} = \left[\mathbf{s}^{(1)},\ \mathbf{s}^{(2)},\ \mathbf{s}^{(3)},\ \ldots,\ \mathbf{s}^{(N_T)}\right]^T \tag{4.30}$$

The frequency selective complex channels between the N_T transmit and N_R receive antennas are modelled by $N_R \times N_T$ band matrixes each of dimension $L_f \times ((L_h - 1) + L_f)$. The channel matrix between the $(n_t)^{th}$ transmit and $(n_r)^{th}$ receive antenna is given by

$$\mathbf{H}^{(n_r,n_t)} = \begin{pmatrix} h_0^{(n_r,n_t)} & 0 & \ldots & & 0 & \ldots & & & 0 \\ h_1^{(n_r,n_t)} & h_0^{(n_r,n_t)} & 0 & \ldots & 0 & \ldots & & & 0 \\ h_2^{(n_r,n_t)} & h_1^{(n_r,n_t)} & h_0^{(n_r,n_t)} & 0 & 0 & \ldots & & & 0 \\ \vdots & & & & \vdots & & & & \\ h_{L_h-1}^{(n_r,n_t)} & \ldots & h_2^{(n_r,n_t)} & h_1^{(n_r,n_t)} & h_0^{(n_r,n_t)} & 0 & \ldots & & 0 \\ 0 & h_{L_h-1}^{(n_r,n_t)} & \ldots & & & h_0^{(n_r,n_t)} & 0 & \ldots & 0 \\ \vdots & & & & & & & \vdots & \\ 0 & 0 & & & & & \ldots & 0 & h_{L_h-1}^{(n_r,n_t)} \end{pmatrix}$$

The received signal vector,

$$\mathbf{r}^{(n_r)} = \left[r_n^{(n_r)}\ r_{n-1}^{(n_r)}\ \ldots\ r_{n-(L_f-1)}^{(n_r)}\right] \tag{4.31}$$

at the $(n_r)^{th}$ receive antenna is,

$$\mathbf{r}^{(n_r)} = \sum_{n_t=1}^{N_T} \mathbf{H}^{(n_r,n_t)}\mathbf{s}^{(n_t)}. \tag{4.32}$$

The signal model for MIMO system can be reformulated to,

$$\mathbf{r} = \mathbf{Hs}\ +\ \mathbf{v}, \tag{4.33}$$

where $\mathbf{s}$ is the transmit symbol matrix from antennas n_t, $\forall n_t = 1, .., N_T$ according to (4.30) and $\mathbf{r}$ is the receive symbol matrix at antennas n_r, $\forall n_r = 1, .., N_R$ according to,

$$\mathbf{r} = \left[\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, \ldots, \mathbf{r}^{(N_R)}\right]^T. \tag{4.34}$$

The overall channel matrix is of dimension $N_R\ L_f \times N_T\ ((L_h - 1) + L_f)$ and becomes,

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}^{(1,1)} & \mathbf{H}^{(1,2)} & \dots & \mathbf{H}^{(1,N_T)} \\ \mathbf{H}^{(2,1)} & \mathbf{H}^{(2,2)} & \dots & \mathbf{H}^{(2,N_T)} \\ \vdots & & & \\ \mathbf{H}^{(N_R,1)} & \mathbf{H}^{(N_R,2)} & \dots & \mathbf{H}^{(N_R,N_T)} \end{pmatrix}$$

and the received noise matrix of size $N_R \cdot L_f$ is,

$$\mathbf{v} = \left[\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N_R)}\right]^T \tag{4.35}$$

4.5.2 MIMO LMMSE chip equalizer

MIMO LMMSE chip equalizer for WCDMA/HSPA is presented in [144]. The equalizer coefficients for reconstruction of the transmitted chip sequence from transmit antenna n_t is the argument vector minimizing (4.36).

$$\arg\min_{\mathbf{g}^{(n_t)}} = E\{\left|\mathbf{g}^{(n_t)^H}\ \mathbf{r} - \boldsymbol{\delta}_D^T\ \mathbf{s}\right|^{\ 2}\} = \mathcal{J}(\mathbf{g}^{(n_t)}). \tag{4.36}$$

The minimization is done by deriving with respect to the equalizer coefficients $\mathbf{g}^{(n_t)*}$.

$$\frac{\partial \mathcal{J}(\mathbf{g}^{(n_t)})}{\partial \mathbf{g}^{(n_t)*}} = (\sigma_s^2\ \mathbf{H}\mathbf{H}^H + \sigma_v^2\mathbf{I})\mathbf{g}^{(n_t)} - \sigma_s^2\ \mathbf{H}\boldsymbol{\delta}_D^{(n_t)} = 0. \tag{4.37}$$

This yields the equalizer coefficients for the chip symbol sequence (n),

$$\mathbf{g}^{(n_t)} = \left(\mathbf{H}\mathbf{H}^H + \frac{\sigma_v^2}{\sigma_s^2}\ \mathbf{I}\right)^{-1}\ \mathbf{H}\boldsymbol{\delta}_D^{(n_t)}. \tag{4.38}$$

Again $\boldsymbol{\delta}_D$ is a zero vector with 1 at delay $((L_h - 1) + L_f)(n_t - 1) + D$, where D is the delay of the equalized signal, $D \geq L_h$. Now, the equalizer coefficients $\mathbf{g}^{(n_t)}$ are given by the receive matrix of all antennas $n_r = 1, .., N_R$ for a transmit signal vector from antenna n_t.

$$\mathbf{g}^{(n_t)} = \left[\mathbf{g}^{(1,n_t)}\ \mathbf{g}^{(2,n_t)}\ \mathbf{g}^{(3,n_t)}\ \dots\ \mathbf{g}^{(N_R,n_t)}\right]^T, \tag{4.39}$$

where the equalizer vector $\mathbf{g}^{(n_r,n_t)}$ of length L_f is,

$$\mathbf{g}^{(n_r,n_t)} = \left[g_0^{(n_r,n_t)} g_1^{(n_r,n_t)} \dots g_{L_f-1}^{(n_r,n_t)}\right] \tag{4.40}$$

and includes the equalizer coefficients for antenna pair (n_r, n_t).

4.6 LTE multiantenna receivers

Multiantenna design for LTE is addressed e.g. in [145]. The signal model of flat fading MIMO channel assuming uncorrelated transmit antennas N_T and receive antennas N_R with precoding is,

$$\mathbf{r} = \mathbf{H}\ \mathbf{F}\mathbf{s} + \mathbf{v}, \tag{4.41}$$

where $\mathbf{r} \in \mathbb{C}^{N_R \times 1}$, $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$, $\mathbf{s} \in \mathbb{C}^{N_T \times 1}$ and $\mathbf{v} \in \mathbb{C}^{N_R \times 1}$. With correlated MIMO channel, represented by $\boldsymbol{\mathcal{H}}$, without loss of generality we get,

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\Sigma}_R^{1/2}\ \mathbf{H}_w\ \boldsymbol{\Sigma}_T^{1/2}, \tag{4.42}$$

where $\boldsymbol{\Sigma}_T$ is the transmit antenna correlation matrix, $\boldsymbol{\Sigma}_R$ is the receive antenna correlation matrix and $\boldsymbol{\Sigma}_T$ and $\boldsymbol{\Sigma}_R$ are independent. $\mathbf{H}_w$ is the random channel matrix with non-correlated, independent and identically distributed (i.i.d.) entries. With stacking the elements of $\boldsymbol{\mathcal{H}}$ to vector $vec(\boldsymbol{\mathcal{H}})$, we can write

$$\mathbf{h} = vec(\boldsymbol{\mathcal{H}}) = \big(\boldsymbol{\Sigma}_T^{1/2}\ \otimes\ \boldsymbol{\Sigma}_R^{1/2}\big)\mathbf{h}_w = \boldsymbol{\Sigma}_h^{1/2}\mathbf{h}_w. \tag{4.43}$$

The Kronecker product $\otimes$ is element by element multiplication. Notation $(\cdot)^{1/2}$ takes the squareroot of Eigenvalues of the diagonal matrices generated from the full correlation expansion, or elementwise squareroot of the power correlation elements. $\Sigma_T^{cmplx} \curvearrowright \left(\Sigma_T^{pow}\right)^{1/2}$ and $\Sigma_R^{cmplx} \curvearrowright \left(\Sigma_R^{pow}\right)^{1/2}$. The Kronecker model neglects the joint spatial structure of correlation [146] and separates the correlation properties independently at the link ends. This implies that the transmitter does not impact the spatial correlation properties of the receiver.

The correlation coefficients (for complex numbers x and y) can be defined as the complex correlation $\rho_{cmplx} = \langle x, y \rangle$, envelope correlation $\rho_{env} = \langle |x|, |y| \rangle$ or power correlation $\rho_{pow} = \langle |x|^2, |y|^2 \rangle$. For Rayleigh distributed signals $\rho_{pow} = \left|\rho_{cmplx}\right|^2$. The full complex correlation (amplitude and phase correlation) is required to properly model the MIMO radio channel [147]. Channel measurements and spatial channel models for MIMO with correlation are widely present e.g. in [148][149][150] and for wideband signals recently in

[151][152]. For the system performance analysis in Chapter 9, the spatial channel model correlation matrixes are given in [61].

4.6.1 FFT equalizer

Since Cholesky decomposition based tap solver requires huge amount of calculation, alternative solutions are favoured. The equalizer coefficients can be solved by the FFT algorithm based on the fact that covariance matrix is first made circulant and then it is diagonalized by the FFT matrix. In case, the covariance matrix is not circulant [138] by nature, its modification to a circulant one may cause performance degradation. For spread signals (WCDMA/HSPA), the covariance matrix is not as likely circulant than it is for the block transformed signals (LTE). Of the FFT method, an inner-product and an outer-product form exists [153]. The covariance matrix $\boldsymbol{\mathcal{R}}$ can only be used in the outer-product solution, whereas the inner-product solution requires matrix inversions of large $N_T \times N_R$ matrices.

The first choice of equalization is to implement the equalizer as a time domain FIR filter and to compensate the estimated channel $\widehat{h}$ by the complex conjugate channel coefficients $\widehat{h}^*$. The channel convolution operation can, however, be alternatively implemented using multiplication in frequency domain for computational savings. As described in Section 3.4.1, the FFT requires element-by-element vector multiplication in frequency domain, which corresponds to the circular convolution in time domain. This circular operation yet has to be made equivalent to the linear convolution in time domain. The FFT equalizer can be implemented e.g. by a computationally simple overlap-add technique, shown as a block diagram in Fig. 4.1. The overlap-add FFT first generates zero-padded signal vectors (the length of the zero padding has to be longer than the channel and the equalizer i.e. $> L_h + L_f - 1$), then executes element-by-element multiplication with weights and finally makes element-by-element addition of the component vectors. After the inverse FFT, the time domain signal consists of the overlap-add signal components, which correspond to the convolution of the channel and the signal equalized by the weight vector $\underline{\mathbf{W}}$. This equalized signal vector corresponds to the transmit signal vector similar to what could be generated by the time domain equalizer that would require many more computational operations per symbol block. The difference of the computational complexity of the tap solver by time domain algorithms and by frequency domain algorithms is analysed e.g. in [138][153]. The complexity increases as a function of the required equalizer length and the

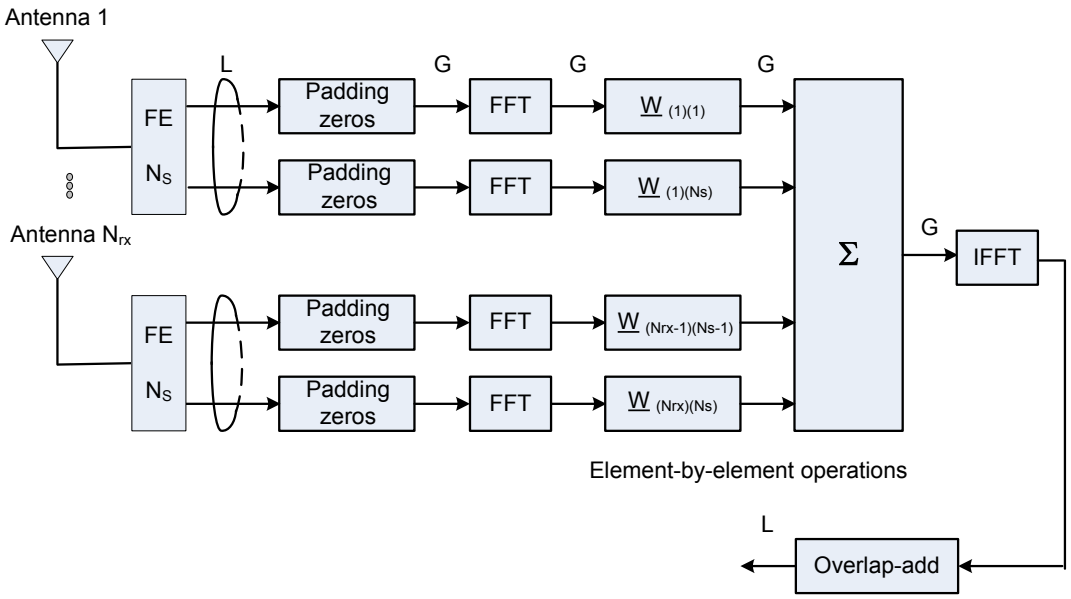


Figure 4.1: Frequency domain equalizer, a block diagram for the FFT based solution.

signal bandwidth. The complexity of the time domain algorithms get evident already in the analysed 5 MHz band, as a function of increasing equalizer length.

4.6.2 QRDM receiver

QRDM receiver is an M-algorithm, which is applied to the modified receive vector and modified channel matrix. QRDM is based on **Q R** decomposition of the channel matrix **H** [142][143].

$$\mathbf{H} = \mathbf{Q}\ \mathbf{R}, \tag{4.44}$$

where **Q** is unitary $N_R \times N_T$ matrix and **R** is upper triangular $N_T \times N_T$ matrix. The received signal of original (4.41) (see also the discussion in Section 3.5.3 from (3.11) to (3.15)) modified by the **Q R** decomposition is,

$$\mathbf{r} = \mathbf{QR}\ \mathbf{Fs} + \mathbf{v}. \tag{4.45}$$

The estimation of the transmitted signal (omitting precoding **F**),

$$\widehat{\mathbf{s}} = \mathbf{Q}^H \mathbf{r} = \mathbf{R}\ \mathbf{s} + \mathbf{Q}^H \mathbf{v}. \tag{4.46}$$

As $\mathbf{Q}$ is unitary, the statistical properties of $\mathbf{Q}^H\ \mathbf{v}$ remain the same as for $\mathbf{v}$. For estimation of the signal transmitted from antenna n_t, we get

$$\widehat{\mathbf{s}} = \mathbf{r}^{(n_t,n_t)}\mathbf{s}^{(n_t)} + \sum_{i=(n_t+1)}^{N_T} \mathbf{r}^{(n_t,i)} \cdot \mathbf{s}^{(i)} + \sigma_v^2\ \mathbf{I}. \tag{4.47}$$

In here, (n_t, n_t) diagonal elements are taken out of the sum and $\mathbf{r}^{(n_t,i)}$ remain upper triangular ($i \geq (n_t + 1)$). For a frequency selective block transmission of LTE, subcarrier frequency indexes could be applied in the equations above. In (4.47), the off-diagonal interference terms of the other transmit antennas may iteratively be detected and at least partly cancelled.

4.6.3 Iterative turbo equalizer

Turbo equalization is an iterative joint-process of equalization and channel decoding, where soft symbol output of the decoder gives feedback to the equalizer for setting the equalization weight vector. Turbo equalizer combines the excellent error correction capability of the turbo decoder and soft symbol replica cancellation by adaptive linear filtering. Turbo equalization for LTE is studied in [154]. The block diagram of Turbo equalizer is shown in Fig. 4.2.

For the received signal in (4.41) the soft symbol replica cancellation actually reduces interference from the received signal along the iterations as seen in (4.48)

$$\widetilde{\mathbf{r}}^{(n_r,n_t)} = \mathbf{r}^{(n_r,n_t)} - \sum_{\substack{i=1 \\ i\neq n_t}}^{N_T} \widehat{\mathbf{h}}^{(n_r,i)} \cdot \widetilde{\mathbf{s}}^{(i)} \tag{4.48}$$

where $\widetilde{\mathbf{s}}$ is the soft estimate of $\mathbf{s}$, $\widehat{\mathbf{h}}$ is the channel estimate and $\widetilde{\mathbf{r}}$ is the soft estimate of the received singal at the current iteration. In (4.48), the transmit and receive antenna indexes (n_r, n_t) are present. If combined signals from antennas $n_r = 1, .., N_R$ are used, we get the soft estimate

$$\widetilde{\mathbf{s}}^{(n_t)} = \mathbf{W}^{H\ (n_t)}.\ \widetilde{\widehat{\mathbf{r}}}^{(n_t)}, \tag{4.49}$$

where $\mathbf{W}^{(n_t)}$ represents the filter coefficients and $\widetilde{\widehat{\mathbf{r}}}^{n_t}$ represents the multiantenna com-

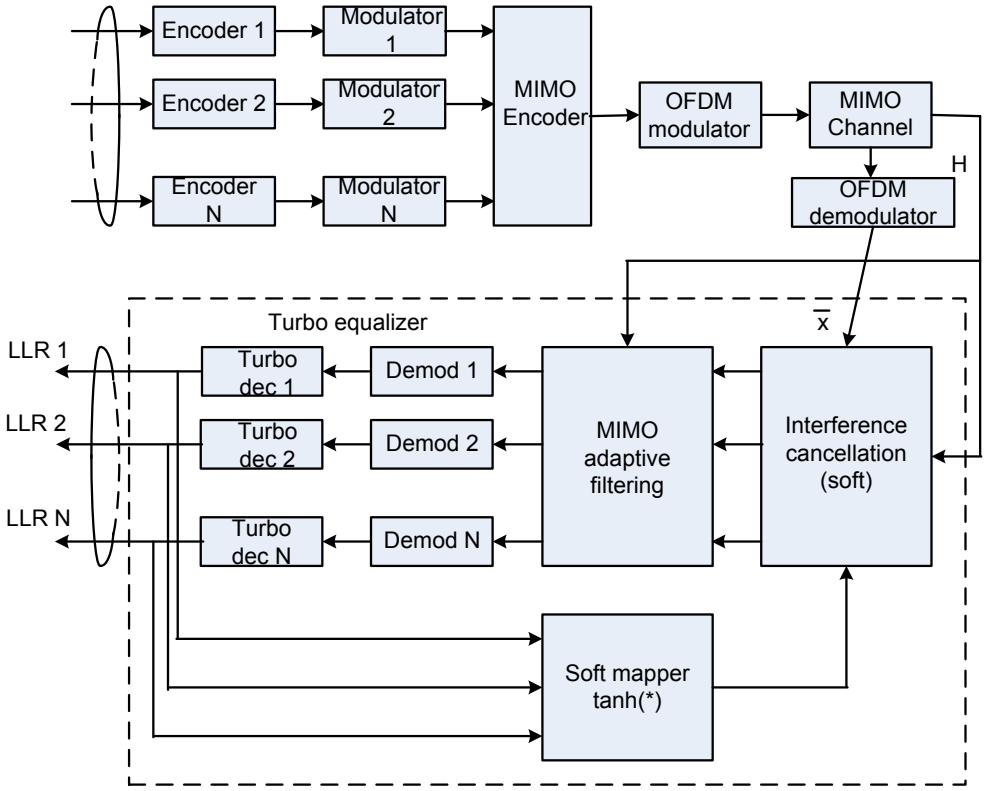


Figure 4.2: Block diagram of the Turbo equalizer receiver for a multistream transmission.

bined receive vector from a single transmit antenna n_t, as given in (4.50).

$$\widetilde{\mathbf{r}}^{(n_t)} = \left[\tilde{\mathbf{r}}^{(1,n_t)}\ \tilde{\mathbf{r}}^{(2,n_t)} \dots \tilde{\mathbf{r}}^{(N_R,n_t)}\right]^T. \tag{4.50}$$

The optimum filter is based on LMMSE, where the filter coefficients are updated for every symbol. The MMSE filter used for the iteration is

$$\mathbf{W}^{(n_t)} = \Big[\sum_{\substack{i=1 \\ i\neq n_t}}^{N_T} \widehat{\mathbf{h}}^{(i)} \cdot \widehat{\mathbf{h}}^{H\ (i)} + \sigma_v^2 \mathbf{I}\Big]^{-1} \cdot \widehat{\mathbf{h}}^{(n_t)} \cdot \boldsymbol{\delta}_D^{(n_t)}. \tag{4.51}$$

As an alternative, a matched filter solution exists by $\mathbf{W}^{(n_t)} = \widehat{\mathbf{h}}^{(n_t)}$ for $\forall n_t = 1, .., N_T$. Soft replicas are created by calulating the LLR for all channel symbols and mapping them to the soft estimates $\widetilde{\mathbf{s}}$ by a hyperbolic tangent function.

Chapter 5

Packets and Internet protocols

5.1 Datagram representation

A packet in information technology is a formatted block of data carried by a packet mode network. A packet consists of two kinds of data: control information and user data also known as the payload. In hierarchical protocols, the payload typically contains a Service Data Unit (SDU) of a higher layer protocol instead of "user data". The payload in a *datagram* presentation may hence contain nested protocol segments inside, where in each protocol layer a header indicates the next following header, which may be the next higher protocol layer or a further options header of the same protocol layer.

Control information provides indications to categorize the user data inside a payload of a packet and indications, where in the network nodes or where in the protocol machines the payload needs to be delivered to. In the datagram of the Internet Protocol, the control information includes at least the Internet source address and the Internet destination address. In telecommunication protocols, the control information includes the channel identity or the identity of the protocol entity for further processing. Typically, the control information additionally includes a sequence number and an error detection code like Cyclic Redundancy Check (CRC) based on polynomial divisors. The delivery of payload to the next processing node is generic in a sense that the next node may be a distant apparatus apart in the network, may be a processing unit (a chip) apart in the same apparatus or may just run in a different software module, task or a function in the same processing unit. Control information is appended to the payload as a header and/or a trailer. Communication protocols differ in their conventions for distinguishing between Information Elements of the protocol and formatting the data contents of the packet.

Exact definitions appear in the standard specifications, e.g. by IETF and 3GPP.

Datagram representation of a packet defines protocol hierarchies in nested format, where each layer consist of headers relevant to that layer (e.g address, segment number, acknowledgement and length), bitfield indicators of packet control including the next header descriptor and encapsulated next layer payload. Datagram presentation allows a protocol machine to operate at each layer very fast interpreting controls of its own layer only and passing the payload blindly to the next engine present in a recognized address. Datagram format enables distributed protocol architectures and distributed processing of a large number of packets per time unit despite of their varying contents.

5.2 Network protocols for packet communications

The thesis studies packet communications that for a wide area, global networking is implemented by the Internet protocols. The Internet protocols and Radio Access are functionally cooperative and intertwined and they together form the convergence of Mobility and the Internet. The studies of fixed networking and radio access mobility may often be separable, and it is possible to omit some of their properties, when studying one or another.

The Internet protocols obviously impact the performance of the Radio Access Network (RAN) (setup delays, transport delays, overhead) and the efficiency of the air interface. It is of debate, how much should RAN see of these protocols. The baseline is of course that the Internet protocols may act for session management, transmission control and for routing in the transport network and are invisible to RAN. However, some of the RAN *bearer* parameters have dependency on the Internet protocols and may need to be derived from them, especially if QoS differentiation is expected. Further, Internet traffic activity, packet sizes and packet overhead are largely dominated by the Internet protocol in question. Say Transmission Control Protocol (TCP) typically generates large packets (TCP segments), small overhead and very high but bursty loading of buffers, whereas User Datagram Protocol (UDP) typically generates nearly constant size packets (small for voice, large for video) with predictable packet arrival rate. On the other hand, if Internet security is in use, the contents of the payload is not visible in the intermediate routing nodes, neither in RAN and decision making may happen based on session level negotiations and packet header fields. Because the packets appear in the traffic buffers of the UE and

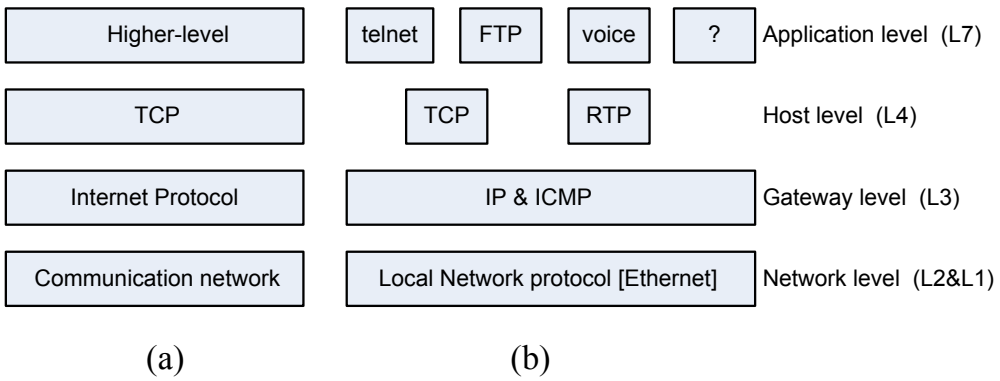


Figure 5.1: The classical TCP model. (a) Protocol hierarchy and (b) the relationship of conventional protocols in different layers.

RAN, and because the radio protocol stack directly sees the Internet packets as its SDUs, the source-destination behaviour of Internet flows need to be covered in relevant parts in this thesis. The impact is visible in loading, packet arrival statistics, SDU size distributions and overhead, as well it sets conditions for the segmentation, windowing, multiplexing and signalling functions. All this finally contributes to the system performance measures.

The classical TCP protocol model in Fig. 5.1 illustrates the place of TCP in the protocol hierarchy and its relation to other protocols. The number of protocols has expanded since the invention of the TCP model and their layered structure is vanishing towards protocol engines interacting with each other. Fig. 5.2 shows a simplified example of a protocol environment experienced today. This thesis is more focused on the radio systems, but selected aspects of the Internet protocols are covered of necessity throughout Section 5.2. The Internet protocols have impact also on the mobile network architecture covered in Chapter 6 and the system performance in Chapter 9, therefore at least session initiation, transmission control, user datagram, Internet and Ethernet protocols are described here, as well as the essence of audio codecs with an impact to the protocol payload formats and their header compression.

5.2.1 Session Initiation Protocol

Session Initiation Protocol (SIP) [155] is a session layer protocol in the Open Systems Interconnection (OSI) reference model. In the TCP/IP reference model it is placed to the application layer. SIP is accepted as the 3GPP signaling protocol in packet domain

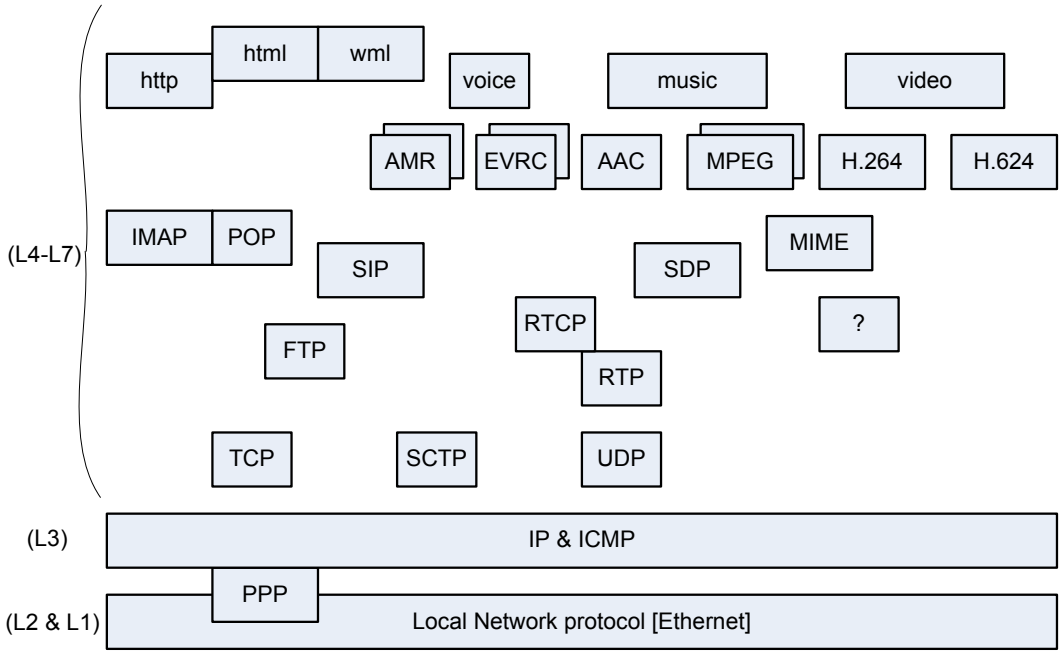


Figure 5.2: An example of a simplified protocol environment today. The layered structure of protocols is vanishing, but there exist a jungle of protocols at least markup languages, source codec protocols, multimedia types, session initation/description protocols, application layer protocols, transmission protocols and lower layer Ethernet protocols. In this simple figure, e.g. authentication, discovery, mobility and domain name server protocols are omitted.

and it is a part of the IP multimedia architecture for the Internet services in cellular systems. SIP clients use TCP or UDP to connect to SIP servers. SIP is primarily used in setting up and terminating voice or video calls, but it can be used for session initation of any application using any combination of media types (Rich media). All voice and video communications are then carried over a separate session layer protocol, typically Real-time Transport Protocol (RTP) [156].

5.2.2 Transmission Control Protocol

Transmission Control Protocol [157] provides a connection-oriented service for the connectionless Internet. It is needed for applications that expect reliable data delivery. TCP offers a robust service to its applications by four major properties. TCP is error-free, because lost segments may be recovered by retransmissions and the likelihood of data loss is very small, except temporarily during congestion. TCP provides assured delivery for the application, so that it gurantees finding the destination by the unique Internet address and a port number. In case, destination cannot be found e.g. due to it is powered off, TCP will

generate a failure indication to the application. TCP generates in-sequence delivery by segmenting application data and windowing the received segments in order to feed them to the application in sequence order. TCP avoids data duplication, because the segments have a unique order by the segment numbers, and segment duplication can be detected inside a large (32 bit) numbering space.

TCP operates end-to-end, so that it recognizes transmission control only at the ports of the original source and at the ultimate destination of data. TCP does not participate at the intermediate routers along the path, which is the duty of the Internet routing protocol instead. This lets TCP adapt to the joint-properties over all intermediate and diverse transmission paths that may exist between the source and the destination.

TCP always starts by connection establishment, which happens by three-way handshaking between the source and the destination, exchanging a Connect Request packet, Connect Confirm Packet and an Acknowledgement Packet. These packets may readily include data payload. After connection establishment, TCP is governed by four algorithms [158]; the slow start, congestion avoidance, fast retransmit and fast recovery. TCP must at minimum implement the slow start and the congestion avoidance algorithms. TCP has two key variables per connection state: the congestion window ($cwnd$) and the receiver advertised window ($rwnd$). The minimum of $cwnd$ and $rwnd$ governs the data transmission. The state variable slow start threshold ($ssthresh$) determines, whether the slow start ($cwnd < sstresh$) or congestion avoidance ($cwnd > sstresh$) algorithm is used. By these algorithms, TCP aggressively manages the bandwidth-delay -product of the transmission path. The conventional TCP algorithms have been developed for further improved performance and faster start to the high capacity regimes, see e.g. Reno and new Reno algorithms [158].

The client/server model has become one of the central ideas of network computing. Most applications today use the client/server model, so does TCP. The client/server model has been used to distinguish distributed computing by smaller dispersed computers from the "monolithic" centralized computing of mainframe (host) computers. Recently, this distinction has largely disappeared though, as even the mainframes have also turned to the client/server model and become part of the network computing architecture. The roles of client and server are briefed below.

A client is known as the sender of the Connect Request packet. The client initiates requests, waits and receives replies and usually connects to a small number of servers at

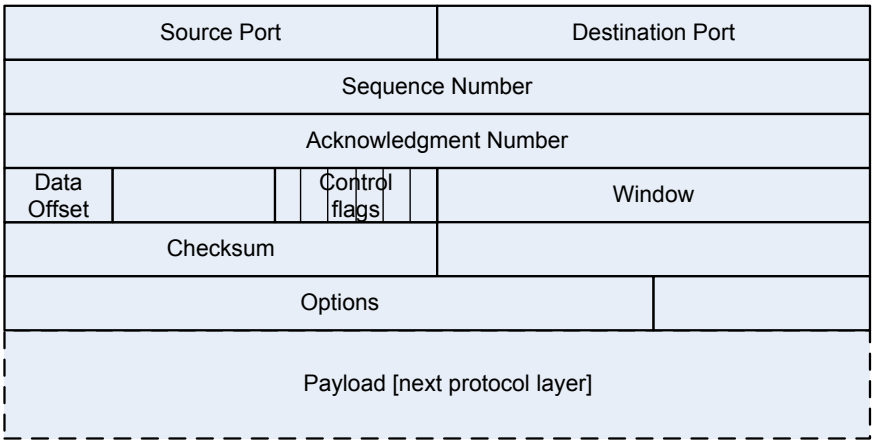


Figure 5.3: TCP header format and the payload.

a time. The client software often directly interacts with the end user device. A server is known as the receiver of the Connect Request packet sent by a client. The server is passive, it waits for requests from clients and upon reception of a request, processes them and serves replies. A server usually accepts connections from a large number of clients, but it does not directly interact with the end user devices. Another type of network architecture is known as peer-to-peer, where each node or instance of the program can simultaneously act as both a client and a server. Both the client-server and peer-to-peer architectures are in wide use today, evidently so in mobile communications.

TCP header format [157] is show in Fig. 5.3 and its most important fields are described below. **Source Port** and **Destination Port** (together with IP Source Address and IP Destination Address) form a universally addressable communication connection for an application. **Sequence Number** gives the index of the first data octet in this segment. **Acknowledgment Number** indicates the next Sequence Number, the sender of the Acknowledgment expects to receive. It thus acknowledges correct reception of all segments up to this Sequence Number. Acknowledgments actually drive the activity of the data sender. **Window** indicates, how many bytes of data the sender of this field, i.e. the data receiver, is willing to accept. **Control flags** are needed for special situations, e.g. to urge delivery of a segment or to avoid silly window syndrome by synchronizing the Sequence Numbers at the transmitter and the receiver. The control flags include; **URG**: Urgent Pointer field significant, **ACK**: Acknowledgment field significant, **PSH**: Push Function, **RST**: Reset the connection, **SYN**: Synchronize sequence numbers, FIN: No more data from sender.

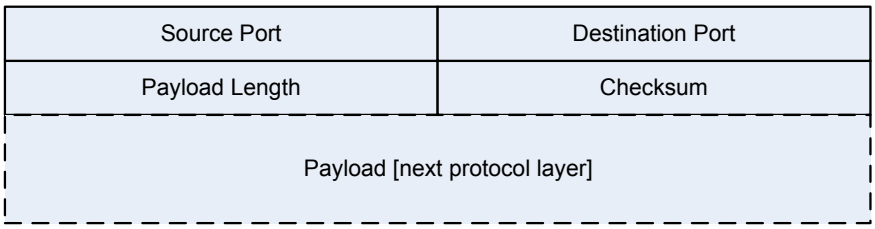


Figure 5.4: UDP header format and the payload.

5.2.3 User Datagram Protocol

User Datagram Protocol [159] has the format shown in Fig. 5.4, and it provides a connectionless service for the application protocols. It optionally includes the application **Source Port** number and the mandatory **Destination Port** number. UDP is transaction oriented, thus delivery and duplicate protection are not guranteed. UDP is often used for real-time applications, especially conversational ones, where reliable delivery by end-node retransmissions is not feasible within the delay constraints anyway. Any retransmissions and acknowledgements by end-nodes would just increase the packet discard probability due to exceeded packet life-time. As well, UDP is often used for high volume real-time media services, where single packet losses are not that critical for the user perception of quality and where the retransmission might unnecessarily congest the network. UDP anyhow is flexibe for variable payload and it includes the segment length field.

5.2.4 Internet Protocol

The Internet Protocol (IP) is the packet-switched communication protocol over any interconnected networks of routers. The Internet nodes worldwide have been using Internet Protocol version 4 (IPv4) [160] for the past three decades, and it is clearly the dominant IP version. IPv6 [161] is designed to add much more, such as nearly unlimited IP address space to connect everyone and everything, stateless auto-configuration, seamless mobility, automated network management, mandated security and new service differentiation levels. Even if IPv6 is commonly assumed in the studies, its deployment is in very early stage and its penetration is below 1% of the Internet traffic in any country, where it is in use.

The originally defined IPv4 had addressing capability limited to 32 bits. IETF decided to upgrade that in the IPv6 protocol for larger address space of 128 bits, which is expected to be sufficient for all the future needs. Further on, IPv6 was enhanced to include QoS

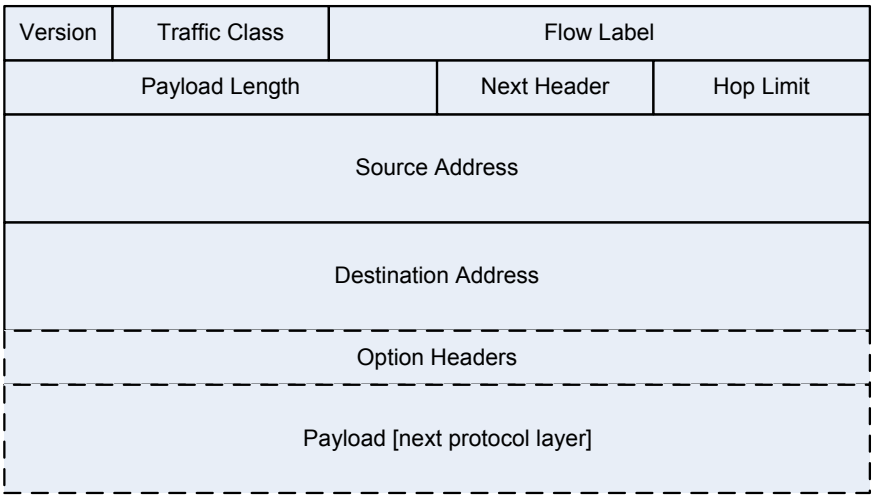


Figure 5.5: IPv6 header format and the payload.

differentiations. For this reason Flow Label and Traffic Class fields were defined. Flow Label serves quality of service differentiation for real-time flows and Traffic Class serves to distinguish between priorities of packets for differentiated services (DiffServ) [162].

In addition to its greatly expanded addressing capability IPv6 provides simplified (mandatory) header format, improved support for extensions and options in nested headers, flow labeling and new authentication/privacy capablilities. The header format of IPv6 datagram is shown in Fig. 5.5. **Payload Length** defines the total length of the IPv6 packet following the header. This length includes all extensions, options and nested protocol headers. **Next Header** identifies the type of the following header in the nested header presentation, which forms the datagram. **Hop Limit** may be set to limit the transmission path. It is decremented by each forwarding node until the packet gets dropped at Hop Limit zero. **Source Address** and **Destination Address** are the universal, global addresses of any node or any device in the network. The Address field consists of the scope fields i.e. subnet scopes.

5.2.5 Internet Control Message Protocol

Internet Control Message Protocol (ICMP) is an integral part of IPv4 and IPv6 and is implemented by every node. It reports errors in packet processing and provides diagnostics. It also includes 'PING' message, which is used in the Internet to test connectivity between network nodes and delay performance. PING-message may carry payload to test loading as well.

5.2.6 Ethernet protocol

Ethernet protocol (originally in IEEE 802.3) is a link protocol in the fixed network that carries Internet packets between network nodes i.e. routers and/or end-user devices having an Ethernet address. Ethernet does not provide routing functionality, but it carries the payload fast with low overhead and efficiently multiplexed from the source to its next destination. The maximum length of the data-field of a packet sent in an Ethernet frame is 1500 bytes. This cut-off value sets the maximum length of an IP datagram sent over the Ethernet, and TCP often experiences that as the Maximum Transmission Unit (MTU) over the networks. Even if Gigabit Ethernet over fiber were locally available having a larger frame size for superior bandwidth-delay -product of transport, TCP anyhow commonly experiences MTU of 1500 bytes between the end nodes.

5.2.7 Real-Time Transport Protocol

Real-Time Transport Protocol [156] is a transport protocol for real-time applications. It may carry payloads of various formats, and therefore RTP headers are defined as a common fixed part and a variable part, where the format of the variable part depends on the application type and is therefore separately standardized for each media encoder, see e.g. RTP for Adaptive Multirate (AMR) and Adaptive Multirate Wideband (AMR-WB) audio codecs in [163]. Plenty of other media types are covered in other specifications. RTP fixed header part includes the **Payload Type (PT)** and the **Sequence Number (SN)**. PT identifies the format of the RTP payload and determines its interpretation by the application. A profile may specify a default static mapping of a payload type to a payload format. An example of a static PT mapping is Enhanced Full-Rate codec (EFR) for GSM. Additional PT codes may be defined dynamically through non-RTP means (e.g. by the Session Description Protocol or the RTP control protocol). A set of default mappings for audio and video is specified in the companion IETF specifications.

RTP source may also change PT during a session. Examples of dynamic PT fields are those for the AMR codecs (see Section 5.3), where several modes are defined and where the modes may adapt during a session. Changing the codec mode is rare though. For the possible change of the codec bitrate, PT value actually need not change, as the bitrate is announced separately for each voice frame much deeper in the nested variable header. The codec mode (including bitrate) is again set by the Frame Type index of the codec,

defined in [163].

RTP Sequence Number increments by one for each RTP data packet and may be used by the receiver to detect packet loss and to restore the packet sequence. In case, the packet sequence cannot be fully restored at the receiver within the delay constraints, the codec applies means to provide a reasonable output based on incomplete information. Network delay variations, out-of-sequence transport and individual packet losses are typically recovered by the RTP protocol (sequence numbers, timestamps), and the remaining packet delay variations are removed by the playback buffers capable of feeding the codec at constant symbol rate despite of packet delivery jitter over the network.

5.3 Audio codecs and transport protocol payload formats

The payload format for packetization of AMR and AMR-WB encoded voice signals into RTP [156] is specified in [163]. The payload format supports transmission of multiple channels, multiple frames per payload, use of fast codec mode adaptation and it provides robustness against packet losses and bit errors.

The Payload Type is negotiated during session initation, and is defined by the Session Description Protocol [164]. AMR or AMR-WB codec type (and their mode sets) is assigned per session. The codec-mode (codec bitrate) is dynamic and adaptive so that it may change for every payload instance. Inside every RTP payload, there exists a Table of Contents for all the voice frames carried inside the RTP payload. Indication of a voice frame happens by the Frame Type (FT) index field as a single entry in the table of contents. FT indicates either AMR or AMR-WB voice coding mode or comfort noise (Silence Descriptor, SID) mode of the corresponding voice frame inside the payload of this RTP packet. According to the IETF standards, it is not allowed to multiplex different media types within a single RTP session. However, multiple RTP-sessions may exist in parallel (as RTP flows) to carry different media types.

Even if ultimate flexibility of dynamic mode adaptation is allowed, a practical convention is to negotiate the codec mode (bitrate) in the beginning of a session and keep it constant. As voice quality is preferred, the highest bitrate mode (12.2 kbps for AMR) is most typically selected. This is known to result a very good voice quality according to the mean opinion score of various listener tests. Changing the codec mode to a lower bitrate may be experienced as reduced voice quality. A lower bitrate mode may still be selected

e.g. in the case of low capacity network. Also temporary congestion or short term lack of capacity may cause codec mode changes to avoid call dropping. Call dropping would lower the voice quality metric much more radically than temporary lowering the codec bitrate. For optimizing the use of radio resources, in an operation point below capacity limits, codec mode adaptation is not recommended, and steady voice quality is preferred instead.

5.4 Header compression

As explained in Section 5.1 and visible in Fig 5.5, the networking headers of IP are present for each and every packet of data delivered over the Internet. Every packet further carries nested headers of multiple protocol layers, typically at least those of the application protocol, session level protocols and TCP/IP or RTP/UDP/IP. This actually provides a huge overhead for the application information over the length of the IP datagram. For fixed networking this overhead is an issue, but it is not a dramatic one, because it is well handled by the large overdimensioning of network capacity that is needed anyway to cope with the bursty behaviour of traffic. In any event, this overhead is a large issue for the radio interface, where overdimensioning is not favoured due to high cost per bit and where the networking headers are actually not needed, because the datagrams (network SDUs) are addressed, transported and sequenced entirely by the radio protocols. For these reasons, header compression mechanisms are specified to act over the radio access network (in WCDMA/HSPA) or the radio interface (in LTE).

Header compression of TCP/IP is fairly straightforward, unlike the compression of RTP/UDP/IP. Classical TCP/IP header compression in [165] will omit delivery of the static fields of the header, which typically allows compression to about 2 to 5 bytes. This reduces the TCP/IP overhead in long term average below 1%, whereas originally the overhead per packet may be several hundred percentages or even far beyond if the application payload happens to be small. (For a concrete example, subsection 7.2.2.1 gives the model of Pareto distributed packet size in (7.11), which is parametrized in Table 9.1 for the analysis in Section 9.3. The mean packet size of 288 bytes would have a conservative 16% long term average overhead assuming IPv4 without header compression.) RTP/UDP/IP compression is much more difficult because of the real-time changes of some header fields. Robust Header Compression (ROHC) [166] framework and its profiles were

specified particularly for the RTP/UDP/IP [167]. More context profiles e.g. for TCP/IP were recently updated to ROHC.

For large RTP packets as multimedia and video media types, the compression efficiency is not outmost critical due to their high bandwidth demand of payload and due to small relative overhead. For small RTP packets as voice media types, the compression efficiency is critical, because the networking header overhead is large (easily over 100%) compared to the small payload of the voice frames, because a large number of simultaneous voice conversations are typically active in the network [168] and because the voice capacity is a critical measure of the system efficiency. However, once IPv6 is not yet in use really and because the overhead of IPv4 is much smaller, ROHC is not that much in use in practical networks. ROHC for a voice media type is described in subsections 5.4.1 and 5.4.2.

The behaviour of voice conversation includes regular periods of voice packets during a voice spurt and long silence periods in between the talk-spurts. This creates moments of well predictable contents in the fields of the headers and other moments when the fields of the headers are less predictable. In order to provide large overhead reduction, ROHC has to operate in different states, depending on the predictability of the header fields. The goal is to fully compress the static fields and to minimize the size of the remaining dynamic fields. The minimized size depends on the regularity/irregularity of the dynamic bit fields. In a network of small Round Trip Time (RTT) variation, low error rate and stable routing tables, RTP Sequence Number often increments regularly by one. In a network of large RTT variation, higher error rate and frequently updating routing tables, RTP Sequence Number may increase and decrease abruptly, which requires a larger window to resolve the irregularity.

5.4.1 Header compression states

ROHC defines three compressor states as the *Initialization and Refresh (IR)*, *First Order (FO)* and *Second Order (SO)* states. The compressor starts in the lowest compression state (IR) for exchanging the full headers and transits gradually to the higher compression states. The compressor will always operate in the highest possible compression state, where the compressor is confident that the decompressor has sufficient header context to decompress the header to the replica of the original. Some fields of the original header may not need to be reproduced at the receiver, and those may be stripped instead. The ROHC states in the reliable operation mode is shown in Fig. 5.6 and described below.

Initialization and Refresh state

The purpose of IR state is to initialize the decompression context at the receiver, or to recover it after a context failure. IR header is always needed in the beginning of a voice session. In IR state, the compressor sends complete header information. This includes all static and non-static fields in uncompressed form plus some additional information, as the assigned ROHC Context Identity (CID) of the compression profile.

First Order state

The purpose of FO state is to efficiently communicate any irregularities in the headers of a packet flow to the decompressor. In FO state, static fields can be fully removed and regular dynamic fields can be compressed to the minimum. Depending on the irregularity of the fields in the header, different sizes of remaining FO header may be needed. Often, FO header size is between 2 to 8 bytes, or up to 40 bytes depending on how the algorithm preferres to discretize. For voice service, FO header is most typically needed always at the beginning of a talk-spurt after a silence, occasionally after a handover or at any irregularity in the packet delivery over the network.

Second Order state

The purpose of SO state is to compress the header to the minimum that is required, when the header is completely predictable given by the least significant bits of the RTP Sequence Number. SO header size is discretized to one byte. The compressor switches from SO state to FO state, whenever the source header no longer conforms to the uniform pattern and could not be decompressed at the receiver from the previous context just by the least significant bits of the RTP Sequence Number. The compressor switches back to SO state from FO state, whenever FO header has sufficiently updated the decompression context and regularity appears again in the accuracy of the least significant bits.

5.4.2 Header compression modes of operation

The ROHC scheme has three modes of operation; Unidirectional, Bidirectional Optimistic and Bidirectional Reliable mode. The optimal mode to operate depends on the characteristics of the network environment, such as feedback capability, error statistics and expected irregularities. Reliable mode is favoured for a conversational voice session over cellular, because feedback mechanisms are available and because it enables more reliable state changes at the compressor due to knowledge of the decompressor state. The feedback is signalled per CID and it includes the ROHC-Acknowledgement and information specific

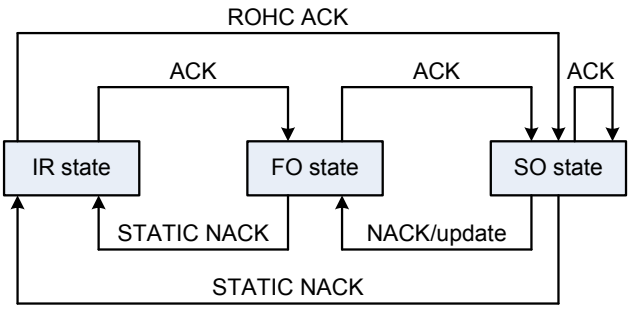


Figure 5.6: ROHC state machine in the reliable birectional mode, which is used for a VoIP conversational session. Omitting ROHC ACKs is possible in the optimistic mode.

to the compression profile.

The positive ROHC-acknowledgement (ACK) is an indication that a fully correct compression context is available at high probability at the decompressor (at a given RTP Sequence Number) and it therefore allows transit to a higher compression state or allows remaining in that higher state. A negative ROHC-acknowledgement (NACK) is partial information to indicate that the dynamic part of the compression context is not correct and several successive packets have failed to be decompressed correctly. Negative ROHC-acknowledgements are also called context requests, which make the compressor transit back to the FO state and deliver more and longer fields of the headers for more complete updates of the decompression context. Negative ROHC-acknowledgement includes RTP Sequence Number of the latest successfully decompressed packet. Incorrect static part of the compression context is negatively ROHC-acknowledged (STATIC-NACK) separately, and it makes the compressor transit back to the IR state.

The reliable mode adds overhead to the reverse flow by adding ROHC-feedback (typically ACKs) to its flow of compressed headers i.e. its SO headers or FO headers. The positive ROHC-acknowledgements need not be frequent, but a fast delivery of negative ROHC-acknowlegement is preferred in order to fast recover of any improper context at the decompressor. Equally well, fast positive ROHC-acknowledgements are preferred at all times, when they enable transit to a higher compression state. In a way, ROHC acknowledgements may drive compression efficiency to the optimum, even though ROHC would be functional without them. As positive ROHC-acknowledgements need not be transmitted frequently, their additional overhead is not large. However, as a consequence, they add variability to the ROHC compressed header sizes and therefore have an impact

to the optimal transport block sizes at the radio interface.

Decisions about the transitions between the compression states are taken by the compressor on the basis of variations in the fields of the packet headers, positive ROHC-acknowledgements from the decompressor and negative ROHC-acknowledgements from the decompressor. Additionally, periodic timeouts for state changes may be used especially in the unidirectional mode that applies for streaming.

Chapter 6

Mobile network architecture and procedures

6.1 Evolution of the network architecture

The networking protocols conventionally operate either in the circuit-switched domain or recently in the packet-switched domain. The circuit-switched domain is implemented by Signalling System 7, where the GSM Mobile Application Part handles the mobility [169]. Call control is handled together with the Public Switched Telephone Network. The focus in this chapter is on the packet-switched domain operation, where the networking is implemented by the IP. The services in the packet-switched domain are provided by the Evolved Packet Core (EPC), which may include the IP Multimedia Subsystem (IMS) [9][170][171]. The architecture provides solutions for the end-to-end operation over the Internet. During the gradual transition from the circuit-switched network technologies to the packet-switched network technologies, interworking solutions will exist in practical implementations so that critical operations such as *voice call continuity* can be provided [172].

The protocol architecture includes 1) networking and transport protocols based on IP and tunneling, 2) convergence protocols to create SDUs for mobile delivery with proper context identity and sequence numbering from the packets of the Internet flows and 3) radio protocols, which control radio transmission, radio network resources and delivery of the payload. Additionally, *non-access stratum* protocols exist for Mobility management, Session management, as well as for security control and authentication. Mobility Manage-

ment includes e.g. Tracking Area Update procedure, idle mode cell selection and camping. Session Management includes Paging procedures, call control for the circuit-switched domain and Session Initiation for the packet-switched domain.

6.2 UMTS reference architecture

In the UMTS architecture [169] shown in Fig. 6.1 on the left, the major evolutionary step was introduction of IP connectivity via the Iu-interface of the Serving GPRS Support Node (SGSN) and via Gn-interface of the Gateway GPRS Support Node (GGSN) to the internet packet domain. 3GPP further extended the Iu-protocol (in standard Release 4) by so called "flex-Iu" solution, which replaced the Iu point-to-point IP connectivity by the flexible point-to-multipoint connectivity. This important change was a true enabler for large changes in the core network implementation to split functions into server architecture. The server architecture enables distributed and light-weight server implementations of all core network functions. The communication between the servers happens by the IP protocols. This new architecture is fault-tolerant (may contain redundant servers) and it can easily be upgraded for increasing load. It can also be upgraded for new media types in the IMS solution [170].

6.3 System Architecture Evolution

UMTS architecture evolution, shown in Fig. 6.1 on the right, is rather called System Architecture Evolution (SAE) or Evolved Packet System (EPS) [9][173]. An overview of the EPS and its network elements is shown in Fig. 6.2. EUTRAN operates fully in the packet domain, it does not have a role for a centralized radio network controller and the eNodeBs route both the control plane and user plane packets via IP tunnels to the EPC. The Mobility Management Entity (MME) operates in the control plane and executes procedures like idle mode mobility, Paging, Authentication and setup of the *EPS bearer*. MME contacts Home Subscriber Server (HSS) for subscriber information and it is capable of storing the UE mobility context with cellular identifiers (SAE Temporary Mobile Subscriber Identity, S-TMSI) and the IP addresses.

In the user plane, the eNodeB packet routing and transfer functions implement the EPS bearer and connectivity to the serving Packet Data Network (PDN) gateways. These gateways have connectivity and routing capability of the Internet, and they contain the

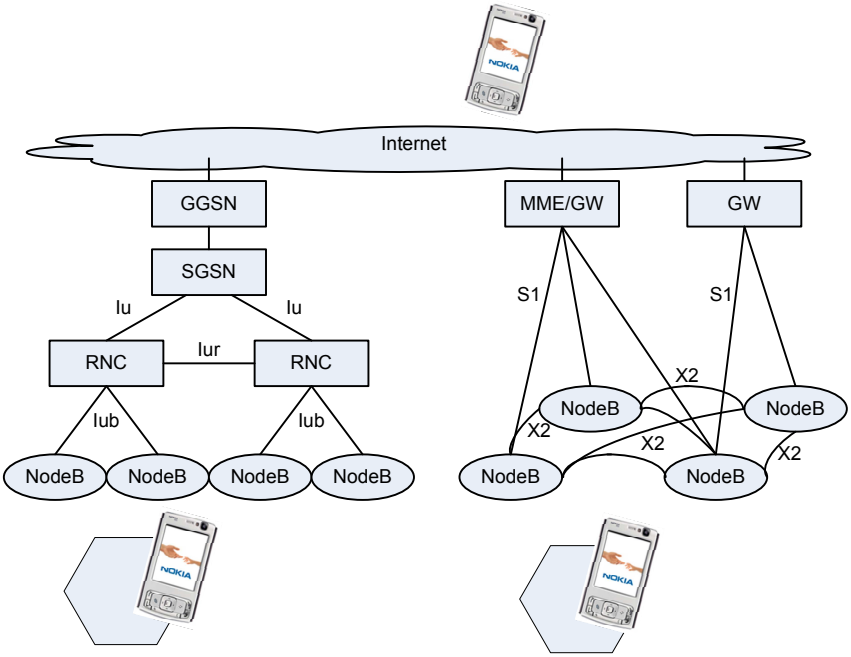


Figure 6.1: UTRAN architecture on the left and EUTRAN architecture on the right. UTRAN supports flexible packet interface over the Iu. EUTRAN supports flexible packet interface over the S1 and X2. The interfaces are specific to the nodes and specific to a UE connection, because IP-tunneling is used as the transport protocol.

Packet Data Protocol (PDP) context that the UE opens and configures for service connectivity and QoS handling. QoS functions are applied on the PDP context ports, where they may provide subscriber differentiation or traffic differentiation by the SAE berarer QoS parameters [174]. The QoS traffic classes are commonly known as the conversational, streaming, interactive and background treatment classes. The QoS parameters of EPS bearer are discussed shortly in Section 6.3.6.

The new interfaces of EUTRAN include S1- and X2-interfaces, which replace the Iub-, Iur- and Iu-interfaces of the UTRAN. S1-MME acts in the control-plane between the EUTRAN and the MME, and S1-U acts in the user-plane interface between the EUTRAN and the serving PDN gateways. The S1 Application Protocol (AP) implements the functions and protocols for EPS bearer management, Context transfer, Mobility, UE capability indication, Paging, *non-access stratum* signalling transport and S1 interface management. X2-interface acts between the eNodeBs by the defined X2 Application Protocol (X2AP), which is vital for mobility. X2 implements the functions and protocols for handover, load management and eNodeB configuration.

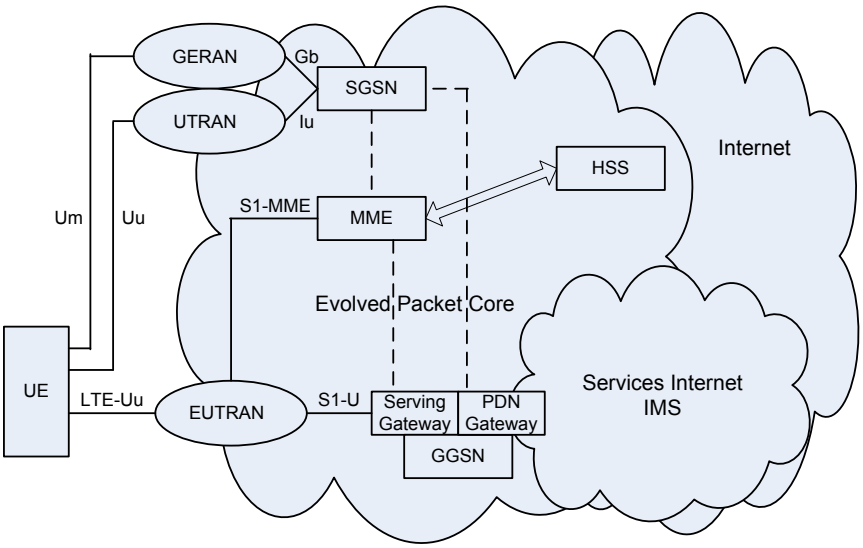


Figure 6.2: System architecture evolution from WCDMA/HSPA UTRAN towards LTE EUTRAN, and the Evolved Packet Core of the Evolved Packet System.

6.3.1 Addressing

Internet addressing is implemented either by IPv4 or IPv6 or both. This is known as the dual-stack approach, and sufficient network support exists to operate with either format. The key principle is always-on IP connectivity of the UE (users), and this is enabled by the establishment of a default EPS bearer already during the Initial Attach.

The default EPS bearer is established, whenever the UE switches from the *LTE_Idle* state to the *LTE_Active* state and connects to a PDN gateway. Any additional EPS bearer that is established to the same PDN is referred to as a dedicated EPS bearer. The PDN gateway selection function uses subscriber information to allocate a PDN gateway that shall provide PDN connectivity for the EUTRAN access. There is no distinction between handling the default and dedicated bearers in respect to the access network.

6.3.2 Attach procedure

The Attach procedure initializes implementation of the QoS, as it creates the Traffic Flow Templates (TFT) based on the subscribed QoS profile, present in HSS. Service differentiation happens by labeling and filtering the DiffServ fields of the IP packets (see Section 5.2). The priorisation, queueing and scheduling of packets (packet flows) in the network nodes happen according to the defined TFTs. The attach procedure is described in Fig. 6.3.

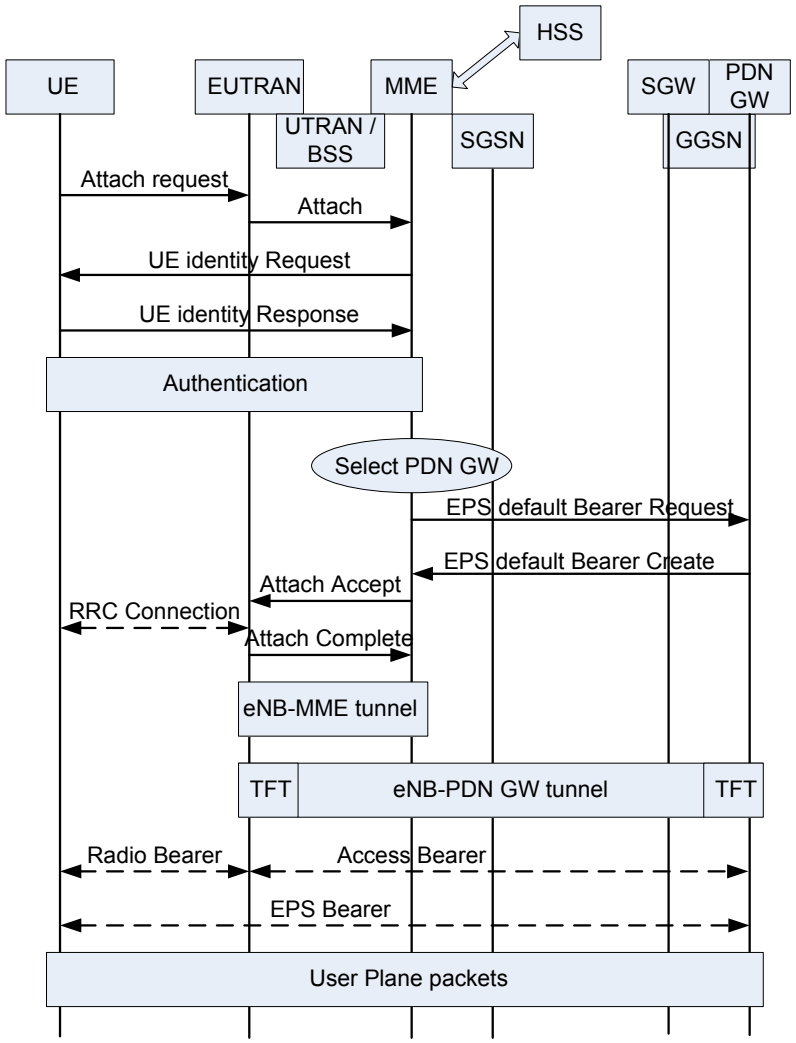


Figure 6.3: The principles of the Attach procedure. EUTRAN network elements are shown above and UTRAN network elements are visible for courtesy.

The Mobility Management Context includes the IP address of a PDN gateway and an Access Point Name (APN) for the subscribed UE. This may actually be a list of all APNs the UE is permitted to access with a default APN indication. In case, several APNs are listed for the UE, MME decides which APN(s) to use and MME constructs the UE context.

If the PDN subscription context does not contain PDN gateway address, the MME selects one of them and allocates an EPS Bearer Identity for the default bearer associated with the UE. MME sends *Create Default Bearer Request* to the selected serving gateway, where the serving gateway creates a new entry in the EPS bearer table, and the UE sends *Create Default Bearer Request* to the PDN gateway.

The Attach procedure may actually trigger one or multiple Dedicated Bearer Establishment procedures for Dedicated EPS bearers. During the Attach procedure, the UE may request for an IP address allocation. In this case, PDN gateway may act as a DHCP server and is able to assign IP addresses that the UE requested for. Fully IETF based mechanisms for IP address allocation are also supported. In this case, these protocols run entirely in the user-plane between the UE and the Internet.

6.3.3 Tunneling

MME sets up a tunnel between the eNodeB and the PDN gateway to serve the UE. The tunnel is recognized by the assigned Tunnel Endpoint Identifiers (TEID) and it may apply IP protocol for routing between the tunnel endpoints. IP tunneling is simply a way to encapsulate the IP datagrams to the payload of another IP datagram, whose Source Address and Destination Address are those of the tunnel endpoints respectively. Hence the tunnel end point identifier is an identifier of a unique IP address. Inside the tunnel, the tunneling headers of the IP packet implement the Differentiated Services QoS. Despite tunneling headers add overhead, they are necessary for easier networking, security and priority weighting.

Tunnel endpoint identifiers are retrieved by the MME and included to the Mobility Management Context of the UE. Tunnel endpoint identifiers let MME communicate with the serving nodes and allow setup or modification of tunnels for the use by those serving nodes, respectively. The Mobility Management Context incudes at least the unique identity of the UE over the S1 interface in the scope of the eNodeB and the unique identity of the UE over the S1 interface in the scope of the MME. These identities of the UE allow the user plane and control plane tunneling at the eNodeB and at the MME respectively. Other important endpoint identities and tunneling addresses are the identity of the MME and the identities of the serving PDN gateways.

6.3.4 PDP context activation

PDP context activation is requested by the UE. At this request, SGSN shall initiate procedures to set up the PDP contexts. The procedure includes subscription checking, APN selection and host configuration. Default PDP context is activated always, when the UE is switched to the LTE_Active state. Secondary PDP context activation reuses the primary PDP context parameters, except the QoS parameters which will be set for quality

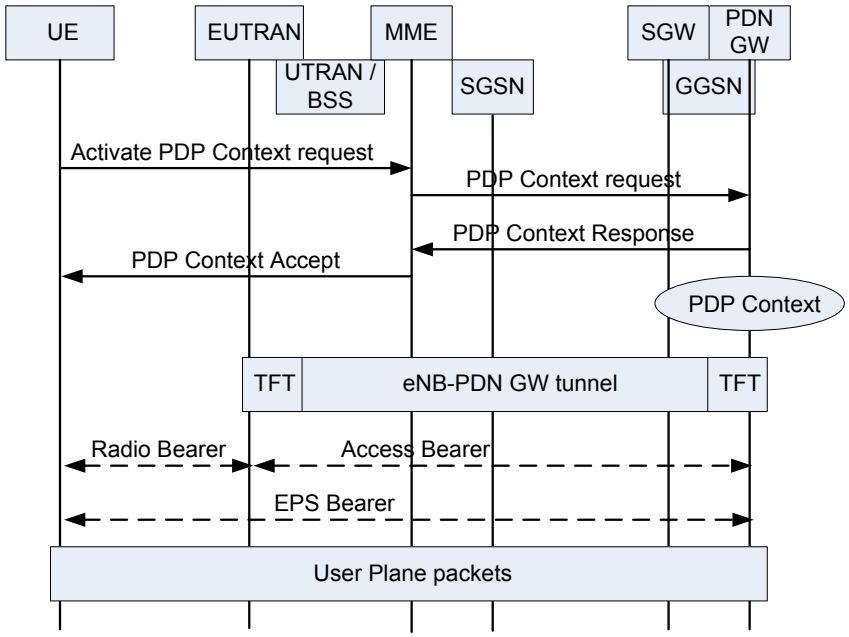


Figure 6.4: The principles of the PDP context activation procedure. EUTRAN network elements are shown above and UTRAN network elements are visible for courtesy.

differentiation (DiffServ). The PDP contexct activation procedure is described in Fig. 6.4.

6.3.5 Session management

After the default EPS bearer, tunnels and PDP context are set up, sessions may be intiated and terminated e.g. by SIP, see Section 5.2.1. An Application Server, like SIP application server, may host and execute services. SIP application server provides Internet Mobility Services so that it establishes sessions on behalf of the service sources. SIP application server communicates with the serving Call State Control Function (CSCF) or a proxy, which is the first contact point of UE in the service network. CSCF or the proxy may be found in the network by the IETF discovery procedures. SIP and SDP protocols will define the sessions and media-types to use in communication by the UE devices and servers. The generic address format is Uniform Resource Indicator (URI), which may be of type Uniform Resource Name (URN) or Uniform Resource Locator (URL). URI enables interaction with representations of the resource over a network. The URI format of resources is shortly, *username@realm*. The Domain Name System may convert such SIP names to IP addresses.

6.3.6 Bearer service model

The EPS Bearer Service model [175], see Fig. 6.5, divides the end-to-end service requirements into three parts, namely to the local bearer service, the UMTS bearer service and the external bearer service. The local bearer service contains the mechanisms to map the end-user service between the UE and the mobile termination of the service in the subscriber module. The UMTS bearer service contains the functionality to implement QoS over the UMTS network consisting of the UTRAN and the EPC. Within UTRAN, the bearer experiences changes as a function of time due to dynamic properties of the radio channels, radio resource sharing and mobility. However, the bearer service should be maintained seamless of those variations [45]. The core network bearer service is more stable due to large capacity and large volumes of delivered data. The core network bearers are gradually replaced by the Internet and the borders of RAN and core network are vanishing to the EPC with server/gateway architecture. The external bearer applies Internet Protocols defined by IETF. Conventional delivery in Internet has QoS, which is provided by so large capacity (overdimensioning) that it is experienced seamless. In a throughput limited network with increasing shares of differing service types, bandwidth-delay-product is not commonly sufficient for all services. Therefore, DiffServ [162] may provide QoS at need. DiffServ allows priorisation of packet flows between any two routing nodes by having Traffic Class field used in the IP packet header. Services and applications communicating in the packet domain over the EPS bearers are shown in Fig. 6.6.

UTRAN Edge Node contributes to DiffServ by acting as an Internet routing node. It preferably gets DiffServ configured as part of the PDP context parameters. DiffServ functions in UTRAN Edge Node may include classifiers, meters, markers, droppers and shapers for packet handling actions on the traffic flows [173]. These functions are implemented in concrete by the Traffic Flow Templates. All PDP contexts of a subscriber are associated with the same Mobility Management context for the subscriber. Mobility Management Context includes the subscribed QoS profiles and TFTs for each of the GPRS bearers, as said in Section 6.3.2.

In LTE, the EPS subscribed QoS profile is retrieved from the HSS. In the LTE_Active state both the MME and the serving PDN gateway will store the EPS bearer QoS profile with the quality parameters for every activated EPS bearer. The EPS bearer QoS parameters are [173], **QoS Class Identifier (QCI)** is a scalar that can hold one value at a time and is used as a reference to the access node specific parameters that control bearer

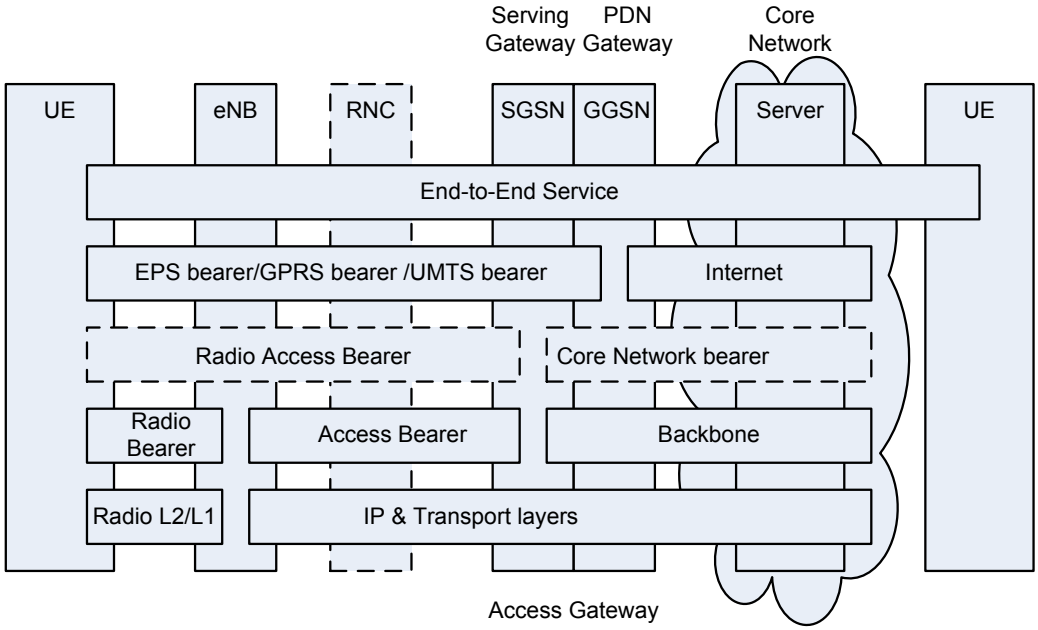


Figure 6.5: End-to-end bearer model of the WCDMA/HSPA UTRAN (dashed) and LTE EUTRAN. The system architecture evolution for LTE greatly simplifies the bearer model.

level packet forwarding treatment (e.g. scheduling weights, admission thresholds, queue management thresholds, link layer protocol configuration). A one-to-one mapping of QCI values to the defined characteristics are standardized by the 3GPP. These characteristics describe the packet forwarding treatment in terms of the resource type (guaranteed or non-guaranteed bit rate), priority, packet delay budget and packet loss (or packet error) rate [48].

Example services of different QCI values include conversational voice, conversational video, streaming, real-time gaming, IMS signalling, interactive gaming, interactive web browsing on Hypertext Transfer Protocol (HTTP), best effort traffic and Rich media[1]. The best effort traffic includes several prevailing applications and protocols running on TCP like; email using Internet Message Access Protocol (IMAP) or Post Office Protocol (POP), chat and File Transfer Protocol (FTP). On the radio interface and on S1, each Protocol Data Unit (PDU) (e.g. RLC PDU) is indirectly associated with one QCI via the bearer identifier. The same applies over the tunneling protocols be-

[1]Originally introduced in R.L. Daft and R.H. Lengel, Information richness: a new approach to managerial behavior and organizational design, *In: (Eds.) L.L. Cummings and B.M. Staw, Research in organizational behavior 6*, pp. 191-233, Homewood, IL: JAI Press, 1984.

tween the servers/gateways of EPS. Thus, each Service Data Flow (SDF) is associated with one and only one QCI. QCI describes the packet forwarding treatment that a service data flow experiences on the EPS bearer between the UE and the PDN gateway. **Allocation and Retention Priority (APR)** is the priority of the allocation and retention that only takes action for the bearer establishment and modification, occasionally also for bearer dropping. Once successfully established, ARP has no impact on packet level forwarding (e.g. scheduling and rate control). **Guaranteed Bit Rate (GBR)** is the bitrate expected to be provided on that bearer and **Maximum Bit Rate (MBR)** limits the bitrate, because excess traffic may be limited (as packet discard) by the rate shaping function. These parameters exist for each EPS bearer with the Guranteed Bit Rate property. **Aggregate Maximum Bit Rate (AMBR)** applies to a group of EPS bearers that share the same PDN connection, are capable of sharing capacity dynamically and provide non-GBR services.

In WCDMA/HSPA architecture [169], the *radio access bearer* is between the UE and the edge node of RAN. The *radio bearer* extends from the UE to the NodeB and further to the RNC. The radio access bearer is realised by a radio bearer and an access bearer over the Iub/Iur- and Iu-interfaces. The access bearer actually runs over the Iub/Iur- and Iu-interfaces, where the communication is separated to the contol plane and user plane. The control plane application protocols define cell configurations, transport channels, radio links, physical resource sharing, system information, measurements and handover. The application protocol over the Iu-interface manages connectivity of UTRAN to the core network, see e.g. [58]. Further, it configures the radio network [40] and transfers *non-access stratum* information. In the user plane, the *frame protocol* runs on top of the fixed network transport layers over the Iub/Iur-interfaces, see Section 6.5. The frame protocol is defined for different transport channel types. The frame protocol includes functions for the transport of MAC PDUs, outer-loop power control, retransmission control and frame quality indication [176].

For specific deployments, 3G implementation may be tailored so that it omits dedicated channels and only allocates shared (HSPA) channels, in addition to the common channels. This implementation avoids soft handover and simplifies the architecture by increasing NodeB functionality. Savings can be observed by simplifications of the radio access bearer, reduction of the signalling procedures and decreased communication delays. Yet another shared channel packet-switching method was found in [42]. Further improvements of HSPA

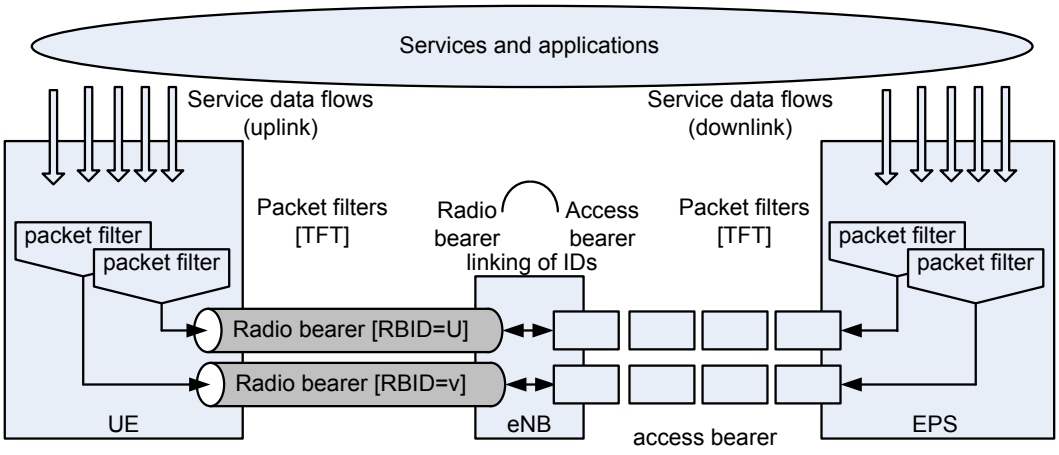


Figure 6.6: Services and applications communicating in packet domain over the Evolved Packet System. Packets of the service data flows are filtered by the Traffic Flow Templates. Two EPS Bearers are shown, each consisting of one radio bearer and one access bearer according to [173].

are addressed in [177], where protocol optimizations for VoIP are proposed.

6.4 Radio protocol architecture

The protocol architecture with the UE, RAN and core network with connectivity to the Internet is shown in Fig. 6.7. Services and applications run between the client and a server or between UEs of the mobile Internet. Mobility and Session management belong to the higher layers of the *non-access stratum*. The radio protocol architectures for WCDMA and WCDMA/HSPA are described in Sections 6.4.1 and 6.4.2.

6.4.1 WCDMA radio protocol architecture

In WCDMA architecture, the air interface includes macro diversity combining receiver as a key technique for improving the performance. This technique, however, causes a large impact to the RAN architecture and protocols. The peer-to-peer relationship of radio protocols in the WCDMA, further including HSPA, is shown in Fig. 6.8.

In downlink, macro diversity combining will require duplication of transport blocks in the RNC and delivery of transport blocks within a delay window to the buffers of all NodeBs belonging to the active set of the receiving UE. This is required so that the transmission of the physical code channel from different NodeBs in the active set may happen with the symbol-level accuracy experienced (over the radio interface) at the receiver

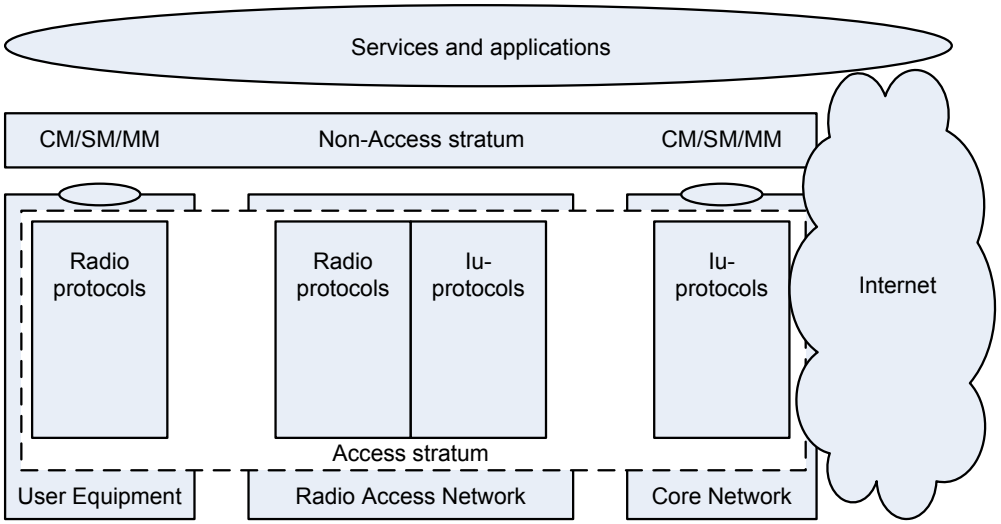


Figure 6.7: Protocol architecture for the *access stratum* and *non-access stratum*. Call Management/Session Management/Mobility Management appear in the higher layers of the control plane.

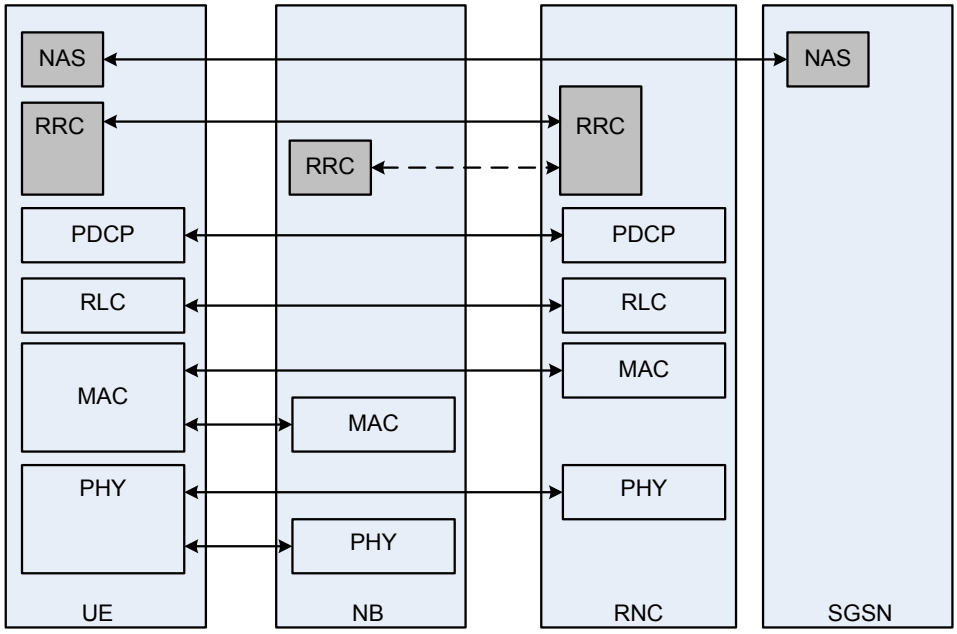


Figure 6.8: Radio protocols and their peer-to-peer relationship in WCDMA/HSPA. The *non-access stratum* protocols are marked as NAS, the others are radio dependent protocols respectively.

within the soft combining window. This required accuracy is much smaller than the frame period, and timing is adjusted in the NodeB by phasing the code channel relative to the local frame timing of the cell. The impact of macrodiversity and protocol termination for the dedicated channels in the UE and RNC are shown in Fig. 6.9.

To obtain the diversity combing gains in uplink, the transport blocks of a UE received

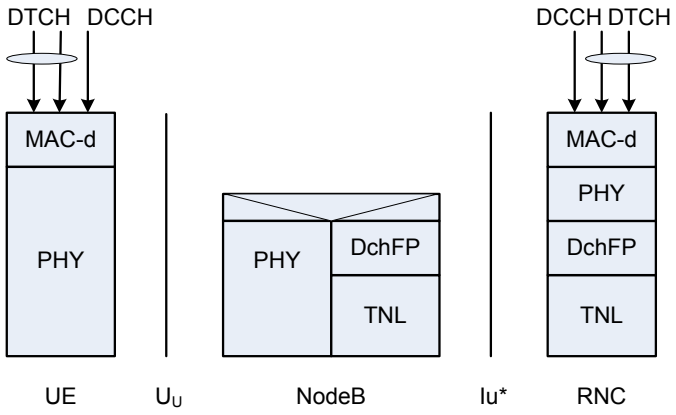


Figure 6.9: WCDMA dedicated channels having macrodiversity and terminating in the UE and RNC. The Frame Protocols for the dedicated channels (DchFP) run over the Iub/Iur/Iu-interfaces (Iu*).

by different NodeBs have to be selection combined in the RNC, before their MAC PDU contents is processed. This will require delivery of all transport blocks to the RNC. The frame protocol of different Iub-interfaces needs to provide the transport blocks of a radio frame to the combining window including the connection frame number, so that correct transport blocks can be properly selected without too long delays. The connection frame number runs commonly for a UE connection (to UTRAN), even if the system frame number over the radio interface in different cells may be uncoordinated.

Having all dedicated channels transmitted from the RNC, plenty of control plane information need to be transferred between the NodeBs and the RNC. Actually, this information needs to be transferred back and forth, because the Information Elements needed in the protocol messages are created and integrity protected in the RNC. Therefore, the information needed from the NodeB first has to be transferred to the RNC, where the actual protocol encoding (abstract syntax notation) and integrity protection happens, and then transmitted back to the NodeB inside the MAC multiplex of the transport channels. As described in Section 3.3 there exists alternatives for Rake receiver technique, e.g. MMSE, that could omit macrodiversity combining. However, UEs are by default expected to have a Rake receiver and therefore the architecture is defined for the macrodiversity.

6.4.2 WCDMA/HSPA radio protocol architecture

In WCDMA/HSPA architecture, the fast dynamic sharing of code-tree resources requires MAC in the NodeB. In the architecture, shown in Fig. 6.10, the serving RNC splits

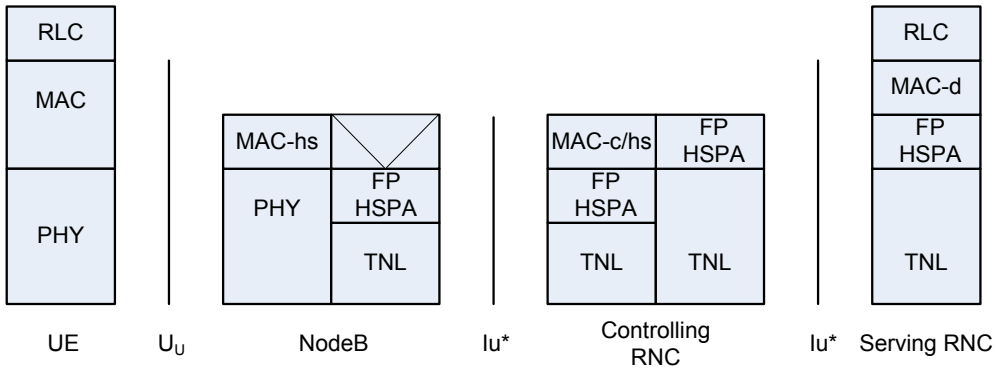


Figure 6.10: WCDMA protocol architecture with downlink HSPA. The serving RNC splits the transport channels for HSPA. The Frame Protocol for the HSPA-channels (FP_HSPA) runs over the Iub/Iur/Iu-interfaces (Iu*).

the transport channels for HSPA acting in the NodeB [178]. The frame protocol for HSPA-channels (FP_HSPA) runs over the Iub-interface to deliver the RLC PDUs without macrodiversity requirement to a single NodeB. The shared MAC (MAC-hs) creates MAC-hs PDUs, actually it creates a single MAC-hs PDU per short (2 ms) Transmission Time Interval (TTI) for a single HSPA transport channel, which typically consists of multicode transmission and achieves high instantaneous throughput, see Section 3.2. The shared channel operation allows scheduling at the NodeB and enables fast transport block creation from the MAC PDUs present in the NodeB buffer. Scheduling, (see Section 8.3) includes user selection, creation of the transport block (per short TTI) and link adaptation [179]. Fast scheduling of HSPA in the NodeB removes the possibility to maintain active set and apply macrodiversity soft combining. Yet, the active set and soft combining may be used in the concurrent dedicated channels as in WCDMA without HSPA.

Even though WCDMA/HSPA is not able to utilize macrodiversity gains on the HSPA channels, multipath diversity gains remain and fast scheduling gains add. This means that slow channel variation can be exploited as Channel Quality Indication (CQI) for the scheduler to optimize user selection and link adaptation algorithms.

The capacity of the HSPA transport channel is typically larger than that of a dedicated channel due to its low spreading factor and due to its multicode transmission. In HSPA, time-multiplexing allows a constant subtree allocation to be efficiently shared by all UEs served on the HSPA transport channels. Multiplexing of dedicated channels require either dynamic allocation of code-tree nodes or allocation of the lowest needed code-tree node for the highest peak data rate. This will lead to code tree resource limitations, which for

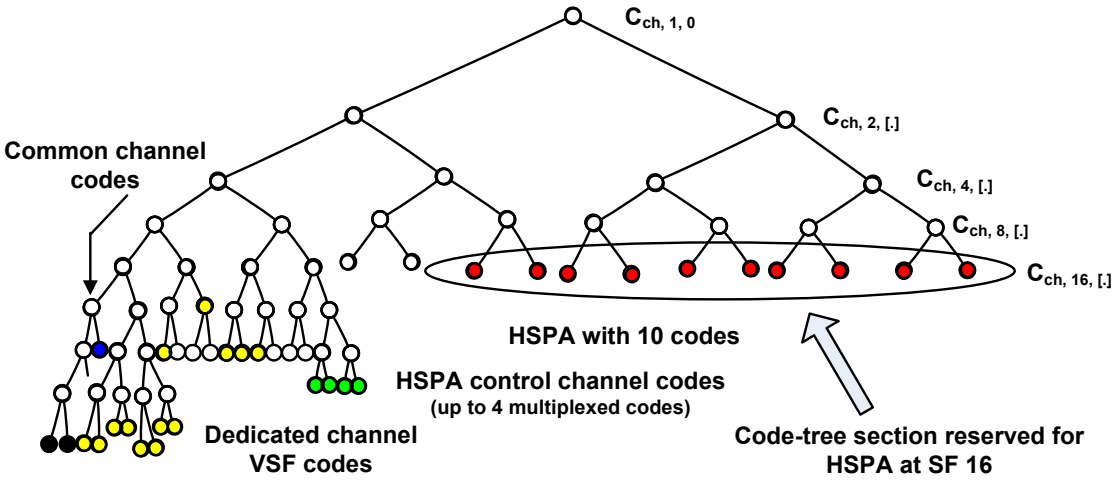


Figure 6.11: WCDMA channelization code-tree with common channels, dedicated channels (Variable Spreading Factor) and shared HSPA channels.

bursty peak rate transmissions will either cause frequent reallocation of code-tree nodes or will cause low spreading factor nodes be allocated permanently for the sake of serving peak rates and under-utilizing those spreading resources during average data rates [16][19]. The reallocation of code-tree nodes actually requires execution of the transport channel reconfiguration procedure (code handover), see Section 6.5.1. An example of the WCDMA channelization code-tree with allocations for the common channels, dedicated channels and shared HSPA channels is shown in Fig. 6.11.

The uplink HSPA is shown in Fig. 6.12, where one scheduling NodeB allocates radio resources and decides the transport format, but macrodiversity combining is still available in the RNC similar to that of the dedicated channels [180]. The frame protocol is specific to HSPA. Also in uplink, HSPA has multicode transmission. Multiple channelization codes may be multiplexed with equal spreading factor, so that they will carry a single transport block per TTI. Due to coverage limit, HSPA uplink may need to exist either in 2 ms or 10 ms TTI formats. Coverage of 2 ms TTI may be reduced by its discontinuous transmission ratio and less coding (higher channel code rate) compared to 10 ms TTI. Transmission on multiple channelization codes provides high instantaneous throughput, and it is a significant source of MAI interference. Therefore, transmission on uplink HSPA channels is typically bursty. In order to control capacity and interference load on short term at the receiver, the scheduler operates in a single serving NodeB per UE instead of

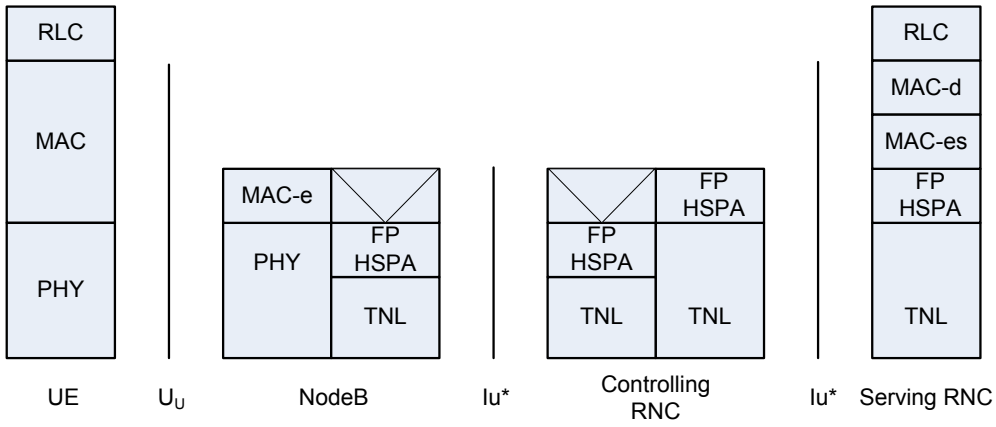


Figure 6.12: WCDMA protocol architecture with uplink HSPA. One scheduling NodeB allocates resources and decides the transport format combination, but macrodiversity selection combining is available from all receiving NodeBs in the serving RNC. The Frame Protocol for the uplink HSPA-channels (FP_HSPA) runs over the Iub/Iur/Iu-interfaces (Iu*).

scheduling in the serving RNC. Yet macrodiversity combining is feasible, because HSPA uplink may be received by multiple NodeBs in the active set of the UE.

6.4.3 LTE radio protocol architecture

LTE architecture is different from WCDMA/HSPA having no RNC and including the Internet packet SDU buffers already in the eNodeB. This architecture, shown for radio protocols in Fig. 6.13, allows coupling of segmentation decisions not only to the packet (SDU) sizes of traffic flows in the Internet but also to the channel state, short term channel dependent scheduling decision and transport format selection. Flexible segmentation tends to minimize the overhead in providing the best fit of the packet sizes in the queue to the best fit of the transport block size. Transport channel switching is neither a problem, as LTE transport resources are shared for all logical channels of a UE. For voice and video packets, segmentation is preferably avoided completely. This creates less protocol overhead and allows strict scheduling that satisfies the packet delay requirements. For data packets, segmentation is preferably tailored to the amount of data in the transmission buffer rather than to the actual packet (SDU) sizes. This creates less overhead per packet (SDU) and allows scheduling by greedy throughput weighting algorithms.

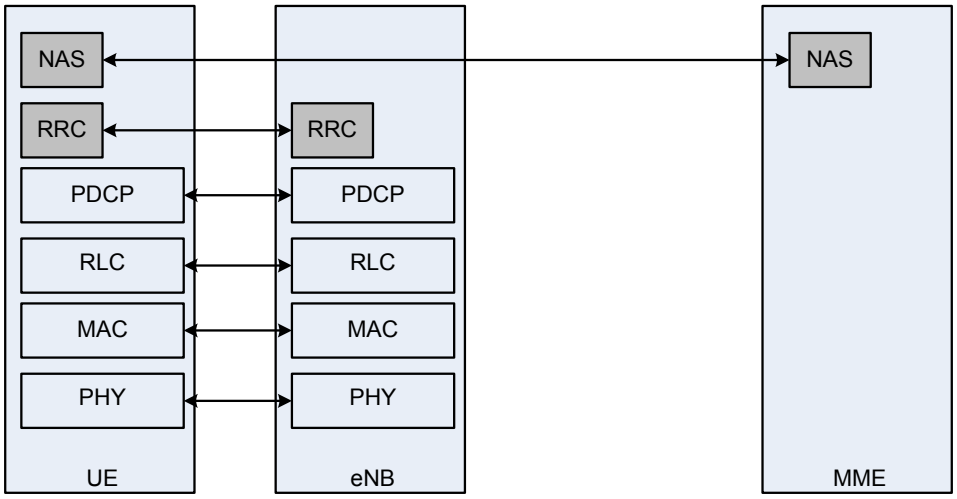


Figure 6.13: Radio protocols and their peer-to-peer relationship in LTE. The *non-access stratum* protocols are marked as NAS, the others are radio dependent protocols respectively.

6.5 Radio protocols

6.5.1 Radio Resource Control

Radio Resource Control (RRC) protocol [181][182] acts in the control plane over the radio interface. The RRC is responsible for procedures like the System Information Broadcast, Paging, Handover, Bearer management, Transport channel reconfiguration and Physical channel reconfiguration. The RRC is also responsible for the UE state handling and UE idle mode mobility.

The circuit-switched dedicated channels are classically signalled by the frequency channel number, timeslot number or channelization code indexes, which are valid unless reconfigured or released. Packet allocations, on the contrary, are signalled for the defined packet allocation period, on short term basis only.

In LTE, the allocation happens by the cell Radio Network Temporary Identity (C-RNTI), which is a unique identity of a UE in the serving cell. By detecting and decoding this identity, UE can confirm that the allocated resources are dedicated to it for the defined time interval. In packet communications, there is no commitment for the future allocations, just those assigned for a defined packet transmission period.

In WCDMA/HSPA packet allocations, C-RNTI is used as well. However, because of the macrodiversity connection over an active set of several cells, and because of several transport channel types, there has to be a family of additional RNTIs that provide a unique identity of UE for allocations on a wider scope than over a single cell. These identities

include HS-RNTI for the HSPA channels, S-RNTI for the serving RNC and U-RNTI for the UTRAN.

6.5.1.1 Radio Resource Control states of the User Equipment

The operation of the UE is primarily determined by the states. In WCDMA/HSPA, the UE state model is complex and state transitions cause latency to the protocol machines. For the operation in the packet-switched domain, LTE has the benefit of reduced number of states, namely the LTE_Idle state and the LTE_Active state. In the LTE_Idle state, the UE Mobility is handled by the MME. The procedures in the idle state include Tracking Area Update, Paging and RRC connection establishment. RRC connection establishment procedure consists of *RRC_ConnectionRequest* and *RRC_ConnectionSetup* messages, which may change the UE from the *RRC_Idle* state to the *RRC_Connected* state and as a consequence from the LTE_Idle state to the LTE_Active state.

In the RRC_Connected state, seamless mobility is provided and the UE has EPS bearer, at least the default EPS bearer, set-up between the UE and the PDN gateway. The necessary tunneling and transport layers are also setup via the serving eNodeB to the MME (control plane) and to the network servers/gateways (control plane/user plane).

LTE provides an efficient discontinuous operation mode (DRX) for UE power saving, although remaining in the RRC_Connected state. This kind of DRX operation avoids state switching and protocol setup delays, which are inherently needed in WCDMA/HSPA. DRX parameters are set and controlled by the RRC procedures, see e.g. [47].

LTE bearer service consists of the EPS bearer and an external bearer. The EPS bearer is implemented by the radio bearer and the access bearer. LTE does not require heavy bearer setup procedures over the air interface with the UE being involved, because the EPS bearer is setup at the request of the UE by the network procedures. The radio bearer (over the air interface) has a one-to-one relationship to a logical channel. The *RRC_ConnectionReconfiguration* message over the air interface between the eNodeB and the UE may actually convey all the information at once for the full radio resource configuration. This configuration includes the radio bearers and further mobility, security and measurment control information.

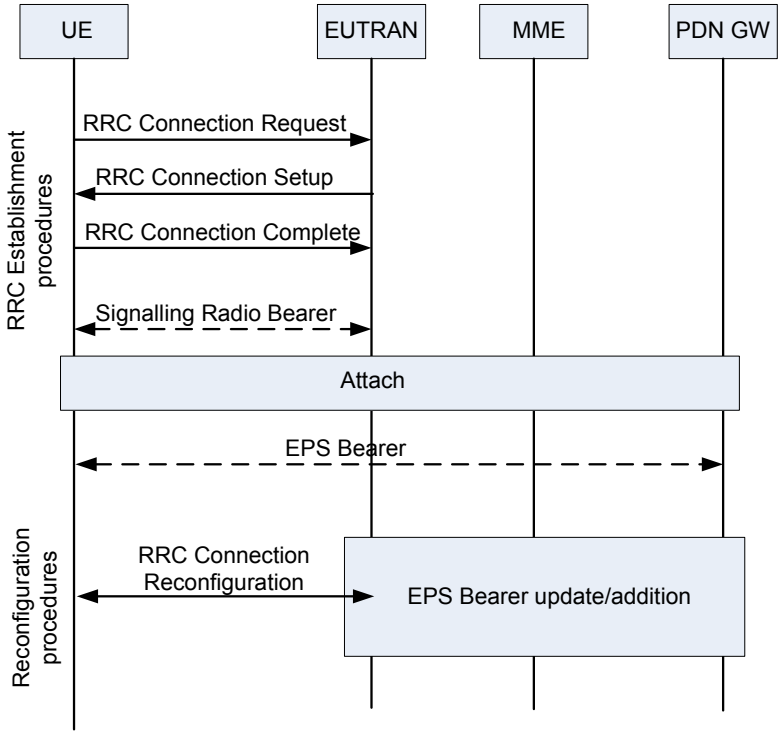


Figure 6.14: The principle of the RRC connection establishment and the RRC connection reconfiguration for the EPS bearer management.

6.5.2 IP packets and segmentation

Packet transmissions in the Internet are characterised by the variable length of packets and by their irregular interarrival processes. Ethernet provides a maximum transmission unit (MTU) and allows packet length variation up to that constraint. Networking protocols are able to discover the MTU along the path of an end-to-end connection during the three-way handshake of a connection setup or by a path MTU discovery procedure. Connection setup may assume a default MTU of 576 bytes or 1500 bytes supported by all Ethernet links.

6.5.3 Packet Data Convergence Protocol

The header compression described in Section 5.2 operates at the Packet Data Convergence Protocol [183][184]. For every packet flow that applies header compression, there is a compressor entity and a decompressor entity. In WCDMA/HSPA architecture, the compression entities of the network are in the RNC, whereas in LTE they are in the eNodeB. The peer compression entity is in the UE for decompression of the downlink packets and for compression of the uplink packets. The peer protocols identify the compression

context [43][57].

6.5.4 Radio Link Control

The function of the Radio Link Control (RLC) protocol is to handle the Internet SDUs (TCP/IP or UDP/RTP/IP packets after the PDCP processing) i.e. to concatenate, segment and create formatted PDUs to appear at the logical channel interface as the MAC SDUs. The RLC controls the flow of PDUs and takes care of retransmissions and reordering, so that the peer-to-peer delivery of the RLC SDUs is as seamless as possible, or meets the required QoS requirements of the bearer. The RLC and TCP are analysed e.g. in [185][186][187].

The RLC entity operates in one of the two modes, *Acknowledged Mode (AM)* or *Unacknowledged Mode (UM)* [188][189]. Further, a *transparent mode* RLC is available without any RLC functionality for the use of the Paging and the System Information messages, which are present on the signalling radio bearer.

The RLC in the Acknowledged Mode includes peer-to-peer protocol with the RLC *state variables*, which are shared and maintained by the transmitter and the receiver. The state variables require peer acknowledgements and allow selective retransmissions and flow control. As the receiver has perfect information about the correctly decoded transport blocks, its duty is to inform transmitter about the needed retransmissions and status of the receiver window. The transmitter maintains all RLC PDUs inside its window until they are positively acknowledged or until their discard timer has triggered. The transmitter window can be dimensioned so that it discards packets already at the transmitter. The transmitter discard function will create a congestion-like situation for the TCP networking protocol. This will cause TCP reduce its segment generation activity to half (as dropped segments will not trigger acknowledgements) and will drop the bitrate of that flow respectively. On the other hand, bitrate increase will happen by linear increase of the segment generation activity by the TCP, after the exponential slow start. The major difference of the receiver discard vs transmitter discard functions is loading of the radio link. The receiver discard applies for low delay and to keep the memory space small, whereas the transmitter discard avoids overloading.

The RLC in the Unacknowledged Mode maintains only local state variables, neither feedback signalling nor retransmissions are supported. Both RLC modes provide segmentation of SDUs and insequence delivery. If the RLC insequence delivery is not set, the

RLC is capable of delivering full SDUs instantaneously to the higher layers in the order they are present in the receiver RLC PDU window despite of missing RLC PDUs. This process may reduce the SDU average delay observed at the RLC layer by increasing the number of small delay observations and decreasing the number of large delay observations. However, it may not reduce the actual delay of the largest delay observations, which are a consequence of missing PDUs causing incomplete SDUs remaining in the window waiting to get completed. On the other hand, if the RLC insequence delivery is not set and even if it may make the RLC average delay smaller, it does not necessarily reduce the end-to-end average RTT. This is because the RLC SDUs may appear faster into the TCP Segment Window, which therefore advances fast and may drop packets that will cause end-to-end retransmissions instead of the RLC retransmissions. It is also part of the memory management design, how much memory is allocated for the radio protocols, how much for the TCP. A good convention is to design memory for the radio protocols according to the radio capabilities, in terms of the supported throughput (as the maximum number of transport channel bits per TTI) and have a larger memory for the TCP machines. This design allows maximal exploitation of the radio capabilities up to a designed throughput category. On the other hand, design like this tends to avoid activation of radical congestion control algorthms of TCP during a session.

6.5.4.1 Segmentation in WCDMA/HSPA

In WCDMA/HSPA, segmentation of the network SDUs is done at the RLC in a node that is distant from the radio interface (i.e. in the RNC). As the channel state is not known, the PDU size cannot be tuned accordingly, and hence the segmentation and concatenation to the fixed size PDUs is defined instead. As the packet sizes in the Internet depend on the type of the application (voice, video, data) and on the networking protocol (UDP/RTP or TCP) it is beneficial to tailor the PDU sizes per traffic flow instead of having a single size PDU defined. This is feasible, because the traffic flows can be multiplexed to different logical channels and each of the logical channels may have a specific RLC parametrization. In WCDMA/HSPA, the RLC PDU sizes may only try to comply to the expected average behaviour of the higher layer protocols, as the channel state is not known at the time of segmentation. Setting the PDU size constant per logical channel allows instantaneous switching of the logical channel between the transport channels without resegmentation. Handling the RLC PDU reception in sequence over a radio link with errors requires a num-

ber of state variables, windows (both at the transmitter and the receiver) and reordering buffers for the assembly of the original SDUs in the receiver.

6.5.4.2 Segmentation in LTE

In LTE, segmentation is done in the eNodeB, where rather accurate knowledge of the channel state is available. The resources allocated for the radio link and the link adaptation format are decided immediately on short term matching both to the data transmission needs, resources available, number of links sharing the common resources and the expected channel quality. On the other hand, all transport channel resources are available as the shared channel. The RLC segment sizes are flexible and the segmentation headers offer minimum overhead. This implies that small packets like VoIP will have no segmentation, and the variable PDU size can accommodate variable length VoIP packets without large padding fields. For variable length large data packets, RLC PDU can include as large number of bytes from the data buffer as possible, without matching to the SDU sizes. In this case, one PDU may carry several concatenated SDUs, even large ones, or it may carry as large segments of a single SDU as possible. The flexibility of the PDU size allows a reasonable instantaneous choice that fits to the selected transport block size without padding. In case of individual SDUs waiting in a transmission buffer, the PDU size and resource allocation may match the SDU size as accurately as the resolution of the transport block size allows. If the channel conditions, resource allocation or link adaptation formats change for the retransmission, the RLC segment size may be changed without resegmentation, because the "segment offset" mechanism [52] may indicate new segments with updated segment length relative to their length and position in the original PDU byte-buffer. All this flexibility [46] was not possible in the fixed size PDU solution of the WCDMA/HSPA.

6.5.5 Medium Access Control and logical channels

MAC layer provides data transfer services on the logical channels. A set of logical channel types is defined for different kinds of data transfer services offered by the MAC. Each logical channel type is defined by what type of information is transferred, see Fig. 6.15.

MAC provides logical channels, manages transport channels and multiplexes logical flows. MAC-d operates on the dedicated channels and MAC-c on the common channels. Dedicated logical channels are Dedicated Control Channel (DCCH) and Dedicated Traf-

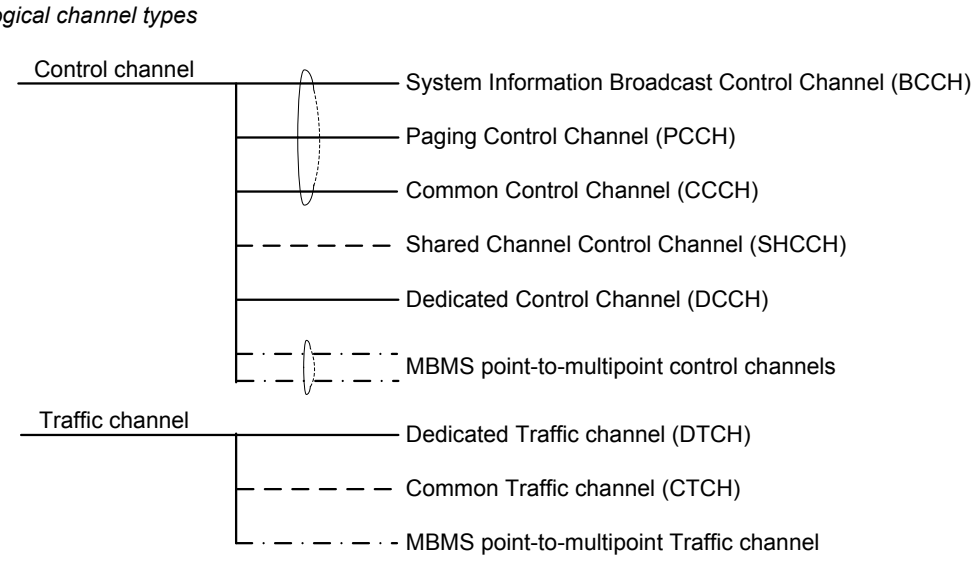


Figure 6.15: Logical channel types for WCDMA/HSPA and LTE.

fic Channel (DTCH). Common logical channels include the Broadcast Control CHannel (BCCH) and Paging Control CHannel (PCCH) in downlink and Common Control CHannel (CCCH) in uplink.

6.5.5.1 WCDMA/HSPA MAC

In WCDMA/HSPA UTRAN, MAC [190] entities are distributed to the NodeB and RNC, by the necessity of the macrodiversity property as described in Section 6.4.1. The MAC functionality in the NodeB for all transport channel types is shown in Fig. 6.16. The MAC functionality for all transport channel types in the UE is shown in Fig. 6.17, however, this functional split remains quite fictive for the real implementation in the UE.

MAC-d handles the dedicated channels and it interacts with the RLC. MAC-d is uniform for dedicated and shared channel types. As Mac-d is distant to the radio interface, it has to assume constant size RLC PDUs per logical channel, which it can effectively pack as a transport block set into the code channel. MAC-d provides transport channel type switching, scheduling and priority handling of queues. MAC-d handles in particular the DCCH and DTCH logical channels. The DCCH carries RRC messages on the signalling radio bearer. The DTCH carries any SDUs available from the user plane of the higher layers, which are optionally compressed by the PDCP and segmented by the RLC.

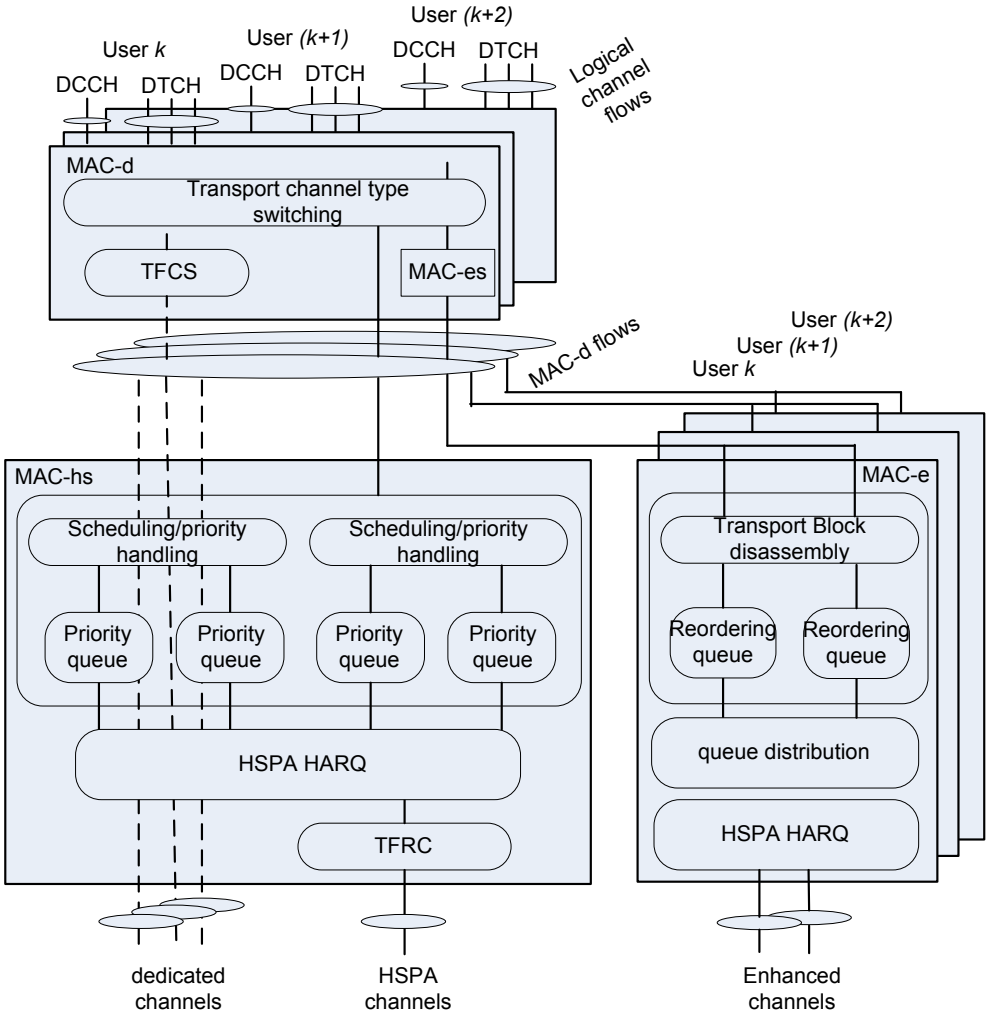


Figure 6.16: An example of MAC functions in the NodeB, HSPA shown for downlink transmission and uplink reception.

MAC-hs/es/e/is/i manage the logical channel transport especially on the shared transport channels in the HSPA downlink and on the enhanced transport channels in the HSPA uplink. MAC-hs/es/e/is/i thus provide scheduling, priority handling, retransmission, reordering and multiplexing functionality to make the shared and enhanced transport channels transparent to the MAC-d. The MAC-d will benefit of the enhanced transport services managed by the MAC-hs/es/e/is/i and scheduled by the NodeB. The distributed MAC architecture is dictated by the macrodiversity combining property in UTRAN.

MAC-c handles the common transport channels, mainly Broadcast Channel (BCH), Paging Channel (PCH), Forward Access Channel (FACH) in downlink and Random Access

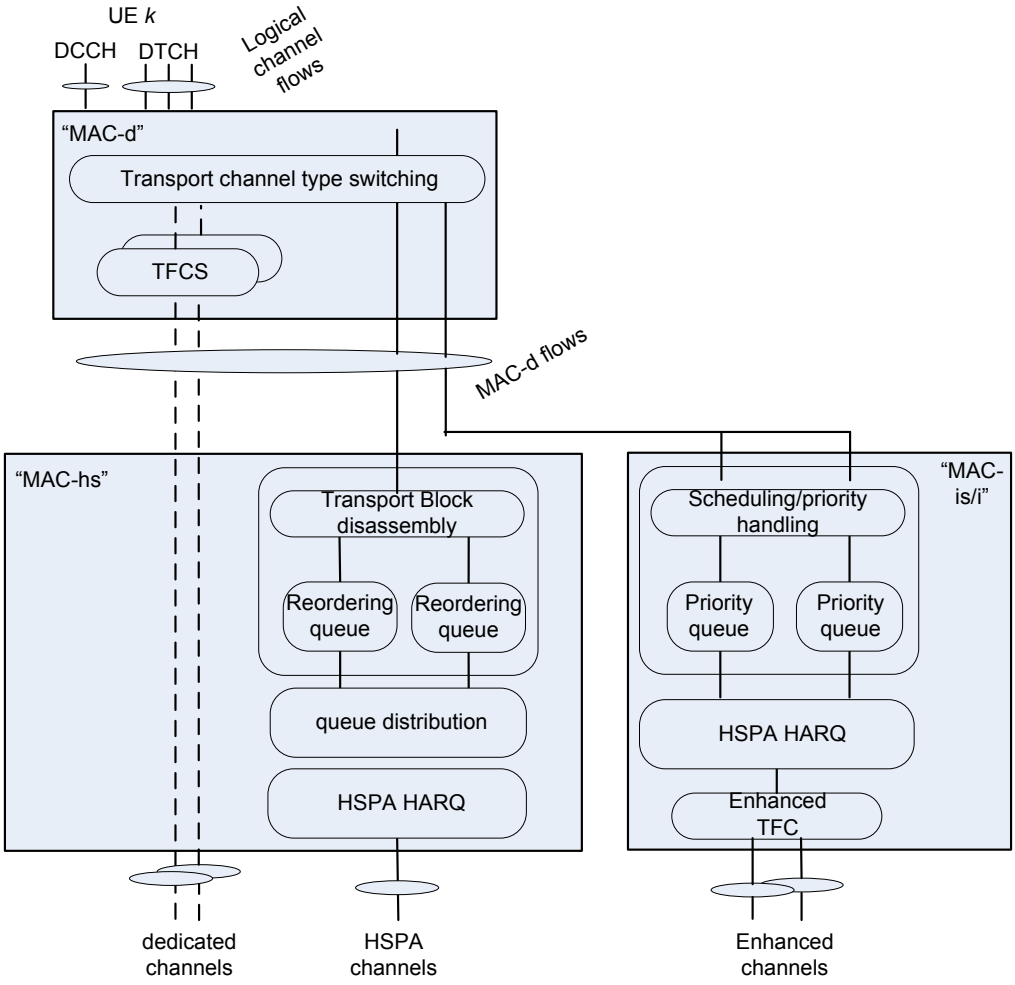


Figure 6.17: An example of MAC functions in the UE, HSPA shown for downlink reception and uplink transmission. Split of MAC to MAC-d/hs/is/i is fictional only and implementation specific in reality.

Channel (RACH) in uplink.

MAC-hs provides HSPA protocol in downlink and multiplexes UEs fast per 2 ms TTI. As the RLC PDU size is fixed per logical channel, MAC-hs can form a transport block per TTI without dynamic interaction to the distant MAC-d. The only concern is flow control between the MAC-hs and the MAC-d so that the MAC-hs window has sufficiently high utilization that avoids starvation of the physical allocations. The MAC-d transmitter keeps the duplicate MAC-d PDUs stored until they are acknowledged by the peer. The MAC-d window update rate is dependent on the MAC-hs activity.

MAC-es provides the HSPA protocol in uplink and it demultiplexes PDUs for MAC-

d. MAC-es is needed in the serving RNC for macrodiversity selection combining and for reordering of the MAC-es PDUs. There is no flow control over the Iub, as the MAC-es PDUs inherently appear into the common receiver window of the receiving MAC-d. The acknowledgements of the HSPA channels from the NodeB to the UE transmitter are fast physical layer signals.

MAC-e is the HSPA protocol for uplink in the NodeB. It boosts the transport channel capacity by using the enhanced HSPA physical channels in uplink.

The MAC in the UE triggers MAC-d PDUs and creates a single MAC-es PDU per logical channel for a TTI. Each MAC-es header carries the *transmission sequence number* for reordering. Multiple MAC-es PDUs from multiple logical channels may be multiplexed to a single MAC-e PDU transmitted as a single transport block on the E-DCH during a TTI. MAC-e header indicates the number of PDUs in the payload and for each PDU its logical channel identity and the MAC-es SDU size. The MAC-e in the NodeB handles scheduling, retransmissions and demultiplexing of the MAC-e PDUs into the MAC-es PDUs. The MAC-d entity in the RNC receives the MAC-d PDUs extracted by the MAC-es that were received from the shared transport channels, in addition to those MAC-d PDUs that were received from the dedicated transport channels.

MAC-is in the UE concatenates multiple MAC-d PDUs into the MAC-is PDUs for each logical channel and multiplexes them into a single MAC-i PDU per TTI. The MAC-is generates the transmission sequence number for each logical channel for reordering at the peer MAC.

MAC-i includes the retransmission entity and provides information about the transport format combination of the E-DCH. The retransmission entity controls retransmissions of the transport blocks, thier redundancy version and the power offset between the channelization codes.

6.5.5.2 LTE MAC

In LTE, MAC [191] network architecture is not distributed, and there exists a single MAC entity in the eNodeB common to all served UEs according to Fig. 6.18 and a single MAC entitity in each UE according to Fig. 6.19. Of course, still the common channel and shared channel functionality can be partly separated. MAC handles the data transfer and resource allocation for the logical channels on the shared transport channels, both in downlink and in uplink. The MAC largely contributes to the transport efficiency by

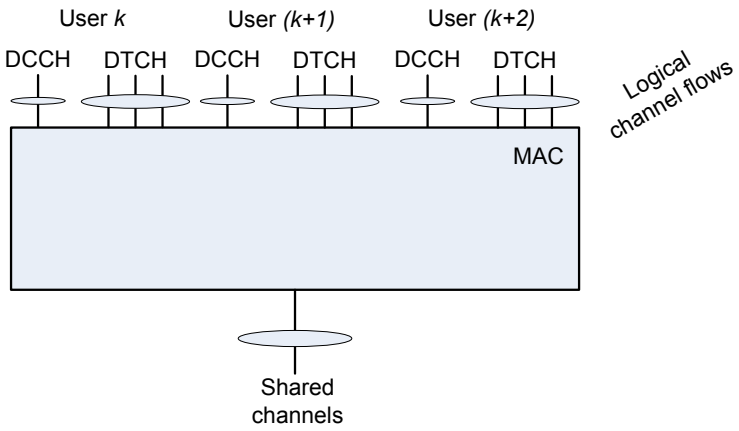


Figure 6.18: LTE MAC in the eNodeB. Fast scheduling and time-frequency multiplexing of shared channels.

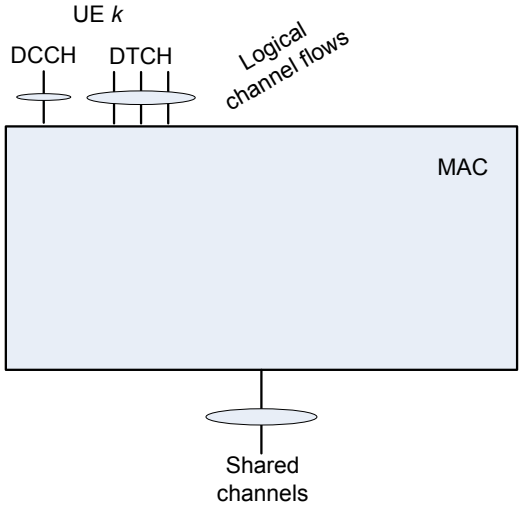


Figure 6.19: LTE MAC in the UE. Fast scheduling by the eNodeB and time-frequency multiplexing of shared channels.

scheduling and priority handling of the queues. Due to the shared transport resources, MAC is fully flexible in creating the MAC PDU (i.e. the transport block) by multiplexing the flexible size SDUs from multiple logical channels into a single transport block. Buffer Status Report (BSR), Power Headroom Report (PHR), Discontinuous Reception mode (DRX) command and Timing Advance (TA) are MAC control elements that may be multiplexed with the MAC SDUs to any transport block. The MAC functions also take care of mapping the logical channels onto the common transport channels namely BCH, PCH and RACH.

The logical channel data flows through MAC are described by the processing structure of the Layer-2 protocol functions in Fig. 6.20 for the eNodeB, and in Fig. 6.21 for the UE.

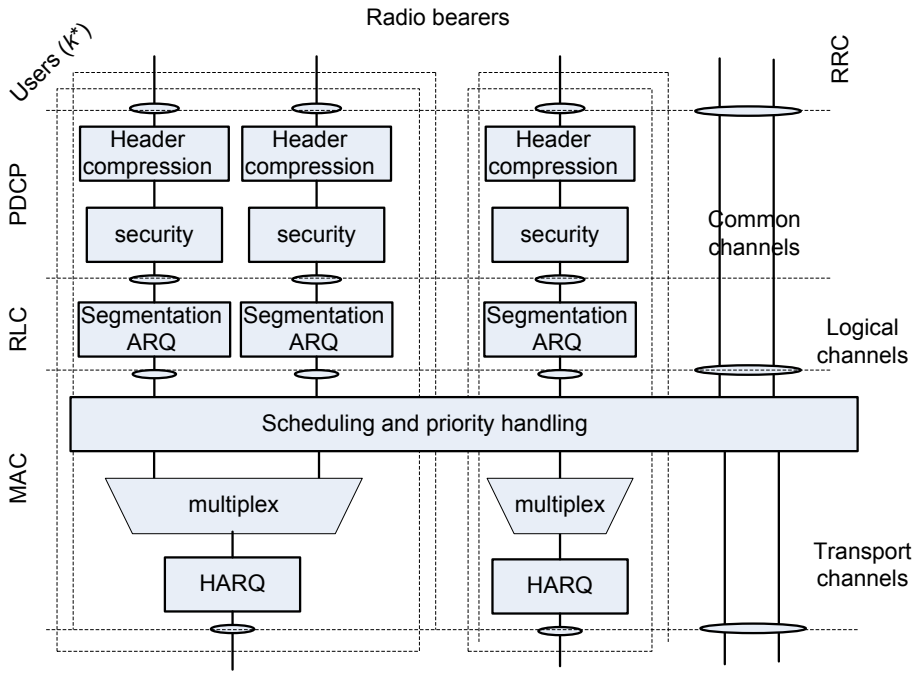


Figure 6.20: Processing stucture of Layer-2 protocols in the eNodeB.

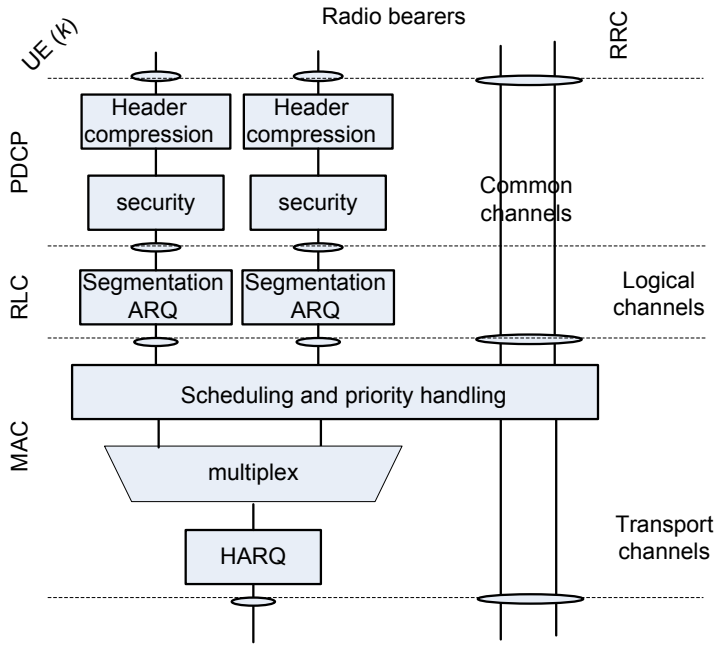


Figure 6.21: Procssing stucture of Layer-2 protocols in the UE.

6.5.6 Transport channels

Transport channels define the interface between the physical layer and the MAC. A transport channel is defined by how and with what characteristics data is transferred over the air interface. Properties of the transport channels contribute largely to the performance in the measures of throughput, spectral efficiency, packet delay, interference, energy consumption, computational complexity and even the architecture [192][193][105].

The transport channels may be classified to common, dedicated and shared channels. The common channels carry cell global or system global information for all receiving UEs, or in the case they carry specific information for a single UE, the common channels will use explicit inband identification of that UE. The common channel formats, at least the system broadcast, need to be generically known from the standard specification. The access to the common channels is provided over the full coverage area. Transport, multiplexing and identifying UEs on the common channels is inefficient, cause delays and consume radio resources more than on the dedicated or shared channels. Therefore, the common channels typically carry small payloads (from single bits to tens of bytes). The common channels are used primarily by the UE to access the network, to initate sessions and to wake-up from inactivity. This enables that the idle mode operation of a UE becomes feasible otherwise, whenever the communication is not truly active. The idle mode operation saves UE power resources and makes the mobility tracking procedures of a UE light for the network.

The dedicated channels have inherent addressing of the UE, i.e. it uses those dedicated resources (e.g. a channel or resource number) that are granted to the UE by a signalling procedure. Grant of the dedicated resources is reliable, but it implies delays. A dedicated resource assignment is valid for a long time (from seconds to tens of minutes). The allocation is either for a determined period or it may be infinite but terminated by a release procedure. The dedicated channels may carry different traffic types that are efficiently multiplexed to result variable amounts of transport data (from tens of bytes to a few kilobytes) for a single UE.

The shared channels use explicit UE addressing, which is present in a shared control channel that immediately assigns resources on the shared data channel for a short time (of milliseconds) without a signalling procedure. The shared control channels are efficient and easy to receive, they save radio resources and UE processing power. The shared channels can assign payload on the shared channel, where the payload may range from small bursts (tens of bytes) to very large bursts (tens of kilobytes) with high granularity. The flexibility

of the shared channel transport is crucial for the efficiency of packet multiplexing.

Yet another class of transport channels may exist for Multimedia Broadcast. These transport channels are known as the broadcast channels and broadcast control channels. They may be included to, or excluded from the system without impact to the other communication channels, thus they are kind of add-ons. The broadcast transport channels serve multiple users with identical data contents, and the receivers may be spread over a large service area (including many cells). The broadcast transport saves radio resources by point-to-multipoint transport instead of providing transport on multiple point-to-point links. The broadcast channels may carry data intensive multimedia services, if sufficient amount of radio resoures is available. Also narrowcast and simulcast are feasible. The broadcast control channels announce broadcast resources and may assist service (stream) selection from the broadcast composition.

WCDMA offers a large set of common and dedicated transport channel types, whereas WCDMA/HSPA in addition offers the shared transport channel types. LTE provides a minimal set of transport channel types, mainly the shared channel types only. This provides large gains in flexibility and efficiency for LTE. The characteristics of different transport channel types are described for the WCDMA in Section 6.5.6.1, for the HSPA in Section 6.5.6.2 and for the LTE in Section 6.5.6.3 respectively. The properties of the transport channels are further depicted in Fig. 6.22.

6.5.6.1 WCDMA transport

In WCDMA, the combination of the transport channels triggered per TTI is called the *coded composite transport channel*, and its contents during the frame period is indicated jointly by the Transport Format Combination Indicator (TFCI) [192].

For triggering each transport channel per TTI, the MAC decides a transport format for every transport channel from the allowed *transport format set* of that transport channel [193]. The *Transport format combination* then uniquely defines the set of transport formats for the combination of all active transport channels during the frame. The transport format combination is indicated to the receiver by the TFCI, which is transmitted on the physical control channels during the frame.

For SDUs, which don't have a fixed timing relation, a variable number of variable size transport blocks may be triggered per frame, depending on the scheduling decisions. For scheduling efficiency, TTI is often short e.g. one frame. For SDUs, which have a regular

interarrival process, as video stream or VoIP, one transport block will be triggered per TTI and its transport Format remains constant over the frames of the TTI. For constant delay, TTI is often long e.g. 4 frames for voice and tens of frames for video.

6.5.6.2 HSPA transport

In HSPA, where the physical channel capacity may be much larger than in WCDMA and where the resources are time multiplexed, only one transport block per TTI is triggered. Thus, TTI is common and fixed for the HSPA transport channel. Otherwise, time domain scheduling would not have the required degree of freedom. One convenience for the HSPA protocol efficiency is the limitation to trigger the logical channels of equal PDU size only into the same TTI.

Allocation and control of HSPA resources

In HSPA, fast scheduling and time multiplexing requires a control channel that operates in phase to the High-Speed Downlink Shared Channel (HS-DSCH), which carries the transport block. This control channel, the High-Speed Shared Control Channel (HS-SCCH), utilizes shared channelization code resources, and it appears time multiplexed for different users similar to the HS-DSCH. The HS-SCCH announces the *transport format and resource combination*, HARQ related information and scheduling control in terms of the allocated UE identity. The signalling overhead of fast scheduling the UEs would increase by the inclusion of the necessary UE identification. However, in HSPA a UE specific scrambling mask is applied on top of the CRC detection check, which adds no scheduling overhead on the HS-SCCH code channel.

UE operation for HSPA reception

For HSPA reception in downlink, the UE will continuously process the HS-SCCH and filter its UE specific scrambling mask. After the correct detection of the UE identification on the HS-SCCH, the receiver knows that a correctly decoded signalling block is usable. The HS-SCCH is leading the HS-DSCH transmission so that the UE can decode the first part of the HS-SCCH to get the demodulation and decoding information for the HS-DSCH. Only if the UE gets a scrambling mask hit on the first part of the HS-SCCH, decoding of the second part of the HS-SCCH is needed and hence the decoding of the HS-DSCH will follow thereafter.

Physical transport and timing

In HSPA, all the signalling for fast scheduling is instantaneously available from the

HS-SCCH that is code-multipexed with the HS-DSCH. Phasing of the HS-SCCH relative to the HS-DSCH will set timing constraints for decoding at the UE receiver. For HSPA, the fast retransmission of code blocks is defined as Hybrid Automatic Repeat reQuest (HARQ) process, where the retransmissions of the code blocks increase decoding performance over the variant channel [194][195][196][197]. The CQI and HARQ feedback transmitted from the UE to be decoded by the NodeB before scheduling the next TTI will set the timing constraints for decoding and scheduling at the NodeB [194][198][6]. This will imply the minimum round-trip-time requirement for the HS-DSCH. The processing time requirements are implicitely mandated by the standard specification, which sets the minimum number of HARQ processes i.e. the minimum time-distance of the HARQ process to 6 TTIs (12 ms). If multiple HSPA transmissions are allowed per 2 ms TTI, each code multiplexed resource has to be controlled by a separate HS-SCCH. This is obvious, because the timing relationship of the HS-SCCH and HS-DSCH is fixed and because a single code channel can only carry a single signalling instance per TTI.

Uplink HSPA operation

For HSPA transport in uplink, Enhanced Dedicated Channel (E-DCH) creates a single transport block per TTI carried on a number of physical HSPA code channels with a mutually equal spreading factor. The UE operation in transmitting the HSPA channels and the NodeB operation in receiving the HSPA channels do not largely differ from the conventional processing of transport channels in other respects than by largely increased processing and buffering requirements. The HSPA transport in uplink may use HARQ retransmissions similar to the downlink HSPA, however for the HARQ in uplink, the NodeB provides the feedback signalling. The HSPA implementation in uplink has a large impact to the MAC protocol, as described in Section 6.5.5.1.

6.5.6.3 LTE transport

In LTE, the transport channels are reduced for the sake of efficiency. The MAC and the physical layer processing take care of almost all of the transport functionality as well. The allocation and control of the transport resources is decided by the scheduler in the eNodeB. Fast signalling of the shared downlink transport is present on the shared physical control channel that is present in the same transmission frame as the actual transport block, see Section 6.5.7.3 and Fig. 8.1. The feedback for the downlink transport is provided in the physical uplink control channel, if not present multiplexed with the uplink data on

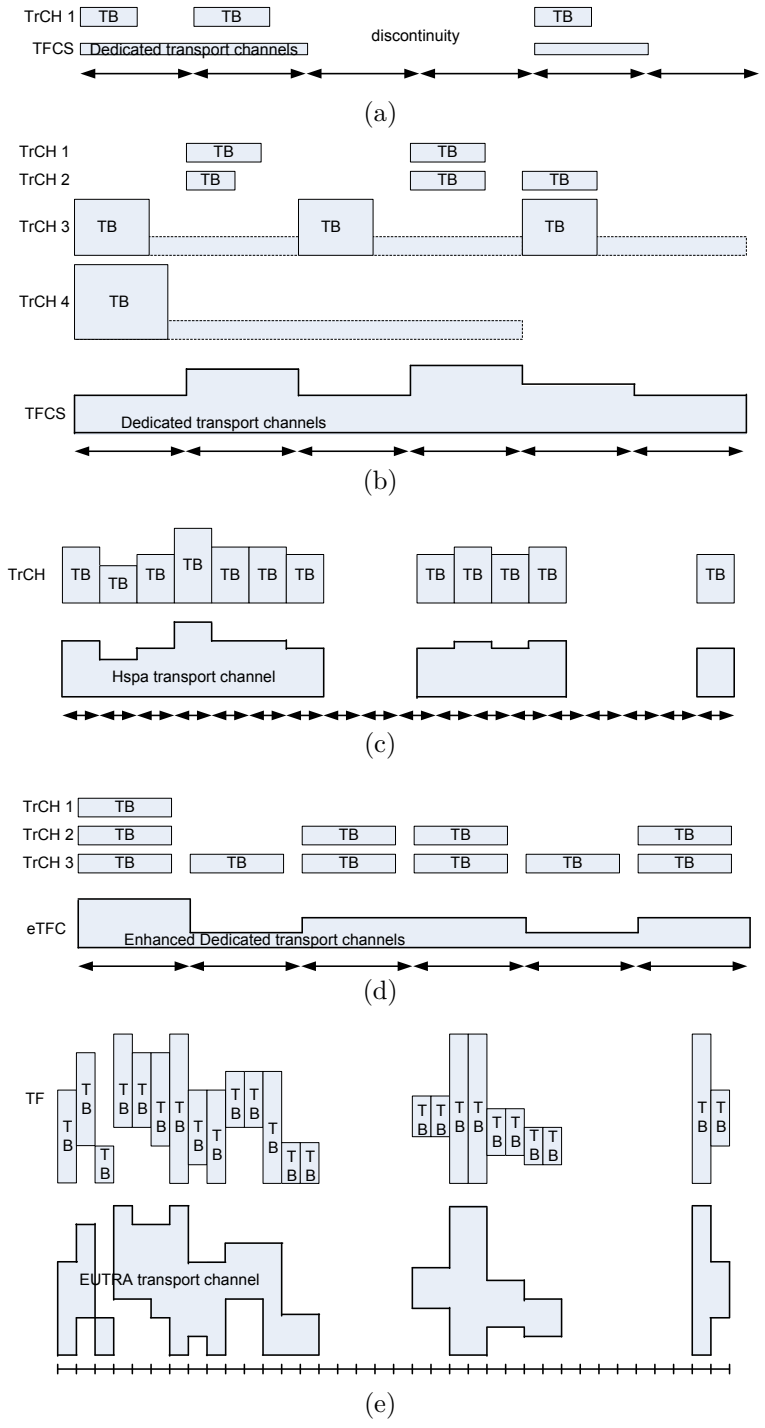


Figure 6.22: Functionality of the transport channels in WCDMA, WCDMA/HSPA and LTE. (a) A constant rate dedicated transport channel in WCDMA. (b) Variable rate dedicated transport channels multiplexed in WCDMA. (c) High speed, variable rate transport channel of WCDMA/HSPA downlink. (d) Enhanced and dedicated transport channels multiplexed in WCDMA/HSPA uplink. (e) Shared transport channels in LTE.

the respective uplink shared transport channel. The feedback has to precede the next downlink allocation by about three frames (3 ms), in order to let the scheduler have time for channel dependent decisions, to let the MAC compose the transport block and yet to reserve time for the encoding and decoding processes. The retransmission processes of LTE are discussed in Section 8.2.5, see also Fig. 8.4.

In LTE, the uplink transport happens in the shared uplink transport channel according to the allocations signalled in the respective shared physical control channel in downlink, see Section 6.5.7.3 and Fig. 8.2. The request for uplink transport resources is provided as the scheduling request in the uplink physical control channel, unless there is already uplink data transport active. When uplink data transport is active, the MAC may signal a Buffer Status Report, see Section 6.5.5.2, multiplexed into the transport block with the data to let the serving eNodeB make efficient scheduling decisions, see Section 8.3.

For LTE reception in downlink, the UE will buffer the control region of the frame and search the physical control signalling from its search space of the control channel candidates. Again, a match of the UE identity in the CRC scrambling mask and correct decoding of the signalling block will indicate that the signalling fields are for this specific UE. The signalling will indicate the size of the transport block and its allocated PRBs. Further information about the transport format as HARQ, AMC and MIMO are present. The physical control signalling for the uplink shared channel resources appears as a separate signalling block. The search procedure of the uplink signalling block from the control region of the frame is very similar to the search of the downlink signalling block. The signalling for an uplink allocation in a given transmission frame refers to the actual uplink resources that are available in a subsequent transmission frame after about three frames (less than 3 ms), in order to let the UE decide the multiplexing of logical flows, to let the MAC compose the transport block and yet to reserve time for the encoding and decoding processes, see Fig. 8.4. The signalling will indicate the size of the transport block and its allocated PRBs. Further information about the transport format as HARQ, AMC and MIMO are present.

6.5.7 Physical channels

At the transmitter, the physical layer processes transport blocks received from the local MAC by the following functions; append of error check code, codeblock segmentation, channel coding, rate matching, interleaving, multiplexing and radio frame segmentation.

At the receiver, the functionality is inverse; the physical layer processes transport channels by demultiplexing, deinterleaving, rate matching, decoding and error checking to deliver the correctly received transport blocks to the local MAC.

The physical layer modulates and makes channelization for the coded symbols. The channelization is done by the spreading operation or by a transform, see Chapter 3. At the output of a physical channel, the symbol stream may optionally be split to multiple transmit antennas. The receiver makes complex baseband processing by the channel estimation, equalization, symbol detection and demodulation functions. The physical code channels carry signals that may be received from multiple transmit antennas and received by multiple receive antennas. The MIMO transmission provides multiple symbol streams into the radio channel(s), which after multipath propagation can be received and estimated as separate signal components, or which are jointly detected in a matrix form. The MIMO transmission adds spatial diversity, and it is a fundamental performance differentiator of a system, as seen in Chapter 9.

The common physical channels appear for the purposes of cell detection, carrier measurements, synchronization, channel estimation and carrier access. The dedicated and shared physical channels are defined for WCDMA and WCDMA/HSPA in [192] and for LTE in [105]. Their properties are shortly covered in this section.

6.5.7.1 Physical channels of WCDMA

Dedicated Physical Data Channel (DPDCH) carries modulated data symbols of the transport channels. The data symbols are split to the periods of slots for fast power control, but a transport block set of the transport channel is allocated per frame period. On a DPDCH, a single cell multiantenna transmission is feasible either for the transmit diversity or for the multistreaming [90][91][93]. In macrodiversity transmission, a single DPDCH may have multiantenna transmission from the multiple cells in the active set.

Dedicated Physical Control Channel (DPCCH) carries modulated symbols for the purpose of physical control. The physical control signals include pilot sequences for the channel estimation and timing estimation, *Transmit Power Control* (TPC) symbols, *Transport Format Combination Indicator* (TFCI) and *Feedback Indication bits* (FBI). The FBI bits assist the choice of the modulation and coding, as well as the choice of the multiantenna transmission mode.

In downlink, the DPDCH and DPCCH channels are time multiplexed in every slot

of each frame. There are m symbols of DPCCH and n symbols of DPDCH in every slot. The amount of DPCCH fields depends on the slot format. The amount of DPDCH symbols depends on the spreading factor of the channelization code selected per frame. The spreading factor is termined by the selection of the transport format combination. The downlink allocation algorithm needs to keep the orthogonality of the channelization code resources, and thus the lowest code tree node for the expected highest data rate needs to be allocated. The spreading factor choice above the lowest node of the code tree is flexible. If a higher transport rate is requested, *TransportChannelReconfiguration* and *PhysicalChannelReconfiguration* procedures (RRC) need to be executed.

In uplink, the DPCCH and the DPDCH are code multiplexed in every slot of each frame. Multiplexing allows continuous transmission of the DPCCH on a high spreading factor channelization code. Thus, the NodeB can do continuous time tracking of the spreading sequence and can apply fast power control accurately. The full code-tree is available for every mobile due to the UE specific scrambling. The spreading factor of the channelization code is variable per frame and reconfiguration procedures are avoided. The DPDCH has full flexibility to select over the large range of transport format combinations.

6.5.7.2 Physical channels of WCDMA/HSPA

High Speed Physical Downlink Shared channel (HS-PDSCH) provides the advantage of a fixed spreading factor, fast scheduling, short TTI of a subframe period and minimized signalling. The resource allocation of HSPA is covered in [178][199]. (Prior to HSPA, other shared channel types were defined but they got later obsoleted by the HSPA.)

High Speed Shared Control Channel (HS-SCCH) is a physical channel that carries modulated symbols for signalling the usage and format of the HS-PDSCH. It provides a pointer to the transport format and resource combination by the *transport format and resource indicator* (TFRI). It also announces the channelization code-set, modulation and HARQ related signalling, as well as the receiving UE identity. The HS-SCCH carries a fixed number of bits on a fixed spreading factor channelization code per subframe.

High Speed Dedicated Physical Control Channel (HS-DPCCH) carries the modulated symbols for the signalling feedback for the transmission on the HS-PDSCH. The HS-DPCCH carriers HARQ acknowledgements, CQI and Precoding Information (PCI). The HS-DPCCH is present on a fixed spreading factor channelization code, and it is code-multiplexed with the DPCCH in uplink. The HS-DPCCH appears only in relation to the

scheduled HS-PDSCH subframes.

Enhanced Dedicated Physical Data Channel (E-DPDCH) carries the modulated data symbols of the HSPA transport channel. It has a short scheduling interval of 2 ms or 10 ms, its transmission can be discontinued and it may have an individual power amplification facor relative to the DPCCH. The fast power control is applied commonly to all uplink physical channels.

Enhanced Dedicated Physical Control Channel (E-DPCCH) carries the modulated symbols to signal the TFCI for the E-DPDCH. As the DPCCH has a constant spreading factor, and it cannot accommodate the transport format control of the E-DCH, there is a need for this separate E-DPCCH code channel, which includes the signalling of the transport block size, the HARQ redundancy version and the short buffer status indication.

6.5.7.3 Physical channels of LTE

Physical Downlink Shared Channel (PDSCH) carries the transport blocks as coded and modulated symbol blocks in the allocated multiples of PRBs during the 1 ms scheduling interval (TTI).

Physical Downlink Control Channel (PDCCH) carries the allocation information for a single UE per scheduling interval. The allocation information is signalled on separate PDCCHs for downlink and uplink of a UE. The signalling formats of the physical allocations differ e.g. based on the UE capability and MIMO schemes. The PDCCH channels of all UEs signalled in a subframe are interleaved and mapped to the control symbol region of a subframe extending from two to four OFDM symbols [56]. The UE receiver searches for a matching PDCCH blindly, where the detection is based on the UE identification mask matched to the C-RNTI [8].

Physical Uplink Shared Channel (PUSCH) carries the transport blocks as coded and modulated symbol blocks in the allocated multiples of PRBs during the 1 ms scheduling interval (TTI).

Physical Uplink Control Channel (PUCCH) carries the physical control information as the downlink acknowledgements, CQI, Precoding Matrix Identity (PMI) and Scheduling Request (SR). Every UE is assigned an orthogonal sequence in the known PUCCH resources to provide orthogonal transmissions of the control information for a large number of UEs.

6.5.8 About the channels and performance

6.5.8.1 Transport channel types

The different transport channel types and physical channel types have unique properties each, and their performance differ significantly. For a careful system analysis, each channel should be separately analysed. The common channels have impact mainly on the cell search times, coverage probability, access times and overhead. The system performance studies address these critical items separately as a prerequisite to show that the system is feasible. Once the feasibility is proven, the actual system performance studies rather focus on the numerics of the control and transport metrics as such, omitting the control channels in other respects than their overhead.

In the dedicated and shared channels the coverage-throughput-delay performance is critical, and their performance is tried to be maximised both from the single user and the system point of view. The coverage requirement sets constraints to the transport block size. The coverage is defined by the smallest packet size delivered within the delay constraint. If the delay constraint is loose, retransmissions can increase the symbol energy over thermal noise (E_s/N_0) for correct decoding and thus a larger transport block size can be provided even in the coverage (or power) limited case. If the delay constraint is tight, E_s/N_0 will be constrained and finite transmission power to the channel is available. The probability of correct decoding can be increased by lowering the channel code rate, which however decreases the number of transport channel bits in the finite sequence of coded symbols provided within the tight delay constraints. The largest coverage is available for a single information bit coded with the lowest channel coding rate limited by the length of the finite symbol sequence. The single bit transmission scheme may apply for a control signal (e.g. a single bit ack/nack), but it does not apply for a packet of a traffic flow.

The minimum transport block size needs to include at least a part of the packet payload and a segmentation header so that a packet can be recreated from its segments at the receiver. An alternative is to include the full minimum packet into a transport block and repeat its redundancy transmissions within the constrained delay. In both techniques, the coverage maximizes for continuous transmission of the symbol sequence. Any discontinuity of the transmission sequence, say due to time duplexing or due to segmentation to partial (parallel) HARQ processes, will lead to degradation of the coverage by the discontinuous transmission factor. Actually, segmenting an information block to multiple HARQ pro-

cesses would not decrease the coverage otherwise, but it will decrease the coverage-limited throughput because of the increased segmentation overhead. For these reasons, LTE may avoid segmentation by proper allocation of the physical resources. In the specifc case of the coverage limited uplink, it is feasible to have a longer TTI by bundling four subframes to a single TTI. This allows extending the code block and carrying a larger information block without discontinuity [54]. This increases the coverage-limited throughput for small payloads, and short LTE subrame will hence not limit the coverage.

6.5.8.2 Retransmissions

HARQ retransmissions are defined for the shared channel types both in HSPA and in LTE, but they do not apply in the dedicated channels of WCDMA. As stop-and-wait HARQ is defined, transport over the physical channel does not provide insequence delivery of transport blocks. For HARQ, reordering of transport blocks is necessary below the MAC. The transmission sequence number of a transport block keeps the transmitter and receiver window in phase and lets the receiver window be read in the correct order [44]. The size of the receiver window depends on the number of HARQ processes and the allowed maximum number of retransmissions. The receiver window follows the latest correctly received transmission sequence number, and it includes a list of transport blocks inside the window. All correctly received transport blocks before the first missing one inside the window are immediately delivered to the MAC. The discard timer is reset and it is running for the first correctly received transport block following the first missing transport block in the window. At the expiration of the timer, the missing transport block(s) is omitted and the correctly received transport blocks are delivered in sequence to the MAC. In the RLC acknowledgement mode (AM), the missing transport block will cause retransmission of the RLC PDU(s). In the RLC unacknowledgement mode (UM), the missing transport block will cause discard of the RLC SDU(s).

Having multiple HARQ processes defined the retransmission processes can be synchronous or asynchronous. For a synchronous process, the timing relation is constant between the original transmission and the retransmissions. The synchronous process is favoured for the simplicity of signalling. For an asynchronous process, only the minimum distance between the transmissions is defined to guarantee sufficient processing time for the transmitter and receiver, otherwise the scheduling instant is free. In LTE, the downlink HARQ processes are asynchronous, since the scheduling decisions are valid flexibly in time

after the feedback is received from the UE. In uplink, the transmissions are synchronous to minimize the UE effect of searching and decoding the next allocation and feedback, and to let the UE enter the power-save (sleep) fast and efficiently.

The HARQ processes may either be adaptive or non-adpative processes. For the adaptive HARQ, the transport format and the allocated frequency resources of a HARQ process may change during the retransmission instances. For the non-adaptive HARQ, both the transport format and the frequency resources are selected only for the first transmission instance of a HARQ process, and they will remain constant since for all the retransmission instances. For the next first transmission of new data, the transport format selection is free again.

6.5.8.3 Multiantenna transmission

MIMO transmission is applied on the physical channels because of its high expected gains [91][92][145]. MIMO can provide diversity gains in the low SINR regime and throughput gains in the high SINR regime, if the propagation channel additionally has large spatial distribution. Any physical channel type may gain of a diversity transmission, which at least increases the coverage probability. Those physical channel types, which target at maximizing the instantaneous throughput efficiency may additionally gain of multistream transmission. However, MIMO transmission may suffer from antenna and RF-chain imperfections, lack of high practical rank in the channel and from low SINR operation point. These factors often decrease the theoretically expected gains in a practical environment. The practical differentiators of the MIMO performance are the number of transmit antennas and the number of receive antennas, the MIMO transmission technique and the MIMO receiver algorithm. The efficiency of MIMO also depends on the Channel State Information (CSI) and Channel Distribution Information (CDI) that are experienced by the receiver, and whether this information is available at the transmitter too. In addition to the reporting overhead, reporting delay and accuracy impact the MIMO efficiency.

6.5.8.4 Multiantenna reception

Multiantenna reception applies for any channel type. It actually is *de facto* receiver technique in the NodeB both for WCDMA and WCDMA/HSPA, and it is expected to become that also in the eNodeB for LTE. The challenge for multiantenna reception capability is larger in the UE, where the antenna placement and the form factor of the product

set tight practical constraints. For WCDMA/HSPA, dual antenna reception is foreseen implementation in the UE, but for LTE it is expected to become a commodity [200].

Chapter 7

Internet services and traffic models

7.1 Session activation

Session activation is traditionally modeled by the call arrival process. Similarly for data services, a session activation process may be defined. A human interaction is typically the trigger of activation, therefore the session arrival process can be modelled by a Poisson process as in (7.1). For each service, there is an independent session activation process, which at a given time instant can be modelled by a two-state event according to (7.2). The traffic characteristics during the session are defined by the traffic models described in Section 7.2 [62][63][64].

The session activation and termination have a two-fold meaning for the system analysis. First, it may define the load in the simulated system and second, it impacts statistic collection. In a snap-shot simulation approach, the drops are independent relative to the channel states and hence the long term statistics over session duration are not available. In dynamic simulations, the statictics over the session duration is available and it may include, in addition to the channel dynamics, also mobility and handover effects. The session activation probability per traffic type allows mixed traffic simulations, where the ratio of traffic types can be deterministic, see [23]. If in addition to session activation, also session termination is defined, it is important to verify that traffic equilibrium is maintained during the collection of statistics. Therefore, for a given session termination process, a respective session arrival process has to be active too. For the simulation purposes, other approaches may be defined like gradually increasing load by the session activation process, until a favoured load level has been reached. The load level is then preserved by denying further session activations and terminations during the simulation run. In this approach,

the traffic fluctuation is solely due to dynamics of the traffic model itself, due to mobility (handovers) and due to actions by the radio resource management algorithms (scheduling).

The session activation process is characterized by the rate parameter λ, which describes the traffic intensity such that the number of session arrivals (ΔN) in a simulation time interval $(t, t + \Delta t)$ follows a Poisson distribution as,

$$P(\Delta N) = \frac{e^{-\lambda(\Delta t)} \cdot [\lambda(\Delta t)]^{\Delta N}}{(\Delta N)!} \tag{7.1}$$

In case of individual and independent two-state (TRUE/FALSE) events, this process can be reformulated to a binomial distribution (or likewise Bernoulli distribution), whose probability mass[1] function f is given by,

$$f\big(k\ ;\ p\big) = p^k \cdot (1-p)^{1-k} \quad , \tag{7.2}$$

where p is the probability of a session activation event (k) becoming TRUE.

7.2 Traffic models

As Internet services are so diverse, it is not reasonable to target the system design to be optimized for service specific solutions, but on the contrary see the services generally as Communications over the Internet Protocol (CoIP), which expands from Voice over the Internet Protocol (VoIP) to text, images, video and other forms of digital context [201], including multimedia and broadcasting. It is about convergence of all forms of communications. This means that new communications modality [202][203] happens in the Internet in a way that is combining e.g. voice, video, chat, data and presence. Converging services are realized in the open Internet, and voice applications are no longer expected to dominate but rather appear among the other modalities of media.

Rich media describes a broad range of digital interactive media. Rich media can be downloadable from the web or may be embedded as objects in a web page. If downloadable, it can be viewed or used offline with media players. The defining characteristic of rich media is that it exhibits dynamic motion. This motion may occur over time, as in video newscast, or in direct response to a user interaction, as an interactive slideshow presentation activated in a web page. Accessible rich media on a web page includes captions, audio and navigation.

[1]The probability mass of discrete random variables is given here instead of the mean of continuous random variables.

Although the services are so diverse, it is still necessary to try to find their critical common and distinctive characteristics, in order to simplify them for the analysis. As a result, all traffic may be categorized to traffic classes having given statistical behaviour and quality requirements. Depending on the actual service-case in a traffic class, the parameters of the traffic class model may still differ. The services are often defined to the background (a.k.a. best effort), interactive, streaming and conversational traffic classes. Services modelled by these traffic classes include; VoIP, file transfer, email, web-browsing, chat, audio/video streaming and gaming. Typical protocols cover SIP, FTP, HTTP, IMAP/POP, RTP, TCP, UDP and IP as introduced in Chapters 5 and 6 among others. The traffic modelling for the system analysis needs to take key the characteristics, parametrisation and quality metrics of these classes into account.

7.2.1 Conversational voice

Voice source model for the simulator depends on the voice codec and on the networking protocols. As voice codecs are very many (AMR, AMR-WB, EFR, EVRC, EVRC-WB, MP3, AAC, etc), it is typical to use otherwise common assumptions of the models, but differentiate their key indications as the bit rate or the packet size (distribution), interarrival time, source coding quality and decoding delay.

The voice codecs are based on spatio-temporal nature of audible sounds. The temporal pressure-wave of a phonem lasts of order 20 ms and the audible frequency reaches 16 Hz to 20 kHz. Thus, it is possible to sample the phonem and with a sampling rate conversion technique provide a voice frame of 20 ms per phonem. Discretizing (or localizing) the phonems in time and frequency allows a rate compatible framing of voice packets so that a generic transmission protocol may be formulated, see Section 5.2. Traditionally, a constant bitrate voice frame of 8 kb/s with packet interarrival time 20 ms is used for voice capacity analysis. Recently, AMR compatible bitrates representing the highest voice quality at rate 12.2 kb/s have dominated the working assumptions. Enhanced Variable Rate Codec (EVRC) at rate 7.95 kb/s is also frequently used in the literature for comparison of cdma2000 systems. The main differentiator of the voice source model in this study according to [62][63][64] compared to a conventional one is handling of the networking headers. The networking headers do not depend on the voice source as such, but often they are combined in the source model for the simulations.

The duration of a voice session is exponential distributed with the expected mean

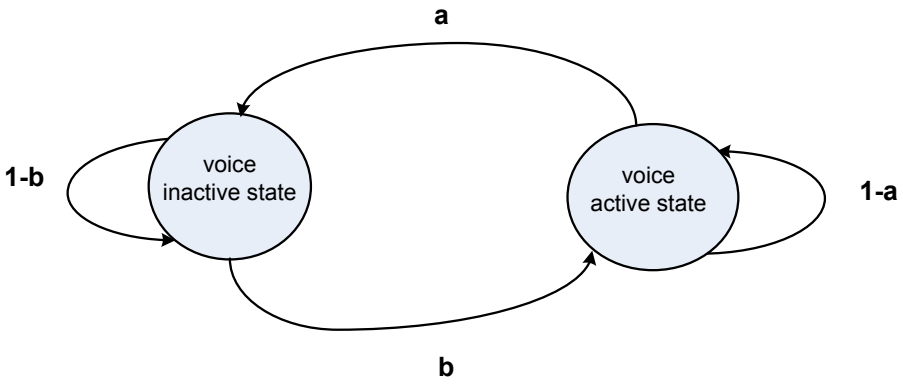


Figure 7.1: Voice activity model.

according to the measurements in live networks. [62][63][64], The reciprocity of voice activity between the caller and the callee is not known to follow a given statistical pattern, but it may be characterized as symmetrical interactivity if averaged over large number of conversational sessions. The activity may be characterized by the voice activity factor, where each voice burst (talk-spurt) is exponential distributed in length with a mean value of order 3 seconds. The voice session can thus be modelled by a two-state model with voice activity parameter and distributed talk-spurt length [62][63].

7.2.1.1 Detailed description of the VoIP traffic model

The voice activity model impacts on VoIP packet creation. A simple model consists of two states, "voice inactive state" and "voice active state", shown in Fig. 7.1[2]. The former represents silence and the latter represents a talk-spurt i.e. a sequence of successively active voice frames. During activity, each voice frame contains the spatio-temporal characteristics of a phonem, encoded by the voice codec and encapsulated to a defined voice frame format.

In the voice activity model, the probability of transition from the active state to the inactive state is equal to a. The probability of transition from the inactive state to the active state is b. The model is updated at the voice codec frame rate $R_{codec} = 1/T_{codec}$, where T_{codec} is the encoder frame duration (20ms for AMR and AMR-WB).The Voice Activity Factor (VAF) λ is given by (7.3),

$$\lambda = \frac{b}{a+b} \tag{7.3}$$

A talk-spurt is defined as the voice active period (τ_{VP}) between entering the active

[2]This two-state model is simplistic but valuable in that it allows a generalization for the regularity of talk-spurt arrival statistics and resource reservation activity

state and leaving the active state. This corresponds to the sequence of phonemes forming words separated by short periods of silence. The probability that a talk-spurt (VP) has a duration (τ) of n voice frames is given by,

$$p_{\tau_{VP}}(n) = a\ (1-a)^{n-1} \tag{7.4}$$

Correspondingly, the probability that a silence period (SP) has a duration (τ) of n voice frames is given by

$$p_{\tau_{SP}}(n) = b\ (1-b)^{n-1} \tag{7.5}$$

In voice traffic models, the voice active period and the silence period are exponential distributed. The mean talk-spurt (μ_{VP}) and the mean silence period (μ_{SP}) in duration of voice frames are given by (7.6). (In the performance analysis of Chapter 9, the mean talk-spurt and the mean silence periods are both about 2 seconds.)

$$\begin{aligned} \mu_{VP} &= E[\tau_{VP}] = \frac{1}{a} \\ \mu_{SP} &= E[\tau_{SP}] = \frac{1}{b} \end{aligned} \tag{7.6}$$

The state transitions between the active state and the inactive state are independent. Therefore, the probability distribution of the time period (in number of voice frames) between successive active states (τ_{AE}) is the convolution of the probability distributions $p_{\tau_{SP}}$ and $p_{\tau_{VP}}$, and is given by (7.7).

$$p_{\tau_{AE}}(n) = \frac{b}{b-a}\ a\ (1-a)^{n-1} + \frac{a}{a-b}\ b\ (1-b)^{n-1} \tag{7.7}$$

The mean time μ_{AE} between successive active states is given by

$$\mu_{AE} = \mu_{VP} + \mu_{SP} \tag{7.8}$$

Accordingly, the mean rate of transitions into the active state is in other words the mean talk-spurt interarrival rate $\overline{R}_{AE}$ and is given by (7.9)

$$\overline{R}_{AE} = \frac{1}{\mu_{AE}} \tag{7.9}$$

7.2.1.2 Networking headers of the VoIP traffic

VoIP is assumed to consist of RTP/UDP/IP headers, whose size varies according to the protocol options. VoIP is present at the radio interface either as uncompressed packets or compressed packets. The packet compression here refers to the compression of the bit fields of the networking headers and not the voice frame itself. VoIP compression techniques are defined e.g. in [165][166] of which ROHC is dominating, see Section 5.2.

7.2.2 Background traffic

Background traffic is a notation for the packet delivery by the best effort algorithms. The best effort delivery dominates in the Internet, where QoS differentiation is traditionally not in use, and TCP protocols drive the performance. Recently, however, QoS differentiation of the Internet has gained more momentum.

Best Effort data can be modeled as infinite packet queue at the transmit buffers, which strives for maximizing the network load. In case, protocol overhead needs to be taken into account or packet delivery (SDU) is the measure instead of the PDU measure, even the infinite queue model needs to be aware of the packet size distribution.

File Transfer Protocol (FTP) is often adopted for the packet size distribution. The traffic source generates files, e.g. by a truncated Pareto distribution (Bradford distribution or alike), where the probability is heavily tailed i.e. the probability of a small file is very large compared to a large file, but yet very large files exist at a non-negligible probability. An example parametrisation creates files of the mean size 2 Mbytes and the maximum size 5 Mbytes. FTP segments the application layer data to segments, so that one FTP segment with further TCP/IP networking headers fits to the Maximum Transmission Unit (MTU) of the transmission paths between the end nodes. The MTU is often constrained by the Ethernet, and the default values of 1500 bytes and 576 bytes therefore dominate the practical MTU observations.

7.2.2.1 Properties of the truncated Pareto distribution

For a delay analysis, a finite queue model is needed instead. The FTP traffic generator is applicable for a finite queue model too, but a further notation of a packet call is necessary. A packet call is a period of packet transfer for a single file in a session, where a given number of files are to be transferred. Between each packet call (file transfer), there exists a non-

active reading time without packet flows of file transfer. In other words, this corresponds to the time interval of file download requests. A FTP session may be parametrized by the session arrival process, the number of packet calls (files) per session, the request time between the packet calls, the number of packets in a packet call (as one file is segmented for the TCP/IP transfer over the Ethernet), the interarrival time between two consequtive packets in a packet call and the packet size distribution. The packet size of the FTP model is trivial, because FTP utilizes the MTU size and packets are predominantly of this size;

$$pkt\_size(i) = \begin{cases} MTU \ , \text{if } i \leq \text{trunc}\{file\_size/MTU\} \\ file\_size - \text{trunc}\{file\_size/MTU\} \cdot MTU, \quad \text{otherwise} \end{cases}, \tag{7.10}$$

where $i = 1, .., \text{trunc}\{file\_size/MTU\} + 1$, $i \in \mathbb{N}$.

Many practical traffic sources are expected to follow a modification of the FTP model, e.g. Pareto distributed packet size models and email models are common. The Pareto distribution in (7.11) yields a heavy tailed packet size distribution, which additionally requires a cutoff value for the maximum transmission unit (for MTU, see Section 5.2.6). Here, the cutoff-value causes samples fold inside the distribution so that if an input sample (x) is below or above the cutoff-value i.e. $x < \text{th}_{min\_size}$ or $x > \text{th}_{max\_size}$, the input sample is discarded and it is not calculated into the actual sample count, but the sample is regenerated from inside the distribution until it satisfies $\text{th}_{min\_size} \leq x \leq \text{th}_{max\_size}$. Typically, the upper bound cutoff value is applied only. The Pareto distributed packet size is given e.g. in [62],

$$\begin{array}{ll} f\big(x \ ; \ \alpha\big) & = \frac{\alpha \ k^{\alpha}}{x^{(\alpha+1)}} \qquad , x \geq k \\ F\big(x \ ; \ \alpha\big) & = 1 - (\frac{k}{x})^{\alpha} \quad , x \geq k \end{array}, \tag{7.11}$$

with Pareto index k. In (7.11), f is the probability density function (pdf) and F is the cumulative distribution function (cdf). This notation is followed throughout Sections 7.2.2 and 7.2.3.

In the email model of [64], the number of transmitted and received emails per packet call follows a log-normal distribution, and the email size follows the Cauchy distribution (7.12) instead of the Pareto distributed file size. The Cauchy distributed email size is,

$$\begin{array}{ll} f\big(x \ ; \ \mu\big) & = \frac{1}{\pi[(x-\mu)^2+1]} \\ F\big(x \ ; \ \mu\big) & = \frac{1}{\pi} \ \arctan(x-\mu) + 1/2 \end{array}. \tag{7.12}$$

For Cauchy distribution μ is the median (because for a discrete Cauchy distribution the mean is not defined) and the height is $1/\pi$, if $\gamma = 1$.

7.2.3 Interactive traffic

7.2.3.1 web-browsing

A more sophisticated traffic model is needed for the interactive traffic like web-browsing that uses the Hypertext Transfer Protocol (HTTP) [204][205][206]. A web-page is encoded by a markup language (e.g. Hypertext Markup Language HTML, eXtensible Markup Language XML, Wireless Markup Language WML or Standard Generalized Markup Language, SGML), which defines the object presentation of the page.

The web-page consists of a main object and several embedded objects (e.g. pieces of text, pictures, advertisements, etc), whose contents may be fetched by opening remote information sources at given URL [155]. After receiving the main object of the page, the web-browser will fetch the embedded objects and will parse the page.

The main parameters to characterize web-browsing are: the size of the main object, the size of an embedded object in a page, the number of embedded objects on a page, the reading time and the parsing time of the page. The object size follows a log-normal distribution in (7.13) truncated with the cutoff-values (th_{max_size}) and (th_{min_size}) so that $\text{th}_{min\_size} \leq x \leq \text{th}_{max\_size}$.

$$\begin{array}{ll} f\big(x\ ;\ \mu,\sigma\big) & = \frac{1}{\sqrt{2\pi}\sigma x} \cdot \exp[\frac{-(\ln x-\mu)^2}{2\sigma^2}] \quad ,x>0 \\ F\big(x\ ;\ \mu,\sigma\big) & = \frac{1}{2}\ \big(1+erf[\frac{(\ln x-\mu)}{\sqrt{2}\sigma}]\big) \quad ,x>0 \end{array}\ , \tag{7.13}$$

where μ and σ are the mean and the standard deviation of the variable's natural logarithm. The number of embedded objects in a page has a heavy tail, and therefore it can also be modelled by a Pareto distribution, as in (7.14). Again, this model may apply an upper bound cutoff value. The expected number of objects per page is five according to practical observations.

$$\begin{array}{ll} f\big(x\ ;\ \alpha\big) & = \frac{\alpha\ k^\alpha}{x^{(\alpha+1)}} \quad ,x \geq k \\ F\big(x\ ;\ \alpha\big) & = 1-(\frac{k}{x})^\alpha \quad ,x \geq k \end{array} \tag{7.14}$$

In the previous models, the reading time and the web-page parse time are exponential distributed according to (7.15).

$$\begin{aligned} f\big(x\ ;\ \lambda\big) &= \lambda\ e^{-\lambda x} &&, x \geq 0 \\ F\big(x\ ;\ \lambda\big) &= 1 - e^{-\lambda x} &&, x \geq 0 \end{aligned}\ , \tag{7.15}$$

where λ is the rate parameter. For interactivity, a HTTP request (*GET* or *HEAD*) triggers the activity from the client to the server, and it impacts on the discontinuous transmission, access payload and wake-up requirements at the air interface.

7.2.3.2 Interactive gaming

Interactive gaming is neither used in the system analysis in any other meaning than analytically calculating the minimum delay for small packets. It is sufficient to show that the minimum delay is below a threshold (e.g. 50 ms) and the probability of exceeding the threshold is sufficiently small (10%). Gaming traffic is seldom exhaustive as the interarrival rate is uniform distributed ($0 < x \leq 40\ ms$) and its packet sizes are small.

7.2.4 Streaming

Streaming traffic is used as a constant bitrate source. This can be used to analyze a bitrate coverage contour. However, such analysis is very heavy in computation. The streaming condition requires that the amount of traffic is in equilibrium i.e. the number of packets transmitted within a delay constraint equals the average bitrate. The delay constraint may be fairly large for streaming (several seconds), and traffic smoothing is required by a playout buffer at the receiver. Any transmit and scheduling delay variation and the consequent receive delay variation observed at the input of the playback buffer will not be present at the output of that buffer. Therefore, these variations do not decrease the user perception.

7.2.5 Video traffic

Video streaming and conversational video telephony are rarely used in the system analysis, even if their models exist in the literature. Specific to the video sources are their nominal bitrate and the requirement for a playback buffer, which has to cope with the transmit and scheduling delay variation. Video codecs typically generate regular frames with full information of the picture and a sequence of intermediate frames with differentially coded information. Each video frame consists of a variable number of variable size packets. The

interarrival process is rather deterministic because of the constant frame rate, but the number and size of packets may vary.

Chapter 8

Radio Resource Management

Radio Resource Management (RRM) is a set of functionality of protocols and algorithms, which targets at efficient use of radio resources in a cellular network. In order to be efficient, RRM has to operate efficiently and with low overhead in all scenarios with channel dynamics and mobility involved.

In a single link scenario, the RRM suffers from imperfect channel information between the protocol peer entities. Channel measurements have SINR dependent inaccuracies and any reporting thereof happens in finite resolution with reporting delay and errors. These imperfections are present, but may be negligible for a low velocity link, whereas for a high velocity link they will remain notable.

In a single cell scenario with multiple links, RRM operates with interlink interference characteristics. Radio resource sharing becomes a dimensional problem as many competing link requirements have to be met simultaneously. On the other hand, with multiple links, multiplexing efficiency increases and the opportunity for scheduling arises.

In a network with multiple cells, the RRM operates in a difficult dimensional situation, where only partial measurements, partial information and partial control between neighbour cells are available. All of the neighbour cell information and control exchange happen with notable delays and they increase the signalling cost on that interface. Of course, for cells that are operated from the same site (often called as sectors), these mutual impacts remain node internal limitations and they may largely vanish. When the cells are controlled by different sites, the aforementioned limitations will more concretely exist.

For intersite operation, the RRM has further characteristics, whether the cells are either synchronous or non-synchronous. Both asynchronous and synchronous operations

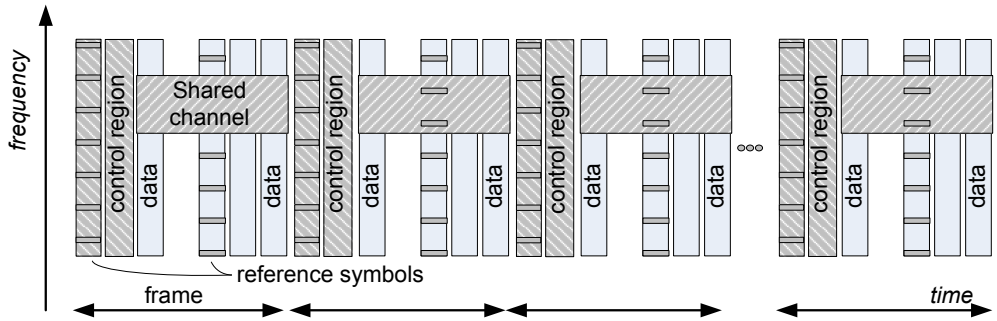


Figure 8.1: The frame, symbol and channel structure of LTE downlink.

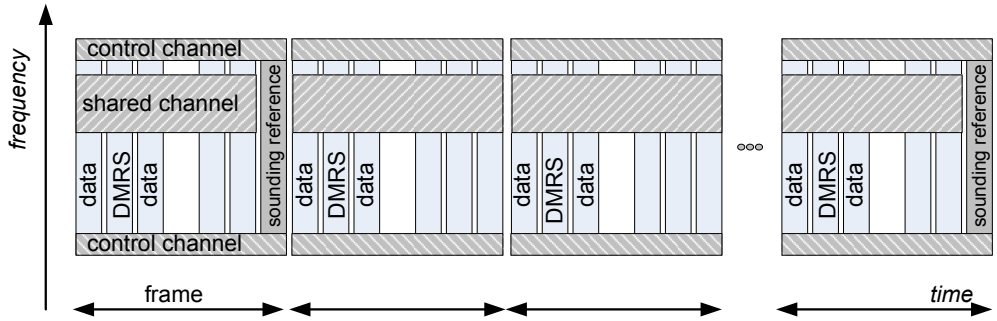


Figure 8.2: The frame, symbol and channel structure of LTE uplink.

have their benefits and drawbacks, and RRM has to cope with both of them. A standard specification defines enablers for asynchronous operation in a general case and provides synchronous cells as an additional opportunity. WCDMA, WCDMA/HSPA and LTE are capable to operate both in asynchronous and synchronous modes due to their pilot channel and reference sequence design.

The RRM is very tightly dependent on the frame and channel structure of the multiple-access schemes, see Chapter 3 and Section 6.5 . The frame, symbol and channel structure of LTE is shown in Fig. 8.1 for downlink and in Fig. 8.2 for uplink.

The frame and channel structures define the physical resources, how scalable they are, in which domains (time, frequency, code, spatial) they apply, what their physical characteristics are and how to allocate them. The RRM functionality often includes those of power control, link adaptation, scheduling, handover, load control and admission control. For this thesis, selected aspects of the RRM are covered.

The Radio Resource Management contributes to all of the performance measures of a cellular system. A RRM concept provides algorithm and protocol means to utlize the multiple-access scheme and transmission-reception techniques in a well-performing, convenient way. Often, the RRM concept is justified in its capability to provide the highest

average capacity and the highest spectral efficiency in several cell scenarios evaluated over a large number of user distributions (drops) over a long duration covering a large number of channel states. However, equally critical measures of a RRM concept is high cell-edge throughput (cell-edge spectral efficiency), high adaptability to the dynamics of single radio links, guarantees of packet delays and fairness in serving the users.

8.1 Power Control

8.1.1 Closed-loop power control

The closed-loop power control is applied in the access network (UTRAN or EUTRAN) to compensate variations of the radio channel for the receiver. The properties of power control depend on the multiple-access it applies for. The principle of the closed-loop power control is to get feedback of the received power levels or expected received power levels from the receiver to the transmitter, in order the transmitter to apply the most convenient transmit power at a time. The power control ultimately targets at maintaining the received bearer quality, and obtain this target by as small amount of power and transmit resources consumed, with as less interference generated, as practical. The commonalities and differencies of the power control principles in WCDMA, WCDMA/HSPA and LTE are addressed in this Section.

In WCDMA, power control is twofold, fast immediate power control to compensate channel transients and a slow outer-loop power control to set the received SINR to meet the expected BLER target at the receiver. The fast power control adapts to the rapid channel variations for a continuous, direct-sequence spread spectrum transmission. Transients are avoided by a chopper like functionality. As the received SINR forms a distribution, the BLER control is not accurate, but will statistically meet the BLER target at a given probability. If decoded BLER is too high compared to the BLER target, the power control target is increased to yield power-up commands by the fast, closed-loop power control. If decoded BLER is unnecessarily low, the power control target is slowly decreased to yield power-down commands by the fast power control. Outer-loop power control operates as a saw-curve, fast-up and slowly-down to control the link quality [6]. The outer-loop power control sets the received SINR target per channelization code that applies to all bearers carried on that code channel. If the bearers multiplexed on the came channelization code have different link quality (BLER) targets, the channel coding and rate matching

algorithms have to realize them assuming close to equal received SINR. This happens by the transport channel processing algorithms. Otherwise, if the SINR target requirements remain different, the bearers need to be multiplexed onto different code channels having different power weight.

In WCDMA, where macrodiversity is applied on the radio links and multiple sites are in control of the physical signalling, TPC feedback is common to all code channels, but TPC from different receiving sites may differ. The UE will always choose a dominating down-command and applies up-command otherwise. The outer-loop power control will tune the received power of the code channels commonly per cell according to the diversity combined signals to meet the BLER target of the transport channels. The BLER measure is feasible, when the decoding probability is known after the selection combining in the RNC. Thus, the outer-loop power control operates locally between the RNC and the NodeB.

UE controls the cell transmit power according to the decoding result of the soft combined signals. Any NodeB that can hear the transmit power control will act respectively. Thus, the strongest signal paths at the receiver have the strongest dominance on the power control, but also weak signal components will contribute positively decreasing the needed transmit power. The outer-loop power control operates locally in the UE to set the received SINR target despite of how many and which cells are in the active set for transmitting signals.

In HSPA, downlink power control is mainly replaced by the adaptive modulation and code rate selection (see Section 8.2). For power control, a share of the total cell transmit power is allocated for the HSPA channelization codes, this share can be slowly changed. In uplink HSPA, the transmit power of the parallel channelization codes (DPCCH, DPDCH, E-DPCCH and E-DPDCH) are jointly controlled by the fast power control, as the codes will experience channel variations equally. Yet, a channelization code specific power offset (β) is defined, and it may boost the received BLER of the code channel lower or higher. Again, the BLER target of the channelization code depends on the bearers (logical channels) that are multiplexed on it.

In LTE, the closed-loop power control applies to a burst transmission per subframe. As several UEs are multiplexed to the same PDSCH, there may be a UE specific power offset between the PDSCH allocated for a given UE and the Reference Symbols (pilots). This power offset is signalled for each downlink allocation, only if necessary, for the proper detection of the amplitude modulated symbols (QAM). In LTE uplink, the power control

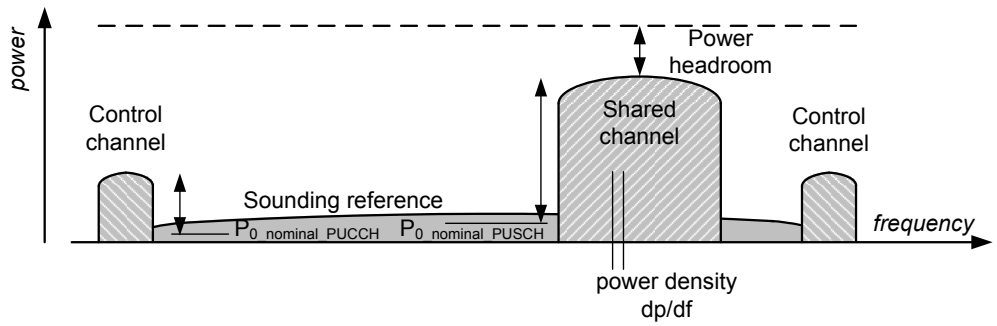


Figure 8.3: The principle of the power control of LTE uplink channels.

of the PUSCH is based on fractional pathloss compensation based on the received signal power measurements and the BLER target [119][207]. Otherwise, the transmit power scales with a cell specific fractional pathloss compensation factor, with the allocated number of resources, and with the instantaneous selection of the Modulation and Coding Scheme (MCS), according to (8.1). The eNodeB may signal for each subframe allocation a transmit power correction relative to the previous transmission. For the UE power control, a power headroom report by the UE is occasionally needed at the eNodeB. The principle of power control for different channels of Fig. 8.2 is shown in Fig. 8.3.

8.1.2 Power control algorithms

The transmit power of the LTE channels is given in [208] by (8.1). The shared data channel transmit power scales by the nominal power (P_{0_PUSCH}) and relative to at least the allocated bandwidth (M_{PUSCH}), transport format (TF), power headroom and fractional pathloss compensation ($\alpha_{PL} \cdot L$). The cell specific and UE-specific control parameters may be defined by the higher layer (RRC) signalling. The short term power adjustment (f_i) may be given inside the downlink shared control channel per allocation. This may be an absolute or an accumulated relative adjustment. The transmit power of the control channel is set absolutely in reference to the nominal power and relative to at least the pathloss and control channel signalling format. The transmit power of Sounding Reference Symbol (SRS) is controlled relative to at least the cell specific shared data channel power reference, bandwidth and fractional pathloss.

$$P_{PUSCH}(i) = min\{P_{max},\ 10\ \cdot \log(M_{PUSCH}(i)) + P_{0\_PUSCH} + \alpha_{PL} \cdot L + \Delta_{TF}(TF(i)) + f(i)\} \quad (8.1)$$

8.1.3 Open-loop power control

In all aforementioned systems, also the open-loop power control applies. It is used, whenever the closed-loop formula is not yet running e.g. in the beginning and after discontinuity of communication. The open-loop power control is applied on the mobile terminal based on its estimate of the pathloss to the receiving site. The open-loop power setting, as the mean power per subframe (slot), may have a fairly large inaccuracy of order $+/-9\ dB$ in normal conditions and $+/-12\ dB$ in extreme conditions [200].

8.2 Link adaptation

The link adaptation algorithm selects both the modulation and channel coding rate that form the most feasible combination to maximize the instantaneous link throughput, or cell throughput in a multiple-access scheme, which is a function of the received signal quality and BLER of the decoded transport block. The selection of the modulation alphabet depends on the received SINR and the selection of the code rate depends on the expected BLER. A high modulation order in a high SINR condition may still do with a high code rate, which allows a large transport block delivered and provides a high instantaneous throughput. On the other hand, a high order modulation in a low SINR condition requires a low rate channel code and even then, higher probability of symbol errors may limit the throughput compared to a lower order modulation with a higher code rate in equal SINR conditions. The coded symbols require channel power and cause interference to the other links, thus the maximized instantaneous throughput is actually a target with as low channel power and as small number of coded symbols as possible, i.e. maximizing the transport block size (see Section 6.5 for segmentation and multiplexing).

The well-known link adaptation curvery consists of instantaneous throughput as a function of the modulation and code rate selection. The increasing modulation order and increasing code rate require increasing SINR. The SINR behaviour actually depends on the multiple-access technique and intercell interference mechanisms, the BLER depends on the signal waveform and the receiver algorithms. The higher the SINR is required, the

smaller is its time-probability in a selective channel and the smaller is the coverage area, where it is experienced. Link adaptation requires knowledge of the channel state at the receiver fed back to the transmitter. In the CQI measurement delay, reporting delay and finite reporting resolution always have an impact.

For a selection of a large transport block, link adaptation is a crucial feature, because it can only be selected for a short transmission time with a notable probability. Also, it is important to be able to select a more robust format with adaptation after the initial selection of a large transport block to provide coverage for a weakening link without excessively increasing the transport block latency.

8.2.1 Adaptive coding

Code rate adaptation provides close to optimal BLER with a maximal Information Block Length (IBL) as a function of channel SINR and a short term selection of the modulation alphabet. Rate matching is necessary to match the IBL to the physical resource grid. In WCDMA/HSPA and in LTE, convolutional codes and Turbo codes are used. Convolutional codes provide higher coding gain for a small IBL like control signalling, whereas Turbo code [209] outperforms for larger IBLs of high rate data. Low Density Parity Codes (LDPC)[1] were studied for LTE, but were not selected due to their rate matching properties [210]. LDPC puncturing to a given rate matching factor often requires a rate compatible design of the code matrixes and their inner interleavers, unlike the Turbo code which simply enables variable rate puncturing of a single mother code.

8.2.2 Adaptive modulation

Adaptive selection of the modulation alphabet is applied with a proper channel coding choice to maximize the instantaneous throughput for a given BLER target [211]. For WCDMA, adaptive modulation set is not supported and it becomes feasible only with WCDMA/HSPA, where the channel state information is available at the scheduler, and fast scheduling enables adaptation to the fast channel variations.

The set of modulation alphabet and coding type is a choice of system design due their requirements on the implementation of the transmitter-receiver chain. The dynamic range, sensitivity and decoding complexity are key issues, as are the requirements for the

[1]Originally by R.G. Gallager, Low-Density Parity-Check codes, IEEE Transactions on Information Theory, Vol.8, Issue 1, pp.21-28, Jan. 1962

linearity, EVM and noise figure of the receiver, see Section 3.5.

For the amplitude modulated (QAM) symbols, the peak power varies depending on the instantenous choices of the modulated symbols. The peak-to-average power ratio behaviour decreases the coverage probability and increases power consumption. In the eNodeB transmitter, the power limiter cuts the power peaks to become wideband noise. The power-probability of the highest power peaks is fairly small due to a large number of modulated subcarriers, thus the power-density of the inband noise remains small. In the UE transmitter, the linear region of the power amplifier limits the coverage and sets constraints for the modulation [212]. In practise, uplink transmissions at least up to 16QAM are feasible, but the role of 64QAM is disputable, because its coverage-area probability remains small in mobile reception.

8.2.3 Adaptive multicode transmission

A specific link adaptation feature designed for the HSPA is multicode transmission, see Section 3.3. Actually, the multicode transmission was already introduced for the WCDMA standard, but it became more applicable in the practical implementations with the HSPA, where short multiplexing frame and full signalling support at MAC also exist for the multicode transmissions.

The multicode transmission targets at exploiting good channel states for increasing bursty data rates, when enough transmit power is available and low expected BLER is possible. In this manner, it is a link adaptation scheme as well as adaptive coding and modulation. In the multicode transmission, the information code block length is increased and a large data block is modulated for a short transmission interval on several parallel channelization code channels.

8.2.4 Rank adaptation

A Modulation set up to 16QAM is feasible in uplink and up to 64QAM in downlink. Typically, 64QAM requires a diversity reception mode, where the modulation symbol energy is summed from two antenna branches before decoding. Constellation errors at the receiver may thus be reduced. If the transmitter has two transmission branches, a better performance compared to 64QAM may be obtained by multistreaming 16QAM modulated symbols. Both 64QAM and multistreamed 16QAM require sufficiently high SINR and low EVM of the transmitter-receiver chain. Multistreaming additionally requires sufficiently

high channel rank up to the number of streams.

As both the SINR and the channel rank are time-variant, short term modulation and coding adaptation provide gains. In WCDMA/HSPA, the choice of adaptation extends to time and channelization code domains. In LTE, the choice of adaptation extends to time and frequency domains. The frequency selectivity of the allocated resources over a wideband channel increases the probability of higher order modulation and higher coding because at each time instant there exist PRBs which have higher SINR than the average wideband SINR; $SINR(PRB_j) > \overline{SINR(i)}$, for selected $PRB_j \subset \{PRB_i\}_{i=1}^{N_{PRB}}$. This allows the short term expected throughput per selected PRBs be higher than the wideband expected throughput per PRB; $\hat{r}(PRB_j) > \overline{\hat{r}(PRB_i)}$. The frequency selectivity does not impact the channel rank as the propagation paths are not that strictly frequency selective, and the scatterers are about the same for different frequency components of the wideband signal. However, frequency selectivity that increases SINR acts as an enabler for utilizing the channel rank. A wideband allocation of a lower SINR will decrease the magnitude of the Eigenvalues over the noise plus interference compared to a narrowband allocation of higher SINR. That is, even if the same relative Eigenvalues would exist in the wideband and narrowband channels, the Eigenvalue-to-Interference-plus-noise ratio remain higher for the narrow bands having a higher SINR. Even if the rank order of the channel is not frequency selective, the precoding information has to be selected per subband, because the relative phase information depends on the subcarrier frequency.

Thus, for the rank adaptive MIMO, the rank order may be wideband information, but precoding matrix selection (PMI) should still be frequency selective. PMI feedback should be provided per subband that is as less frequency selective as possible i.e. ideally per subcarrier. However, practical limitations of the amount of feedback signalling favours PMI information per larger subband than a single subcarrier, that subband yet is smaller than the coherence bandwidth of the channel. This PMI feedback could be in the order of two to four PRBs, as used in the system performance analysis in Chapter 9. The PMI selection over a bandwidth smaller than the coherence bandwidth includes the multipath propagated delay spread signal components and provide good gains of precoding selection [92][213][145]. The PMI selection over a bandwidth larger than the coherence bandwidth does not include all multipath propagated components. In this case, the precoding selection would miss partial energy of the signal components, and were not optimal. Therefore, the frequency selective LTE may benefit more of precoding than the wideband spread

WCDMA/HSPA.

8.2.5 Adaptive retransmissions

Adaptive retransmission techniques like HARQ can also be seen as a link adaptation means, even if it is a functionality of its own. The HARQ adapts the *block probability* of correct decoding by transport block retransmissions and by combining the transport blocks. The HARQ is discussed from protocol point of view in Section 6.5. The automatic repeat request (signalling) by the receiver from the transmitter is applied to retreive incorrectly decoded packets and receive them correctly. If the original transmission is channel coded (Hybrid ARQ), the repeated transmissions will add redundancy and increase time diversity [214][215]. In case, the retransmissions are scheduled to different frequency resources apart the coherence band (possible in LTE), frequency diversity is also increased. Several HARQ types are defined according to the many variants of the block combining techniques e.g. in [195][196][197][216].

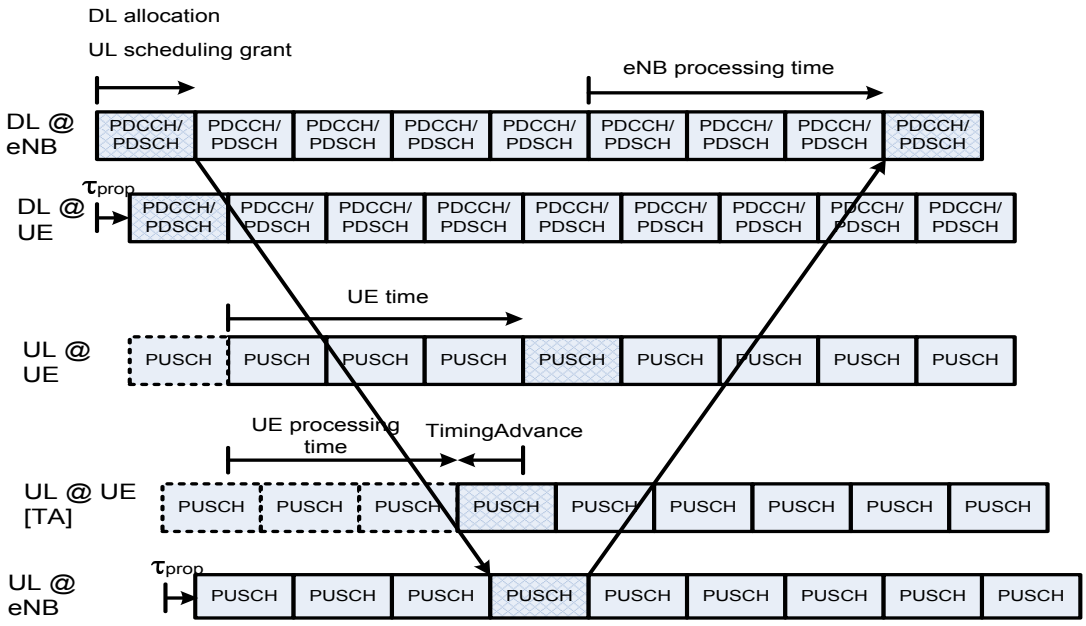


Figure 8.4: The principle of retransmissions in LTE.

The HARQ provides link adaptation by adding channel power and redundancy as small increments of retransmission. This lets BLER be met at close to the optimal adaptive channel power with a small transmission delay. The HARQ efficiently uses channel power, adds time and frequency diversity and adds soft symbol combining gains. The HARQ allows the

initial transmission at a higher nominal BLER and integrates channel power for correct decoding. Symbol gains are present by combining the symbol vectors of the retransmitted replica already before decoding. The transmitted replica can be modified, e.g. the bit to the modulation symbol mapping can be changed for incremental redundancy, or the code rate may be changed, which further adds redundancy. For any HARQ combining scheme, the information block (i.e. the transport block) of transmitted instances is mandated to be bit-exact, even if the symbol mapping or code rate would adapt. The design choices of WCDMA/HSPA and LTE include both Chase combining and Incremental Redundancy with a defined set of redundancy versions [216][217][218][219]. The Chase combining is favourable for its property of self-decodable transport blocks and minimum delay. The incremental redundancy is favourable for its adaptive redundancy and maximal throughput. For LTE, the number of redundancy versions was reduced for simplicity compared to WCDMA/HSPA without a net impact to the performance. The short round-trip-time of LTE HARQ is shown in Fig. 8.4.

8.2.6 Adaptive transmission bandwidth

Adaptive Transmission Bandwidth (ATB) technique, which is feasible in LTE but not in WCDMA/HSPA also contributes to the link adaptation. Because of the UE transmission power constraints, the transmission power occasionally gets limited, and when the coverage is critical, the bandwidth has to be decreased to reach sufficient E_s/N_o at the receiver. Condensing the transmit power to a narrow subband evidently increases the power spectral density compared to a wideband transmission. (WCDMA, on the other hand, increases the received Es/No of a wideband transmission by increasing the processing gain.) In LTE, the ATB provides gains especially, when the transmission resources can be scheduled into the subbands having good channel state and a higher than average SINR.

8.3 Scheduling

The scheduler targets at scheduling all active users in a multiuser environment using the physical resources in the best possible way. What the best possible way is, depends on many arguments. Often it is the way that also maximizes the use of radio resources, such schedulers are of type channel dependent schedulers.

Classical schedulers operate in a *Round Robin* manner or according to bearer (logical

channel) priority principles. In these, Round Robin is blind to the channel, to the achieved throughput and to the bearer priority. Time domain scheduling targets at user selection based on the priority, QoS or fairness.

A *Proportional Fair* scheduler is an example of a well performing scheduler, which weights the scheduler decisions according to a priority metric that can be set to most efficiently use the best radio resources for the selected users, but it also serves the users that experience low throughput expectations due to difficult channel conditions (cell edge), hence it is fair for users in proportions of the throughput.

For a scheduling function, combinations of many arguments may be formed, and often the knowledge of the channel state and possibly SINR of the candidate resources are also used by the scheduler, whose primary constraints are the bearer priority, QoS attributes and delay constraints given by the higher layer protocols.

In general, the scheduler[2] should select user $\mathbf{m}'$ among the active candidate users $\mathbf{m}$ at a scheduling period to the resources n, for whom;

$$\mathbf{m}' = \arg\max_{\mathbf{m}}\{P_{\mathbf{m}}(n)\}, \tag{8.2}$$

where P(n) depends on the utility function of the scheduler, which includes the priority metrics.

The scheduled user $\mathbf{m}'$ among the candidate users $\mathbf{m}$ with the highest priority metric according to the utility function is the one that actually gets scheduled to the candidate physical resources (n) during this period [220].

The utility function may consist of soft algorithms (linear, logarithm, time weighting, throughput weighting), but it may further include tight cut-offs, say if a bearer priority is strict, a higher priority bearer would always be scheduled instead of a lower priority one. Such strict priorities inside a utility function are feasible only if the traffic is not exhaustive to block other users traffic.

As said, the scheduler targets at using the physical resources in the best possible way. WCDMA/HSPA allows time scheduling during scheduling periods, which applies on the resolution of a frame (10 ms for WCDMA) or per subframe (2 ms for HSPA). On the transport channels, the transport format is selected per TTI, which is a multiple of 10 ms frames in WCDMA, a single 2 ms subframe in HSPA and a single 1 ms subframe in LTE.

[2]Theory shown by Kelly in European Trans. Telecommunications, vol.8, 1997 and again by J. MO and J. Walrand in IEEE Trans. on Networking, vol.8, no. 5, Oct. 2000

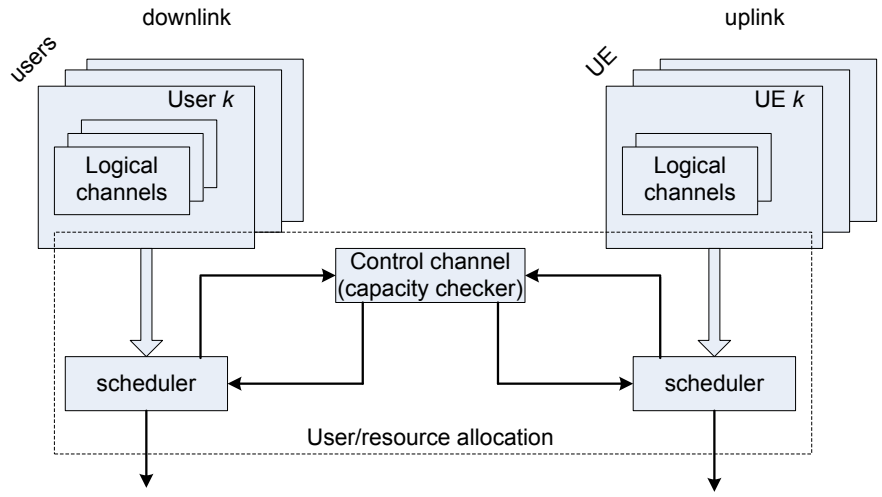


Figure 8.5: Scheduling of the logical channels.

Due to the multiplexing of logical channels on to the physical codes, the transport format combination is formed per frame in WCDMA/HSPA and per one subframe in LTE. A transport block set (WCDMA) or a single transport block (HSPA) is triggered once per TTI. In WCDMA, stable interference behaviour is appreciated and scheduling typically acts on much longer term (about 100 to 400 ms periods), whereas HSPA acts fast on the short term variations. Long term scheduling is done in WCDMA, because the schedulers act in the RNC, which causes delays (relative to the channel state information), and the channel variation is compensated by the other RRM algorithms [221]. Fast scheduling is done in HSPA, because the schedulers act in the NodeBs, which make them fast relative to the channel, and the channel variation is not compensated by the other RRM algorithms, but is exploited in the scheduler and link adaptation formats instead.

LTE creates a flexible size transport block per TTI and the scheduler acts both in time and frequency domains, see Fig. 8.5, in a very high resolution of one subframe (1 ms) and one PRB (180 kHz). The LTE schedulers may provide the highest expected gains. There may be lack of accurate channel state information though, which reduces the scheduling gains, but even then, fast blind scheduling (e.g. block frequency hopping) may gain.

8.3.1 Channel dependent scheduler

WCDMA applies variable spreading for the data rate adaptation, and therefore adaptive modulation is not feasible. WCDMA scheduling operates in time domain as all channelization codes perform equally and frequency diversity over the full bandwidth is always

available. Time scheduling is based on mutual SINR equations on long-term, because the fast power control is expected to compensate the fast channel state variations, and the outer-loop power control sets the BLER target per channelization code for the quality control (see QoS). In downlink, interference is experienced uniquely by each UE depending on its location in the cell. Close to the NodeB, the intracell interference is small due to small dispersion, which maintains the orthogonality property of the transmitter sequences quite will in the receiver as well. Also intercell interference may be fairly small due to large distance to the interfering cell. At the cell edge, the intracell interference increases because the channel dispersion gets larger and the code orthogonality at the transmitter is partly lost in the receiver. The intercell interference is also larger because of the reduced propagation distance to the interfering cell. Macrodiversity combining, again, converts the interfering resources to a useful signal, but that on the other hand causes system performance loss, because duplicated data has to be transmitted from multiple cells. In uplink, interference is experienced commonly at the cell receiver as a sum of the scrambling code correlation noise i.e. the noise rise as the Interference over the Thermal noise (IoT), see Section 8.6.

For downlink, the network selects the spreading factor in the OVSF code-tree per UE so that it can transmit the scheduled transport format combination set on a channelization code. The channelization code is allocated in a dedicated manner, and the spreading factor is selected for rate adaptation. Actually, the network selects a transport format from the transport format set for each logical channel, and after multiplexing it forms the transport format combination set to be transmitted on a channelization code during a frame. Signalling procedures as *TransportChannelReconfiguration* and *PhysicalChannelReconfiguration* may change these settings (see Section 6.5.6).

For uplink, the network provides UE a transport format combination set that the UE may freely use in the selection of the transport formats per TTI. In case, the signal quality in terms of BLER and QoS becomes a problem at the receiver, the network needs to update the transport format combination set respectively. The most robust transport formats are always present in the transport format combination set, but the highest formats may be selectively masked e.g. due to causing excessive load. The HSPA performance in uplink is studied in [222].

For HSPA downlink transmission, network provides a transport block per TTI. Uplink CQI is available as a MCS proposal by the UE. The UE reports the highest MCS that it

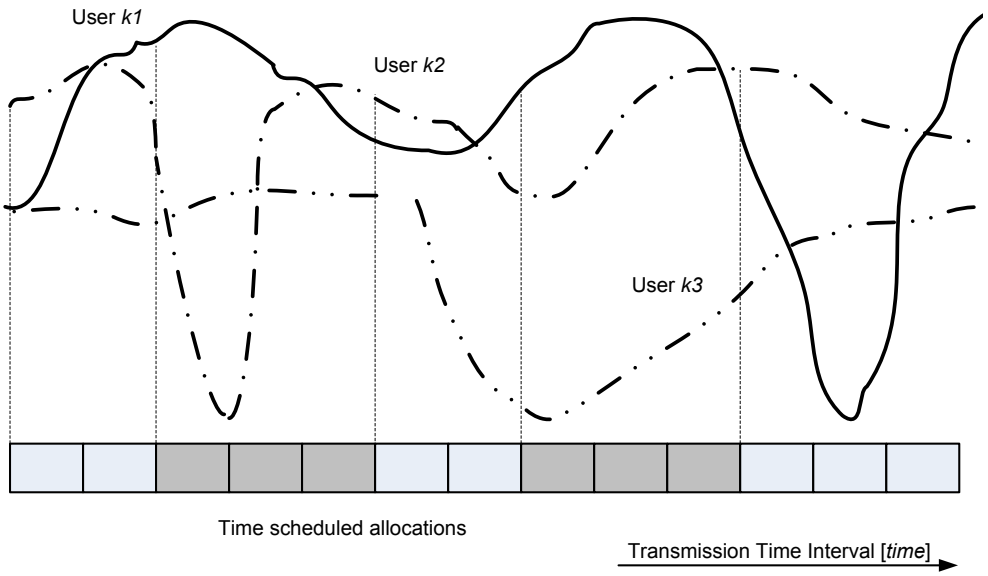
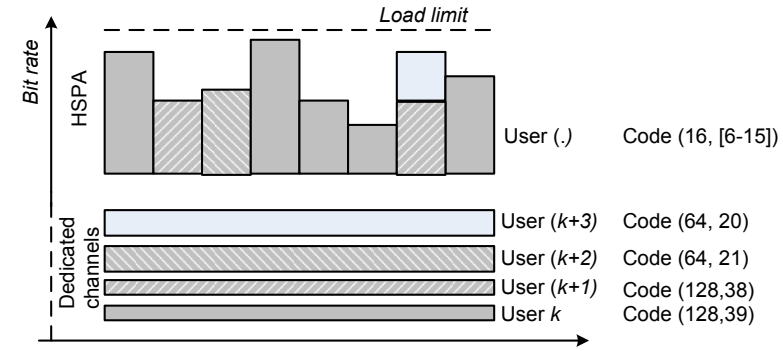


Figure 8.6: Time scheduled allocations of HSPA in a time-variant channel.

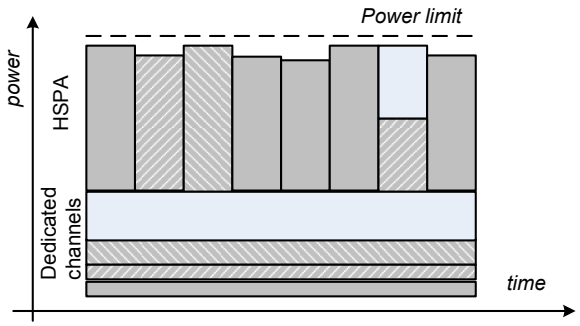
could receive with the target quality, if the recently measured channel conditions remain. However, the network may choose a more robust format e.g. due to a small transport block available, or it may risk the quality by choosing a less robust format operating at higher BLER. The UE may report MCS proposals based on the experienced target quality including SINR measurements, BLER or actually experienced QoS. The receiver algorithms contribute largely to the expected BLER as a function of SINR. For the advanced receiver algorithms, see Chapter 4, a lower SINR threshold may already lead to a proposal of a higher MCS compared to a basic receiver in the same channel state with the same expected SINR. In the case of multiantenna transmission, in addition to the SINR (CQI report) the channel rank information (PCI) had to be taken into account for the selection of the transport format including the MIMO formats. The WCDMA/HSPA-scheduler is shown in Fig. 8.6 and its time scheduling principle in Fig. 8.7.

For HSPA uplink transmission, UE provides a transport block per TTI. The HSPA additionally offers MCS of 4PAM modulation, which increases symbol rate compared to WCDMA. In HSPA, uplink may use macrodiversity signal combining in the network, but the uplink resources are always controlled by a single (master) cell, and therefore scheduling is much faster than in WCDMA.

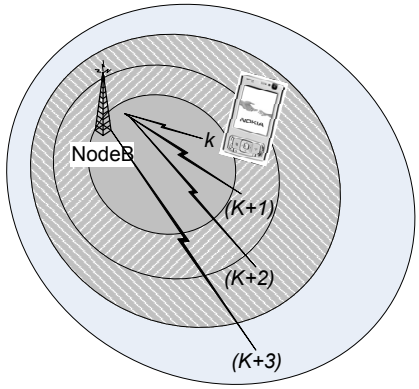
In LTE, the physical resources are defined in time and fractional frequency of the system bandwidth. Due to the frequency selectivity, scheduling may happen in both time and fre-



(a)



(b)



(c)

Figure 8.7: Time scheduling principle for HSPA. (a) Allocation of the dedicated channelization code resources and HSPA channelization code resources and the achieved bit rate. (b) Allocation of the dedicated and HSPA channelization code resources and the cell transmit power. (c) The impact of transmit power and bit rate to the coverage. The achieved bit rate depends on the location and channel conditions, total NodeB transmit power is close to constant.

quency domains. The CQI report needs to provide information in the time-frequency resolution of physical resources, in order to allow efficient scheduling also in frequency domain. As the eNodeB allocates uplink resources too, it needs to have frequency selective information available also during the periods of no uplink transmission. In LTE, this is arranged by channel sounding, where the UE transmits periodic, fractional bandwidth sounding reference symbols (SRS) for the purposes of eNodeB measurements on the channel quality. The frequency domain scheduling in LTE is studied e.g. in [220][223][224][225][226].

For the selection of MIMO formats, the uplink CQI reports containing frequency selective SINR need to be complemented by the rank indicaton and frequency selective precoding selection. A long CQI report (CQI, RI, PMI) can be transmitted during the uplink transmissions time-multiplexed onto the PUSCH. During the periods of no uplink transmission, a small PUCCH payload is available for short CQI reports. Even if the PUCCH payload is fairly small, it can be aggregated for a longer report, if necessary.

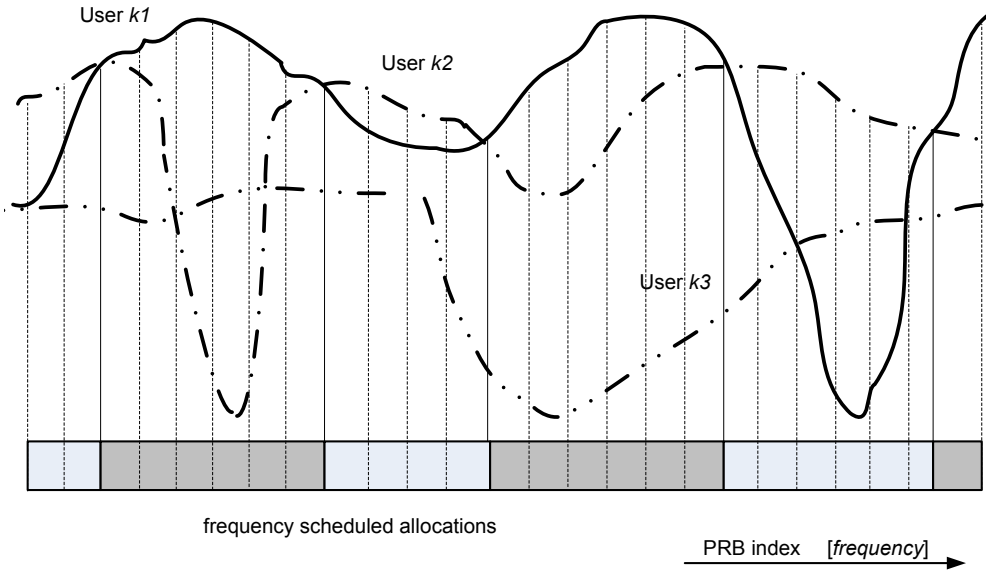


Figure 8.8: Frequency scheduled allocations of LTE in a frequency selective channel.

In all reporting, the channel coherence time is important, because the reports outside the coherence time are useless, and therefore fairly frequent reporting is necessary. The shorter is the reporting period and the higher is the reporting accuracy, the higher is the reporting overhead. The channel coherence time is a function of the UE velocity, and the scheduler gains a lot of the CQI reports at low velocity (3 km/h to 15 km/h), whereas at high velocity (120 km/h) the scheduling gains are lost and blind adaptive algorithms

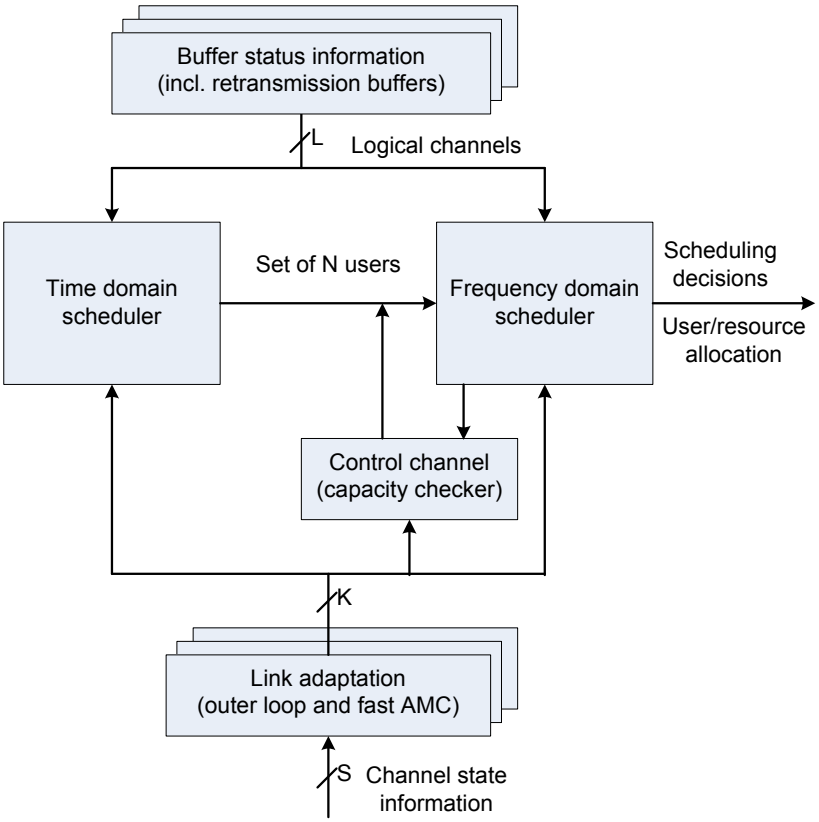


Figure 8.9: Time-frequency channel dependent scheduler of LTE.

are feasible instead. The channel is most difficult around at velocity 30 km/h, because reporting is still feasible and the scheduler may provide gains, but only if fast and accurate CQI reports are available, which increases the overhead and loading of the reverse link.

In LTE, the frequency selectivity of the channel additionally provides a very important gain factor for the scheduler. The CQI reports outside of the coherence band are useless, and therefore narrowband reporting is needed. The narrower is the reporting band and the higher is the reporting accuracy, the higher the reporting overhead. The impact of velocity to the channel selectivity is otherwise not that large, but in order to be valid, the frequency reporting has to happen to all the frequency components within the coherence time, which is short for a high velocity channel and which therefore largely adds the reporting overhead. Again, for a very high velocity channel (120 km/h), blind adaptive scheduling is feasible, or a distributed transmission may apply instead of the channel selective scheduling. The LTE scheduler is shown in Fig. 8.8 and its time-frequency principle in Fig. 8.9. The Quality of Service scheduling includes priority weighting of the packet flows, which is visualized with the buffer management in Fig. 8.10 [41][227].

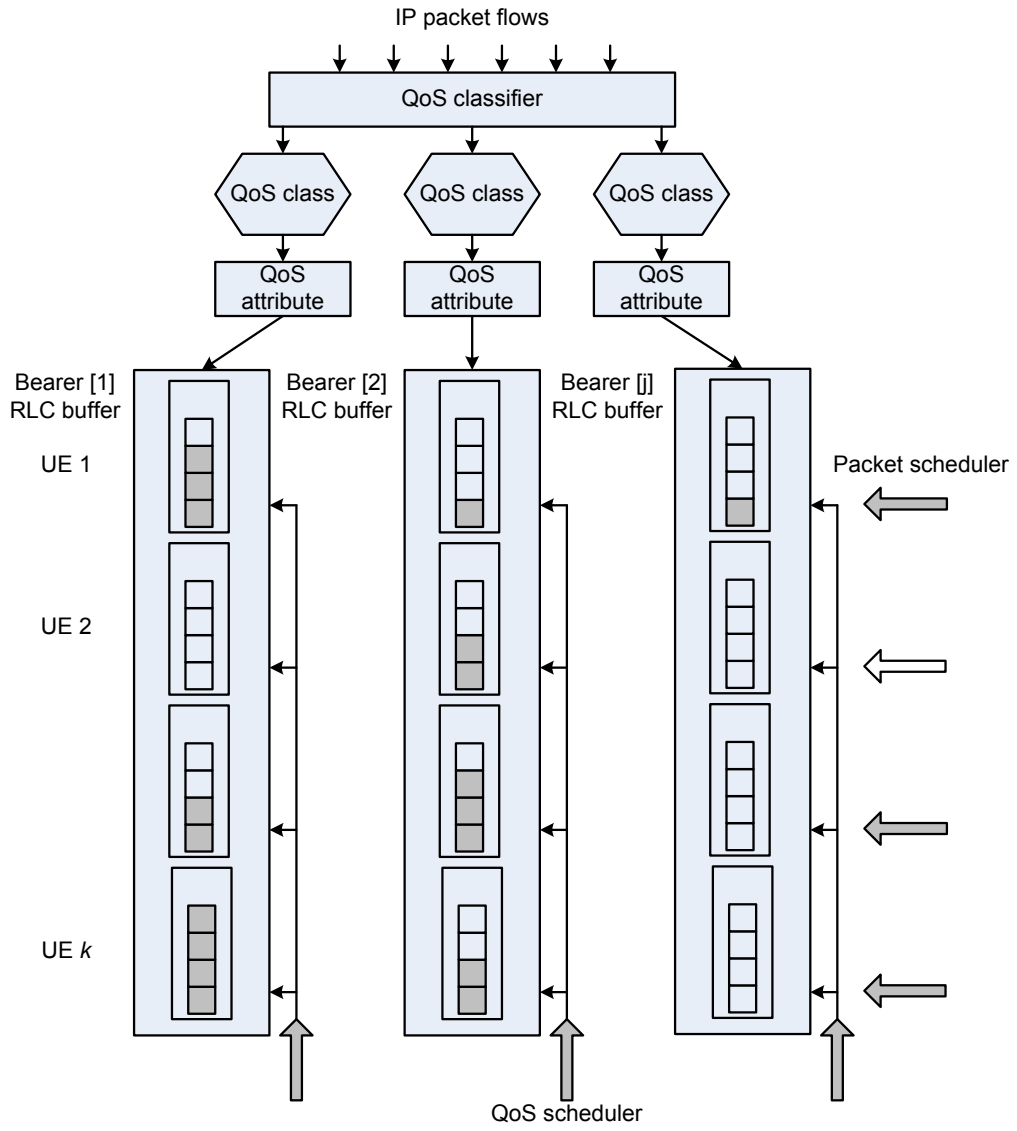


Figure 8.10: The principle of scheduling packets of the packet bearers. QoS classification and RLC buffer management of packet flows for the logical channels is described.

8.3.2 Scheduler algorithms

The schedulers target at maximizing the link throughput of each user and simultaneously maximize the cell throughput and capacity of the system. However, more complex than that, the schedulers have to include variety of priority metrics like weighting of users, weighting of traffic flows, weighting of perceived quality (e.g. delay) and fairness of resource use. The mutual importance and aggressiveness of each criterion is very subjective and can be given as a set of scheduling policies.

The scheduler may operate on a selected principle without weighting (Best Effort) or with weighting (Quality of Service). The scheduler may further combine several principles and have different weighting policy for each metric. The scheduling criteria may be gathered to a utility function, or a cost function, which returns the scheduling decisions. Often, channel dependent properties are inherently included in the weights, so that the link throughputs are obtained preferably in the good channel state resources, rather than in the poor channel state resources. In addition to the channel state, the interference conditions may be included in the scheduling decisions. This is visible e.g. when the instantaneous throughput maximization is used as the scheduling criterion. There, both the expected channel state and interference form the SINR argument of the throughput estimate. This is particularly useful in heavily loaded networks and at the cell edge, where low SINR is typical. Coordination of transmissions with low mutual interference to the co-channel resources is a benefit over scheduling highly interfering transmissions to the co-channel resources [12][31]. In practise, however, interference control requires measurements with fairly long averaging time and their control with dynamic packet traffic is not that well predictable. Therefore, interference averaging or interference randomization may be targetted instead of interference control, and other scheduling metrics will dominate. In packet communications, recently, delay sensitive criterion have begun to gain more attention.

As described earlier in Section 8.3, it can be noted that there exists a scheduling function $P_{\mathbf{m}}(n)$, which calculates the scheduling metric according to the function P for each user $\mathbf{m}$ for the scheduling resources n. Here, the scheduling resources may be a scheduling interval for a time-domain scheduler, a scheduling subband for a frequency-domain scheduler or a combination thereof. Common to all schedulers, the scheduler returns the user $\mathbf{m}'$ for the resources n, where $\mathbf{m}'$ maximizes the value of the scheduling function (i.e. the set of scheduling criteria) as already given in (8.2), and when additionally diversified for the frequency resource b, we get (8.3).

$$\mathbf{m}' = \arg\max_{\mathbf{m}} \{P_{\mathbf{m},b}(n)\} \tag{8.3}$$

For LTE, the frequency resources are PRBs in a subframe, hence $b \subseteq \{PRB_i\}_{i=1}^{N_{PRB}}$ and $N_{PRB} = N_{BW}/\ N_{sc}^{PRB}$. In this Section, some typical schedulers are shortly described by their scheduling functions in (8.4) to (8.16).

Time-domain proportional fair scheduler maximizes the expected instantaneous throughput $\hat{r}(n)$ for the scheduling resources n relative to the long-term mean throughput $\overline{R_{\mathbf{m}}}$ experienced by each user $\mathbf{m}$ before the pending scheduling decision. $\hat{r}(n)$ is evidently a function of SINR. It can be noted that optionally the long-term mean may be the weighted mean for the more recent scheduling events over the long-time past decisions [228]. The scheduling function of the time-domain proportional fair scheduler is,

$$TD-PF; \quad P_{\mathbf{m}}(n) = \frac{\hat{r}_{\mathbf{m}}(n)}{\overline{R_{\mathbf{m}}}} \tag{8.4}$$

The scheduling function of the frequency-domain proportional fair scheduler (with frequency resource index b) is,

$$FD-PF; \quad P_{\mathbf{m},b}(n) = \frac{\hat{r}_{\mathbf{m},b}(n)}{\overline{R_{\mathbf{m}}}} \ . \tag{8.5}$$

Frequency domain adaptive scheduler is studied in [224].

The scheduling function of the guaranteed rate proportional fair scheduler (8.6) includes coefficients $\underline{\alpha}$ and $\underline{\beta}$ to determine, how aggressively the current scheduling decision should target at meeting the guranteed rate. When the mean exceeds the guaranteed rate, aggressiveness may be less and smooth, but when the mean drops below the guranteed rate, high aggressiveness is needed.

$$GPF; \quad P_{\mathbf{m}}(n) = \frac{\left[\hat{r}_{\mathbf{m},b}(n)\right]^{\underline{\alpha}}}{\left[\ \overline{R_{\mathbf{m}}}\ \right]^{\underline{\beta}}} \tag{8.6}$$

with $\overline{R_{\mathbf{m}}(n)} = (1-\frac{1}{a})\overline{R_{\mathbf{m}}(n-1)}+\frac{1}{a}\ R_{\mathbf{m}}(n-1)$, where the forgetting factor a is proportional to the averaging time. This equation represents the weighted average and it is updated for every scheduling period. In case, user $\mathbf{m}$ was not scheduled, $R_{\mathbf{m}}(n-1) = 0$ and in case, user $\mathbf{m}$ was served with its expected instantaneous throughput, $R_{\mathbf{m}}(n-1) = \hat{r}_{\mathbf{m}}(n-1)$.

The scheduling function of the time-domain maximum throughput scheduler is,

$$TD-MT; \quad P_{\mathbf{m}}(n) = \hat{r}_{\mathbf{m}}(n) \ . \tag{8.7}$$

The scheduling function of the time-domain blind equal throughput scheduler is,

$$TD-BET; \quad P_{\mathbf{m}}(n) = \frac{1}{\overline{R_{\mathbf{m}}}} \ . \tag{8.8}$$

The scheduling function of the frequency-domain maximum equal resource scheduler is,

$$FD-MER;\quad P_{\mathbf{m}}(n)=\hat{r}_{\mathbf{m},b}(n)\Big|\big\{\textstyle\sum_i b_i(n)=\sum_j b_j(n)\big\},\forall\ i,j\ \ ;\ i\neq j\ . \tag{8.9}$$

The scheduling function of the frequency-domain maximum throughput-to-average scheduler is (8.10), where the scheduling decision maximizes the expected throughput obtained in a frequency selective candidate subband b over the wideband, instantaneous, frequency-averaged throughput.

$$FD-MTA;\quad P_{\mathbf{m}}(n)=\tfrac{\hat{r}_{\mathbf{m},b}(n)}{\hat{r}_{\mathbf{m}}}\ . \tag{8.10}$$

The QoS aware weighted proportional fair scheduler is,

$$QwPF;\quad P_{\mathbf{m}}(n)=\tfrac{\hat{r}_{\mathbf{m}}(n)}{\overline{R_{\mathbf{m}}}}\cdot w_{\mathbf{m}}(s)\ , \tag{8.11}$$

where $w_{\mathbf{m}}(s)$ is the QoS weight factor of a service flow s that is active for user $\mathbf{m}$.

Other more complex utility functions relative to the guaranteed bit rates may be defined as (8.12) or by added fairness in (8.13). The guaranteed bit rate realizes over a long term average as $\overline{R_{GBR}}$. In (8.12) and (8.13), the frequency domain scheduling index could as well be included.

$$GBR;\quad P_{\mathbf{m}}(n)=\hat{r}_{\mathbf{m}}(n)\cdot\big[1+\underline{\beta}\cdot\exp^{\ -\underline{\beta}(\overline{R_{\mathbf{m}}}-\overline{R_{GBR}})}\big] \tag{8.12}$$

$$GBR-F;\quad P_{\mathbf{m}}(n)=\hat{r}_{\mathbf{m}}(n)\cdot\big[\tfrac{1}{\overline{R_{\mathbf{m}}}}+\underline{\beta}\exp^{\ -\underline{\beta}(\overline{R_{\mathbf{m}}}-\overline{R_{GBR}})}\big] \tag{8.13}$$

With the frequency domain allocations, the instantaneous throughput may be estimated per subband (or PRB) with $\hat{r}_{\mathbf{m},b}(n)$ instead of $\hat{r}_{\mathbf{m}}(n)$.

Yet, other schedulers can be designed including a delay criterion. Such examples are given in (8.14), (8.15) and (8.16). Delay weighting is important for real-time services and QoS, as introduced in [229][230]. Delay weighting in the form of the modified largest weighted delay first is studied for HSPA in [231] and for LTE in [232]. The scheduling function of the modified largest weighted delay first scheduler is,

$$M-LWDF;\quad P_{\mathbf{m}}(n) = -\log(\delta_{\mathbf{m}})\frac{\hat{r}_{\mathbf{m}}(n)}{\overline{R_{\mathbf{m}}}} \cdot \frac{\max\{d_{\mathbf{m}}\big|B_{\mathbf{m}}(n)\}}{d_{req,s}} \ , \tag{8.14}$$

where $\delta_{\mathbf{m}}$ is the aggressiveness of the delay criteria set for user $\mathbf{m}$, $d_{req,s}$ is the maximum delay constraint for a service flow s (served for user $\mathbf{m}$), $d_{\mathbf{m}}$ is the delay of a packet in the transmit buffer $B_{\mathbf{m}}$ of user $\mathbf{m}$ and $\max_d\{d_{\mathbf{m}}\big|B_{\mathbf{m}}(n)\}$ is the longest delay (d) of a packet for user $\mathbf{m}$ in buffer $B_{\mathbf{m}}$ at the scheduling time n. This delay is also called the head of line packet delay. In (8.14), the first factor is the QoS differentiation term (a.k.a the desired maximum violation probability [230][3], the second factor is the proportional fair property and the third factor is the delay weight.

The scheduling function of the channel dependent earliest due date scheduler given in (8.15) is similar to M-LWDF in (8.14), however with a different formulation of the QoS term.

$$CD-EDD;\quad P_{\mathbf{m}}(n) = a_s \cdot \frac{\hat{r}_{\mathbf{m}}(n)}{\overline{R_{\mathbf{m}}}} \cdot \frac{\max\{d_{\mathbf{m}}\big|B_{\mathbf{m}}(n)\}}{d_{req,s}} \ , \tag{8.15}$$

where a_s is the QoS term for the service flow s of user $\mathbf{m}$.

The scheduling function of the exponential rule scheduler is,

$$ER;\quad P_{\mathbf{m}}(n) = a_s \cdot \frac{\hat{r}_{\mathbf{m}}(n)}{\overline{R_{\mathbf{m}}}} \cdot \exp\left(\frac{a_s \cdot \max\{d_{\mathbf{m}}\big|B_{\mathbf{m}}(n)\} - \frac{1}{N}\sum_{\mathbf{m}=1}^{N} a_{s,\mathbf{m}} \cdot (\max\{d_{\mathbf{m}}\big|B_{\mathbf{m}}\})}{1+\sqrt{\frac{1}{N}\sum_{\mathbf{m}=1}^{N} a_{s,\mathbf{m}} \cdot (\max\{d_{\mathbf{m}}\big|B_{\mathbf{m}}\})}} \right) \ , \tag{8.16}$$

where a_s is as in (8.15) and $s \in \mathbf{S}_{QoS}$. $\mathbf{S}_{QoS}$ is the QoS weight-set of the service flows. In other words, s is the relative priority of a traffic flow in the set of QoS classes $\mathbf{S}_{QoS}$, see Subsection 6.3 for QoS class identifier (QCI). According to (8.16), the scheduler attempts to equalize the weighted delays[4] of the users, when their mutual differences are large. Say, if a user has clearly larger delay than the average delay of all users relative to the average delay of all users, that user will get a high scheduling metric because of the exponent function. When all users have delays close to the average of all users, the exponent term approaches unity and the scheduler equals Proportional Fair among that QoS class.

The delay dependent schedulers may cause unwanted side effects by forcing un-

[3]The QoS differentiation term, here, is logarithmic due to the delay is inverse proportional to the channel bitrate as Shannon capacity. For simplicity, the base of logarithm may be ten instead, because it as well preserves a unique, monotonic mapping to the scheduling priority.

[4]The delay considered here, is the maximum packet delay (head of line delay) relative to the delay criterion.

favourable decisions at a critical moment. For example, M-LWDF in (8.14) is throughput optimal, but is not fair, unless modified according to [233]. The impact of the SDU discard timer to the throughput and outage was studied for HSPA, and it was concluded that (8.14) only sees the number of flows not meeting the QoS, which it tends to minimize without a resource concern. Hence, [233] proposes to let QoS of the flows degrade in the most resource consuming links and instead let a higher number of other flows meet their QoS requirements. On the contrary, it may be seen that if the QoS guarantee is highly valued, the optimal cell throughput and capacity may let be degraded for the sake of meeting that QoS requirement of the flows.

Other convenient scheduling choices may be set with lower than exponential aggressiveness, e.g. to simply monitor and measure the delay distribution and have a soft impact on the scheduling priorities, if the delay distribution seems to become poor relative to the delay thresholds. This of course requires measurements and sufficient reference statistics for the decision making.

In any case, voice packet traffic has specific charcteristics, and voice quality is often highly preferred over the other traffic classes. Therefore, voice will always require specific scheduling mechanisms similar to presented in [24]. In any of the scheduler equations above, the channel dependent property based on measurements can be included in the form of the CQI-report. It inserts the factor $\max\{CQI_{\mathbf{m},b}(n)\}/\ \overline{CQI_{\mathbf{m}}}$ to the scheduling function. The CQI applied to the frequency domain proportional fair scheduler with delay weights can be written,

$$P_{\mathbf{m}}(n) = \frac{\hat{r}_{\mathbf{m},b}(n)}{\overline{R_{\mathbf{m}}}} \cdot \frac{\max\{CQI_{\mathbf{m},b}(n)\}}{\overline{CQI_{\mathbf{m}}}} \cdot \frac{\max\{d_{\mathbf{m}} \big| B_{\mathbf{m}}(n)\}}{d_{req,s}}, \tag{8.17}$$

where $CQI_{\mathbf{m},b}(n)$ is the CQI report of frequency resources (b) for user $\mathbf{m}$, that is available at the scheduling event (n). $\overline{CQI_{\mathbf{m}}}$ is the CQI reported for the resources on average. In a time domain CQI based scheduler (index b is omitted), this metric weights the selection of a user, whose channel quality is expected to be the highest compared to its average channel state. In a frequency domain CQI based scheduler (index b included), this metric weights the selection of a user, whose channel quality in frequency resources (b) is expected to be the highest compared to its wideband average.

8.4 Handover

As the efficiency requirements are very high, the target is to design schemes which operate in frequency reuse one networks. This means that all frequency components of the allocated bandwidth are available in all cells. This is unlike in the conventional TDMA, where the co-channel interference is so dominant that the cells using the same frequency had to be at sufficient distance. For continuous coverage, geographically overlapping cells had to be arranged by increasing the frequency reuse. This decreases the spectral efficiency by a factor of 1/reuse. Again, when high throughput is required, the system bandwidth has to be large and frequency reuse is not tolerable for the spectral efficiency at all.

In WCDMA and WCDMA/HSPA, frequency reuse one is provided by the spreading and scrambling operations. Spreading provides robustness against narrowband interference (see Section 3.3). Scrambling avoids intersequence (channelization code) interference, because scrambling creates cross correlation noise that is alike Gaussian white noise due to long pseudo-noise sequences in use. In downlink, channelization codes are (ideally) orthogonal per cell. For a frequency reuse one network with the same tree of channelization codes in use in all cells, the cell specific scrambling provides protection against the intercell interference. In uplink, the channelization codes are non-orthogonal due to the arbitrary signal phase between the transmitted signals from different UEs. Hence, the uplink signals need to be protected by the UE specific scrambling both against the intracell and intercell interference.

Soft handover is inherent for Rake receivers, which receive individual signal components by correlators and combines them before decoding. Hence, the complexity of combining is not largely different whether the signals are transmitted from the same cell or from different cells. Of course, the timing of the physical signals to be combined has to be adjusted to appear within a combining window at the receiver. Also, the information bits and modulation symbols of the combined channels have to be equal, even if they are transmitted from different cells (or sites). Besides, the cell-specific control channels are not defined to operate in macrodiversity, but they need to be separately received for all cells in the active set. The soft handover is not feasible with the MMSE receivers in practise, and it would hence require multiplication of the receivers in the UE. The hard handover with the MMSE receiver may on the contrary perfom well, if the channel dispersion is not large and if the multipath propagation is not severe.

For WCDMA with HSPA, the cell edge coverage may be satisfied by switching logical channels from the HSPA channels to the dedicated channels having an active set in the soft handover. However, HSPA channels themselves do not execute a soft handover and at the cell edge, the neighboring cell behaves as an interferer and does not combine constructively. This sets a challenge for the HSPA only transmission scenario in other than isolated cells. For a soft combining receiver, the cell edge is assumed at about 0 dB SINR. For a MMSE receiver, the cell edge may be far below 0 dB.

For WCDMA, soft handover operates by adding radio links to the active set, which includes all cells (cell sectors) that transmit identical signals (cell specific scrambling is different though) to be combined at the receiver. The active set is adaptive, and it is set by the network (*ActiveSetUpdate* procedure) per UE assisted by the UE measurements. The active set applies for those physical channels, whose signals on the channelization codes can be combined. Naturally, the active set is equal for all channels of the UE operating in macrodiversity. Soft combining for the downlink signal is arranged in the network (NodeB and RNC) and executed in the UE receiver. The uplink signal combining is done in the network (NodeB and RNC) so that the NodeB may do soft combining and the RNC will do selection combining.

In LTE, the OFDMA signal combines the multipath propagated signals inherently, as the channel dispersion is mitigated by the guard interval and equalization algorithms. The soft combining of signals from different cells is not practical as that would require independent channel estimation and independent receiver for signals transmitted from different cells.

In LTE, for a hard handover in a frequency reuse one network, the signal conditions to be received reach strongly negative SINR [234]. Typically, the cell edge reception is assumed at about -5 dB SINR and due to handover delays and margins reception at about -7 dB is still possible. This is feasible by the cell specific scrambling, cell-specific independently fading frequency components, link adaptation and frequency domain scheduling. LTE handover is controlled by the network (*HandoverCommand* procedure) and is assisted by the UE measurements.

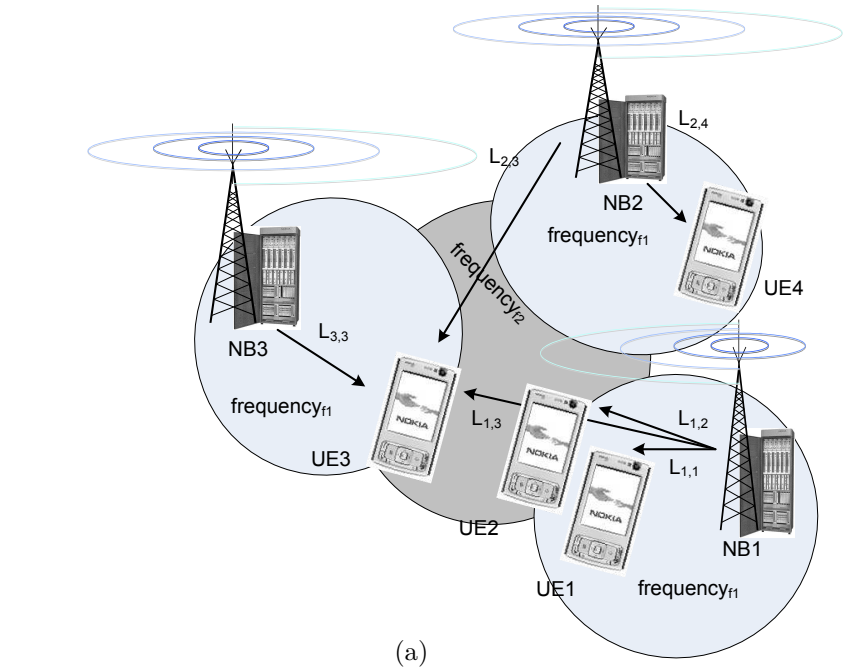
8.5 Algorithmic collaboration

RRM is a manifold set of functionalities that contributes largely to the system performance. Not a single algorithm alone can be crucial, but each of them may provide gains (or losses). Well-performing system can only be achieved, if RRM algorithms collaborate positively and the concept can be sufficiently supported by the protocols [211][235][236][237]. If any algorithm in the RRM algorithm set is neglected or is misfunctional, the system performance will decrease or radically drop even [238]. In a specific case, a high performance gain may be shown by careful tuning. However, more important is to apply soft algorithms and adaptation that can work commonly in dynamic environments in all channel states and in all network scenarios at all times. This is the favour of generic (standardised) protocol machines that deliver universal message contents, where parameters and values serve tuning to each operational state. For a proper justification, feasibility of the RRM concept shall therefore be shown in scenarios including large number of channel states, propagation conditions, interference (load) situations and their dynamics. This is contrary to the achievable gains reported in a single specific scenario with a set of algorithms well tuned just for that.

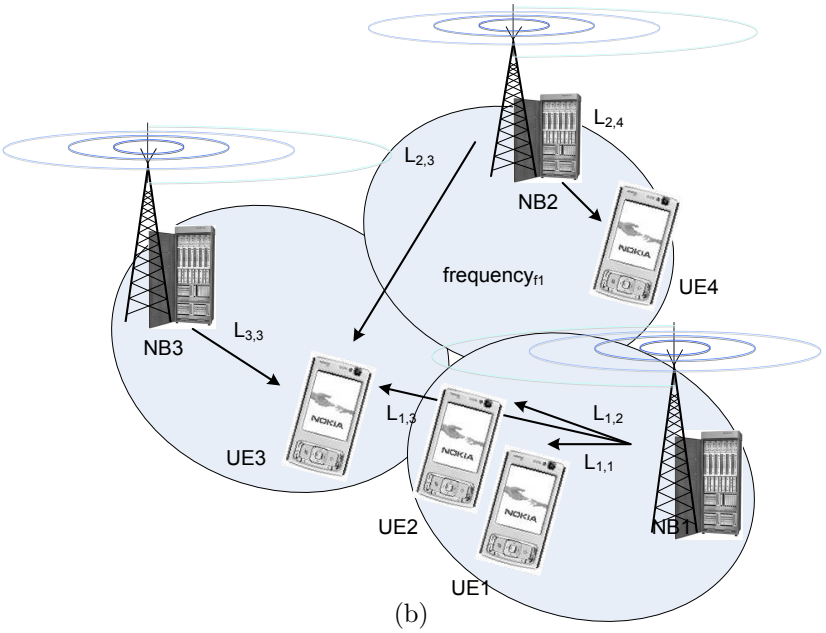
8.6 SINR and load

The great challenge of WCDMA, WCDMA/HSPA and LTE is to operate on frequency reuse one, which is contrary to the 2G systems that require large distance to the co-channel interferer, and hence introduce large frequency reuse factors, which significantly decrease the spectral efficiency. Frequency reuse one is possible in WCDMA/HSPA by the spreading and scrambling operations. Frequency reuse one is possible in LTE by frequency localized selective scheduling and scrambling. Fig. 8.11 visualizes the impact of frequency reuse one to the SINR equations.

In WCDMA, the wideband SINR in downlink at the UE is given by (8.18) and the wideband SINR in uplink at the NodeBs is given by (8.19) respectively. In downlink, p_{rx} is the received wideband power transmitted to the receiver. $I_{rx,j}$ is all the transmitted power from the serving cells in the active set that is received as interference at the receiver and that is not perfectly orthogonal to the desired code channels. Non-orthogonality of the code channels is defined by the (average) orthogonality factor $(1 - \alpha)$. $I_{rx,k}$ is the received interference power from the cell neighbourhood that does not belong to the active



(a)



(b)

Figure 8.11: Intercell interference (a) in a frequency reuse network and (b) in a frequency reuse one network

set. I_{ACI} is the total adjacent channel interference power received after filtering from the adjacent frequency carriers. In downlink, the adjacent channel selectivity of a UE receiver dominates the adjacent channel power leakage property of a NodeB transmitter.

$$SINR^{(UE)} = \frac{P_{rx,i}^{(UE)}}{(1-\alpha)\cdot \sum_{\substack{j\subset J_{AS} \\ j\neq i}} I_{rx,j}^{(UE)} + \sum_{k\subset K_{NS}} I_{rx,k}^{(UE)} + I_{ACI}^{(UE)} + \eta}, \tag{8.18}$$

where J_{AS} is the set of serving cells in the active set (AS), and K_{NS} is the set of NodeBs transmitting in the neighbourhood of the serving cell that generate interference at the UE receiver. $(1-\alpha)$ is the orthogonality factor, η is the thermal noise power. For the HSPA channels, the active set includes a single serving cell and thus j is the sum of interfering radio links in that cell only. All other cells in the neighbourhood generate non-orthogonal interference.

In uplink, p_{rx} is the received wideband power transmitted by a single UE to the network that is received at the NodeB. $I_{rx,j}$ is all the received interference power transmitted by the other UEs served in the same cell. $p_{rx,k}$ is all the received interference power from the UEs transmitting to the other cells in the neighbourhood. I_{ACI} is the total adjacent channel interference (ACI) power received after filtering from the adjacent frequency carriers [13]. In uplink, the property of the adjacent channel power leakage of a UE transmitter dominates the adjacent channel selectivity of a NodeB receiver.

$$SINR^{(NB)} = \frac{P_{rx,i}^{(NB)}}{\sum_{\substack{j\subset J_{UE} \\ j\neq i}} I_{rx,j}^{(NB)} + \sum_{k\subset K_{UE}} I_{rx,k}^{(NB)} + I_{ACI}^{(NB)} + \eta}, \tag{8.19}$$

where J_{UE} is the set of transmitting UEs served by the same serving NodeB, and K_{UE} is the set of transmitting UEs served by other NodeBs in the neighbourhood that generate interference. η is the thermal noise power.

The wideband SINR is given in (8.18) and (8.19) above. Multiplying (8.18) and (8.19) by the processing gain $G_{p,j}$ for a desired code channel j, yields the narrowband signal on the code channel over the wideband interference. The processing gain of the spreading operation has a large impact to the baseband link budget, ranging from about 12 dB for a HSPA code channel to over 20 dB for a low rate code channel.

The link capacity $\mathbf{c}$ (see Section 8.8) is defined by the SINR, as each code channel (j) can receive bit rate R_j with power $p_{rx,j}$ for the target $(E_b/\ N_0)_j$ according to

$$SINR_j = \left(\frac{E_s}{I_0}\right)_j = G_{p,j}\cdot \frac{P_{rx,j}}{I_{total}+\eta} = \frac{\mathbf{W}}{R_j\cdot \lambda_j}\cdot \frac{P_{rx,j}}{I_{total}+\eta} \tag{8.20}$$

and

$$\left(\frac{E_b}{N_0}\right)_j = \mathsf{C}\left(\frac{E_s}{I_0}\right)_j \tag{8.21}$$

E_b/N_0 is obtained from E_s/I_0 by the channel coding gain (C) after the receiver algorithms including symbol combining and interference rejection have acted. $\mathbf{W}$ is the chip rate and R_j is the channel coded bit rate carried on the modulation alphabet. λ_j is the activity factor. The effective bit rate is $R_j(\text{eff}) = R_j/\text{code rate}$ (see Entropy coding in Section 8.8.3) with residual BER proportional to $\left(E_b/N_0\right)_j$.

In downlink, WCDMA/HSPA capacity is limited by the total transmit power to maintain SINR for the required Eb/No and BLER targets at the receivers. The downlink load factor is defined either as the transmit power load (8.22) or the throughput load (8.23) as a function of bit rate.

$$\mathcal{L}_{DL} = P_{tx,total}/\max\{P_{tx}\}, \tag{8.22}$$

or

$$\mathcal{L}_{DL} = \sum_{j=1}^{N} R_j \frac{(\frac{E_b}{N_0})_j \cdot \lambda_j}{\mathbf{W}}[(1-\alpha_j) + i_j], \tag{8.23}$$

where the load observation is different per each UE, so that the orthogonality and the other cell interference observations are specific to a UE with the other cell interference to the serving cell interference ratio (i). If we study a receiving UE (j), we can note that it has processing gain both over the non-orthogonal serving cell interference and over the other cell interference. The mean downlink load may be given by (8.24) with the notation of the mean orthogonality and the mean other cell to the serving cell interference ratio ($\overline{i}$). However, the mean values in here hide the important property of the UE location dependent geometry factor $\mathbf{G}$ given by (8.25).

$$\overline{\mathcal{L}_{DL}} = [(1-\overline{\alpha}) + \overline{i}] \cdot \sum_{j=1}^{N} \frac{1}{\frac{\mathcal{W}}{(\frac{E_b}{N_0})_j \cdot R_j \cdot \lambda j}} \tag{8.24}$$

$$\mathbf{G} = \frac{1}{\overline{i}} = \frac{\sum_{j \subset J_{AS}} P_{rx,j}}{\sum_{k \subset K_{NS}} P_{rx,k} + \eta} \tag{8.25}$$

In uplink, WCDMA capacity is typically limited by the interference, but for very

large cells with small number of users, also thermal noise limit may exist. The higher the interference load in the network, the higher transmit power is needed to maintain the SINR for the required Eb/No and BLER target at the receiver. Therefore, WCDMA network load is defined as the wideband noise rise, which is the integrated interference power over the thermal noise at the receiver according to,

$$\mathcal{L}_{UL} = 1 - \frac{\eta}{I_{total}}, \tag{8.26}$$

where the noise rise is the interference plus thermal noise power load relative to the thermal noise power, i.e. I_{total}/η.

A differential increase of the interference load will cause the need to increase the transmit power by many UEs, therefore the load increase is cumulative and shows a well-known exponential curve having a pole capacity. The network can safely be operated as long as the load control algorithm can predict the impact of a differential increase of load to the mean noise rise. Typically, this load may reach up to about 0.4 to 0.7 of the pole capacity [14][6].

Actually, in WCDMA the load estimation is based on the wideband interference power [6], because it behaves rather steadily. The wideband power estimation keeps the coverage probability within the planned limits. The coverage probability is the second order momentum of BLER a.k.a as the outage probability. The capacity of the network depends on both the load in the serving cell and the load in the cell neighbourhood, so that the inverse of the serving cell to the other cell interference ratio (i) impacts on the soft capacity. The higher the other cell interference relative to the serving cell interference, the smaller is the capacity of the serving cell for a given load. The alternative load definition in terms of the throughput may be given as,

$$\mathcal{L}_{UL} = (1+i) \cdot \sum_{j=1}^{N} \mathcal{L}_j = (1+i) \cdot \sum_{j=1}^{N} \frac{1}{1+\mathcal{L}_j}, \tag{8.27}$$

where the load factor of UE_j is $\mathcal{L}_j = \mathbf{W}/\Big(\big(E_b/N_0\big)_j \cdot R_j \cdot \lambda_j\Big)$. In the equations above $P_{rx,j}$ is the received power in downlink (8.28) and in uplink (8.29);

$$P_{rx,i}^{(UE)} = \sum_{j \subset J_{AS}} P_{tx,j}^{(NB)} \cdot L_j^{(UE)} \cdot \Omega_j^{(UE)} \cdot \mathfrak{f}_j^{(UE)}, \tag{8.28}$$

$$P_{rx,i}^{(NB)} = \sum_{j \subset J_{UE}} P_{tx,j}^{(UE)} \cdot L_j^{(NB)} \cdot \Omega_j^{(NB)} \cdot \mathfrak{f}_j^{(NB)}. \tag{8.29}$$

where

$$\begin{aligned} L_j^{(UE)} &= L_j^{(NB)} \\ \Omega_j^{(UE)} &= \Omega_j^{(NB)}. \\ \mathfrak{f}_j^{(UE)} &\neq \mathfrak{f}_j^{(NB)} \end{aligned} \tag{8.30}$$

$P_{tx,j}$ is the transmit power from the BS_j in the active set to the UE_j, L_j is the pathloss from the BS_j to the UE_i, Ω_j is the shadow fading and $\mathfrak{f}_j$ is the sum fast fading power of the signal components transmitted from the BS_j to the UE_i.

$P_{tx,j}$ is the transmit power from the UE_j to the BS_i, L_j is the pathloss from the UE_j to the BS_i, Ω_j is the shadow fading and $\mathfrak{f}_j$ is the sum fast fading power of the signal components transmitted from the UE_j to the BS_i. Further, $L_j(UE) = L_j(BS)$ and $\Omega_j(UE) = \Omega_j(BS)$, but $\mathfrak{f}_j(UE) \neq \mathfrak{f}_j(BS)$.

The transmitted and received signal on a channelization code c_{ch} is

$$\begin{aligned} T &\rightarrow c_{ch,SF,m} \cdot m \cdot \cos(\omega t) \\ R &\rightarrow m(t - \tau_{\,l}) \cdot \textstyle\sum_l \alpha_l \cdot \;\; c_{ch,SF,m}(t - \tau_{\,l}) \cdot \cos(\omega_t + \omega_D t + \Theta_l) \end{aligned} \quad , \tag{8.31}$$

where l is the path index, h_l is the amplitude of the complex channel coefficient of path l, τ_l is the path delay of the channelization code c_{ch}, m is the modulated symbol and $\cos(\cdot)$ is the channel phase distortion, i.e. $\omega(t)$ is the transmitted symbol phase, ω_D is the Doppler shift and θ_l is the angular shift or phase of the complex channel coefficient of path l. The receiver synchronizes to the carrier in time, phase and frequency, it tracks and correlates the channelization code c_{ch} at phase τ_l and detects the (amplitude and) phase modulated symbol stream e.g. by a maximum likelihood detector, see Section 4.2.1.

For LTE, the link SINR is calculated per set of subcarriers that belong to the allocation of a single UE. Both downlink and uplink are orthogonal, if ideal filtering is assumed.

$$SINR^{(UE)} = \frac{\sum_{sc=sc[a]}^{sc[b]} P_{rx,sc}^{(UE)}}{\sum_{k \subset K_{NS}} \sum_{sc=sc[a]}^{sc[b]} \mathcal{I}_{sc,k}^{(UE)} + \eta}, \tag{8.32}$$

$$SINR^{(eNB)} = \frac{\sum_{sc=sc[a]}^{sc[b]} P_{rx,sc}^{(eNB)}}{\sum_{k \subset K_{UE}} \sum_{sc=sc[a]}^{sc[b]} \mathcal{I}_{sc,k}^{(eNB)} + \eta}, \tag{8.33}$$

where $\mathcal{I}_{sc,k}^{(\cdot)}$ is the frequency selective interference received at the receiver $(\cdot)$ on subcarrier sc from all interfering sources k transmitting on that subcarrier.

The total wideband SINR extends the subcarrier indexing over the full system band of PRBs i.e. $sc = 1, .., N_{sc}^{PRB} \cdot N_{PRB}$. In a multiantenna system, the received SINR may be observed per single antenna branch in the receiver, or as the sum of SINR of all receiver antenna branches, depending on the combining technique (see Chapter 4). On the other hand, the use of multiple transmit antennas has to split the total transmit power for the antenna branches. The received signal gains may still be observed even if the transmit power remains the same. This is due to the spatial diversity of the propagation paths and due to precoding of the transmitted symbols.

8.7 Coverage probability

The fundamental difference in WCDMA, WCDMA/HSPA and LTE coverage probability exist in the specifics of SINR. In WCDMA, the coverage probability is increased by the signal gains of macrodiversity combining in the active set of the serving cells. In downlink, the HSPA channels do not experience the active set gains and it may experience slightly worse coverage probability compared to WCDMA and WCDMA/HSPA (which may switch to the dedicated channel transport in coverage limited situations). The capacity gains of the HSPA links may on longer term average compensate the slightly lower coverage probability of a given throughput.

LTE gains a lot in coverage probability due to the frequency selective nature of SINR. Depending on the allocation (scheduling) decisions, an allocation to the subband resources of high signal power and low interference clearly increases the coverage probability compared to an allocation into the average wideband resources. Another gain of LTE in the coverage probability is the capability of adaptive subband allocations (see Section 8.2.6), which allows higher transmit power density per subband compared to a wideband transmission with the same total power.

The coverage probability may be separately calculated for the downlink and uplink, however, the more vulnerable link at a time will dominate the coverage [239]. As the pathloss and shadowing are equal for the two link directions, the multiple-access signal equations and the power dimensioning will determine, how vulnerable the link is. Often, the practical dimensioning of links follow the principle "downlink dies first". This is

beneficial to initiate handovers based mainly on the downlink measurements. Though, the SINR probability function behaves in a different way per multiple-access and though their statistical observations are different, the coverage probability as a function of SINR is commonly given by (8.34),

$$P_r(\gamma_0) = \frac{\int_{\gamma_0}^{\infty} p(x)\ d(x)}{\int_{-\infty}^{\infty} p(x)\ d(x)}, \tag{8.34}$$

where $p(\cdot)$ is the probability function of the argument SINR and $P_r(\cdot)$ is the cumulative distribution function of p(SINR) i.e. $P_r(\gamma) = \int_{\gamma}^{\infty}\ p(SINR)\ d(SINR)$ integrated over the high SINR regime. The cumulative distribution function $P_r(\gamma)$ is tested for being larger than γ_0, and γ_0 is so called protection ratio.

The argument SINR in the probability function may be the instantaneous SINR for a single observation, but rather it is a short term averaged SINR observation over an allocation period, e.g. a frame. It is also notable that the threshold for the outage (inverse of the coverage probability) is typically a low fractile value, which represents a small tail of the probability density distribution, and it is clearly different from the actual SINR target [240]. Regards formula (8.34), it can be seen that a multiantenna receiver increases the coverage probability compared to a single antenna receiver. First, the instantaneous SINR of a multiantenna receiver may be higher in comparable channel conditions due to antenna combining gains and second, the SINR distribution of a multiantenna receiver may show a shorter tail. Further, a net SINR gain may be experienced, in case the multiantenna reception algorithm is capable of interference rejection (see Chapter 4).

The coverage probability of a differentially small area element $p_c(x, y) = P_r(\gamma_0)$ at a location (x, y) of a given distance from the base station (NodeB or eNodeB) can be integrated over a wide area network $\mathcal{A}$ (or otherwise over a sufficiently representative area, say a simulation area) to yield a realistic measure of the coverage probability as the coverage reliability area, where the SINR level is larger than γ_0. The area integration with a base station (NodeB or eNodeB) grid of three sectors per site is shown in Fig. 8.12. Hence,

$$P_c^{\mathcal{A}} = \int \int_{\mathcal{A}} p_c(x, y)\ dx\ dy \tag{8.35}$$

In an analytical method, the SINR distribution is not known and it is proposed e.g. in [241] to be replaced by a log-normal or Gaussian equivalent distribution of random

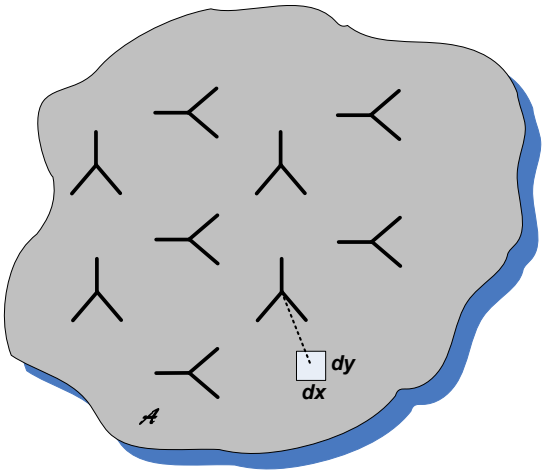


Figure 8.12: Coverage probability over a wide area $\mathcal{A}$ with multiple cells.

variables, whose mean and standard deviation are assumed. In the simulated methodology, the SINR is truly calculated over the allocated resources and a true distribution is available. Therefore, a simulated coverage analysis can be expected to give a more realistic statistical measure. On the other hand, even if the log-normal equivalent distribution will deviate from the actually simulated SINR distribution, it can be argumented that its approximation is not that inaccurate, when considering the probability density tail up to the fraction γ_0 only.

The coverage probability definition exists in terms of SINR (8.34), but as well it can be given in terms of the Bit Error Probability (BEP) and the Packet Error Probability (PEP), as in (8.36) and (8.37) respectively. BEP coverage probability and PEP coverage probability may be defined as the second momentum of BEP and PEP respectively i.e. the outage is the probability density of BEP and PEP for not meeting the given threshold BEP_{th} or PEP_{th} respectively. The long term statistics of BEP and PEP are controlled appropriately by the RRM to avoid outage. Dropping of connections i.e. having BEP or PEP above the threshold on longer term is considered worse than blocking new sessions being initiated. In packet communications, however, blocking new sessions is rarely applied and the coverage is actually a soft property that is met at the expense of increasing the packet delay rather than increasing the BEP or PEP as such. Regards PEP, it is not justified to consider the outage purely as a function of PDU BLER probability density, but rather it should be the air interface SDU discard probability or even a TCP segment discard probability, after the air interface retransmissions (HARQ and optional RLC retransmissions). The likelihood of SDU disgard and TCP segment discard are typically very

low (compared to PDU BLER) and very bursty, which make them rare events outside of a discard burst.

$$P_r = \left[BEP \leq BEP_{th}\right] \tag{8.36}$$

$$P_r = \left[PEP \leq PEP_{th}\right] \tag{8.37}$$

8.8 Capacity

Channel capacity is often given in the measures of Shannon capacity, ergodic capacity or capacity vs outage probability. Shannon capacity is the bound of the static capacity as a function of the channel power i.e. SNR. The ergodic capacity may include the channel state information and channel distribution information (see Section 6.5.8). The obtained capacity depends on the channel information available at the receiver and at the transmitter, the channel signal to noise ratio and the complex correlation of the channel on each antenna element [242][243]. The receiver obtains channel information relative to a transmitter by channel estimation from the reference symbols (pilots). The transmitter may obtain channel information expected at the receiver by the receiver signalling feedback to the transmitter. In case, the channel state information is not available or is expected to be inaccurate, the transmitter may apply blind-to-channel transmissions by precoding e.g. for space-time diversity or by creating cyclic delay diversity. In LTE, these means can additionally apply in a frequency selective manner, because resource allocation is possible both in time and frequency domains.

8.8.1 Ergodic channel capacity

The ergodic capacity of a random channel for a single transmit antenna single receive antenna configuration, where the channel matrix $\mathbf{H} = H_{11} = \mathbf{h}$, is given in (8.38). The transmit power limit of signal $\mathcal{S}$ is P_T such that $E\{|\mathcal{S}|^2\} \leq P_T$. The ergodic capacity of channel $\mathbf{h}$ includes the complex channel coefficients, and becomes,

$$\mathcal{C} = E_{\mathbf{h}}\{\log_2(1 + \rho \cdot \mid \mathbf{h} \mid^2)\}, \tag{8.38}$$

where ρ is the mean SNR, i.e. $\frac{E\{|\mathcal{S}|^2\}}{\sigma_v^2}$ and $E_{\mathbf{h}}$ is the expected value over all channel realizations. If $\mathbf{h}$ is Rayleigh distributed $|\ \mathbf{h}\ |^2$ follows a Chi-squared distibution [244], and (8.38) becomes,

$$\mathcal{C} = E_{\mathbf{h}}\{\log_2(1 + \rho \cdot \chi_2^2)\}, \tag{8.39}$$

where χ_2^2 is a Chi-squared random variable with two degrees of freedom (notation $(\cdot)_2$). If $\mathbf{h}$ is Additive White Gaussian distributed with zero mean and $\sigma_v^2 = 1$ instead, the expected value of the channel coefficient is 1 and (8.38) simplifies to,

$$\mathcal{C} = \{\log_2(1 + \rho)\} \tag{8.40}$$

8.8.2 Ergodic MIMO capacity

The ergodic capacity of a random channel for a MIMO system requires calculation of the covariance matrix [244][245][246]. For the transmit signal vector $\mathbf{s} = [\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \mathbf{s}^{(3)}, \ldots, \mathbf{s}^{(N_T)}]^T$ with zero mean, the covariance matrix Φ is

$$\Phi = E\{\mathbf{s}\ \mathbf{s}^H\}, \tag{8.41}$$

whereas the covariance matrix of the received complex Gaussian vector $\mathbf{r}$ in the receive antenna n_r is,

$$E^{(n_r)}\{\mathbf{r}\ \mathbf{r}^H\} = \mathbf{H}\Phi\mathbf{H}^H + \sigma^2\mathbf{I}^{(n_r)}. \tag{8.42}$$

For the uncorrelated noise in each receiver branch ($n_r = 1, .., N_R$), the ergodic MIMO capacity without channel state information becomes

$$\mathcal{C} = E_{\mathbf{H}}\big\{\log_2\big[\det(\mathbf{I}^{(N_R)} + \frac{P_T}{\sigma^2 N_T} \cdot \mathbf{H}\mathbf{H}^H)\big]\big\}, \tag{8.43}$$

where $\frac{P_T}{\sigma^2} = \rho$ is the mean SNR. The transmit power constraint of all antennas is $E\{\sum_{n_t=1}^{N_T} \big|\mathbf{s}^{(n_t)}\big|^{\ 2}\} \leq P_T$. A study for the MIMO capacity with partial knowledge of the channel state information can be found in [247].

After a brief introduction to the theory of the channel capacity, where the amount of knowledge of the channel state information and channel distribution information at the transmitter and at the receiver are important performance factors, -and where many open

research questions still exist-, it can be concluded that in a practical system analysis, simulations are among the utmost important methodologies, because they actually compute the channel from the propagation models, and they thus include the properties of coloured interference and frequency selectivity of the multiple-access technique that typically remain assumed in an analytical framework. In the system simulations, the measurement imperfections and the feedback delays can be included to the actual simulation model, so that their impact cumulates directly to the running calculations of the output statistics.

8.8.3 System capacity

The system capacity is often given in the simulated measures over a multilink, multicell analysis. The long-term sum of the link throughputs per cell yields the cell throughput. The average of the cell throughputs over a simulated network area yields the system mean throughput. The convention is to scale the system throughput by the system bandwidth per cell resulting capacity in the units of [b/s/MHz/cell]. For a voice source, having a constant source codec rate, the system capacity is given in the units of [number of calls/MHz/cell]. The WCDMA/HSPA system capacity is analysed in [199][248] and the LTE system capacity in [24][249][250][251].

The simulated system capacity takes into account both the short term (small scale) and long term (large scale) variability of the propagation environment, and particularly it includes the capability of the RRM algorithms to adapt to these dynamic changes. For these adaptations, the operation point target setting, link adaptation, power control and handover thresholds, transport format selection criteria, measurement imperfections and signalling feedback each have an impact that is not visible in the analytical capacity calculations. With the advanced algorithms, adaptation to these dynamics may happen efficiently, but on the other hand, they will increase computation and signalling load (overhead) respectively. Therefore, a tradeoff maximizing the capacity in dynamic conditions is often a heuristic process, where the greed algorithms are sacrified for the sake of saving measurements, computing and signalling load. These dependencies are yet hard to commensurate, as often a reverse link impact is present. The method for a system analysis is to simulate both link directions with compatible parameters and interdependencies, so that a consistent statistical set of results is created. For a system analysis, all the parameter bounds and thresholds have to be honoured simultaneously and the outage in any of them need to have an impact to the results.

In practice, the simulation methodology is a choice of high importance, because it has to include sufficient statistics of the changing channel states. In order the adaptive algorithms to operate, a snapshot analysis is favoured so that within a single snapshot the channel is continuous, and the schemes with feedback may also apply. In order to cover a large set of channel states, a large number of independent random drops needs to be generated.

For the measure of the system capacity, full load has to be generated. Other important studies with fractional load may be motivated though [224]. In both cases, the traffic models and the RRM algorithms play an important role. The traffic models impact the result, as they have differing characteristics to load the cells. For the circuit-switched connections, the capacity is different from that of the packet-switched connections. With the packet traffic, the capacity of VoIP [252] is different from that of the Best Effort data [253]. For the Best Effor data, the capacity differs for the variety of buffering or queueing models. Especially for the packet traffic, good schedulers provide capacity gains. Other important gains of the (simulated) system capacity may be achieved by utilizing the link properties efficiently by the precoding, channel estimation and receiver algorithms. The reference symbol overhead (synchronization symbols, pilots, channel estimation sequences etc), the feedback (CQI and PMI) and the overall control overhead also contribute to the system capacity results. All these factors differ per transmission system and per transmission scheme.

The objective of the system capacity analysis is to maximize the system information rate as a function of the bandwidth resources integrated over a unit-area. The information rate (R) of a single traffic source is,

$$R = \mathbf{E}(s) \cdot \frac{T_c}{T_s}, \tag{8.44}$$

where $\mathbf{E}$ is the entropy rate of the symbol source s, T_c and T_s are the channel encoder output rate and input symbol rate respectively. (If s represents the set of all source symbols, the entropy rate is equal to the average code block length with an optimal channel code.) Let $\mathbf{c}$ be the mean system capacity per unit bandwidth of a differentially small area element, the wide area capacity is;

$$\mathcal{C}^{\mathcal{A}} = \int \int_{\mathcal{A}} \mathbf{c}\ (x, y)\ dx\ dy \tag{8.45}$$

For VoIP, the number of satisfied users per unit bandwidth per cell is a sufficient capacity measure, as it includes the probability of a user being at a cell edge.

For the Best Effort data, the mean capacity over a unit area is not a very descriptive measure alone, because the user perception changes a lot depending on, where the user locates in the area. Therefore, a full throughput distribution per unit bandwidth per cell is valued instead. Once numeric quantization is expected, it shall at least include the mean and the 5 percentile. Often, the peak throughput is given for further information. The 5 percentile measure describes the critical cell edge performance.

Chapter 9

Selected results on the system performance

In Chapter 9, the system performance of WCDMA/HSPA and the system performance of LTE are covered based on the articles by the author [12][13][14][15] [16][17][18][19][20][21][22][23] [24][26][27][28][30]. Section 9.1 begins with early 3G studies, which gave first indications for the benefits of packet-switched communications on the channel utilization. Next, the changes of the interference characteristics and challenges of scheduling due to packet transport were studied. The WCDMA studies began from the greedy knapsack packet scheduling, code multiplexing and capacity studies using simple snapshot simulators or a dynamic network simulator. Next, a protocol simulator was developed, where the characteristics of the packet delay was studied based on full statistical distributions, and where the critical window and timer settings of the MAC and RLC protocols were analysed. These studies included heavy load from the packet traffic sources, which generate multiplexed bearers on a radio link. Further, this study was brought to a single cell multiuser scenario, where the benefits of the HSPA shared channels were proven in Section 9.3.4. In addition, an adjacent channel interference study (see Section 9.3.5) was carried out, because HSPA lets us expect largely increased load to the cell neighbourhood with different algorithmic impact than the WCDMA. This study showed that the adjacent channel interference requirements of the WCDMA devices need not change for HSPA.

Analysis of LTE begins in Section 9.4 with the design of channel structures and with an analysis of their basic performance properties, like expected gains related to the better exploitation of frequency dimension compared to WCDMA/HSPA. The primary analysis

of LTE in Section 9.5 covers widely the commonly agreed 3GPP [1][61], IEEE [64] and NGMN [59][63] simulation scenarios, traffic models and parameters to conclude the key performance indications of LTE, for the VoIP and Best Effort traffic. The system performance results are obtained by including extensive models of the protocols and algorithms, -or their impact from several detailed component technology studies-, that are brought to a common simulator platform. The methodology and measures are those recommended in documents by the 3GPP, IEEE and NGMN. The results summarize the VoIP capacity and the Best Effort spectral efficiency as the LTE system throughput for several reference scenarios with several transmitter-receiver techniques, radio resource management algorithms and propagation environments. The mobility impact on the system performance is prospected in a scenario of large cell sizes and varying velocities of the served UEs.

9.1 Analysis of early 3G

In [32] prior 3G technology was studied. The studies showed that the offered traffic intensity and channel utilization of the GSM system can be increased by the high-speed circuit-switched data technique as a function of increasing number of channels allocated per user. Only modest number of users is possible to accommodate without increasing the blocking probability of voice calls. It was shown that in a system of 40 channels, the channel utilization measure reached almost 0.8 without increasing the blocking probability of voice calls above 2%. In here, the classical traffic intensity[1] of all voice terminals was 0.625 (25/40) and the classical traffic intensity for data ranged from zero to 0.5 (20/40). The results were obtained for five data users having allocations up to six channels each. When the total number of channels in the system increased, the channel utilization got even higher because of the increased multiplexing opportunities for voice and data.

In [12] a set of algorithms were studied for high density mobile networks, where the frequency reuse factor was squeezed down to three or one including co-channel interference. The method had efficient SINR-packing of packets by centralized channel assignment and power control algorithms with coordination of transmissions from several cell sites. The utility function allowed increase of the scheduling gains to maximize the throughput. The experimented utility function was compared to the well-known interference averaging technique. It was observed that the throughput changed as a function of the SINR (carrier to

[1]Classical definition of traffic intensity is a measure of the average occupancy of a resource during a period of time.

interference ratio) target. Actually, the utility function allowed the increase of throughput, and the spectral efficiency reached about 0.7 with the 6 dB SINR target. When the SINR target was increased, the throughput decreased but the relative gains of scheduling by the utility function compared to the interference averaging increased. Even if the simulation assumptions here were idealistic in that having expected SINR information available in multiple sites at the same time without delays, it led to an important conclusion that sorting and classifying the co-channel interfered resources and scheduling by minimizing the mutual interference has potential gains from the minimum of order 28% up to over 100%. These magnitudes of throughput gains are observed in many later intercell interference coordination studies also.

In [31] the simulation scenario and assumptions were similar to [12] with the difference that the multiple access scheme was variable spreading factor CDMA instead of TDMA, and the utility function was formed for the CDMA equations [39]. The observations showed that the relative gains of interference coordination were lower for CDMA compared to TDMA. It was further noticed that the throughput maximizing utility functions tend to sacrificy fairness, which means that even if the mean system throughput can be increased, some users will experience very low throughput. Yet another important outcome of the study was that the throughput results were reached with much lower power with interference coordination than without it.

Load control of WCDMA was studied in [14] in a multicell interfered environment with different orthogonality factors. The load was increased gradually in the observed cell to obtain the load curve. The downlink load was expressed as the relative transmit power, and the uplink load was the received interference power that could have converted directly to the noise rise relative to the known noise bandwidth. The novelty in [14] included differentiation of the traffic to a high rate connection and a low rate connection that were selected per UE. In downlink, it was possible to admit more traffic, because the ortogonality of the codes is partly preserved. The intercell interference seemed to scatter the required total transmit power from the ideal load curve quite a lot. In the literature, at this time, interference coupling between cells were often assumed constant, which may approximately hold for the circuit-switched voice connections. However, in the presence of different service types and depending on the position of the UE in the cell (the geometry factor experienced at each UE receiver), the interference coupling may considerably change. This shows as the load observations deviating and scattering above

the expected (smooth) load curve. This phenomenon sets challenges particularly to the scheduler and the load control algorithms studied later, see descriptions in Chapter 8.

9.2 Analysis of WCDMA

In [16][30] throughput analysis was conducted simultaneously with the packet delay analysis. The results were obtained in a simulator, which included the actual protocol models and traffic models in the respect they affect system load and performance. The channel models and network interference calculations were simplified and intracell interference was expected to dominate over the other cell interference. The code-tree allocation algorithm (see Section 6.4.2) was real and had an impact to the queueing delay. In this model, the load increased as a function of increased number of terminals, but the nominal bit rate of each terminal decreased due to the limitations of the system parameters (chip rate and spreading factor) and finally due to the lack of code-tree resources.

The results showed that in the reference case of dedicated channels serving traffic at bit rate 128 kb/s to 512 kb/s, the mean delay and its standard deviation increased significantly, when the number of terminals was increased from ten to 75. The relative throughput increased modestly as a function of increased number of terminals because of slightly higher multiplexing efficiency. According to the simulations, the mean packet delay on the dedicated transport channels increased from 400 to 900 ms. While the mean delay increased, the standard deviation of the delay distribution also increased. The distribution showed positive skewness with a long tail. The analysis in [30] was done for the downlink shared channel transport serving traffic at bit rate 512 kb/s. The code-tree resources of the shared transport channel were not as limited as for the dedicated channels, as explained in Sections 6.4.1 and 6.4.2. Compared to the dedicated channel reference, it was possible to serve many more data terminals (up to 150) on the shared channel, while still maintaining the mean delay low. For the shared channel, the mean delay ranged from 400 ms to 650 ms. The shared channel delay distribution showed lower standard deviation at high loads compared to the dedicated channels, but its throughput variation was larger. Changing the selected BLER target from 20% to 12% and further to 7% reduced the mean delay. The simulations in [30] did not yet include HSPA features like the short TTI, short scheduling interval, adaptive modulation and turbo code rate selection based on the CQI report and diversity transmission schemes. This article however gave an early

indication for gains of dynamically sharing the code-tree resources. This analysis also lacked of network statistics and measures for the actual throughput, instead of the relative ones, which remained hidden by the traffic model. Another related study in [17] included downlink shared channel multiplexing for VoIP packets with header compression.

The dynamic selection of modulation and coding was studied in [18]. It was analyzed, how the physical link layer statistics vary over all channel realizations and by what kind of modulation and coding set they could be efficiently covered. The Turbo coded Eb/No curves are very steep and the modulation and code rate selection for the expected received Es/No resulted a distribution of received BLER. The selection simply targets at maximizing the throughput for the expected Es/No, but the selection threshold includes uncertainty, whether to risk with a higher transport format and possibly more retransmissions, or to select a lower transport format possibly underutilizing the instantaneous throughput but still saving in retransmission probability and reducing the delay. The selection of the modulation and coding was verified relative to the analytical calulations. In [18], yet another domain of adaptation was set for the number of code channels (i.e. multicode transport), which is also enabled on the HSPA channels. Without adaptive change of the number of allocated code channels, the modulation and code rate selection would not have covered all the channel realizations but the probability observations at both ends would have been emphasized. On the other hand, in the middle of the channel state distribution, several equally good choices are available, and the actual selection of the transport format may depend on, whether the power resources or the code-tree resources are more limiting at the time of allocation.

The studies on HSPA were extended to a dynamic network simulator in [19] without the protocol models. It was shown that HSPA can provide large gains by introducing Turbo codes and adaptive modulation and coding (see Section 8.2). In the comparative analysis, it was shown that it is advantageous to multiplex HSPA users not only in time but also in code domain because NodeB has excessive power resources available while serving too few users, e.g. only one user on the HSPA transport channel. Because of the high Es/No-target for the high order modulation that requires high transmit power, it is not an advantage to code-multiplex too many HSPA users. This is because code-multiplexing many users adds multiplexing overhead, reduces the probability of higher order modulations (due to power sharing) and additionally requires multiple HSPA control channels. It is typical assumption that up to about four users may be code-multiplexed to the same TTI in a

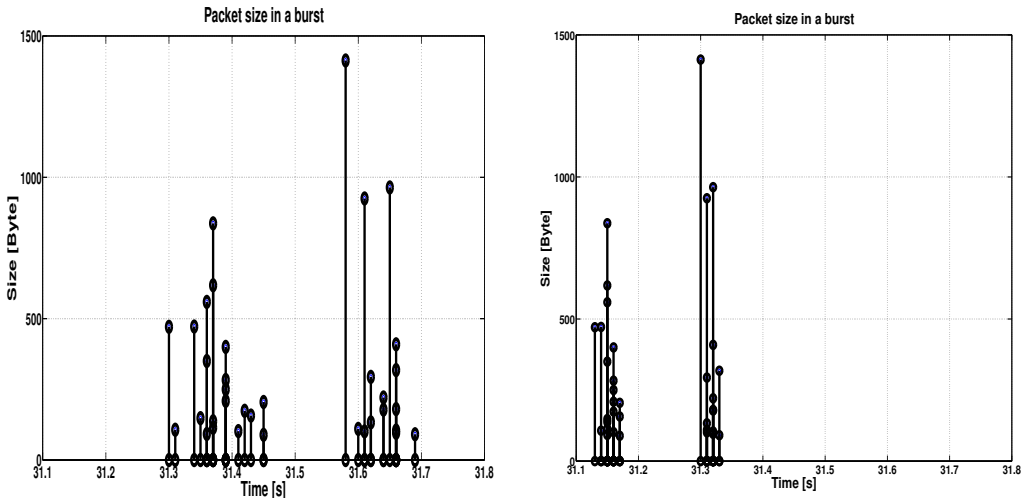
cell, and the other UEs are multiplexed and scheduled in time-domain (see Section 6.5.6). The UE capability classes defined in [254] still motivates code-multiplexing of many users. The reception of parallel HSPA code-channels is namely categorized up to {5, 10 or 15} code channels [254], and therefore serving the UEs of the category limit five code channels typically would not be sufficient to utilize the HSPA power and code-tree resources most efficiently, without code-multipexing many users to the same TTI.

9.3 Analysis of WCDMA/HSPA

9.3.1 A protocol simulator for the WCDMA/HSPA radio interface

The HSPA protocol, described in Section 6.4.2, was studied by simulations in a WCDMA cell. The results are presented in [26][27]. The simulator contains selected features of the RRC, RLC and MAC protocols [181][188][190][192]. The simulator operates with the WCDMA parameters given in Table 9.2. The physical performance is available in these simulations, given by the instantaneous BLER as a function of SINR from the lookup tables [6] and from the Jakes propagation model [83]. The traffic is generated here by a packet traffic model. The sessions are activated by a Poisson process, and the traffic model generates packet calls, where the size of each packet is randomly selected from the Pareto distribution having a specified cut-off value for the maximum size of IP datagrams in an Ethernet network. The traffic source is bursty having reading times [62] between the packet calls, which allows the RLC buffers emptied so that the data does not continuously accumulate to the transmit buffers. The parameters of the traffic source are given in Table 9.1. The function of the traffic source is shown for two packet calls at rate 384 kb/s and 1,500 kb/s in Fig. 9.1. Each bar of the histogram is drawn at the packet-arrival time and its height indicates the packet (SDU) size. The statistics are collected over all packet calls in all sessions during a long simulation (1000 s).

The RRC protocol model includes simple procedures to setup the radio bearers and simple algorithms to schedule packets. The RLC protocol takes care of processing the PDUs, i.e. it includes the polling schemes, retransmissions and selective acknowledgements. The MAC-d schedules the logical channels and selects the transport channel. The MAC-hs schedules the HSPA transmissions, selects the transport format and creates the transport block for each TTI. Fast physical layer signalling is modelled on the HS-SCCH channel in downlink and on the HS-DPCCH channel in uplink with randomly distributed



(a) (b)

Figure 9.1: The traffic source parametrised for (a) 384 kb/s and (b) 1,500 kb/s. A snapshot including two packet calls shows the packet arrival process and the packet sizes.

Table 9.1: Summary of the traffic source parameters.

Traffic source parameters	
Session arrival process	Poisson
Mean number of packet calls in a session	500
Mean number of packets in a packet call	{**25**, 50}
Packet source rate	{384, **1,500**, 7,000} kb/s
Packet inter-arrival time	{ 6.0, **1.536**, 0.329} ms
Reading time	{2, **5**, 15} s
Datagram size distribution	Pareto
Mean / maximum packet size	288 / 1,500 bytes
(α, k) of Pareto distribution	(1.1, 81.5)

signalling errors [2]. The RLC/MAC parameters are summarised in Table 9.2.

The transport on the HSPA channels is modelled with the fast HARQ processes, adaptive modulation, coding and multicode transmission (see Sections 6.5.6 and 8.2). Packet data is available in the NodeB buffers as MAC-d PDUs, which are assembled to a transport block after the transport format is selected. Thus, exactly one transport block per TTI is triggered for the transmission. The HS-SCCH and the HS-DPCCH control information are formed respectively. The transmission sequence number is attached to every transport block for reordering in the receiver. Chase combining is implemented, and the retransmitted symbols are soft combined with the already received symbols to increase

[2]This is justified as these control channels have fast power control, whereas HS-PDSCH does not have.

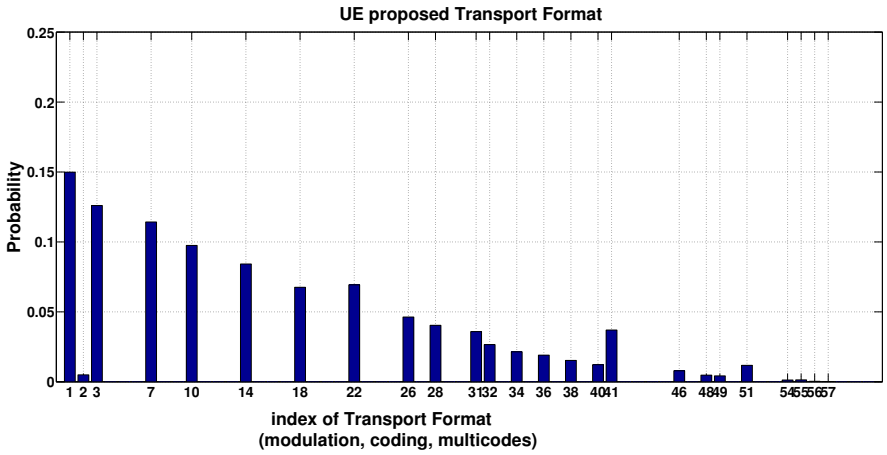
Table 9.2: Summary of the WCDMA/HSPA protocol and transport parameters.

WCDMA/HSPA	
RLC parameters	***Settings***
scheduler	Round Robin
RLC mode	Acknowledgement mode, max 10 transmissions
RLC PDU size	440 bit
RLC window size	{**128**, 256, 512} PDU
Polling schemes	periodic timer, poll timer, poll prohibit timer
MAC parameters	***Settings***
MAC-hs window size	{4, 16, **32**} Transport Blocks
MAC-hs timer	{76, 90, **150** ms }
Iub delay	Distributed [0, 30] ms [3]
TrCH parameters	***Settings***
DCH bit rate, spreading factor	variable
DCH bit rate (while HSPA transport)	30 kb/s
HSPA spreading factor	16
Transmission Time Interval	2 ms
Channelization codes	1 ... max{5, 10, 15}
Modulation and channel coding	{QPSK, 16QAM}, Turbo code rate {1/2, 3/4}
Physical layer transmissions	max {**4**, 6, 8}
FHARQ channels and processes	6, {1..6} per UE
Channel parameters	***Settings***
HS-PDSCH BLER ; HS-SCCH BLER	50% for the first transmission ; 1%
HS-DPCCH p{N|A}, p{A|N}	1e-2, 1e-4
UE velocity	3.6 km/h Jakes channel
Simulation time	1000 s

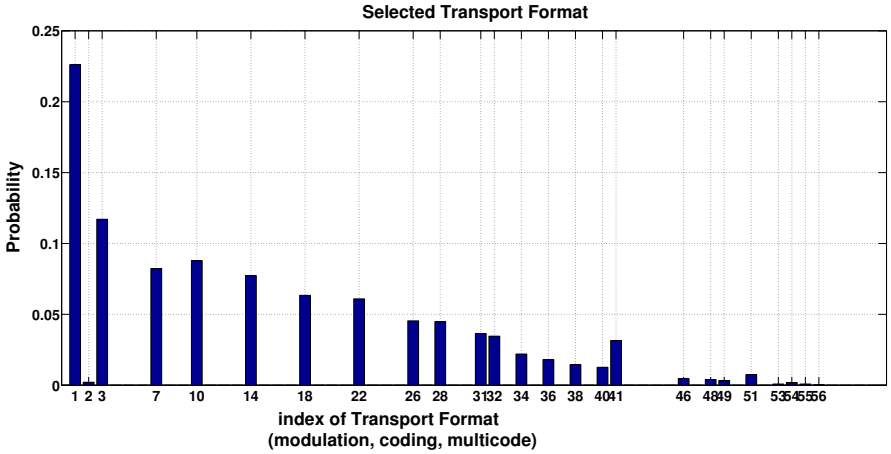
[3]Iub delay was set to 10 ms for the reverse link RLC STATUS PDU.

the probability of correct decoding for the transport block [6]. The transport channel parameters are given in Table 9.2.

The transport format is an ordered set of the modulation, code rate and number of HSPA channelization codes with the increasing symbol energy (Es/No) and increasing transport block payload. The statistics of the transport format proposal by the UE and the transport format selection by the NodeB are shown in Fig. 9.2. In these channel conditions, the UE was proposing more often lower transport formats. With higher power allocation and better average channel conditions (smaller cell size), the probability of proposing higher transport formats would increase. The selected transport format was occasionally lower than proposed, because there was not always enough data available at



(a)



(b)

Figure 9.2: The probability of a given transport format (modulation, code rate, number of multicodes) (a) proposed by the UE and (b) selected by the NodeB, for Best Effort traffic source of 1,500 kb/s with the mean of 25 packets during a packet call.

the moment of transmission to create as large transport block as were feasible to receive. The probability of selecting a higher transport format was increased, when the source rate was increased i.e. the packet interarrival time was decreased at the same time when the mean number of packets in a packet call was increased [26]. [4]

9.3.2 Analysis of packet delay and the MAC-hs window

In this section, the impact of MAC-hs function (see Section 6.5.5) and its parameters to the packet delay distributions are analysed. The MAC-hs window occupancy (memory size) is also compared for each case. The packet delay here is the peer-to-peer delay for a network

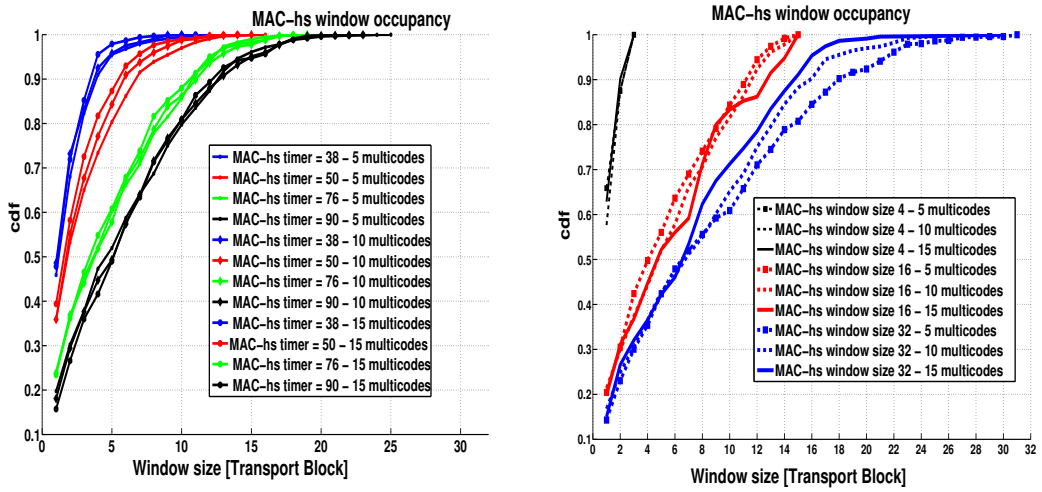
[4] If the source rate is increased and still the amount of generated packets is kept very small, the transport block distribution does not necessarily show increased probability for larger transport block size.

SDU (e.g. a TCP segment in an IP datagram) from the time, when the SDU was created at the traffic source in the network till it was received, correctly decoded and reassembled as an identical SDU at the UE receiver. Core network delays are not included in the presented delay statistics, because they are not that relevant from the WCDMA/HSPA analysis point of view and because they change according to the assumed distance and transport network load.

The distributions of the number of physical layer transmissions and RLC transmissions are shown in [27], from where it can be seen that four physical layer transmissions was typically enough and the probability of RLC retransmissions was about 5%, which is much lower than e.g. on the dedicated channel operating at 12 to 20% BLER. Increasing the allowed maximum number of physical layer transmissions may make the probability of RLC retransmissions even lower. However, the distribution of the RLC transmissions seem to have a long tail, meaning that when the channel dynamics is slow (long coherence time) and the transmissions occur during poor channel conditions, it takes a long time and many retransmissions before high enough Es/No is available for the correct decoding of the transport block. If even more physical layer transmissions are allowed, the fast HARQ-channel is unnecessarily occupied for a longer time, still not decreasing the RLC retransmission probability notably. The properties of protocol retransmissions as a function of slow channel dynamics would argument gains for other than Round Robin types of schedulers. With Proportional Fair schedulers (see Section 8.3), for example, the channel is allocated to the receiver with favourable (better than average) channel conditions, which gains in the instantaneous throughput. In this case, the channel does not remain occupied during the poor channel state, as were the case with the Round Robin scheduler.

The results of the MAC-hs window occupancy in the number of transport blocks is shown as a function of the MAC-hs timer setting in Fig. 9.3(a) and as a function of the MAC-hs window size limit in Fig. 9.3(b). The modulation alphabet and channel code rate were set as in Table 9.2 and the maximum number of HSPA channelization codes was limited to {5, 10, 15}. In both cases, the transport blocks in the MAC-hs window may be of different size because of the link adaptation algorithm (transport format selection).

The packet delay distributions are shown for the timer limited MAC-hs in Fig. 9.4(a) and for the window-limited MAC-hs in Fig. 9.4(b) respectively. The delay distributions show that when the timeout value of the MAC-hs timer increased, the memory requirement of the MAC-hs window got higher and the packet delays are allowed to increase. For a

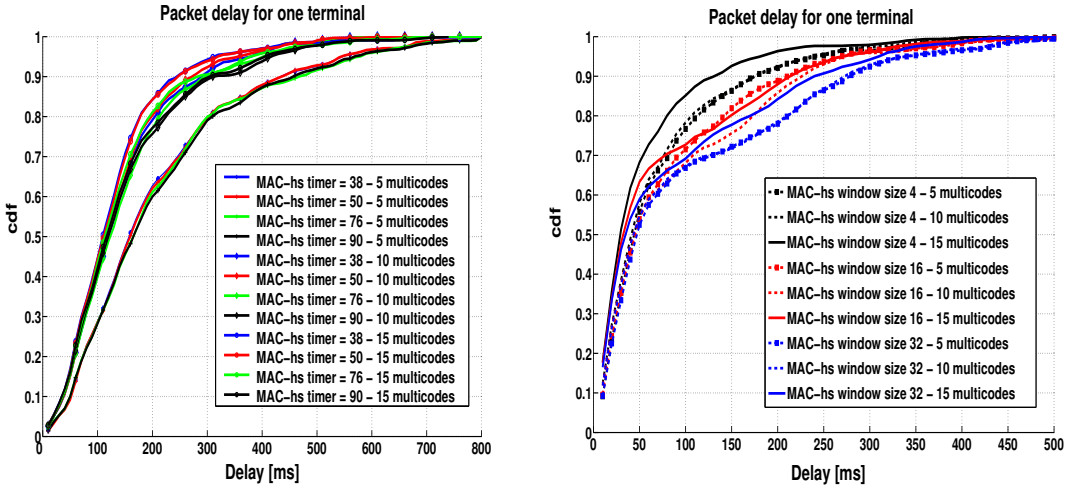


(a) (b)

Figure 9.3: MAC-hs window occupancy (cdf) as a number of Transport Blocks for (a) MAC-hs timer settings and (b) MAC-hs window size limit, for different transport format limits.

given MAC-hs timer value, the MAC-hs window size does not largely depend on the given transport format set. However, buffering larger transport blocks will take more memory both in the NodeB transmitter and in the UE receiver. The packet delay distributions do not show large differences for different MAC-hs timer settings. For a given transport format with 5 multicodes, the probability of getting higher packet delay is higher than with 10 or 15 multicodes, which do not have a large mutual difference. Next, the MAC-hs window size was limited to {4, 16, 32} transport blocks respectively. With a small window, the memory requirements are low and the forced packet delays are small. Here, a transport format set with 15 multicodes makes a difference in the packet delay, as even larger transport blocks can be formed during good channel conditions. With the window limit set to 16, the window was rarely full and having higher transport format yields lower packet delays. With the window limit set to 32, the window was not observed to be full, and it already showed the maximum requirement for the window memory in the UE. With a lower transport format set of 5 multicodes, larger window size is requested. However, the size distribution (in [byte]) of the transport blocks may still remain smaller than that for 10 or 15 multicodes, which allow larger transport blocks, if the channel conditions are very good.

In the MAC-hs timer limited case of Fig. 9.4(a), the probability of having larger packet delays is larger for the transport format up to 5 multicodes, but the mutual difference for



(a) (b)

Figure 9.4: Packet delay (cdf) as a function of (a) MAC-hs timer and (b) MAC-hs window size with transport format limited to a number of multicodes for the Best effort traffic source of 1,500 kb/s.

the transport formats of 10 and 15 multicodes is very small. In the MAC-hs window-limited case of Fig. 9.4(b), where the window is set to the smallest value (four), the transport format of 15 multicodes gains lower packet delay compared to 5 and 10 multicodes. When the window is larger, and it is not that strictly limiting the number of physical layer retransmissions, the packet delay distributions are not drastically different for any of the transport format sets.

The MAC-hs parameters have an impact to the HSPA performance by allowing a high physical channel BLER operation point e.g. of order 30-50% and still making the RLC protocol BLER very low e.g. of order 5%. The example results are shown for the selected packet traffic model and channel conditions including also the Iub transport delay distribution. The packet delay was observed to depend on the transport format set and on the MAC-hs parameter settings.

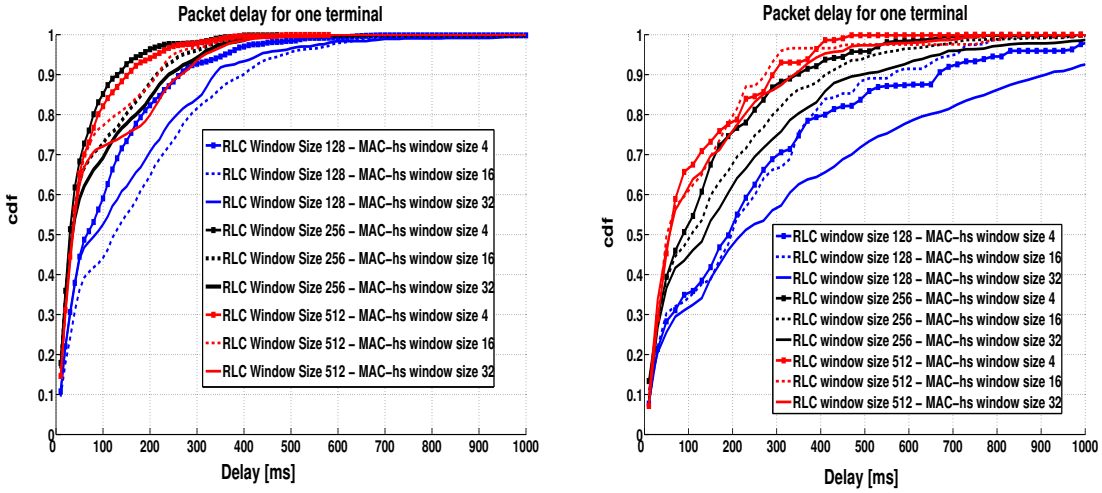
9.3.3 Analysis of packet delay and the RLC window

For high traffic activity and large amount of data to be transmitted on HSPA, it is expected that the RLC window may actually become the bottleneck of throughput and may cause dominant SDU delay instead of the MAC. This behaviour was analysed in [26] by setting the RLC window size to {128, 256, 512} RLC PDUs respectively.

The packet delay on a radio bearer (one UE) is shown in Fig. 9.5. The delay is dominated by the MAC-hs window size if set to a small value. A small MAC-hs window may be motivated by targeting at low delays, but it may also cause decreased probability of correct decoding at the physical layer because of the limited number of retransmissions. In a multiprocess HARQ, where one process had a decoding error, other processes may still continue receiving correctly decoded transport blocks, which cumulate into the receiver window to be reordered according to their *Transmission Sequence Numbers*. It is possible that the receiver window soon gets full, and because it follows the leading edge of the transmission sequence numbers, it will start to discard transport blocks. This will prevent further soft combining in the physical layer and will lead to more RLC retransmissions, which will cause larger SDU delays. If in this case also the RLC window is small, it may become the bottleneck for PDUs, which have to wait in a transmit buffer before adopted to the transmitter window. Increasing the RLC window in this case will release the bottleneck and will allow decreased delays again. This is visible in Fig. 9.5, where the SDU delays decreased significantly when the MAC-hs window was set to 4 and the RLC window was changed from 128 to 256. Increasing the RLC window even further from 256 to 512 did not impact SDU delays notably any more under the assumptions in this example.

When the MAC-hs window was set to a larger value as 16 or 32, it very rarely got full under the assumptions of this example. The delays are now allowed to increase as already seen in Fig. 9.4. In this case, it is clearly possible that the RLC window is the bottleneck again, and by increasing the RLC window size from 128 to 256 and further to 512, smaller SDU delays as a function of increased RLC window size can be observed. The difference of delay degradation with the RLC window size 256 compared to 128 for the MAC-hs window size 16 is clearly visible. Setting both the MAC-hs window size and the RLC window size large, 32 and 512 respectively, neither window hits the maximum. In this case, the SDU delay performance is floating and driven by other parameters in the system. Such other factors include the source rate behaviour, Iub delay distribution, RLC PDU size, transport format set and channel statistics. The packet (SDU) delay as a function of the RLC window size and the MAC-hs parameters is summarized in Table 9.3. The statistics include the mean delay and additionally the 75%, 95% and 99% fractiles to show the heavy tail of the delay distribution.

Fig. 9.5 also shows the packet delay for one UE when the source rate was further increased close to the very ultimate, i.e. from 1,500 kb/s of Fig. 9.5(a) to 7,000 kb/s of



(a) (b)

Figure 9.5: Packet delay distribution (cdf) for one UE as a function of RLC window size and MAC-hs parameters, for the Best Effort traffic source (a) 1,500 kb/s with the mean 25 packets in a packet call and (b) 7,000 kb/s with the mean 50 packets in a packet call.

Table 9.3: Packet delay as a function of RLC window and MAC-hs parameters for a Best Effort traffic source of 1,500 kb/s.

RLC window [TB]	**MAC-hs window [TB]**	**Mean delay [ms]**	**Delay fractile [ms]**		
			75%	**95%**	**99%**
128	4	116	163	364	532
	16	172	264	490	686
	32	151	232	442	740
256	4	58	73	186	346
	16	83	116	274	410
	32	91	134	314	410
512	4	65	81	219	358
	16	78	87	254	390
	32	95	152	322	424

Fig. 9.5(b). The mean amount of packets created during a packet call was also increased, from 25 to 50. For this high source rate, the SDU delays increased for all parameter settings compared to the 1,500 kb/s source rate. Again, the packet delay of a radio bearer (one UE) is dominated by the MAC-hs window size, if set to a small value. A small MAC-hs window will typically result in small delays, but it may also result in decreased probability of correct decoding at the physical layer transmissions, and that will cause more RLC retransmissions respectively. If in this case, the RLC window was also small and became the bottleneck, increasing the RLC window allows decreased delays again. This is visible

in Fig. 9.5, where the SDU delays decrease significantly when the MAC-hs window size was set to 4 and the RLC window size was changed from 128 to 256. Increasing the RLC window size further from 256 to 512 does not impact the SDU delays notably any more. As a function of the increased MAC-hs window size, the delays are now allowed to increase as already seen in Fig. 9.4. In the case of a higher source rate, it is clearly possible that the RLC window is the bottleneck and by increasing the RLC window size from 128 to 256 and further to 512, smaller SDU delays can be observed. Here, with the ultimately high source rate of 7,000 kb/s, the difference of delay degradation with the RLC window size 256 compared to 128 and further 512 compared to 256 for the MAC-hs window size 16 and 32 are clearly visible in all these cases. Thus, the higher source rate and higher amount of data in the buffers will cause the RLC window make a clear difference in the delay, whereas for the source rate 1,500 kb/s the difference was visible for the RLC window size 128 and 256, but not for 512 any more.

The traffic source is a bursty model (see Section 7.2.3.1) having a long reading time, which allows buffers to be more deeply emptied times to times so that the packet calls are not entirely accumulating to the RLC transmit buffers. However, the source rate is not the only critical issue, also the increased number of packets generated during a packet call will impact the delay, because more packets will accumulate to the transmit window waiting to be triggered to the MAC-hs in the NodeB. Increasing the mean number of packets in a packet call will make more data available both in the RLC window and in the MAC-hs window resulting increased probability of higher transport format, as shown in [26]. Still, it is possible that the channel conditions do not allow use of higher transport formats and the delays would hence increase.

For a radio bearer with tight delay requirement, e.g. real time streaming, it seems viable to make the MAC-hs window small, disable the RLC retransmissions (RLC set to Unacknowledged Mode) and set the BLER operation point low. For such settings, it would be necessary to define, how high SDU dropping probability is tolerable. In the example case of the non-real time traffic, 1,500 kb/s source rate is shown in Fig. 9.5 and setting the delay limit to 100 ms, the MAC-hs window size to 4 and the RLC window size to 256 would cause the SDU discard (dropping) probability of about 12%, which were far too high for the equivalent real time traffic. Better settings for the real time traffic could be observed by setting lower BLER target e.g. about 5% instead of 30% to 50% simulated here.

In [26][27], it was shown that the mean delay decreased as a function of decreasing MAC-hs window size, however, there remain a high proportion of packets, which will experience very large delays. The tails of the SDU delay distributions are long. For a given MAC-hs window size, the mean delay and even more clearly the upper fractile of delays decreased as a function of the increasing RLC window size, if it was the bottleneck for the delays. When the RLC window is large enough relative to a given source behaviour, it is not the bottleneck, and the SDU delays remain driven by the other factors of the system.

9.3.4 Analysis of packet delay in a loaded cell

In [26], the performance is analyzed with the MAC-hs operation as a function of the RLC window size in a loaded cell with 20, 40 and 80 active UEs for source rates 384 kb/s and 1,500 kb/s. The source rate and its burstiness are critical, but also the higher amount of packets generated during a packet call will impact the delay, because with longer packet calls more packets may accumulate to the RLC transmit windows to be triggered to the MAC-hs. Also in [26], the measure is the peer-to-peer delay for a network SDU, such as a TCP segment, from the time when the SDU was created by the traffic source in the network till it was correctly received, decoded and reassembled as an identical SDU at the UE receiver. Any core network delays are excluded from these SDU delay statistics.

The distribution of the physical layer transmissions and the RLC transmissions was studied for single links in [26]. In the loaded cell, the mean RLC throughput was above 0.88 with the variance of 10^{-3} and the mean physical layer throughput was above 0.63. and it varied more than for a dedicated channel, because there is no fast power control nor outer-loop power control on HSPA. The HSPA protocol however maintained the relative RLC throughput, so that the RLC PDU BLER was below 12%, which is comparable to the RLC operated on a dedicated transport channel. The deviation of the RLC throughput as a function of increasing number of active UEs remained small. The RLC SDU residual error probability with 12% RLC PDU BLER is $\ll 10^{-3}$ after four RLC transmissions and $\ll 10^{-6}$ after seven RLC transmissions. Allowing larger number of RLC retransmissions increases delay at the radio interface and reduces the probability of TCP retransmissions.[5]

The impact of the MAC-hs timers and the MAC-hs window size was studied in [27], the impact of the RLC window size and different traffic source rates were added in [26].

[5] In QoS guidelines, BER $\ll 10^{-6}$ often appears as a requirement for data services over the radio interface ($BER \ll BLER$).

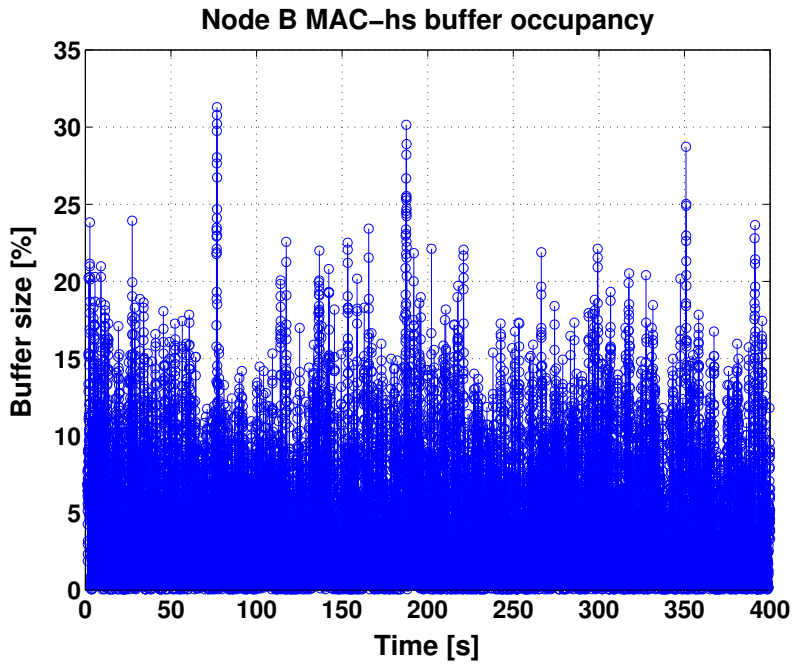


Figure 9.6: MAC-hs buffer load in the HSPA NodeB for 80 UEs with the Best Effort traffic source of 384 kb/s.

Fig. 9.2 shows the probability of selecting a given transport format over all packet calls in all the sessions during a simulation.

The WCDMA/HSPA analysis of Sections 9.3.2 and 9.3.3 was repeated with increasing cell load i.e. by increasing the number of active UE in the cell, Here, a UE is active, when it had a packet session activated but not yet terminated. Because of the bursty nature of the traffic source, adding active UEs does not contribute a linear increase of the load. The probability of active RLC buffers for a given number of active UEs was mean 4 (range $0 \ldots 11$) for 20 UEs, mean 7 (range $0 \ldots 17$) for 40 UEs and mean 11 (range $0 \ldots 21$) for 80 UEs respectively.

The buffer occupancy will depend on the session arrival and departure processes. In order to collect the statistics in a steady-state, the session arrival process was modified to meet the data equilibrium still meeting practical simulation times. Say, if the session arrivals would happen 5 sessions per hour per UE [62], the simulations causing high traffic volume and balanced buffers would be unpractically long, as the protocols and allocation periods are in 2 ms accuracy relative to the channel states. The buffers analysed in this article are in balance by the modified session arrival rate as visible in Fig. 9.6, but their short-term burstiness remain high.

The physical layer BLER operation point was set to 30% to 50% for the first transmission to significantly gain from soft combining of the HARQ retransmissions. The link throughput varies around 30 to 50% with a larger variance than for a dedicated channel, as there is no fast power control nor outer-loop power control applied to the HSPA channel. The mean throughput at the physical layer is about 0.48. The deviation of the physical layer throughput is anyhow rather small because of the high granularity in the transport format set and because of well-tuned link adaptation algorithms. The physical layer throughput does not significantly change from the single link throughput as a function of increasing number of active UEs set to 20, 40 and 80. The RLC throughput, obtained with such a low physical layer BLER, is still moderate reaching the mean below 12%, which is comparable to a typical dedicated channel BLER for the RLC PDUs. Again, the deviation of the RLC throughput is small, and it did not change much with the increasing number of active UEs.

The packet delays are shown in Fig. 9.7, where the sole packet delay is shown as a cdf for every packet despite of the bearer (of the UE) it was transmitted on. The mean packet delay obtained on each bearer (per each active UE) is shown in Fig. 9.8, when the traffic source was set to 384 kb/s and 1,500 kb/s. The results are summarized for the analysis in Table 9.4. The analysis showed that increasing the cell load by increasing the number of active UEs per cell did not cause the delay distributions change largely for the source rate of 384 kb/s, whereas the change is larger for the source rate of 1,500 kb/s. Increasing the source rate from 384 kb/s to 1,500 kb/s will make the mean delays clearly higher, and the highest fractiles of the distributions will extend to larger deviation from the mean. Say, for 80 active UEs, the mean packet delay increased from 105 ms (for 384 kb/s) to 193 ms (1,500 kb/s) and the standard deviation from 74 ms (384 kb/s) to 155 ms (1,500 kb/s). The mean delays observed by all the served UEs remained rather low. Say, for 80 active UEs, the 75% fractile of UEs (60 UE) observed mean delays of 120 ms (384 kb/s) and 200 ms (1,500 kb/s), which are lower than 75% fractiles of the individual packet delays, 130 ms (384 kb/s) and 260 ms (1,500 kb/s) respectively. This result indicates that the multiuser scheduling provides also channel dependent delay gains in a loaded system.

It can be remarked that the parametrisation in this study was used to show the extreme behaviour of the delay observations in a loaded cell. The high BLER operation point, the limited maximum number of physical layer transmissions and the Round Robin scheduler emphasized the delay mechanisms of the protocol stack including the physical layer, the

Table 9.4: Packet delay as a function of number of active UEs per cell.

Traffic Source bit rate [kb/s]	Number of active UEs	Mean delay [ms]	Delay fractile [ms]		
			75%	95%	99%
384	20	98	122	228	320
	40	104	131	244	324
	80	105	130	260	352
1,500	20	175	237	449	630
	40	185	252	473	682
	80	193	260	495	738

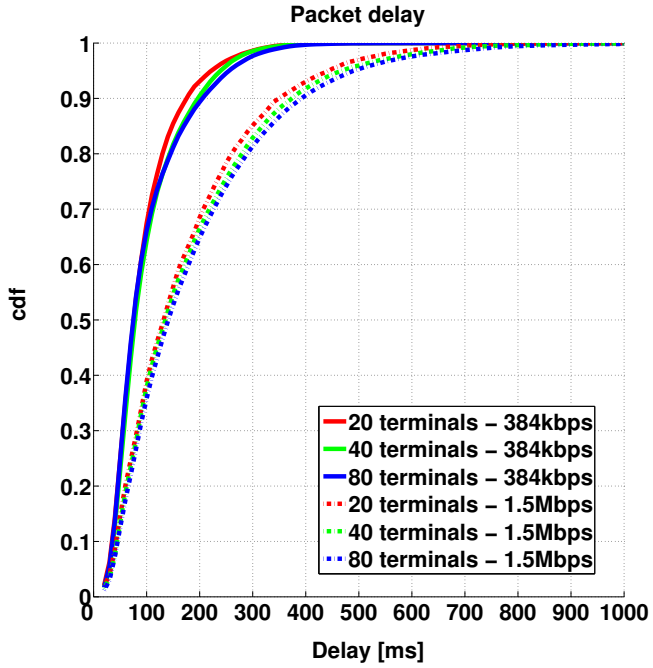


Figure 9.7: The packet delay (cdf) as a function of number of active UEs in a WCDMA/HSPA cell, for the Best Effort traffic source of 384 kb/s and 1,500 kb/s.

MAC-hs and the RLC for the experience of the SDU delay statistics. In a practical network, much more efficient channel dependent and Proportional Fair schedulers are available, and also the parametrisation can be more efficient. Even then, the RLC protocol delays will be visible as the highest delay fractiles. For the RLC protocol, a combination of polling schemes was proposed to avoid individual packets queueing unnecessarily long time in the transmit buffer.

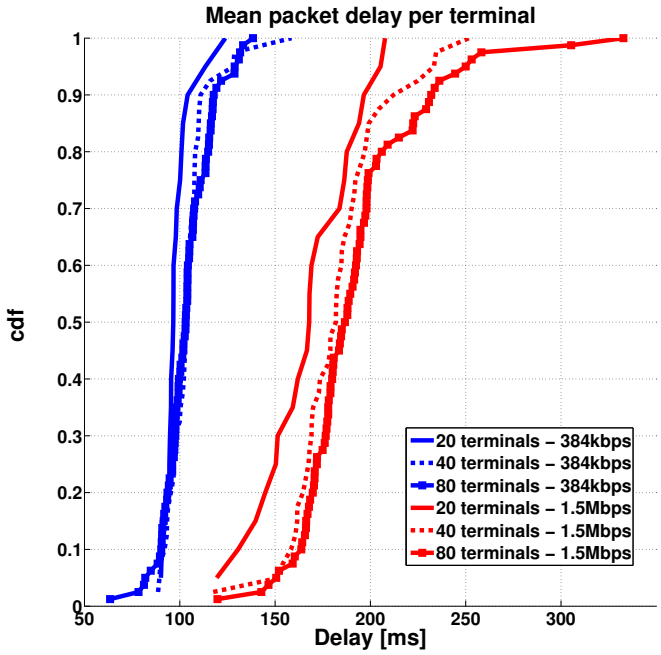


Figure 9.8: The mean packet delay per UE (cdf) as a function of number of active UEs in a WCDMA/HSPA cell, for the Best Effort traffic source of 384 kb/s and 1,500 kb/s.

9.3.5 Analysis of the adjacent channel interference load

The adjacent channel interference load (see Section 8.6) due to the HSPA transport of WCDMA/HSPA was analysed in a macrocell scenario. The simulated network had 33 cells in a configuration, where each NodeB site had three cells and the maximum cell power was 20 W (43 dBm) out of which 14 W (70%) was allocated to HSPA. The UEs are evenly distributed on the simulation area in the beginning of a simulation, and they may move according to the defined mobility model. Fig. 9.9 shows the distribution of the served UE for three cell sectors of one NodeB site. HSPA was modelled so that the physical transport was heavily active in periods of short TTI, and the link adpation algorithms as HARQ, adaptive modulation and coding were active. The performance metrics were the cell throughput and the packet transfer delay as a function of the interference load. The traffic source activity was high for creating high load, so that the packet queues increased, and it was possible to schedule packets for transmission at all times. These exaggerated settings caused HSPA be active for over 93% of TTIs during a simulation.

First, the load is shown Fig. 9.10 as a function of the increasing number of UEs in the simulation area for a constant ACI ratio. The load distributions are rather wide, because

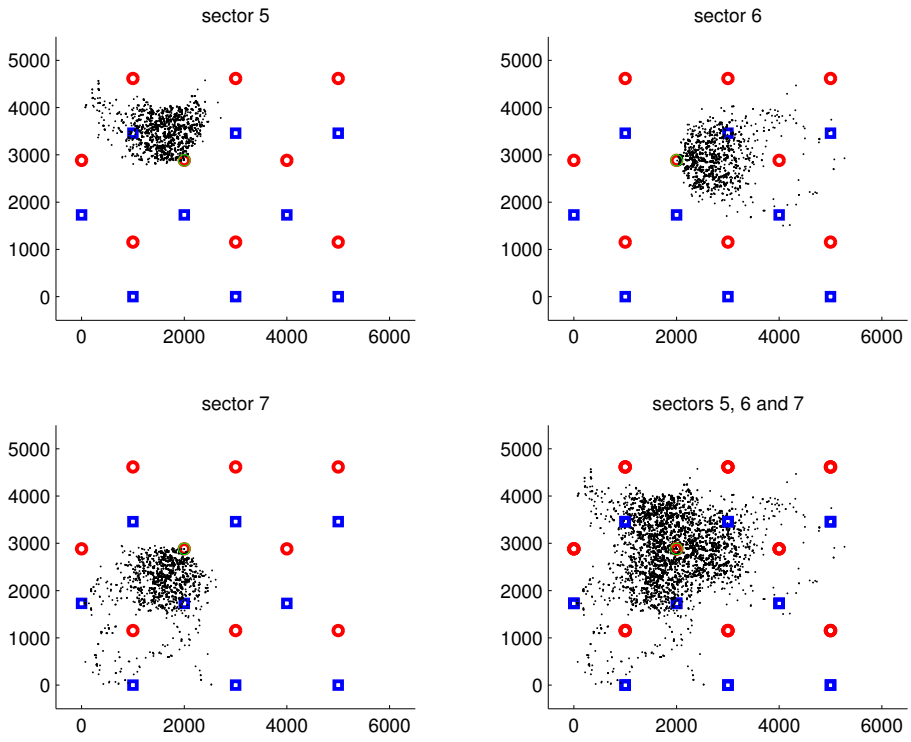


Figure 9.9: Coverage areas of three cells and one NodeB site with 14 W power allocation for HSPA.

the receiving UEs experience different interference power according to their position in the cellular area (geometry factor). The observed load distribution did not change much as a function of increasing number of UEs, because the HSPA power allocation remained constant. The noise floor is at -99 dBm and the UEs seem to have received power levels between $-90 \cdots -40$ dBm. The impact of the interference load due to ACI was analysed more in details. The ACI is dominantly impacted by the adjacent channel selectivity of the UE receiver and the adjacent channel power leakage of the NodeB, whose net impact is announced as the ACI ratio, studied already e.g. in [13]. The interference was observed with different ACI ratios of {20, 30, 40, 100} dB for the increased number of users. The impact of the ACI ratio was analysed as a relative throughput decrease compared to the ideal case, as shown in Fig. 9.11. Table 9.5 again summarizes the relative throughput loss in the system due to the ACI. The results indicate that even for a very high load, the ACI ratio well above 30 dB keeps the impact to the total received interference small and the throughput deviates less than about 5% from the ideal. The ACI ratio below 30 dB would seem to be adequate for low load, but when the load is very high, the interference becomes dramatic and makes throughput losses far too high ($\gg 10\%$).

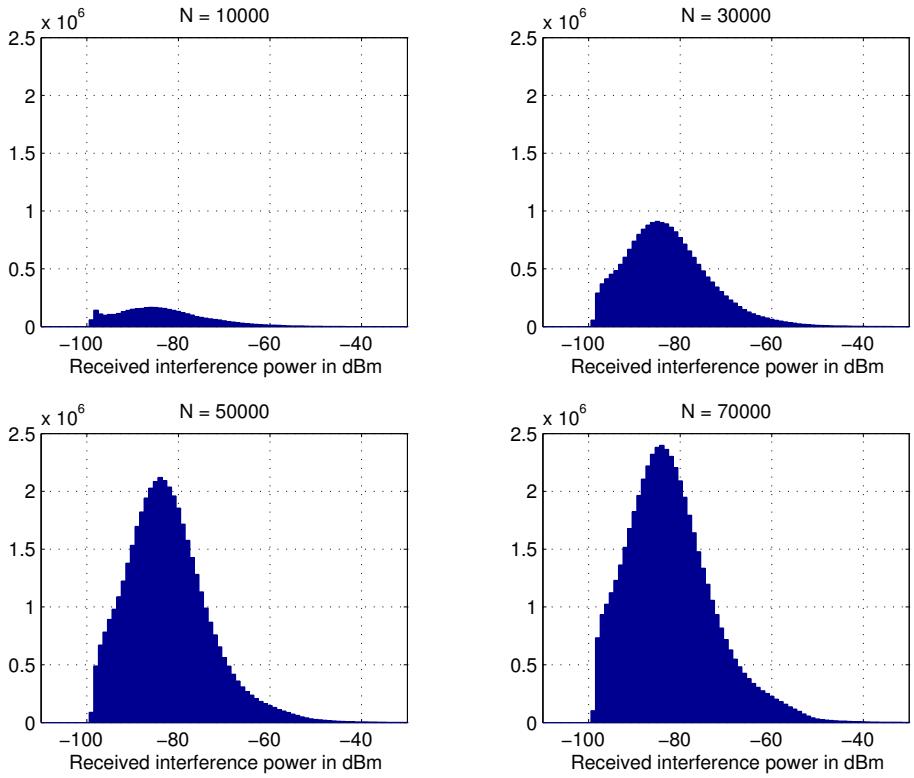
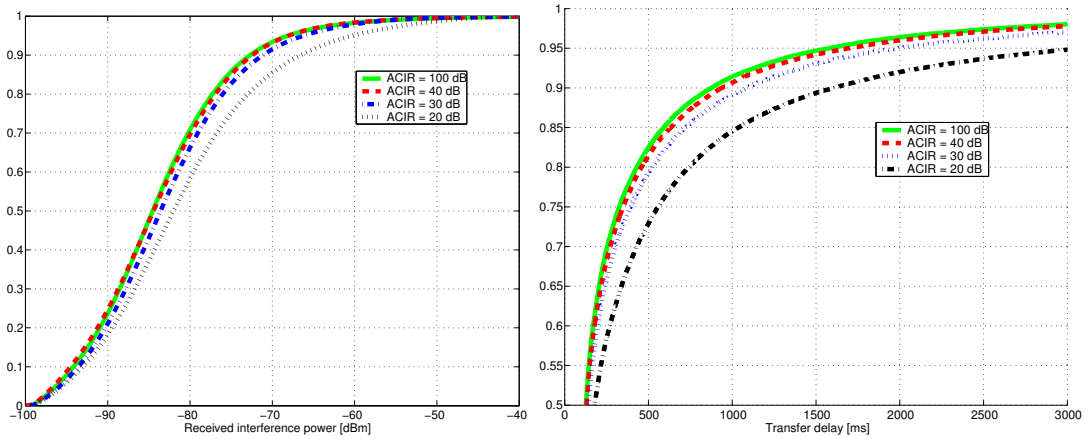


Figure 9.10: Wideband interference load observed by all UEs in the simulation area.



(a) (b)

Figure 9.11: Impact of ACI ratio to the (a) interference load and (b) packet delay (highest 50 percentile shown).

Some further results in Fig. 9.11 show the increase of packet delay, when the number of subscribers was increased and the packet queues got longer. Again, it is visible that the ACI ratio above 30 dB causes a small impact to the total received interference, whereas the ACI ratio below 30 dB has a clear effect. We can remark that the absolute packet delays

Table 9.5: Relative decrease of throughput as a function of ACI ratio compared to the ideal case.

Throughput, relative decrease [%]				
Number of UEs	ACI ratio [dB]			
	20	30	40	100
10,000	0.1	0.3	0.6	0
30,000	2.2	0.7	0.2	0
50,000	23.0	7.3	4.1	0
70,000	27.7	8.8	3.4	0

are very large here due to the extreme load conditions, and that may not be experienced in practical networks. Here, the aim was just to define the worst case relative impact of the ACI.

As a summary of these studies, it can be concluded that even in the extreme load conditions of HSPA, the ACI ratio slightly above 30 dB is sufficient and increasing it further would just require larger filters in the UE front-end without notable reduction in the interference. Also, having smaller cells or low power class NodeBs will make the ACI smaller compared to the macrocell case. On the contrary, operating in a small cell, higher interference load could be reached, but in that case the impact of ACI relative to the total interference including higher multiple access interference would also vanish.

The study was further repeated for unequal load of the adjacent carriers in ratios 50/50, 30/70 and 10/90 while keeping the total number of UEs in the simulation area constant. In that case, the more loaded carrier will generate more ACI to the receivers of the less loaded carrier, but the number of (victim) UE receivers suffering from that increased interference is smaller, therefore having rather small relative impact to the cell throughput. On the other hand, the less loaded carrier will generate less ACI to the more loaded carrier, where more (victim) UE receivers would suffer. This, however is not strongly the case, because the generated ACI power is already so small. The results of unequal carrier loading in the worst case was comparable to the equal load case in the region of interest, where the percentual decrease of throughput due to the ACI was sufficiently low anyway. Based on this study, it was concluded that HSPA does not set any further requirements to the ACI ratio (adjacent channel selectivity and adjacent channel leakage ratio) than already decided for the WCDMA without HSPA. (The minimum requirements are set in the standard as follows; the UE adjacent channel selectivity is 33 dB, the UE adjacent channel leakage

power ratio is 33 dB for the first adjacent carrier and 43 dB for the second adjacent carrier, the NodeB adjacent channel leakage power ratio is 45 dB.) Later, it was specificed for WCDMA/HSPA and LTE coexistence on the adjacent carrier frequencies that the existing requirements are sufficient also in that case. For the wider bandwidth of LTE, the requirement of the ACI ratio may decrease from 33 dB to 27 dB at 20 MHz. The standard does not require the ACI measurements directly, but the methodoglogy was adopted so that the ACI is verified relative to a standardised test case, where the throughput with the ACI shall be above 95% of the throughput without it.

9.4 Analysis of early LTE

Early design of the LTE channel structures is addressed in [20][61]. Many details of these proposals got modified during the standardisation process, but the main principles still hold (see Section 6.5). In [20],the importance of frequency diversity, transmit antenna diversity and time diversity were studied for the physical broadcast channel on 1.25 MHz system information band. Frequency diversity was shown to gain slightly more than 2 dB at BLER 1% when having PRB level distribution (scattering) over the frequency band in 3 PRBs, and by 3 dB when distributing in 6 PRBs over the full system information band, compared to no scattering i.e. without frequency diversity. Transmit antenna diversity by two antennas was shown to gain slightly over 3 dB at BLER 1% compared to a single antenna transmission. This gain is further increased by 1.8 dB when soft combining is applied for two consequent channel blocks, which in addition to the combining gain increases time diversity. The transport means to capture these gains are actually included in the LTE standard [105][208], which allows the reception of the physical broadcast channel block combined over four bursts. This block combining is possible, because its information contents remain constant for that period of time. Multiantenna transmission is allowed in LTE for the broadcast channel too, because the cell detection algorithm includes tests for the multiantenna transmission candidates. Soft combining of code blocks has the advantage that in good channel states, the correct decoding can be fast (correct decoding from the first received block) and in poor channel states, Es/No of the code block can be gradually increased by soft combining a number of consequent code blocks with the expense of added decoding delay [50]. Soft combining of the first two blocks reaches nearly 3 dB gains, and soft combining yet the next code block reaches about 1.8 dB additional

gain at BLER 1%.

Another important result for the downlink shared channel was introduced, namely allocating PRBs in a flexible manner to the frequency band based on the CQI (see Section 8.3.2). A relative throughput increase up to 60% was obtained in a slow, frequency selective channel compared to the allocation of adjacent frequency PRBs. This gain is a function of the number of allocated PRBs, and it increases very rapidly for both allocation types. However, the gains soon saturate for the adjacent PRB allocations, which extend wider than the coherence band of the good, scheduled channel states to the poor channel states also. Increasing the number of allocated PRBs in a flexible manner utilizes the good channel states in all the scheduled frequencies following the order of the highest CQI first [22].

In [21], downlink performance and uplink performance were analyzed as a function of the schedulers and radio resource control setpoints in the macrocell scenario #1 and #3 respectively [61]. These scenarios are widely in use for the analysis, because they capture the most relevant and challenging properties of a wide-area network deployment. Hence, they are considered sufficient to characterize the performance of the WCDMA/HSPA and LTE systems. Shortly, both macrocell scenarios #1 and #3 define 10 MHz bandwidth at 2.0 GHz center frequency for a UE moving 3 km/h in a propagation environment, which additionally has 20 dB penetration loss. The radical difference of these scenarios is the deployment with intersite distance (ISD) 500 m for macrocell #1 and ISD 1732 m for macrocell #3. This difference often causes macrocell #1 deployment become interference limited, whereas macrocell #3 deployment may remain power or noise limited instead. The macrocell scenario #3 seems to be especially difficult for the cell edge users because of the small signal margin to the noise floor.

The scenarios and metrics of the LTE performance analysis are present in [61][63][64]. The LTE performance has been analysed several times along the standardization process and summary reports including results from several companies have been accepted in the 3GPP. Several publications appear in the literature that include results for the LTE, or for its selected technology components. In Section 9.5, the LTE performance summary is mainly collected from the recent publications by the author of this thesis [21][24].

9.5 Analysis of LTE

9.5.1 Traffic scenario and simulation methodology

The article in [24] shows the performance numbers of LTE for the most typical traffic profiles, VoIP (see Sections 7.2.1 and 5.2) and Best Effort data (see Sections 7.2.2 and 5.2). The traffic scenario and simulation assumptions are summarized in Tables 9.6, 9.7 and 9.8.

Table 9.6: Summary of the most important assumptions for the system simulation scenario.

Description	***Settings***
Number of cells	19 sites (or 7 sites), 3 cells per site
Carrier center frequency	Frequency Division Duplex (FDD) 2.0 GHz
intersite distance (ISD)	500 m, 1000 m, 1500 m
System bandwidth	VoIP 5 MHz, Best Effort 10 MHz
Effective bandwidth	VoIP 4.5 MHz, Best Effort 9.0 MHz
Subcarrier spacing	15 KHz
Frequency reuse	frequency reuse 1
User distibution in the area	uniform
Number of active users per cell	VoIP, variable number of users (acc. 2% outage criteria) Best Effort, 10 users
Cell selection	best cell selected per snapshot, 0 dB margin
Path loss	See [61][64], Minimum Coupling Loss 70 dB
Shadow fading	log-normal distribution, std 8 dB correlation between sites 0.5 correlation between cells of a site 1.0 correlation distance for UE 50 m
Penetration loss	20 dB
fast fading profile	UE velocity dependent Jakes fading
UE velocity	{3, 30, 120} km/h
Channel models	3GPP Typical Urban (DL 20 taps, UL 6 taps) SCM-C with path correlation for MIMO ITU modified PedB and VehA with correlation
Channel estimation	Realistic estimation two dimensional (time-frequency) Wiener filter
Link-to-system mapping	EESM with realistic channel estimation AVI with realistic channel estimation

9.5.1.1 Voice over IP scenario

For VoIP, a simple two-state voice activity model presented in Section 7.2 was used. During the voice-active state, a talk-spurt is generated with a sequence of voice packets having

Table 9.7: Summary of the most important assumptions for the downlink simulations.

Description	***Settings***
Multiple access	OFDMA
Cell transmit power	43 dBm for 5 MHz, 46 dBm for 10 MHz
Reference Symbol (pilot) overhead	symbol positions according to [105] {4.55%, 9.09%, 12.12%} for 1-tx, 2-tx, 4-tx
Transmission schemes	1x2, 2x2, 4x2, 4x4 precoded MIMO up to dual codeword with rank adaptation [6].
Receivers	MRC or IRC for single codeword fallback LMMSE for dual codeword MIMO
Power Control	Equal transmit power per PRB
Receiver power dynamics, C/η limit	22 dB
Control signalling (Layer1/Layer2) overhead	1 to 3 symbols (12-21%) per subframe 3 symbols reserved during simulations [7].
Allocations	PRB resolution (12 subcarriers)
CQI	measurement error, realistic Gaussian i.i.d measurement accuracy, zero mean, std 1 dB measurement window (filter length) 5 ms reporting resolution 2 PRB reporting delay 2 ms
Modulation constellation	{QPSK, 16QAM, 64QAM}
Coding	Turbo code, variable rate
Transport format	QPSK{1/5,1/4,1/3,2/5,1/2,3/5,2/3,3/4} 16QAM{2/5,9/20,1/2,11/20,3/5,2/3,3/4,4/5,5/6} 64QAM{3/5,5/8,2/3,17/24,3/4,4/5,5/6,7/8,9/10}
Link Adaptation	Fast, CQI based transport format selected per TB Best Effort, outer-loop BLER target $\sim$ 20%
HARQ	8 HARQ channels (stop and wait operation) asynchronous HARQ processes frequency scheduled HARQ transport format adaptive HARQ Chase Combined retransmissions maximum 4 transmissions
Packet Scheduler	Round Robin scheduler Proportional Fair, channel dependent scheduler

[6]As an exception, 4x4 MIMO schemes were simulated with ideal CQI resolution in 3GPP TU channel.

[7]In downlink, the worst-case overhead was simulated and the realistic overhead is taken into account by postprocessed scaling. The overhead reduced in average by a factor of 1.103.

a constant interarrival time and a constant size after the header compression. For the voice inactive state, the Silence Descriptor (SID) packets are generated. The assumptions for VoIP and protocol models are given in Table 9.9. More detailed description of the

Table 9.8: Summary of the most important assumptions for the uplink simulations.

Description	***Settings***
Multiple access	SC-FDMA (DFT-s-OFDMA)
UE transmit power limit	24 dBm
Reference Symbol (pilot) overhead	2 DMRS symbol blocks per subframe 1 SRS once per subframe (full bandwidth) [8].
Transmission schemes	1x2 with ideal MRC at the eNB 1x4 with ideal MRC at the eNB
Receivers	LMMSE for each antenna branch MRC combining for antenna branches Frequency Domain Equalizer
Power Control	open loop power control according to (3) fractional path loss compensation [0.6, 0.7]
Receiver power dynamics, $C_{rx,max}^{(UE_k)}/C_{rx,min}^{(UE_{k'})}$	17 to 20 dB
Control signalling (Layer1/Layer2) overhead	2 PRB PUCCH for VoIP at 5 MHz (8%) 4 PRB PUCCH for Best Effort at 10 MHz (8%)
Allocations	PRB resolution (12 frequency bins)
Channel Sounding	measurement error, realistic Gaussian i.i.d measurement accuracy, zero mean, std 1dB (SINR[dB] for each UE allocated PRB set)
Modulation constellation	{QPSK, 16QAM, 64QAM}
Coding	Turbo code, variable rate
Transport format	QPSK {1/10,1/6,1/4,1/3,1/2,2/3,3/4} 16QAM {1/2,2/3,3/4,5/6} 64QAM {2/3,3/4,5/6}
Link Adaptation	Fast, channel sounding based Transport format selected per TB outer-loop adjustment of thresholds
ATB	minimum bandwidth is fixed 6 PRB
HARQ	8 HARQ channels (stop and wait process) synchronous HARQ processes frequency scheduled HARQ non-adaptive transport format during HARQ Chase Combined retransmissions maximum 4 transmissions for ISD 500m maximum 8 transmissions for ISD 1500m
Packet Scheduler	Round Robin scheduler Proportional Fair, channel dependent scheduler

[8]In uplink, the worst-case SRS was simulated and the realistic overhead is taken into account by postprocessed scaling. For a wideband SRS transmitted once per 10 ms, the scaling is 1.0818. The scaling for once per 2 ms is 1.0454. Fractional bandwidth SRS would reduce the overhead even more.

assumptions can be found in [61][63][64].

As described in Chapter 5, VoIP packet contains large networking headers compared to the relative size of the voice payload. The RTP/UDP/IP protocol overhead decreases the efficiency and impacts the capacity and coverage of the radio access. However, in LTE transmission this is not a major problem particularly if the network load is low, because the offered capacity is large and the communication flows of different traffic types can be efficiently multiplexed. The Robust Header Compression (ROHC) [166] is assumed in this analysis to achieve maximal efficiency, see the discussion in Section 5.4.

Furthermore, the scheduler impacts efficiency, because it may decide to bundle individual VoIP packets within the delay constraints to the same transport block, and it may optimize physical resource allocations for that transport block size. As voice frames have a fairly deterministic interarrival process, which is moderately jittered by the Internet, the resource allocation scheme may use persistent knowledge of the scheduling events as an alternative of fully dynamic scheduling [21][24][255]. Among such persistent allocation schemes, talk-spurt persistency seems to be the most appropriate choice [24].

Table 9.9: Summary of the most important assumptions for the VoIP model and protocols

VoIP	
Voice codec	AMR 12.2 kb/s, AMR 7.95 kb/s
Networking protocols	RTP/UDP/IP
Radio protocols	PDCP/RLC/MAC/PHY
Header Compression	ROHC, profile RH-0
Payload with overhead	40 bytes for AMR 12.2 28 bytes for AMR 7.95 2x{40,28} with packet bundling
SID overhead	15 bytes once per 160 ms
Voice activity	50%
Talk-spurt duration	exponential distribution, mean 2.0 s

9.5.1.2 Best Effort scenario

For Best Effort traffic, as presented in Section 7.2, the infinite buffer model was applied with a log-normal truncated file size distribution. The assumptions for Best Effort traffic and protocol models are given in Table 9.10. More detailed descriptions can be found in

[61][63][64].

The TCP/IP networking headers are large, but they can be compressed to the minimum at the radio access without dynamic context complexity similar to the RTP protocol. The Best Effort performance depends largely on the packet scheduling algorithms, see Section 8.3. Two types of schedulers were analyzed, namely a Round Robin (RR) scheduler and a Proportional Fair (PF) channel dependent scheduler. The Round Robin scheduler allocates resources equally in time-frequency sequential order. The Proportional Fair scheduler (see Section 8.3) allocates resources in the frequency resolution of a PRB and in the time resolution of a subframe, so that each UE gets resource allocations according to a defined fairness criterion. Such a fairness criterion may be tuned to gain in the average cell throughput, to boost the cell edge throughput, to balance the throughput experienced by the served UEs or to unequally weight the throughput of the logical channel flows.

9.5.1.3 Best Effort scenario with MIMO

The downlink performance of LTE may further be improved by the MIMO transmission technology (see Sections 3.5 and 4.6). This is especially feasible for the Best Effort traffic, which enables scheduling of large transport blocks, when the channel SINR and rank allow. (VoIP traffic with small packets and tight delay requirement obviously does not have this kind of opportunity.)

The MIMO schemes include Per-Antenna Rate Control (PARC) [91][92] and precoded MIMO [105]. A single codeword rank 1 precoding scheme provides a fallback mode, whenever multistream transmission is occasionally not feasible and diversity transmission is needed. For diversity transmission, the maximum ratio combining receiver or interference rejection combining receiver is applied. For the PARC and precoded MIMO, the Linear Minimum Mean Squared Error (LMMSE) receivers represented in Chapter 4 are used separately for the independent symbol streams, according to the transport schemes described in Section 3.5 (see Fig. 3.19).

The CQI feedback per fractional bandwidth allows both rank adaptation and independent selection of the transport format per codeword [105]. The studies in [24] primarily used large cell sizes, where rank adaptation is an essential feature. In a microcell scenario, where the SINR is higher, MIMO could provide even larger gains, because the probability of multiple transport streams gets higher compared to the macrocell. In the macrocell scenario, transmit diversity is typical and multiple transport streams are feasible only

occasionally.

Table 9.10: Summary of the most important assumptions for the Best Effort model and protocols

Best Effort	
Source model	Full buffer data, infinite queue
Networking protocols	TCP/IP (included as payload)
Radio protocols	PDCP/RLC/MAC/PHY
Payload including overhead	maximum Transport Block size exact fit to the PRB allocation

9.5.1.4 Methodology

The simulation methodology uses a quasi-static approach with intercell interference generated by full load packet traffic in a simulated area. The statistics is collected from a set of center cells of the simulation area or alternatively from all cells with wrap-around to ensure rich interference characteristics. The frequency reuse factor is one, which makes scheduling and resource allocation a challenging task. The quasi-static approach allows a large number of snapshot simulations, yet each snapshot has a continuous time model to include a sufficient amount of channel state changes for link adaptation, scheduling and transport format selection. These channel state changes include frequency selective fast fading but also changes of the correlation and phase information of the complex channel. Also interference changes during a snapshot. Link adaptation and the transport format selection algorithms target at the largest instantaneous throughput in the allocated resources as a function of SINR and the expected BLER. The outer-loop of link adaptation may locally tune the transport format selection thresholds and the expected BLER target of the first transmission of a transport block, in order to increase the cell throughput.

The control channel resource utilization is analyzed relative to the resource allocations by calculating the PDCCH formats [49][208] and their coded aggregations on average [53][55][256]. For VoIP, this is taken into account as control channel limitations as the number of PDCCHs available in a subframe. For Best Effort, the worst case overhead is assumed during the simulations and scaling for the real expected overhead is done by postprocessing. The overhead scaling is in the resolution of the OFDM symbols, and its short term variations are not taken into account in [24]. The uplink signalling is calculated

as an overhead that is present in the PUSCH, or it appears as the resource reservation for the PUCCH so that the CQI-reports, downlink acknowledgements and scheduling requests can be transmitted for all the served UEs.

For power control in the downlink simulations, the transmit power is equally shared between the PRBs. In the uplink simulations, the power control (see Section 8.1) is an ideal open-loop algorithm, which assumes path loss measurements with inaccuracy and reporting delay, both given in Tables 9.7 and 9.8. The power setting is accurate, however having fractional path loss compensation. The power setting at a transmit time is a function of the selected transport format, it controls the power spectral density, and it scales with the allocated bandwidth. The power factor P_0, see Section 8.1.2 and (8.1), is varied as a parameter, and it has an impact on the cell edge to the cell mean efficiency ratio as shown in Fig. 9.18 and even more clearly in Fig. 9.23 for the larger ISD. In practice, the open-loop power control is very inaccurate after a discontinuity, and the closed-loop power control correction may additionally be given at every instance of uplink allocation. This closed-loop power correction may be an absolute or an accumulated control word, that will result to an accurate enough power setting, because the HARQ retransmissions will anyway provide further adaptation to the channel dynamics.

The propagation characteristics of a physical link, as the distance dependent path loss, shadowing and frequency selective fast fading are modeled in the system simulator, whereas the link performance is fed to the system simulator either through the Exponential Effective SNR Mapping (EESM) interface [257] or through the Actual Value Interface (AVI) [6]. The typical Urban channel model (3GPP_TU) with 20-taps was applied for training the EESM interface in the subcarrier symbol resolution. Further, the Spatial Channel Model (SCM-C) [258][64] was applied including spatial antenna correlations that were calculated beforehand to the channel covariance matrix.

9.5.2 Performance metrics

The performance metrics for the analysis are set according to [61][63][64]. The number of served VoIP users per cell is analyzed to provide comparable results to the circuit-switched voice capacity in conventional systems. The VoIP capacity is a measure of the number of users per cell according to the Satisfied User Criterion (SUC), when 98% (95%) of the users are satisfied. A user is defined satisfied, if 98% of the VoIP packet delay distribution of that user is below 50 ms. This definition lets us assume end-to-end delay below 200 ms

that is required for seamless mobile-to-mobile communications [60][63].

For Best Effort data, the statistics are created as a full distribution (cdf), from where the mean cell throughput and the cell edge (5 percentile) throughput were calculated and scaled to the spectral efficiency values. For scheduling, a fairness criterion was set in [63]. The fairness bound to exceed was given at points {0.1, 0.2, 0.5} of the normalized throughput by the probability of {0.1, 0.2, 0.5} respectively. This fairness bound is shown as a probability line in the normalized pdf of Figs. 9.13(b), 9.15(a), 9.16(b), 9.19(b) and 9.21(b). The system results of LTE obtained in [24] clearly exceed the fairness bound.

As the velocity of the mobile UE is expected to impact the scheduling and feedback, a study was conducted to draft the performance degradation in the case, where the mixed-velocity links are present during a simulation. In here, a set of modified ITU models (PedB, VehA) [64] were used. The modified model refers to the wideband properties of the channel that provides a large number of taps, frequency selectivity and correlation, unlike the conventional narrowband model. The actual scenario had a random choice of a link property for each radio link so that the velocity was 3 km/h at 60%, 30 km/h at 30% and 120 km/h at 10% probability. This scenario further had to be extended to a large intersite distance (ISD 1,500 m), where high velocity may appear in practice. The larger ISD had a large impact to the results already as such without the velocity impacts.

9.5.3 LTE system performance for VoIP

For VoIP, LTE capacity is summarized in Table 9.11 for downlink and in Table 9.12 for uplink. The results were obtained both for a *fully dynamic* (FD) scheme and for a *semi-persistent* (SP) scheme. The fully dynamic scheme relies on the CQI feedback and efficient scheduling of the frequency resources. The tolerable time constraints of scheduling are bound by the satisfied user criterion (50 ms). The semi-persistent scheme has non-scheduled, blind, constant allocations for every first transmission of a packet. The transport format of the first transmissions is set constant for the duration of a talk-spurt, but the packet retransmissions are scheduled and their transport format is selected similar to the fully dynamic allocations. This obviously saves control channel resources.

Besides scheduling and allocation schemes, VoIP capacity depends largely on the LTE features and properties such as the control channel overhead and signaling efficiency. The capacity can instantaneously be limited either by the shared data channel or by the control channel resources. These limitations may be due to the power or symbol resources. The

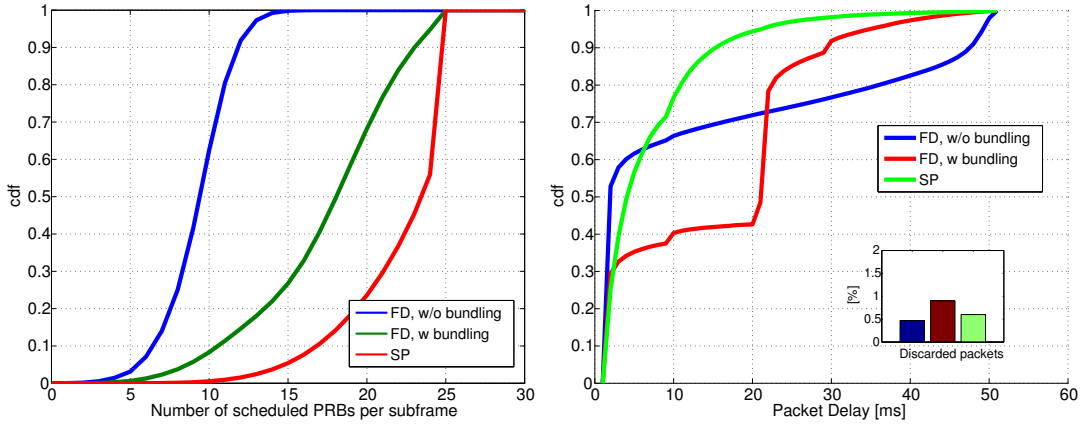


Figure 9.12: VoIP distributions with fully dynamic (FD) and semi-persistent (SP) schedulers; (a) the number of scheduled PRBs per subframe and (b) the packet delay. The results with the fully dynamic scheduler are shown without (w/o) and with (w) packet bundling.

Table 9.11: VoIP capacity in downlink as the number of satisfied users per cell. Evaluation includes a fully dynamic and a semi-persistent scheduler.

VoIP traffic, AMR 12.2 kb/s			
Downlink capacity [SUC, #users per cell]			
Control channel overhead	2 symbol	3 symbol	
#PDCCH channels	4	6	8
without bundling	-	203	272
FD scheduler	-	375	488
SP scheduler	403	366	-

control channel capacity is flexible upto the upper bound of three full OFDMA symbols. The consumption of the control channel resources varies as a function of the scheduler. Packet bundling was applied here, optionally, for downlink and therefore the control signaling need was slightly unbalanced having about 4 to 6 signalled allocations in a downlink subframe and about 4 to 8 signalled allocations in an uplink subframe. Comparing the schedulers, we note that the performance of the FD scheduler gets limited by the control channel resources. The results indicate that VoIP capacity with FD scheduler increases almost linearly as a function of increasing number of the PDCCH channels. The control channel limitation to the VoIP capacity with the FD scheduler is also visible in Fig. 9.12, which contains the distribution (cdf) of the scheduled PRBs per subframe with six

PDCCH channels. Due to the control channel limitation, the average PDSCH utilisation with the FD scheduler is only 40%. However, the control channel limitation can partly be avoided by packet bundling, which increases the average PDSCH utilisation to 70%, because larger transport blocks can be allocated into small physical resources during good channel conditions. This is enabled by the fast CQI based link adaptation. On the other hand, the average PDSCH utilisation for the SP scheduler gets even higher, above 90%, which shows that the VoIP capacity with the SP scheduler tends to be limited mainly by the PDSCH resources.

Table 9.12: VoIP capacity in uplink as the number of satisfied users per cell. Evaluation includes a fully dynamic and a semi-persistent scheduler. (Uplink mean IoT was 13 to 14 dB.)

VoIP traffic, AMR 12.2 kb/s			
Uplink capacity [SUC, #users per cell]			
Control channel overhead	2 symbol	3 symbol	
#PDCCH channels	4	6	8
FD scheduler	-	206	248
SP scheduler	256	259	236-260

The packet delay distribution (cdf) using different schedulers is shown in Fig. 9.12. Due to severe control channel limitations, the amount of packets having delays close to the delay bound (50 ms) is clearly higher for the FD scheduler. With packet bundling, the control channel limitations are relaxed and delay critical packets can be scheduled earlier. Bundling itself adds delays in cumulation of 20 ms. For the SP scheduler, the control channel limitations mainly vanish, because the first packets are scheduled directly into the reserved resources and the packet delay distribution is smooth. In all these schemes, the retransmissions are scheduled at the earliest convenience, after 8 ms HARQ cycle.

The VoIP capacity of LTE extends beyond 350 users in downlink and beyond 230 users in uplink, for ISD up to 1000 m. This is a considerable increase compared to the reference system reported in [259][260][261]. Increasing the ISD from 500 m to 1000 m shows that the downlink performance remains the same, whereas the uplink capacity may decrease over 10% depending on the scheduler.

9.5.4 LTE system performance for best effort traffic

For Best Effort traffic, the results are created with a full set of RRM features (see Chapter 8). The mean cell throughput and the cell edge throughput were calculated from the full statistics, and they were scaled to the spectral efficiency values. Earlier studies of this kind appear in [21][249][250][262] and a numerical summary is available in [1]. The specification and simulation models are updated since, and new results are given in Fig. 9.13, Fig. 9.15 and Table 9.13 for downlink and in Fig. 9.16 and Table 9.14 for uplink. In these tabular results, the downlink control channel overhead was modelled according to the realistic expectations, and the uplink SRS overhead was moderated to the practical values.

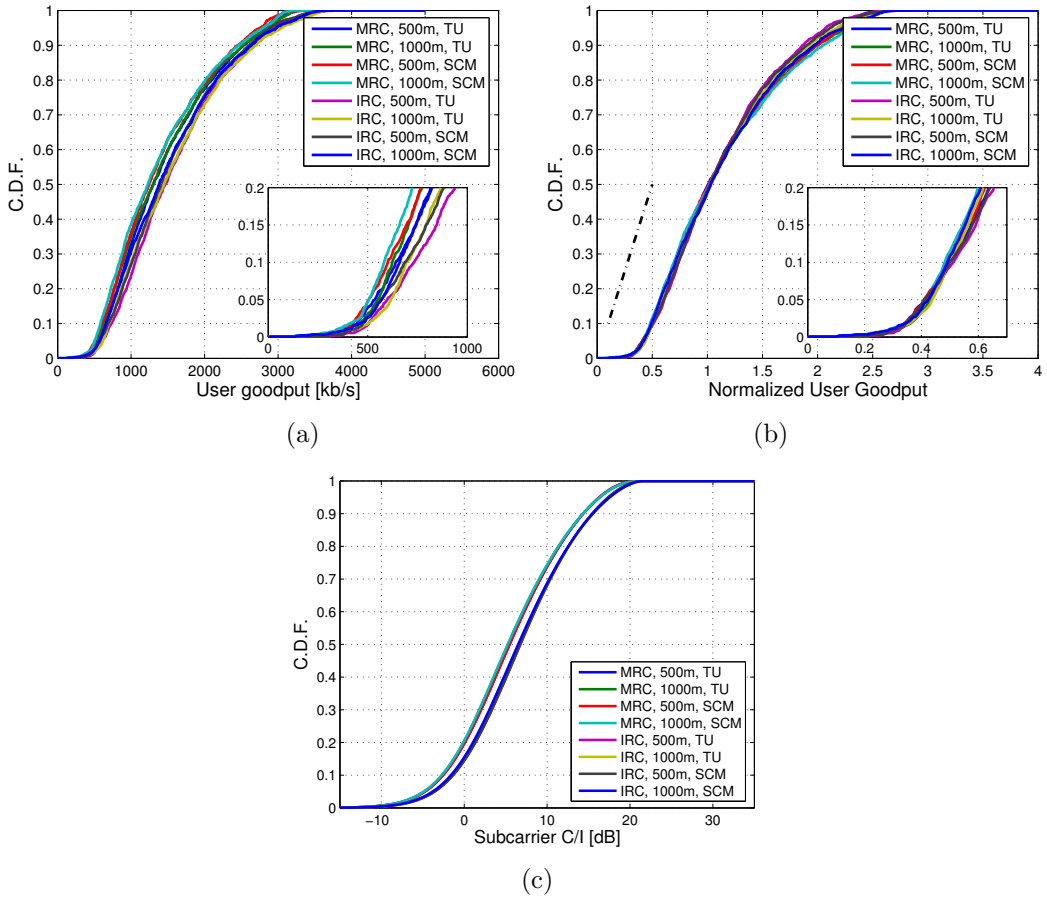


Figure 9.13: Downlink 1x2 results (cdf) with Proportional Fair channel dependent scheduler with MRC and IRC receivers; (a) user goodput with the lowest 20 percentile zoomed, (b) user goodput normalized to the geometric mean and (c) received subcarrier C/I.

For downlink, the baseline results are shown for a two-antenna receiver in Fig. 9.13 and for the selected MIMO transmission schemes of Sections 3.3 and 3.5 in Fig. 9.15.

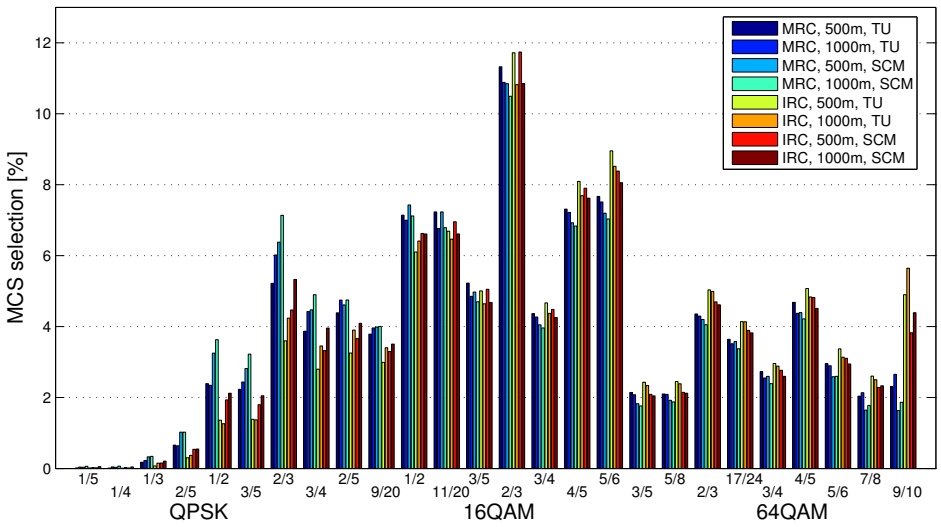
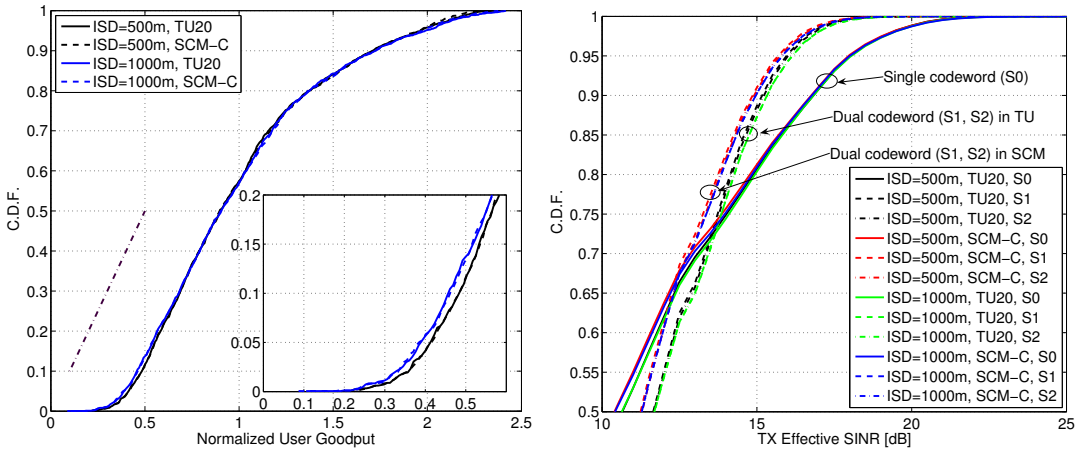


Figure 9.14: Downlink 1x2 results with Proportional Fair channel dependent scheduler with MRC and IRC receivers, modulation and coding statistics.

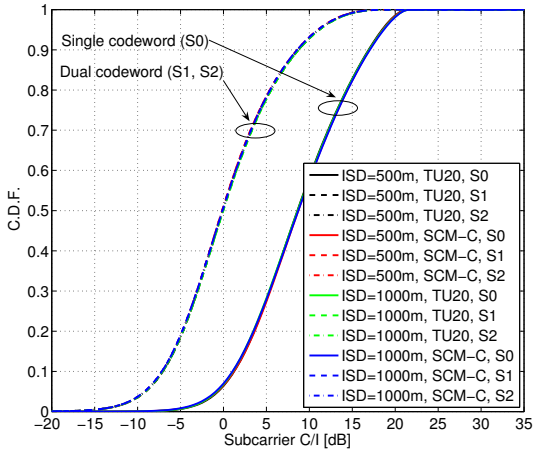
The fairness bound is shown with the normalized throughput curves as a probability line that needs to be exceeded. (The throughput distributions, cdf, are on the right side of the fairness bound.) Looking at the throughput statistics, we see that the IRC receiver algorithms gain particularly in the low SINR regime, which is critical for the cell edge throughput. As regards the channel models, the received subcarrier power to the interference power ratio in Fig. 9.13 is slightly lower over the full range for the SCM model compared to the TU model, which converts to slightly lower reported throughput values when using the SCM model. The full distributions show high throughput observations for all users, and the fairness bound is clearly exceeded. High throughput is achieved primarily by the receiver algorithms in Section 4.6, schedulers in Section 8.3 and adaptation algorithms in Sections 8.1 and 8.2. According to Fig. 9.14, link adapatation seems to provide transport format selection over a large set of transport formats.

When comparing the LTE baseline performance to the WCDMA/HSPA reference in [259], gains of order 250% to 300% were obtained, as summarized in Table 9.13. As stated, the performance may further be improved by MIMO. However, in the selected large cell scenario, where the amount of users, who may have any benefit from the multistream transmission, was modestly around 30%, and therefore MIMO provided only marginal average gains. This is visible in Fig. 6b, where the effective received SINR of the scheduled PRBs indicate the probability of dual codeword being actually in use for the selected



(a)

(b)



(c)

Figure 9.15: Downlink 2x2 results (cdf); (a) user goodput normalized to the geometric mean with the lowest 20 percentile zoomed. (b) The effective received SINR for the scheduled PRBs with the selected transport format shown for the highest SINR regime (above 50 percentile), when the dual codeword transmission was actually applied. (c) Received subcarrier C/I, classified for the single codeword and dual codeword transmissions.

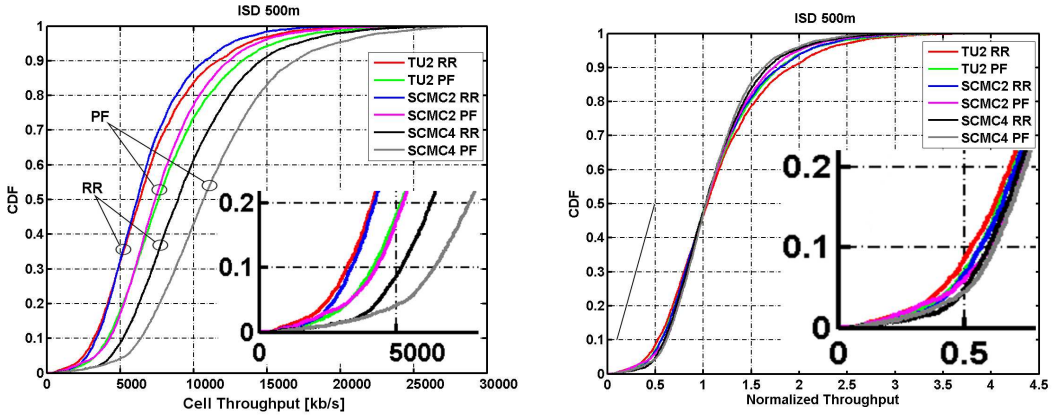
transport format. In these studies, the 4x4 MIMO scheme had more ideal assumptions than the other schemes, and it can be seen as the upper bound of performance in Table 9.13. Anyway, in a microcell scenario, where a larger proportion of users will benefit of the multistream transmissions due to higher SINR distribution, the average gains would be much higher.

For uplink, the baseline performance results are shown for a single antenna transmitter and for a two or four antenna receiver in Table 9.14, see Fig. 9.16. The operation regime was selected following the guidelines of loading in [63] that is not to exceed the Interference

Table 9.13: Downlink spectral efficiency at the mean and at the cell edge for various antenna configurations, MRC/IRC receivers and RR/PF schedulers.

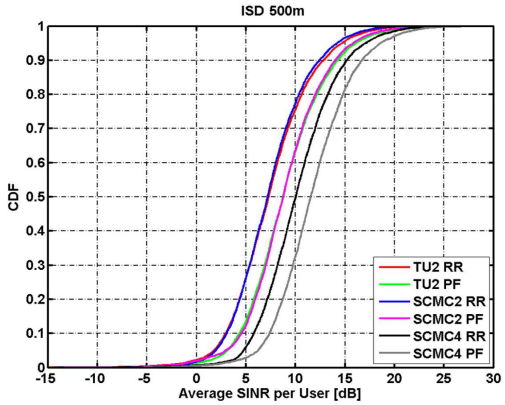
Best Effort traffic		
Feature	spectral efficiency cell mean [b/s/Hz/cell]	spectral efficiency cell edge [b/s/Hz/cell]
HSPA [259]	0.53	0.020
1x2 MRC RR	1.03	0.029
1x2 IRC RR	1.18	0.036
1x2 MRC PF	1.54	0.056
1x2 IRC PF	1.69	0.063
2x2 MRC PF	1.64	0.061
2x2 IRC PF	1.82	0.076
4x2 MRC PF	1.73	0.065
4x4 MRC PF	3.03	0.110

over the Thermal noise (IoT) level 20 dB by a probability larger than 0.05. For this criterion, we observed the mean IoT of about 14 dB with less than 3 dB deviation. The statistics of the transport format selected per transport block are shown in Fig. 9.17 for the adaptive modulation and coding. It was observed that by tuning the algorithms and transport format selection thresholds (outer-loop link adaptation), it is feasible to increase the cell edge performance with a minor impact to the mean. The power control reference setting impacts the mean and the cell edge shown in Fig. 9.18. For a small ISD (we had P_0 close to -60 dB), this is still not as dramatic as it gets for the large ISD shown later (see the mixed velocity case in Section 9.5.5 and Fig. 9.23). There, the cell edge performance may easily be sacrificed by inappropriate parametrisation. Typically, the uplink cell edge efficiency is one of the most critical measures of the system performance, and gains are favored especially in that regime. Throughout the cell throughput results, the Proportional Fair channel dependent scheduling gains are significant. Further, the receiver diversity gains achievable by the increase of the number of antenna elements in the eNodeB are notably large. When comparing the LTE baseline performance to the WCDMA/HSPA reference in [260], gains of order 260% to 370% were obtained at the mean and even 400% to 640% at the cell edge, as summarized in Table 9.13.

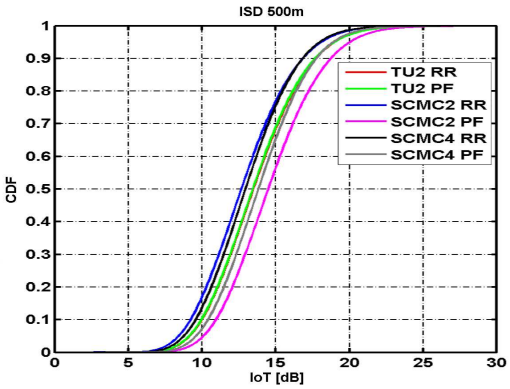


(a)

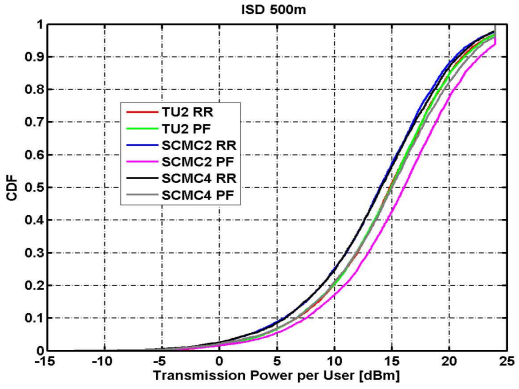
(b)



(c)



(d)



(e)

Figure 9.16: Uplink 1x2 and 1x4 results (cdf) with Round Robin and Proportional Fair channel dependent schedulers; (a) cell throughput, (b) cell throughput normalized to the geometric mean, (c) received SINR averaged per UE, (d) IoT, and (e) UE transmit power.

9.5.5 LTE system performance for best effort traffic in a mixed velocity scenario

The LTE performance for the Best Effort traffic is further analysed in a mixed velcocity scenario to find possible throughput losses due to mobility. A subset of transmission

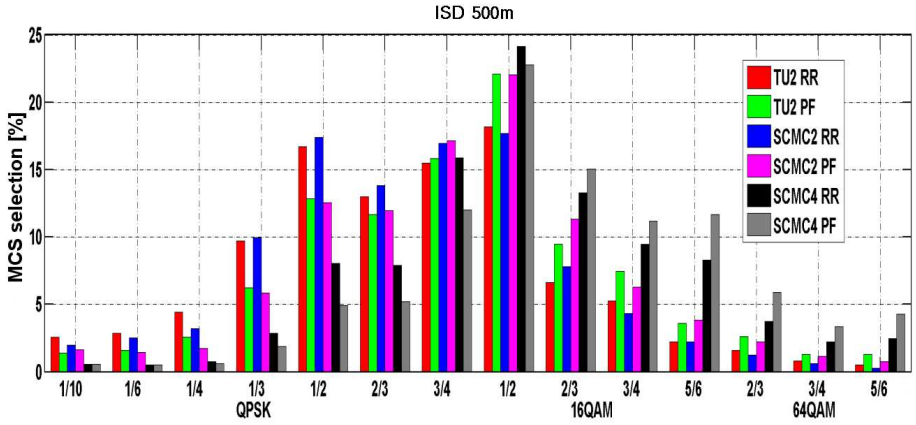


Figure 9.17: Uplink 1x2 and 1x4 results with Round Robin and Proportional Fair channel dependent schedulers, modulation and coding statistics.

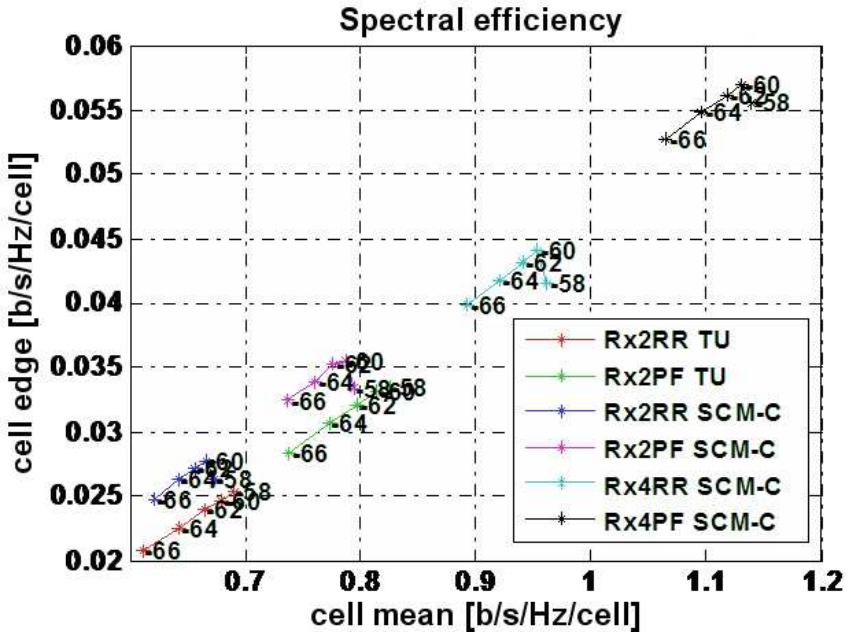


Figure 9.18: Uplink spectral efficiency, the cell edge to the cell mean ratio as a function of power control parameter P_0.

schemes and receivers, used in Section 9.5.4, are analysed here to show the relative performance loss. The goal in this study was not to reach the highest possible performance in the mixed-velocity scenario but rather to study the potential relative losses. Therefore, the multistream MIMO schemes were omitted in this comparison. The results are shown in Fig. 9.19 for downlink and in Fig. 9.21 for uplink. Link adaptation is shown for comparison to Section 9.5.4 in Figures Fig. 9.20 and Fig. 9.22.

The performance at 3 km/h still dominates these results, because the scheduling gains

Table 9.14: Uplink spectral efficiency at the mean and at the cell edge for various antenna configurations, FDE/MRC receivers and RR/PF schedulers.

Best Effort traffic		
Feature	spectral efficiency cell mean [b/s/Hz/cell]	spectral efficiency cell edge [b/s/Hz/cell]
HSPA [260]	0.33	0.009
1x2 RR	0.72	0.030
1x2 PF	0.86	0.036
1x4 RR	1.03	0.045
1x4 PF	1.22	0.058

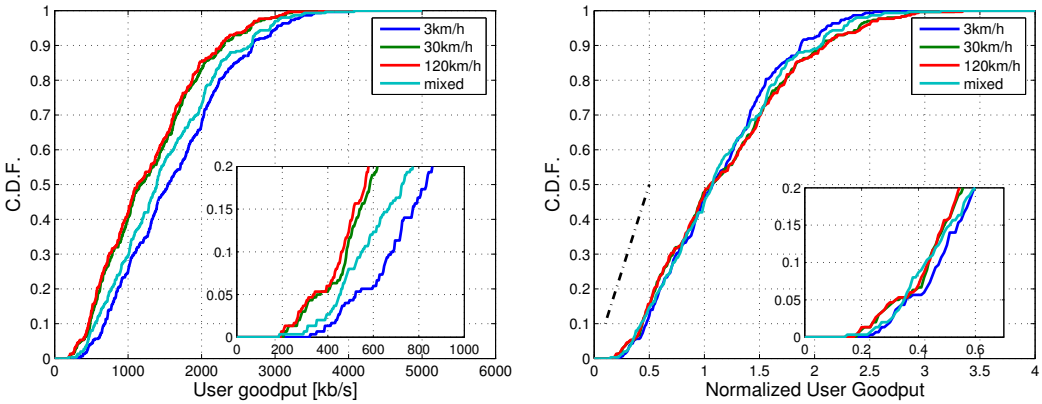


Figure 9.19: Downlink 1x2 results (cdf) in the mixed velocity scanario with Proportional Fair channel dependent scheduler, MRC and IRC receivers; (a) user goodput with the lowest 20 percentile zoomed, (b) user goodput normalized to the geometric mean and (c) received subcarrier C/I.

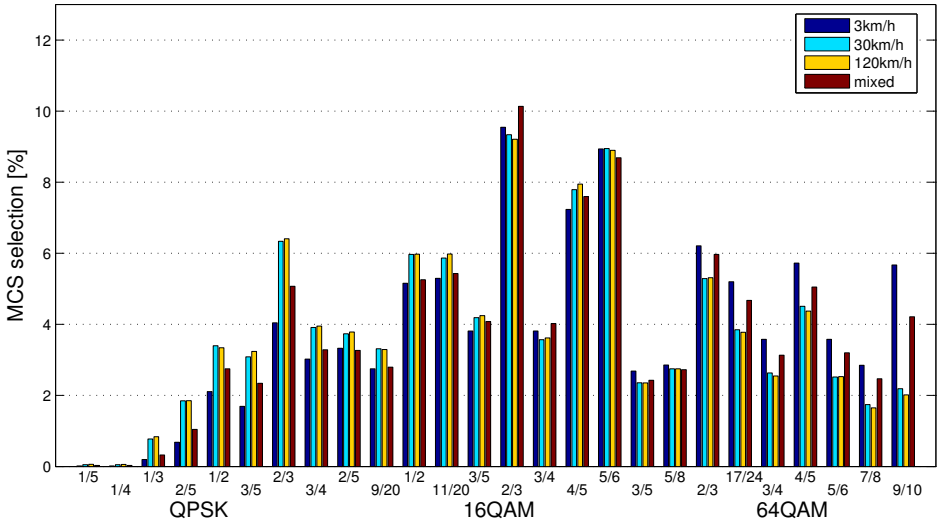


Figure 9.20: Downlink modulation and coding statistics.

at that velocity are significant, as discussed earlier. The large intersite distance set for the mixed velocity analysis will as such imply severe limitations to the observed SINR and hence to the throughput. The performance is critical at about 30 km/h, as the channel dependent scheduling may still gain over a blind diversity transmission. However, the scheduling gains can only be obtained by fairly accurate and fast CQI feedback (for downlink) or accurate channel sounding measurements (for uplink). Thus, the increased overhead (in uplink) of the feedback (for downlink), as well as the inaccuracy of the reports will partly mitigate the achievable gains. For 120 km/h, the channel coherence time is so short that it is not reasonable to try to do channel dependent scheduling at all. On the contrary, it is better to save in feedback and schedule blindly for the maximal frequency diversity. This is doable by allocating PRBs with large frequency spacing according to the flexible FDM allocation scheme, or to frequency shift (hop) PRB allocations between the consequent transmission instances. The latter mechanism applies for the uplink as well. Further in uplink, the power control set-point clearly impacts the cell edge to the cell mean ratio, as shown in Fig. 9.23 for the large intersite distance. Setting P_0 larger will cause larger IoT (the mean IoT was about 6 dB), which allows increase of the cell mean efficiency with the expense of the cell edge. However, in the case of larger cells, decreasing the power control set-point is well motivated in order to generate lower IoT (the mean IoT was about 2.2 dB), which significantly increased the cell edge efficiency with only a moderate impact to the cell mean.

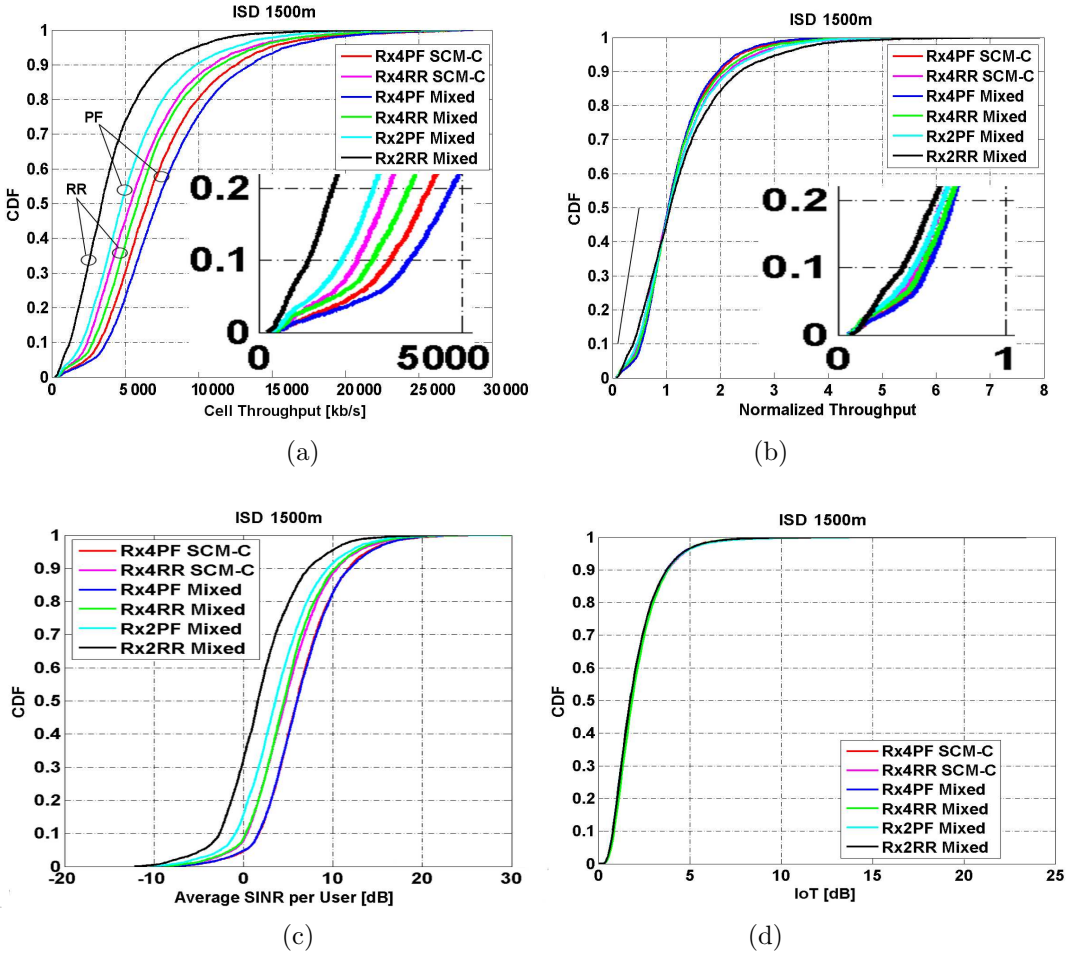


Figure 9.21: Uplink 1x2 and 1x4 results (cdf) in the mixed velocity scenario (large ISD) with Round Robin and Proportional Fair channel dependent schedulers; (a) cell throughput, (b) cell throughput normalized to the geometric mean, (c) received SINR averaged per UE, and (d) IoT.

In the overall mixed-velocity results, the dominance of 3 km/h links kept the losses of the spectral efficiency small (from 2.6% to 9%) in downlink. In uplink, the losses were larger (around 30%) due to non-optimal sounding feedback. During these studies, the performance versus overhead tradeoff was not yet optimized for different velocities, neither was the ATB algorithm (see Section 8.2.6) exploited to the full extent. These means have further potential to improve the results, particularly for the challenging mixed velocity scenario.

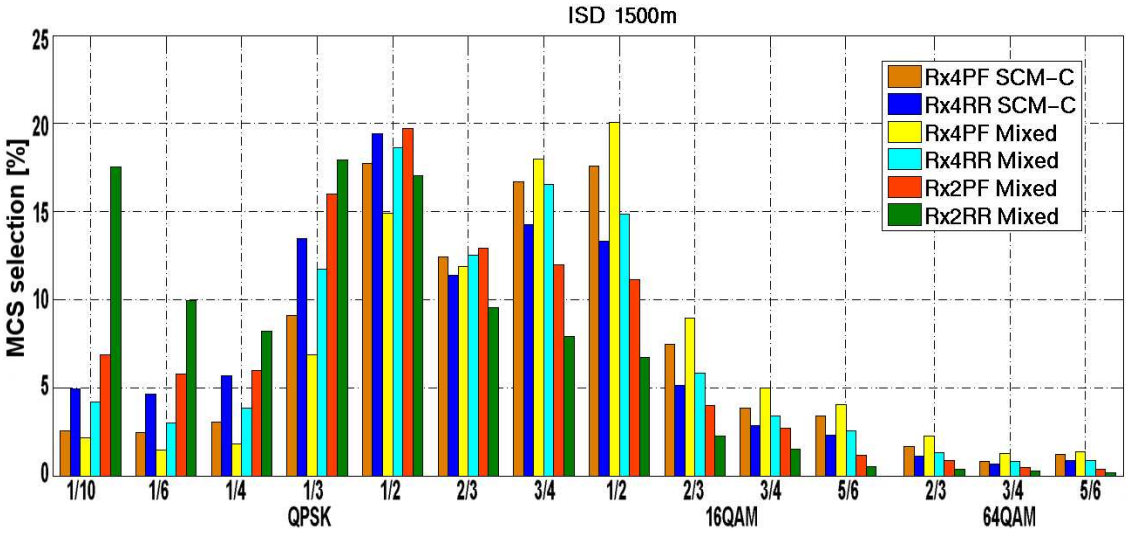


Figure 9.22: Uplink modulation and coding statistics.

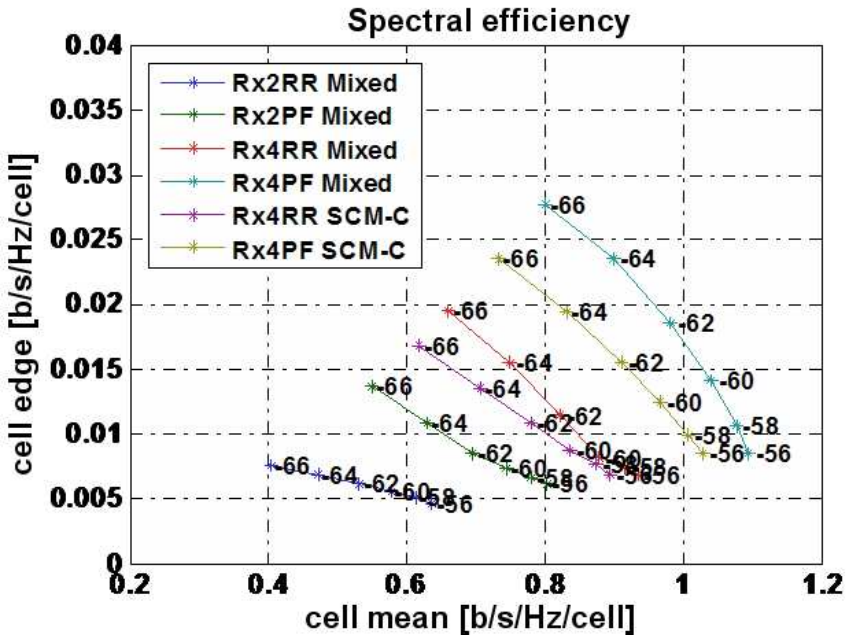


Figure 9.23: Uplink spectral efficiency, the cell edge to the cell mean ratio as a function of power control parameter P_0 for a large ISD.

Chapter 10

Summary

The thesis addresses topical issues of system performance for the converged mobile networks. The research results and system design aspects of the third generation technologies, WCDMA, WCDMA/HSPA and LTE were covered with the focus on radio communications, transmitter-receiver techniques and protocols.

The architecture evolution is motivated from two sides, first by the changes of the air interface technology and second by the requirements for the server based IP networking. The radio protocols have many similarities in their functional principles along the evolution, even if they differ in details due to the multiple-access technique.

The WCDMA/HSPA delay analysis is considered indicative of the relative packet delay performance of the radio protocols. It is notable that smaller packet delays may be observed in practise, when the network load is low and when the link adaptation thresholds rather favour small delay than large capacity. The air interface delays have traditionally been a concern, and making the delays smaller will remain a target for improvements in the practical implementations.

The main results of this thesis are represented in [24], where the VoIP and Best Effort Internet traffic are analysed for LTE in reference to WCDMA/HSPA. The VoIP capacity (in 5 MHz bandwidth) is shown to extend beyond 350 users (70 users/MHz/cell) in downlink and 230 users (46 users/MHz/cell) in uplink, which is a considerable increase compared to the reference. The spectral efficiency of LTE is shown to meet the 3GPP targets (for the IMT), and it provides impressive gains of order 2.5 to 3 times the reference. The spectral efficiency of LTE exceeds the mean of 1.5 b/s/Hz/cell in downlink and 0.8 b/s/Hz/cell in uplink for a baseline antenna configuration in urban macrocells with

20 dB penetration loss. In some scenarios, the gain potential is at least over four times the reference.

The thesis avoids direct numerical comparisons among the 3G technologies, but instead describes the theories, analyses and designs, which lead to the evolution and convergence of the mobile networks rather than their competition. The long term change is driven and motivated by the high level target setting and technical requirements, which strive for satisfying the demand of mobile Internet services that materialize to high compound annual growth rate of packet data volumes. The conversational voice, web-browsing and messaging services are expected to dominate, even though applications in the enterprise and entertainment segments indicate great new momentum as well.

Bibliography

[1] "3GPP," http://www.3gpp.org.

[2] "3GPP2," http://www.3gpp2.org.

[3] "WiMAX Forum," http://www.wimaxforum.org.

[4] "WiFi," http://www.wi-fi.org.

[5] D. Goodman and R. Myers, "3G cellular standards and patents," in *Proc. IEEE International Conference on Wireless Networks, Communications and Mobile Computing*, Jun. 2005, vol. 1, pp. 415–420.

[6] H. Holma and A. Toskala, *WCDMA for UMTS: HSPA Evolution and LTE*, John Wiley, Fourth Edition, 2007, 539 pages.

[7] H. Holma and A. Toskala, *LTE for UMTS: OFDMA and SC-FDMA based radio access*, John Wiley, 2009, 433 pages.

[8] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, Academic Press, Second edition, 2008, 448 pages.

[9] P. Lescuyer and T. Lucidarme, *Evolved packet System (EPS): The LTE and SAE Evolution of 3G UMTS*, John Wiley, 2008.

[10] H. Myung and D. Goodman, *Single Carrier FDMA: A New Air Interface for Long Term Evolution*, Wireless Communications and Mobile Computing. John Wiley, 2008.

[11] "International Telecommunication Union, Radio Communication Bureau; Invitation for submission of proposals for candidate radio interface technologies for the terrestrial components of the radio interface(s) for IMT-Advanced and invitation to

participate in their subsequent evaluation," Circular Letter 5/LCCE/2 (2008), Mar. 2008.

[12] J. Laakso, T. Koljonen, and M. Rinne, "Radio resource knapsack packing for WTDMA air interface," in *Proc. IEEE International Conference on Universal Personal Communications ICUPC 1998*, Florence, Oct. 1998, vol. 1, pp. 255–260.

[13] M. Rinne, S. Hämäläinen, and H. Lilja, "Effects of adjacent channel interference on WCDMA capacity," in *Proc. IEEE International Conference on Telecommunications, ICT 1999*, Cheju, Korea, Jun. 1999, pp. 127–132.

[14] S. Pöyhönen and M. Rinne, "Cell loading in a multi-service CDMA network with inter-cell interference," in *Proc. Fourth CDMA International Conference CIC by ETRI*, Seoul, Korea, Sep. 1999, pp. 87–92.

[15] T. Buot and M. Rinne, "Performance of WCDMA with radio resource management," in *Twelvth Annual International Conference on Wireless Communications, Wireless 2000*, Calgary, Alberta, Canada, Jul. 2000, pp. 203–209.

[16] R. Kwan and M. Rinne, "Performance analysis of the downlink shared channel in a WCDMA network," in *Proc. IEEE International Conference on Telecommunications, ICT 2001*, Bucharest, Romania, Jun. 2001, vol. 3, pp. 379–384.

[17] A. Mate and M. Rinne, "Performance of voice over IP on the WCDMA radio interface with the robust header compression scheme," in *Proc. International Society of Optical engineering SPIE Voice over IP Technology, Information Technology and Communications ITcom 2001*, Denver, USA, Aug. 2001, vol. 4522, pp. 148–156.

[18] R. Kwan, P. Chong, and M. Rinne, "Analysis of the adaptive modulation and coding algorithm with the multicode transmission," in *Proc IEEE Fifty-Sixth Vehicular Technology Conference, VTC 2002-Fall*, Sep. 2002, vol. 4, pp. 2007–2011.

[19] R. Kwan, P. Chong, E. Poutiainen, and M. Rinne, "The effect of code-multiplexing on the high speed downlink packet access (HSDPA) in a WCDMA network," in *Proc. IEEE Wireless Communications and Networking, WCNC 2003*, New Orleans, LA, USA, Mar. 2003, vol. 3, pp. 1728–1732.

[20] T. Kashima, M. Rinne, T. Roman, S. Visuri, S. Jarot, J. Kähtävä, A. Pokhariyal, T. Kolding, and O. Tirkkonen, "Design of channel structures for the Evolved-UTRA

downlink," in *Technical Report of Japan Society of Simulation Technology, JSST-MM2006*, Tokio, Japan, Oct. 2006, pp. 11–17.

[21] M. Rinne, K. Pajukoski, M. Kuusela, K. Pedersen, J. Ojala, E. Tuomaala, I. Kovács, H. Wang, P. H. Michaelsen, C. Rosa, and J. Michael, "Evaluation of the recent advances of the evolved 3G E-UTRA for the VoIP and best effort traffic scenarios," in *Proc. IEEE Eighth Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2007*, Helsinki, Finland, Jun. 2007, pp. 1–6.

[22] T. Kashima, M. Rinne, I. Kovács, M. Kuusela, P. Michaelsen, K. Pedersen, C. Rosa, and E. Tuomaala, "Performance evaluation of Evolved-UTRA data rates and throughput," in *Technical Report of Japan Society of Simulation Technology, JSST-MM2007-21*, Tokio, Japan, Oct. 2007, pp. 80–85.

[23] J. Puttonen, N. Kolehmainen, T. Henttonen, M. Moisio, and M. Rinne, "Mixed traffic packet scheduling in UTRAN long term evolution downlink," in *Proc. IEEE Nineteenth International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2008*, Cannes, France, Sep. 2008, pp. 1–5.

[24] M. Rinne, M. Kuusela, E. Tuomaala, P. Kinnunen, I. Kovács, K. Pajukoski, and J. Ojala, "A performance summary of the evolved 3G (E-UTRA) for voice over internet and best effort traffic," *IEEE Transactions on Vehicular Technology*, vol. 58, issue 7, pp. 3661–3673, Sep. 2009.

[25] R. Kwan and M. Rinne, "A comparison of WCDMA network performance results with frame vs. slot resolution simulations," in *Proc. Fifth CDMA International Conference CIC 2000 by ETRI*, Seoul, Korea, Nov. 2000.

[26] G. Manuel and M. Rinne, "Analysis of the transmission window for the delay performance of the high speed downlink packet access protocol," in *Proc. IEEE International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2003*, Split/Dubrovnik/Ancona/Venice, Croatia/Italy, Oct. 2003, pp. 566–571.

[27] G. Manuel and M. Rinne, "Performance of the medium access control protocol for the high speed downlink packet access," in *Proc. IASTED International Association*

of Science and Technology for Development, Communication Systems and Networks, CSN 2003, Benalmádena, Spain, Sep. 2003, pp. 42–47.

[28] J. Leino, J. Kurjenniemi, and M. Rinne, "Analysis of fast alpha switching for closed loop mode 1 transmit diversity with high speed downlink packet access," in *Proc. IEEE Sixtieth Vehicular Technology Conference, VTC 2004-Fall*, Sep. 2004, vol. 6, pp. 4466–4470.

[29] J. Äijänen, M. Rinne, J. Kurjenniemi, and T. Ristaniemi, "UTRA TDD intra-frequency handover performance," in *Proc. Global Mobile Congress by CIC/IEEE*, Shanghai, China, Oct. 2004, pp. 227–231.

[30] A. Mate, C. Caldera, and M. Rinne, "Performance of the packet traffic on the downlink shared channel in a WCDMA cell," in *Proc. IEEE International Conference on Telecommunications, ICT 2001*, Bucharest, Romania, Jun. 2001, vol. 3, pp. 385–390.

[31] J. Laakso, R. Jäntti, M. Rinne, and O. Salonaho, "Radio resource knapsack packing for WCDMA air interface," in *Proc. IEEE Ninth International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 1998*, Boston, MA, Sep. 1998, vol. 1, pp. 183–187.

[32] J. H. Sarker, S. J. Halme, and M. Rinne, "Performance analysis of GSM traffic channel capacity with(out) high speed circuit switched data," in *Proc. IEEE Fifty-Second Vehicular Technology Conference, VTC 2000-Fall*, Boston, MA, Sep. 2000, vol. 4, pp. 1603–1609.

[33] Y. Yang, T. H. Stitz, M. Rinne, and M. Renfors, "Mitigation of narrowband interference in SC transmission with filter bank equalization," in *Proc. IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2006*, Singapore, Dec. 2006, pp. 748–751.

[34] Y. Yang, M. Rinne, and M. Renfors, "Filter bank based frequency-domain equalization with noise prediction," in *Proc. IEEE Seventeeth International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2006*, Helsinki, Sep. 2006, pp. 1–5.

[35] Y. Yang, T. Ihalainen, M. Renfors, and M. Rinne, "Noise predictive turbo equalization for a filter bank based receiver in a SC transmission system," in *IEEE Sixty-Fifth Vehicular Technology Conference, VTC 2007-Spring.*, Dublin, Apr. 2007, pp. 2389–2393.

[36] T. Ihalainen, T. H. Stitz, M. Rinne, and M. Renfors, "Channel equalization in filter bank based multicarrier modulation for wireless communications," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, article id 49389, 18 pages, 2007.

[37] Y. Yang, T. Ihalainen, M. Rinne, and M. Renfors, "Frequency-domain equalization in single-carrier transmission: Filter bank approach," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 10438, 16 pages, 2007.

[38] A. Viholainen, T. Ihalainen, M. Rinne, and M. Renfors, "Localized mode DFT-S-OFDMA implementation using frequency and time domain interpolation," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 750534, 9 pages, May 2009.

[39] J. Laakso, T. Koljonen, M. Rinne, and O. Salonaho, "Method and system for optimal utilization of the data communication capacity in a cellular radio system," patent no, US6456605, Sep. 2002.

[40] M. Rinne and L. Laitinen, "Method and system for controlling radio communications network and radio network controller," patent no, US6574473, Jun. 2003.

[41] M. Rinne, "Basic quality of service (QoS) mechanisms for wireless transmission of IP traffic," patent no, US6845100, EP1314281, Feb. 2005.

[42] A. Mate and M. Rinne, "Method and device for downlink packet switching," patent no, US6970438, Nov. 2005.

[43] M. Rinne and J. Kalliokulju, "Method for informing layers of a protocol stack about the protocol in use," patent no, US6990107, Jan. 2006.

[44] G. Manuel and M. Rinne, "A delay-reduced stall avoidance mechanism for reordering a transport block," patent application, international publication number WO2006030312, US2006062223, Mar. 2006.

[45] M. Rinne, "Method, apparatus, data structure, computer program and system for providing appropriate QoS end-to-end for data crossing boundaries," patent application, international publication number US20050185651, Aug. 2005.

[46] F. Frederiksen, P. Mogensen, J. kahtava, and M. Rinne, "Method, apparatus and computer program to dynamically adjust segmentation at a protocol layer such as the medium access control (MAC) layer," patent application, international publication number WO2006117613, US20060245452, Nov. 2006.

[47] M. Rinne and O. Tirkkonen, "Discontinuous transmission/reception in a communications system," patent application, international publication number US20060195576, WO06114710, Aug. 2006.

[48] M. Rinne and J.-P. Kermoal, "Apparatus, method and computer program product to configure a radio link protocol for internet protocol flow," patent application, international publication number WO2007023366, US20070064608, Mar. 2007.

[49] M. Rinne, O. Tirkkonen, T. Kashima, S. Jarot, and J. Kahtava, "Unified entry format for common control signalling," patent application, international publication number WO2007023379, Mar. 2007.

[50] M. Rinne, Ü. Parts, and J. Korhonen, "A multicarrier pilot structure for reliable frame detection," patent application, international publication number WO2007052107, US20070098053, May. 2007.

[51] M. Rinne, O. Tirkkonen, and K. Hugl, "Pilot structure for multicarrier transmissions," patent application, international publication number, US20070070944, WO07036787, Mar. 2007.

[52] T. Kashima, M. Rinne, J. Ranta, and P. Purovesi, "Flexible segmentation scheme for communication systems," patent application, international publication number WO2007077526, Jul. 2007.

[53] J. Kahtava, M. Rinne, and O. Tirkkonen, "Apparatus, method and computer program product providing partitioned downlink shared control channel having fixed and variable component parts," patent application, international publication number WO2007132329, US20070265016, Nov. 2007.

[54] K. Pajukoski, E. Tiirola, M. Rinne, and P. Kinnunen, "Transmission time interval allocation for packet radio service," patent application, international publication number US20080080465, Apr. 2008.

[55] M. Rinne, F. Frederiksen, T. Kolding, and S. Visuri, "Shared control channel structure," patent application, international publication number WO2008081004, US20080159323, Jul. 2008.

[56] M. Rinne, O. Tirkkonen, and F. Frederiksen, "Apparatus, methods and computer program structures providing a common signaling entry for a modular control channel structure," patent application, international publication number WO2008084422, Jul. 2008.

[57] M. Rinne and J. Kalliokulju, "Method for informing layers of a protocol stack about the protocol in use," patent no, US7505444, Mar. 2009.

[58] M. Rinne and L. Laitinen, "Control of radio communication network having plural radio network controllers including an anchor controller," patent no, US7454210, Nov. 2008.

[59] H. Akhavan, V. Badrinath, T. Geitner, H. Lennertz, Y. Sha, T. Utano, and B. West, "A White paper by board of NGMN Limited: Next generation mobile networks, beyond HSPA & EVDO," *Next Generation Mobile Networks (NGMN)*, Dec. 2006, 60 pages.

[60] ITU-R, "Guidelines for evaluation of radio transmission technologies for IMT-2000," Recommendation M.1225, 1997.

[61] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Physical layer aspects for evolved universal terrestrial radio access UTRA (Release 7)," *3GPP TR 25.814 v7.1.0 (2006-09)*, 2006.

[62] "Universal Mobile Telecommunications System (UMTS); selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03)," ETSI TR 101 112 v3.2.0 (1998-04), 1998.

[63] R. Irmer, G. Liu, S. Xiadong, J. Krämer, S. Abeta, T. Sälzer, E. Jacks, A. Buldorini, and G. Wannemacher, "A White paper by the NGMN alliance: NGMN radio access

performance evaluation methodology," *Next Generation Mobile Networks (NGMN)*, Jan. 2008, 177 pages.

[64] R. Srinivasan, J. Zhuang, L. Jalloul, R. Novak, and J. Park, "Evaluation methodology document," *IEEE 802.16 Broadband Wireless Access Working Group. IEEE 802.16m*, Oct. 2008, 177 pages.

[65] J.-H. Park, "Wireless internet access for mobile subscribers based on the GPRS/UMTS network," *IEEE Communications Magazine*, vol. 40, issue 4, pp. 38–49, Apr. 2002.

[66] A. Viterbi, *CDMA Principle of Spread Spectrum Communication*, Wireless Communication. Addison-Wesley, 1995.

[67] Y.-B. Lin and I. Chlamtac, "Heterogeneous personal communications services: Integration of PCS systems," *IEEE Communications Magazine*, pp. 106–113, Sep. 1996.

[68] R. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*, Artech House, 2000.

[69] I. Koffman and V. Roman, "Broadband wireless access solutions based on OFDM access in IEEE 802.16," *IEEE Communications Magazine*, vol. 40, issue 4, pp. 96–103, Apr. 2002.

[70] Y. Yaghoobi, "Scalable OFDMA physical layer in IEEE 802.16 wireless man," *Intel Technology Journal*, vol. 8, issue 3, pp. 201–212, Aug. 2004.

[71] "Universal Mobile Telecommunications System (UMTS); requirements for the UMTS terrestrial radio access system (UTRA)," ETSI TR 101 111 v3.0.1 (1997-11), 1997.

[72] R. Carsello, R. Meidan, S. Allpress, F. O'Brien, J. Tarallo, N. Ziesse, A. Arunachalam, J. Costa, E. Berruto, R. Kirby, A. Maclatchy, F. Watanabe, and H. Xia, "IMT-2000 standards: radio aspects," *Proc. IEEE Personal Communications [see also IEEE Wireless Communications]*, vol. 4, issue 4, pp. 30–40, Aug. 1997.

[73] P. Mogensen, "High speed downlink packet access (HSDPA) -the path towards 3.5G," in *IEEE Workshop on Signal Processing Systems, SiPS 2001*, Antwerp, Belgium, sep. 2001.

[74] H. Ekstrom, A. Furuskär, J. Karlsson, M. Meyer, S. Parkvall, J. Torsner, and M. Wahlqvist, "Technical solutions for the 3G long-term evolution," *IEEE Communications Magazine*, vol. 44, issue 3, pp. 38–45, Mar. 2006.

[75] A. Furuskär, T. Jonsson, and M. Lundevall, "The LTE radio interface - key characteristics and performance," in *Proc. IEEE Nineteenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2008*, Cannes, Sep. 2008, pp. 1–5.

[76] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN) (Release 8)," *3GPP TR 25.913 v8.0.0 (2008-12)*, 2008.

[77] "International Telecommunication Union, radio communication study groups, working party 8F; Requirements related to technical performance for IMT-Advanced radio interface(s)," report ITU-R M.2134 (2008), 2008.

[78] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Requirements for further advancements for evolved universal terrestrial radio access (E-UTRA) (LTE-Advanced)," *3GPP TR 36.913 v8.0.1 (2009-03)*, 2009.

[79] D. Astely, E. Dahlman, P. Frenger, R. Ludwig, M. Meyer, S. Parkvall, P. Skillermark, and N. Wiberg, "A future radio-access framework," *IEEE Journal on Selected Areas in Communications*, vol. 24, issue 3, pp. 693–706, Mar. 2006.

[80] M. Roberts, *Mobile Network forecasts: Future mobile traffic, base stations and revenues*, Informa Telecoms & Media, 2008, 238 pages.

[81] M. Roberts, D. Mavrakis, and R. Jesty, *Future mobile broadband: HSPA & EV-DO to LTE networks, devices & services*, Informa Telecoms & Media, Third Edition, 2009, 328 pages.

[82] A. Hoikkanen, "A techno-economic analysis of 3G long-term evolution for broadband access," in *Proc. IEEE Sixth Conference on Telecommunication Techno-Economics CTTE 2007*, Helsinki, Jun. 2007, pp. 1–7.

[83] W. Jakes, *Microwave Mobile Communications*, IEEE Press, New York, 1974, 656 pages.

[84] R. Peterson, R. Ziemer, and D. Borth, *Introduction to Spread-Spectrum Communications*, Prentice Hall, 1995.

[85] F. Adachi, M. Sawahashi, and H. Suda, "Wideband DS-CDMA for next-generation mobile communications systems," *IEEE Communications Magazine*, vol. 36, issue 9, pp. 56–69, Sep. 1998.

[86] E. Dahlman, B. Gudmundson, M. Nilsson, and J. Sköld, "UMTS/IMT-2000 based on Wideband CDMA," *IEEE Communications Magazine*, vol. 36, issue 9, pp. 70–80, Sep. 1998.

[87] T. Ojanperä and R. Prasad, "An overview of air interface multiple access for IMT-2000/UMTS," *IEEE Communications Magazine*, vol. 36, issue 9, pp. 82–95, Sep. 1998.

[88] L. Hanzo and T. Keller, *OFDM and MC-CDMA: A Primer*, John Wiley & Sons, 2006, 411 pages.

[89] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Physical layer procedures (FDD) Release 8," *3GPP TS 25.214 v8.5.0 (2009-03)*, 2009.

[90] R. J. Heath, S. Sandhu, and A. Paulraj, "Antenna selection for spatial multiplexing systems with linear receivers," *IEEE Communications Letters*, vol. 5, pp. 142–144, Apr. 2001.

[91] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, Cambridge, UK, 2003, 276 pages.

[92] A. Hottinen, O. Tirkkonen, and R. Wichman, *Multi-antenna Transceiver Techniques for 3G and Beyond*, John Wiley, 2003, 326 pages.

[93] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, issue 8, pp. 1451–1458, Oct. 1998.

[94] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, Mar. 1998.

[95] R. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *Bell Labs Technical Journal*, vol. 45, pp. 1775–1796, Dec. 1966.

[96] B. Saltzberg, "Performance of an efficient parallel data transmission system," *IEEE Trans. Communications*, vol. 15, pp. 805–811, Dec. 1967.

[97] A. Vahlin and N. Holte, "Optimal finite duration pulses for OFDM," *IEEE Trans. Communications*, vol. 44, pp. 10–14, Jan. 1996.

[98] A. Bahai and B. Saltzberg, *Multicarrier Digital Communications, Theory and Applications of OFDM*, Kluwer Academic, 1999.

[99] B. Priyanto, C. Rom, C. Navarro, T. Sorensen, P. Mogensen, and O. Jensen, "Effect of phase noise on spectral efficiency for UTRA Long Term Evolution," in *Proc. IEEE Seventeenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2006*, Helsinki, Sep. 2006, pp. 1–5.

[100] B. Priyanto, T. Sorensen, O. Jensen, T. Larsem, T. Kolding, and P. Mogensen, "Assessing and modeling the effect of RF impairments on UTRA LTE uplink performance," in *Proc. IEEE Sixty-Sixth Vehicular Technology Conference VTC 2007-Fall*, Baltimore, MD, Sep.–Oct. 2007, pp. 1213–1217.

[101] O. Can and H. Liu, "Phase noise suppression for OFDM systems over fast time-varying channels," in *Proc. IEEE International Conference on Communications ICC 2007*, Glasgow, Jun. 2007, pp. 4346–4350.

[102] I. Kaur, K. Thakur, M. Kulkarni, D. Gupta, and P. Arora, "Adaptive OFDM vs single carrier modulation with frequency domain equalization," in *Proc. IEEE International Conference on Computer Engineering and Technology ICCET 2009*, Singapore, Jan. 2009, vol. 1, pp. 238–242.

[103] M. Batariere, K. Baum, and T. Krauss, "Cyclic prefix length analysis for 4G OFDM systems," in *proc. IEEE Sixtieth Vehicular Technology Conference, VTC 2004-Fall*, Sep. 2004, vol. 1, pp. 543–547.

[104] L. Greenstein, V. Erceg, Y. Yeh, and M. Clark, "A new path-gain/delay-spread propagation model for digital cellular channels," *IEEE Transactions on Vehicular Technology*, vol. 46, issue 2, pp. 477–485, May 1997.

[105] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Evolved universal terrestrial radio access (E-UTRA) physical channels and modulation (Release 8)," *3GPP TS 36.211 v8.6.0 (2009-03)*, 2009.

[106] Y. G. Li and N. R. Sollenberger, "Clustered OFDM with channel estimation for high rate wireless data," *IEEE Transactions on Communications*, vol. 49, issue 12, pp. 2071–2076, Dec. 2001.

[107] U. Sorger, I. De Bröck, and M. Schnell, "Interleaved FDMA – a new spread-spectrum multiple-access scheme," in *Record of IEEE International Conference on Communications ICC 98*, Atlanta, GA, Jun. 1998, vol. 2, pp. 1013–1017.

[108] T. Frank, A. Klein, E. Costa, and E. Schulz, "IFDMA - a promising multiple access scheme for future mobile radio systems," in *Proc. IEEE Sixteenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2005*, Berlin, Sep. 2005, vol. 2, pp. 1214–1218.

[109] T. Frank, A. Klein, and E. Costa, "IFDMA: A scheme combining the advantages of OFDMA and CDMA," *IEEE Wireless Communications*, vol. 14, pp. 9–17, Jun. 2007.

[110] Y. Goto, T. Kawamura, H. Atarashi, and M. Sawahashi, "Variable spreading and chip repetition factors (VSCRF)-CDMA in reverse link for broadband wireless access," in *Proc. IEEE fourteenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2003*, Sep. 2003, vol. 1, pp. 254–259.

[111] Y. Goto, T. Kawamura, H. Atarashi, and M. Sawahashi, "Investigations on packet error rate of variable spreading and chip repetition factors (VSCRF)-CDMA wireless access in reverse link multi-cell environment," in *Proc. IEEE Sixtieth Vehicular Technology Conference VTC 2004-Fall*, Sep. 2004, vol. 2, pp. 944–948.

[112] N. Maeda, Y. Kishiyama, K. Higuchi, H. Atarashi, and M. Sawahashi, "Experimental evaluation of throughput performance in broadband packet wireless access based on VSF-OFCDM and VSF-CDMA," in *Proc. IEEE Fourteenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2003*, Sep. 2003, vol. 1, pp. 6–11.

[113] H. Su, J. Zhang, and P. Zhang, "A comparative study of two receiver schemes for interleaved OFDMA uplink," in *Proc. IEEE Sixty-Third Vehicular Technology Conference VTC 2006-Spring*, Melbourne, Vic., May 2006, vol. 3, pp. 1426–1430.

[114] E. Akay and E. Ayanoglu, "Bit interleaved coded modulation with space time block codes," in *Proc. IEEE Sixtieth Vehicular Technology Conference, VTC 2004-Fall*, Los Angeles, USA, Sep. 2004, pp. 2477–2481.

[115] S. Nagaraj, "Generalized BICM for block fading channels," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 4404–4410, Nov. 2008.

[116] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "User equipment (UE) radio transmission and reception (Release 8)," *3GPP TS 25.101 v8.7.0 (2009-05)*, 2009.

[117] A. Skrzypczak, P. Siohan, and J. P. Javaudin, "Power spectral density and cubic metric for the OFDM/OQAM modulation," in *Proc. IEEE International Symposium on Signal Processing and Information Technology*, Vancouver, BC, Aug. 2006, pp. 846–850.

[118] H. G. Myung, J. Lim, and D. J. Goodman, "Peak-to-average power ratio of single carrier FDMA signals with pulse shaping," in *Proc. IEEE Seventeenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2006*, Helsinki, Sep. 2006, pp. 1–5.

[119] T. Lunttila, J. Lindholm, K. Pajukoski, E. Tiirola, and A. Toskala, "EUTRAN uplink performance," in *Proc. IEEE Second International Symposium on Wireless Pervasive Computing ISWPC 2007*, San Juan, Feb. 2007.

[120] A. Czylwik, "Comparison between adaptive OFDM and single carrier modulation with frequency domain equalization," in *Proc. IEEE Fourty-Seventh Vehicular Technology Conference*, Phoenix, AZ, May 1997, vol. 2, pp. 865–869.

[121] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Communications Magazine*, vol. 40, issue 4, pp. 58–66, Apr. 2002.

[122] D. Galda and H. Rohling, "A low complexity transmitter structure for OFDM-FDMA uplink systems," in *Proc. IEEE Fifty-Fiftieth Vehicular Technology Conference, VTC 2002-Spring*, May 2002, vol. 4, pp. 1737–1741.

[123] J. Zhang, C. Huang, G. Liu, and P. Zhang, "Comparison of the link level performance between OFDMA and SC-FDMA," in *Proc. First International Conference on Communications and Networking in China ChinaCom '06*, Beijing, Oct. 2006, pp. 1–6.

[124] E. Virtej, M. Kuusela, and E. Tuomaala, "System performance of single-user MIMO in LTE downlink," in *Proc. IEEE Nineteenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2008*, Cannes, Sep. 2008, pp. 1–5.

[125] K. Van Acker, G. Leus, M. Moonen, O. Van de Wiel, and T. Pollet, "Per tone equalization for DMT-based systems," *IEEE Trans. Communications*, vol. 49, pp. 109–119, Jan. 2001.

[126] A. Bury, J. Egle, and J. Lindner, "Diversity comparison of spreading transforms for multicarrier spread spectrum transmission," *IEEE Transactions on Communications*, vol. 51, issue 5, pp. 774–781, May 2003.

[127] H. Malvar, *Signal Processing with Lapped Transforms*, Artech House, 1992.

[128] T. Wiegand and N. Fliege, "Equalizers for transmultiplexers in orthogonal multiple carrier data transmission," in *Proc. European Signal Processing Conference*, Trieste, Italy, Sep. 1996, vol. 2, pp. 1211–1214.

[129] L. Vandendorpe, L. Cuvelier, F. Deryck, J. Louveaux, and O. Van de Wiel, "Fractionally spaced linear and decision-feedback detectors for transmultiplexers," *IEEE Trans. Signal Processing*, vol. 46, pp. 996–1011, Apr. 1998.

[130] P. Siohan, C. Siclet, and N. Lacaille, "Analysis and design of OFDM/OQAM systems based on filterbank theory," *IEEE Trans. Signal Processing*, vol. 50, pp. 1170–1183, May 2002.

[131] T. Karp and N. J. Fliege, "Modified DFT filter banks with perfect reconstruction," *IEEE Trans. Circuits and Systems II*, vol. 46, pp. 1404–1414, Nov. 1999.

[132] B. Farhang-Boroujeny, "Multicarrier modulation with blind detection capability using cosine modulated filter banks," *IEEE Transactions on Communications*, vol. 51, pp. 2057–2070, Dec. 2003.

[133] B. Farhang-Boroujeny, "Analysis of post-combiner equalizers in cosine-modulated filterbank-based transmultiplexer systems," *IEEE Transactions on Communications*, vol. 51, pp. 3249–3262, Dec. 2003.

[134] J. Alhava and M. Renfors, "Adaptive sine-modulated/cosine-modulated filter bank equalizer for transmultiplexers," in *Proc. European Conference on Circuit Theory and Design ECCTD*, Espoo, Finland, Aug. 2001, vol. III, pp. 337–340.

[135] J. Alhava and M. Renfors, "Complex lapped transforms and modulated filter banks," in *Proc. International Workshop on Spectral Methods and Multirate Signal Processing SMMSP*, Toulouse, France, Sep. 2002, pp. 87–94.

[136] A. Viholainen, T. Hidalgo Stitz, J. Alhava, T. Ihalainen, and M. Renfors, "Complex modulated critically sampled filter banks based on cosine and sine modulation," in *Proc. IEEE International Symposium on Circuits and Systems ISCAS 2002*, Scottsdale, USA, May 2002, pp. 833–836.

[137] D. Stein, "Detection of random signals in gaussian mixture noise," *IEEE Transactions on Information Theory*, vol. 41, pp. 1788–1801, Nov. 1995.

[138] J. Proakis, *Digital Communications, 3rd edition*, McGraw-Hill, 1995.

[139] H. Myer, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing*, John Wiley, 1998.

[140] X. Li, C. Yin, and G. Yue, "Soft decision equalization of multiple antenna systems over dispersive channels via max-log-map sphere decoder," in *Proc. IEEE Seventeeth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2006*, Helsinki, Sep 2006, pp. 1–5.

[141] T. Krauss, M. Zoltowski, and G. Leus, "Simple MMSE equalizers for CDMA downlink to restore chip sequence: comparison to zero-forcing and Rake," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2000*, Istanbul, Jun. 2000, vol. 5, pp. 2865–2868.

[142] J.-K. Zhang, A. Kavcic, and K. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection," *IEEE Transactions on Information Theory*, vol. 51, issue 1, pp. 154–172, Jan. 2005.

[143] K. Alimgeer, A. Naveed, M. Chaudhry, and I. Qurashi, "A comparison of interference cancellation scheme based on QR decomposition and based on MMSE for space time block code in multi-rate multi-user systems," in *Proc. IEEE Multitopic Conference, INMIC 2006*, Islamabad, Dec. 2006, pp. 79–83.

[144] C. Mehlführer and M. Rupp, "A robust MMSE equalizer for MIMO enhanced HSDPA," in *Proc. Fortieth Asilomar Conference on Signals, Systems and Computers ACSSC '06*, Pacific Grove, CA, Oct.–Nov. 2006, pp. 129–133.

[145] A. Ghosh, X. Weimin, R. Ratasuk, A. Rottinghaus, and B. Classon, "Multi-antenna system design for 3GPP LTE," in *Proc. IEEE International Symposium on Wireless Communication Systems ISWCS 2008*, Reykjavik, Iceland, Oct. 2008, pp. 478–482.

[146] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Transactions on Wireless Communications*, vol. 5, issue 1, pp. 90–100, Jan. 2006.

[147] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, issue 6, pp. 1211–1226, Aug. 2002.

[148] L. Schumacher, J. Kermoal, F. Frederiksen, K. Pedersen, A. Algans, and P. Mogensen, *MIMO Channel Characterisation*, IST-1000-11729 METRA, 2001, 57 pages.

[149] D. Reed, J. Smith, A. Rodriguez, and G. Calcev, "Spatial channel models for multi-antenna systems," in *Proc. IEEE Fifty-Eighth Vehicular Technology Conference VTC 2003-Fall*, Oct. 2003, vol. 1, pp. 99–103.

[150] A. Taparugssanagorn, T. Jasma, and J. Ylitalo, "Spatial correlation and eigenvalue statistics investigation of wideband MIMO channel measurements," in *Proc. IEEE Seventeenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC*, Helsinki, Sep. 2006, pp. 1–5.

[151] M. Narandzic, P. Kyosti, J. Meinila, L. Hentila, M. Alatossava, T. Rautiainen, Y. de Jong, C. Schneider, and R. Thoma, "Advances in "Winner" wideband MIMO

system-level channel modelling," in *Proc. The Second European Conference on Antennas and Propagation EuCAP 2007*, Edinburgh, Nov. 2007, pp. 1–7.

[152] P. Kyösti, J. Meinilä, L. Hentilä, X. Zhao, T. Jämsä, C. Schneider, M. Narandzi, M. Milojevi, A. Hong, J. Ylitalo, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. de Jong, and T. Rautiainen, *Channel models*, IST-4-027756 Winner II, 2008, 82 pages.

[153] E. Lee and D. Messerschmitt, *Digital Communication, 2nd edition*, Kluwer Academic Publishers, 2000.

[154] G. Berardinelli, B. E. Priyanto, T. B. Sorensen, and P. Mogensen, "Improving SC-FDMA performance by Turbo equalization in UTRA LTE uplink," in *Proc. IEEE Vehicular Technology Conference, VTC Spring-2008*, Singapore, May. 2008, pp. 2557–2561.

[155] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "Session Initiation Protocol," *IETF. RFC 3261*, Jun. 2002.

[156] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "A Transport Protocol for Real-time applications (RTP)," *IETF. RFC 3550*, Jul. 2003.

[157] J. Postel, "Transmission Control Protocol, (TCP)," *IETF. RFC 0793, USC/Information Sciences Institute*, Sep. 1981.

[158] M. Allman, V. Paxson, and W. Stevens, "TCP congestion control," *IETF. RFC 2581, USC/Information Sciences Institute*, Apr. 1999.

[159] E. J. Postel, "User Datagram Protocol (UDP)," *IETF. RFC 0768, USC/Information Sciences Institute*, Aug. 1980.

[160] J. Postel, "Internet Protocol, (IP)," *IETF. RFC 0791, USC/Information Sciences Institute*, Sep. 1981.

[161] S. Deering and R. Hinden, "Internet Protocol, version 6 (IPv6)," *IETF. RFC 2460, USC/Information Sciences Institute*, Dec. 1998.

[162] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W.Weiss, "An architecture for differentiated services," *IETF. RFC 2475*, Dec. 1998.

[163] J. Sjoberg, M. Westerlund, A. Lakaniemi, and Q. Xie, "RTP payload format and file storage format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) audio codecs," *IETF. RFC 4867*, Apr. 2007.

[164] M. Handley, V. Jacobson, and C. Perkins, "Sdp: Session Description Protocol," *IETF. RFC 4566*, Jul. 2006.

[165] M. Degermark, B. Nordgren, and S. Pink, "IP header compression," *IETF. RFC 2507*, Feb. 1999.

[166] C. Bormann, C. Burmeister, M. Degermark, H. Fukushima, H. Hannu, L.-E. Jonsson, R. Hakenberg, T. Koren, K. Le, Z. Liu, A. Martensson, A. Miyazaki, K. Svanbro, T. Wiebke, T. Yoshimura, and H. Zheng, "Robust Header Compression (ROHC)," *IETF. RFC 3095*, Jul. 2001.

[167] K. Le, C. Clanton, Z. Liu, and H. Zheng, "Efficient and robust header compression for real-time services," in *Proc. IEEE Wireless Communications and Networking Conference WCNC 2000*, Chicago, IL, Sep. 2000, vol. 2, pp. 924–928.

[168] K. Svanbro, H. Hannu, L. E. Jonsson, and M. Degermark, "Wireless real-time IP services enabled by header compression," in *Proc. IEEE Fifty-First Vehicular Technology Conference VTC 2000-Spring*, Tokyo, May 2000, vol. 2, pp. 1150–1154.

[169] H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian, and V. Niemi, *UMTS Networks: Architecture, Mobility and Services*, John Wiley, 2001, 302 pages.

[170] L. Skorin-Kapov, M. Mosmondor, O. Dobrijevic, and M. Matijasevic, "Application-level QoS negotiation and signaling for advanced multimedia services in the IMS," *IEEE Communications Magazine*, vol. 45, issue 7, pp. 108–116, Jul. 2007.

[171] I. Siomina and S. Wanstedt, "The impact of QoS support on the end user satisfaction in LTE networks with mixed traffic," in *Proc. IEEE Nineteenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2008*, Cannes, France, Sep. 2008, pp. 1–5.

[172] M. El Barachi, R. Glitho, and R. Dssouli, "Context-aware signaling for call differentiation in IMS-based 3G networks," in *Proc. IEEE Twelvth Symposium on Computers and Communications, ISCC 2007*, Aveiro, Jul. 2007, pp. 789–796.

[173] 3rd Generation Partnership Project; Technical Specification Group Services and S. Aspects, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) (Release 8)," *3GPP TS 23.401 v8.5.0 (2009-03)*, 2009.

[174] F. Poppe, D. De Vleeschauwer, and G. Petit, "Guaranteeing quality of service to packetised voice over the UMTS air interface," in *2000 Eighth International Workshop on Quality of Service IWQOS 2000*, Pittsburgh, PA, Jun. 2000, pp. 85–91.

[175] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Evolved Universal Terrestrial Radio Access (E-UTRA) and evolved Universal Terrestrial Radio Access Network (E-UTRAN); overall description; Stage 2 (Release 8)," *3GPP TS 36.300 v8.8.0 (2009-03)*, 2009.

[176] M. Rinne and G. Manuel, "Outer-loop power control with transport block diversity transmission," patent application, international publication number EP1774676A,WO 2006016230, Feb. 2006.

[177] J. Peisa, S. Wager, M. Sagfors, J. Torsner, B. Goransson, T. Fulghum, C. Cozzo, and S. Grant, "High speed packet access evolution - concept and technologies," in *Proc. IEEE Sixty-Fifth Vehicular Technology Conference VTC 2007-Spring*, Dublin, Apr. 2007, pp. 819–824.

[178] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "High speed downlink packet access (HSDPA) overall description; stage 2," *3GPP TS 25.308 v8.5.0 (2009-03)*, 2009.

[179] M. Nakamura, Y. Awad, and S. Vadgama, "Adaptive control of link adaptation for high speed downlink packet access (HSDPA) W-CDMA," in *Proc. IEEE International Symposium on Wireless Personal Multimedia Communications, WPMC 2002*, Honolulu, Hawaii, USA, Oct. 2002, vol. 2, pp. 382–386.

[180] S. Parkvall, J. Peisa, J. Torsner, M. Sagfors, and P. Malm, "WCDMA enhanced uplink - principles and basic operation," in *Proc. IEEE Sixty-First Vehicular Technology Conference VTC 2005-Spring*, May–Jun. 2005, vol. 3, pp. 1411–1415.

[181] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Radio resource control (RRC) protocol specification (Release 5)," *3GPP TS 25.331 v8.6.0 (2009-03)*, 2009.

[182] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Evolved universal terrestrial radio access (E-UTRA) radio resource control (RRC) protocol specification (Release 8)," *3GPP TS 36.331 v8.5.0 (2009-03)*, 2009.

[183] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Packet data convergence protocol (PDCP) specification (Release 7)," *3GPP TS 25.323 v8.4.0 (2009-03)*, 2009.

[184] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Evolved universal terrestrial radio access (E-UTRA) packet data convergence protocol (PDCP) protocol specification (Release 8)," *3GPP TS 36.323 v8.5.0 (2009-03)*, 2009.

[185] R. Bestak, P. Godlewski, and P. Martins, "RLC buffer occupancy when using a TCP connection over UMTS," in *Proc. IEEE Thirteenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2002*, Sep. 2002, vol. 3, pp. 1161–1165.

[186] H. Lin and S. K. Das, "Performance study of link layer and MAC layer protocols to support TCP in 3G CDMA systems," *IEEE Transactions on Mobile Computing*, vol. 4, issue 5, pp. 489–501, Sep.–Oct. 2005.

[187] A. Sang, X. Wang, and M. Madihian, "Differentiated TCP user perception over downlink packet data cellular systems," *IEEE Transactions on Mobile Computing*, vol. 6, issue 3, pp. 252–263, Mar. 2007.

[188] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Radio link control (RLC) protocol specification (Release 8)," *3GPP TS 25.322 v8.4.0 (2009-03)*, 2009.

[189] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Evolved universal terrestrial radio access (E-UTRA) radio link con-

trol (RLC) protocol specification (Release 8)," *3GPP TS 36.322 v8.5.0 (2009-03)*, 2009.

[190] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Medium access control (MAC) protocol specification (Release 7)," *3GPP TS 25.321 v7.12.0 (2009-03)*, 2009.

[191] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Evolved universal terrestrial radio access (E-UTRA) medium access control (MAC) protocol specification (Release 8)," *3GPP TS 36.321 v8.5.0 (2009-03)*, 2009.

[192] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Physical channels and mapping of transport channels onto physical channels (FDD) Release 8," *3GPP TS 25.211 v8.4.0 (2009-03)*, 2009.

[193] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Services provided by the physical layer (Release 8)," *3GPP TS 25.302 v8.2.0 (2008-12)*, 2008.

[194] R. Cam and C. Leung, "Multiplexed ARQ for time-varying channels," *IEEE Transactions on Communications*, vol. 46, issue 1, pp. 41–51, Jan. 1998.

[195] S. Hara, A. Ogino, M. Araki, M. Okada, and N. Morinaga, "Throughput performance of SAW-ARQ protocol with adaptive packet length in mobile packet data transmission," *IEEE Transactions on Vehicular Technology*, vol. 45, issue 3, pp. 561–569, Aug. 1996.

[196] S. Falahati and A. Svensson, "Hybrid type-II ARQ schemes with adaptive modulation systems for wireless channels," in *Proc. IEEE Fiftieth Vehicular Technology Conference VTC 1999-Fall*, Amsterdam, Sep. 1999, vol. 5, pp. 2691–2695.

[197] E. Malkamäki, D. Mathew, and S. Hämäläinen, "Performance of hybrid ARQ techniques for WCDMA high data rates," in *Proc. IEEE Fifty-Third Vehicular Technology Conference VTC 2001-Spring*, Rhodes, May 2001, vol. 4, pp. 2720–2724.

[198] R. Cam and C. Leung, "Throughput analysis of some ARQ protocols in the presence of feedback errors," *IEEE Transactions on Communications*, vol. 45, issue 1, pp. 35–44, Jan. 1997.

[199] T. Kolding, F. Frederiksen, and P. Mogensen, "Performance aspects of WCDMA systems with high speed downlink packet access (HSDPA)," in *Proc. IEEE Vehicular Technology Conference, VTC 2002-Fall*, Vancouver (BC), Canada, Sep. 2002, vol. 1, pp. 477–481.

[200] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Evolved universal terrestrial radio access (E-UTRA) user equipment (UE) radio transmission and reception (Release 8)," *3GPP TS 36.101 v8.5.1 (2009-03)*, 2009.

[201] H. Chen and P. Mohapatra, "CAM: a context-aware transportation protocol for HTTP," in *Proc. Twenty-Third International Conference on Distributed Computing Systems Workshops ICDCS 2003*, May 2003, pp. 922–927.

[202] N. Elouazizi and Y. Bachvarova, "On cognitive relevance in automatic multimodal systems," in *Proc. IEEE Sixth International Symposium on Multimedia Software Engineering*, Dec. 2004, pp. 418–426.

[203] T. Thang, Y. Jung, and Y. Ro, "Modality conversion for QoS management in universal multimedia access," in *IEE Proceedings on Vision, Image and Signal Processing*, Jun. 2005, vol. 152, pp. 374–384.

[204] B. Mah, "An empirical model of HTTP network traffic," in *Proc. IEEE Sixteenth Annual Joint Conference of the Computer and Communications Societies INFOCOM '97*, Kobe, Apr. 1997, vol. 2, pp. 592–600.

[205] M. Lucas, D. Wrege, B. Dempsey, and A. Weaver, "Statistical characterization of wide-area IP traffic," in *Proc. Sixth International Conference on Computer Communications and Networks*, Las Vegas, NV, Sep. 1997, pp. 442–447.

[206] J. Heidemann, K. Obraczka, and J. Touch, "Modeling the performance of HTTP over several transport protocols," *IEEE/ACM Transactions on Networking*, vol. 5, issue 5, pp. 616–630, Oct. 1997.

[207] A. Simonsson and A. Furuskär, "Uplink power control in LTE - overview and performance: Principles and benefits of utilizing rather than compensating for SINR variations," in *Proc. IEEE Sixty-eighth Vehicular Technology Conference VTC 2008-Fall*, Calgary, BC, Sep. 2008, pp. 1–5.

[208] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Evolved universal terrestrial radio access (E-UTRA) physical layer procedures (Release 8)," *3GPP TS 36.213 v8.6.0 (2009-03)*, 2009.

[209] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Conference record IEEE International Conference on Communications ICC*, May 1993, vol. 2, pp. 1064–1070.

[210] H. Wen, L. Zhou, and C. Fu, "On the performance of incremental redundancy hybrid ARQ schemes with rate compatible LDPC codes," in *Proc. International Conference on Communications, Circuits and Systems ICCCAS 2006*, Guilin, China, Jun. 2006, vol. 2, pp. 731–734.

[211] N. Ericsson, "Adaptive modulation and scheduling of IP traffic over fading channels," in *Proc. IEEE Fiftieth Vehicular Technology Conference VTC 1999-Fall*, Amsterdam, Sep. 1999, vol. 2, pp. 849–853.

[212] C. Rosa, D. Villa, C. Castellanos, F. Calabrese, P.-H. Michaelsen, K. Pedersen, and P. Skov, "Performance of fast AMC in E-UTRAN uplink," in *IEEE International Conference on Communications ICC '08*, Beijing, May 2008, pp. 4973–4977.

[213] C. B. Ribeiro, K. Hugl, M. Lampinen, and M. Kuusela, "Performance of linear multiuser MIMO precoding in LTE system," in *IEEE Third International Symposium on Wireless Pervasive Computing ISWPC 2008*, Santorini, May 2008, pp. 410–414.

[214] A. Das, F. Khan, A. Sampath, and S. Hsuan-Jung, "Performance of hybrid ARQ for high speed downlink packet access in UMTS," in *Proc. IEEE Vehicular Technology Conference, VTC 2001-Fall*, Atlantic City (NJ), USA, Oct. 2001, vol. 4, pp. 2133–2137.

[215] F. Frederiksen and T. Kolding, "Performance and modeling of WCDMA HSDPA transmission-HARQ schemes," in *Vehicular Technology Conference, 2002. VTC 2002-Fall. 2002 IEEE*, Vancouver (BC), Canada, Sep. 2002, vol. 1, pp. 472–476.

[216] K. C. Beh, A. Doufexi, and S. Armour, "Performance evaluation of hybrid ARQ schemes of 3GPP LTE OFDMA system," in *Proc. IEEE Eighteenth International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2007*, Athens, Sep. 2007, pp. 1–5.

[217] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with Chase combining and incremental redundancy for HSDPA," in *IEEE fifty-Fourth Vehicular Technology Conference VTC 2001-Fall*, Atlantic City, NJ, Oct. 2001, vol. 3, pp. 1829–1833.

[218] R. Love, B. Classon, A. Ghosh, and M. Cudak, "Incremental redundancy for evolutions of 3G CDMA systems," in *Proc. IEEE Vehicular Technology Conference, VTC 2002-Spring*, Birmingham (AL) USA, May 2002, vol. 1, pp. 454–458.

[219] A. Ghosh, K. Stewart, R. Ralasuk, E. Buckley, and R. Bachu, "Incremental redundancy (IR) schemes for W-CDMA HS-DSCH," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2002*, Lisboa, Portugal, Sep. 2002, vol. 3, pp. 1078–1082.

[220] A. Pokhariyal, T. Kolding, and P. Mogensen, "Performance of downlink frequency domain packet scheduling for the UTRAN long term evolution," in *Proc. IEEE Seventeenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2006*, Helsinki, Sep. 2006, pp. 1–5.

[221] M. Sagfors, R. Ludwig, M. Meyer, and J. Peisa, "Queue management for TCP traffic over 3G links," in *Proc. IEEE Wireless Communications and Networking WCNC 2003*, New Orleans, LA, USA, Mar. 2003, vol. 3, pp. 1663–1668.

[222] J. Peisa, H. Ekstrom, H. Hannu, and S. Parkvall, "End-to-end performance of WCDMA enhanced uplink," in *Proc. IEEE Sixty-First Vehicular Technology Conference VTC 2005-Spring*, May–Jun. 2005, vol. 3, pp. 1432–1436.

[223] J. Lim, H. G. Myung, K. Oh, and D. J. Goodman, "Proportional fair scheduling of uplink single-carrier FDMA systems," in *Proc. IEEE Seventeenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2006*, Helsinki, Sep. 2006, pp. 1–6.

[224] A. Pokhariyal, *Downlink Frequency-Domain Adaptation and Scheduling -A case study based on the UTRA Long Tern Evolution*, Ph.D Thesis, Aalborg University, Denmark, Aug. 2007.

[225] K. Pedersen, G. Monghal, I. Kovács, T. Kolding, A. Pokhariyal, F. Frederiksen, and P. Mogensen, "Frequency domain scheduling for OFDMA with limited and noisy

channel feedback," in *Proc. IEEE Sixty-Sixth Vehicular Technology Conference VTC 2007-Fall*, Baltimore, MD, Sep.–Oct. 2007, pp. 1792–1796.

[226] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, "Dynamic packet scheduling performance in UTRA long term evolution downlink," in *Proc. IEEE Third International Symposium on Wireless Pervasive Computing ISWPC 2008*, Santorini, May 2008, pp. 308–313.

[227] G. Monghal, K. I. Pedersen, I. Z. Kovács, and P. E. Mogensen, "QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution," in *Proc. IEEE Vehicular Technology Conference VTC 2008-Spring*, Singapore, May 2008, pp. 2532–2536.

[228] T. Kolding, "Link and system performance aspects of proportional fair scheduling in WCDMA/HSDPA," in *Proc. IEEE Fifty-Eight Vehicular Technology Conference VTC 2003-Fall*, Oct. 2003, vol. 3, pp. 1717–1722.

[229] J. Kim, D. Lee, G. Hwang, and C. Oh, "A new scheduling algorithm for data services in HDR system: weight-gap first scheduling," in *Proc. International Conferences on Info-tech and Info-net ICII 2001*, Beijing, Oct.-Nov. 2001, vol. 2, pp. 329–334.

[230] A. Foronda, C. Ohta, and H. Tamaki, "An enhanced proportional fair algorithm to provide QoS over a shared wireless link," in *Proc. International Conference on Advanced Technologies for Communications ATC 2008*, Hanoi, Oct. 2008, pp. 106–109.

[231] S. R. Husain, N. Nasser, and H. Hassanein, "Efficient delay-based schedulig scheme for supporting real-time traffic in HSDPA networks," in *IEEE Thirty-Second Conference on Local Computer Networks LCN 2007*, Dublin, Oct. 2007, pp. 685–691.

[232] G. Liu, J. Zhang, J. Zhu, and W. Wang, "Channel and queue aware scheduling for real- time service in multiuser MIMO OFDM system," in *Proc. Asia-Pacific Conference on Communications APCC 2007*, Bangkok, Oct. 2007, pp. 509–512.

[233] P. Gutiérrez, *Packet Scheduling and Quality of Service in HSDPA*, Ph.D Thesis, Aalborg University, Denmark, Oct. 2003.

[234] A. Racz, A. Temesvary, and N. Reider, "Handover performance in 3GPP long term evolution (LTE) systems," in *IST Sixteenth Mobile and Wireless Communications Summit*, Budapest, Jul. 2007, pp. 1–5.

[235] I. Kovacs, K. Pedersen, T. Kolding, A. Pokhariyal, and M. Kuusela, "Effects of non-ideal channel feedback on dual-stream MIMO-OFDMA system performance," in *Proc. IEEE Sixty-Sixth Vehicular Technology Conference VTC 2007-Fall*, Baltimore, MD, Sep.–Oct. 2007, pp. 1852–1856.

[236] I. Kovács, M. Kuusela, E. Virtej, and K. Pedersen, "Performance of MIMO aware RRM in downlink OFDMA," in *Proc. IEEE Vehicular Technology Conference VTC Spring-2008*, Singapore, May 2008, pp. 1171–1175.

[237] M. Anas, C. Rosa, F. Calabrese, K. Pedersen, and P. Mogensen, "Combined admission control and scheduling for QoS differentiation in LTE uplink," in *Proc. IEEE Sixty-Eight Vehicular Technology Conference, VTC 2008-Fall*, Calgary, BC, Sep. 2008, pp. 1–5.

[238] W. Xiao, R. Ratasuk, A. Ghosh, R. Love, Y. Sun, and R. Nory, "Uplink power control, interference coordination and resource allocation for 3GPP E-UTRA," in *Proc. IEEE Sixty-Fourth Vehicular Technology Conference VTC 2006-Fall*, Montreal, Que., Sep. 2006, pp. 1–5.

[239] V. Koshi, "Coverage uncertainty and reliability estimation for microcellular radio network planning," in *Proc. IEEE Fifty-First Vehicular Technology Conference VTC 2000-Spring*, Tokyo, May 2000, vol. 1, pp. 468–472.

[240] J.-P. Linnartz and J. Davis II, "Outage probability in digital cellular radio networks," in *1992 Conference Record of The Twenty-Sixth Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct. 1992, pp. 88–92.

[241] B. Schroder, B. Liesenfeld, A. Weller, K. Leibnitz, D. Staehle, and P. Tran-Gia, "An analytical approach for determining coverage probabilities in large UMTS networks," in *Proc. IEEE Fifty-Fourth Vehicular Technology Conference VTC 2001*, Atlantic City, NJ, Oct. 2001, vol. 3, pp. 1750–1754.

[242] J. H. Kotecha and A. M. Sayeed, "On the capacity of correlated MIMO channels," in *Proc. IEEE International Symposium on Information Theory*, Jun.–Jul. 2003, pp. 355–355.

[243] M. Kang and M. S. Alouini, "Capacity of correlated MIMO rayleigh channels," *IEEE Transactions on Wireless Communications*, vol. 5, issue 1, pp. 143–155, Jan. 2006.

[244] H. Shin and J. H. Lee, "Closed-form formulas for ergodic capacity of MIMO rayleigh fading channels," in *Proc. IEEE International Conference on Communications ICC '03*, May 2003, vol. 5, pp. 2996–3000.

[245] G. Shen, S. Liu, S. hyun Jang, and J.-W. Chong, "On the capacity of MIMO-OFDM systems with doubly correlated channels," in *Proc. IEEE Sixty-Sixth Vehicular Technology Conference VTC 2007-Fall*, Baltimore, MD, Sep.–Oct. 2007, pp. 1218–1222.

[246] Z. Lin, B. Vucetic, and J. Mao, "Ergodic capacity of LTE downlink multiuser MIMO systems," in *Communications, 2008. ICC '08. IEEE International Conference on*, Beijing, May 2008, pp. 3345–3349.

[247] R. Saxena and R. Kumar, "Analysis of MIMO capacity for spatial channel model with partial CSI knowledge," in *Proc. Fourth International Conference on Wireless Communication and Sensor Networks WCSN*, Allahabad, India, Dec. 2008, pp. 187–191.

[248] R. Love, A. Ghosh, R. Nikides, L. Jalloul, M. Cudak, and B. Classon, "Throughput of high speed downlink packet access for UMTS," in *Proc. IEEE Vehicular Technology Conference, VTC 2001-Spring*, Rhodes, Greece, May. 2001, vol. 3, pp. 2234–2238.

[249] Y. Ofuji, T. Kawamura, Y. Kishiyama, K. Higuchi, and M. Sawahashi, "System-level throughput evaluations in evolved UTRA," in *Proc. IEEE Tenth International Conference on Communication systems ICCS 2006*, Singapore, Oct. 2006, pp. 1–6.

[250] H. Lei., L. Zhang, X. Zhang, and D. Yang, "System level evaluation of 3G Long Term Evolution," in *Proc. IEEE Eighteenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC*, Athens, Sep. 2007, pp. 1–5.

[251] J. J. Sanchez, D. Morales-Jimenez, G. Gomez, and J. T. Enbrambasaguas, "Physical layer performance of long term evolution cellular technology," in *IST Sixteenth Mobile and Wireless Communications Summit*, Budapest, Jul. 2007, pp. 1–5.

[252] S. Choi, K. Jun, Y. Shin, S. Kang, and B. Choi, "MAC scheduling scheme for VoIP traffic service in 3G LTE," in *Proc. IEEE Sixty-Sixth Vehicular Technology Conference VTC 2007-Fall*, Baltimore, MD, Sep.–Oct. 2007, pp. 1441–1445.

[253] T. Chahed, A.-F. Canton, and S.-E. Elayoubi, "End-to-end TCP performance in W-CDMA / UMTS," in *Proc. IEEE International Conference on Communications ICC 2003*, May 2003, vol. 1, pp. 71–75.

[254] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "UE radio access capabilities (Release 8)," *3GPP TS 25.306 v8.6.0 (2009-03)*, 2009.

[255] D. Jiang, H. Wang, E. Malkamäki, and E. Tuomaala, "Principle and performance of semi-persistent scheduling for VoIP in LTE system," in *Proc. IEEE International Conference on Wireless Communications, Networking and Mobile Computing WiCom 2007*, Shanghai, Sep. 2007, pp. 2861–2864.

[256] N. Miki, Y. Kishiyama, K. Higuchi, and M. Sawahashi, "Investigation on optimum coding and multiplexing schemes for L1/L2 control signals in OFDM based evolved UTRA downlink," in *Proc. IEEE Eighteenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2007*, Athens, Sep. 2007, pp. 1–6.

[257] K. Brüninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link performance models for system level simulations of broadband radio access systems," in *Proc. IEEE Sixteenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2005*, Berlin, Sep. 2005, vol. 4, pp. 2306–2311.

[258] D. S. Baum, J. Hansen, and J. Salo, "An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM)," in *Proc. IEEE Sixty-First Vehicular Technology Conference, VTC 2005-Spring*, May-Jun. 2005, vol. 5, pp. 3132–3136.

[259] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Physical layer aspects of UTRA high speed downlink packet access (Release 4)," *3GPP TR 25.848 v4.0.0 (2001-03)*, 2001.

[260] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Feasibility study for enhanced uplink for UTRA FDD (Release 6)," *3GPP TR 25.896 v6.0.0 (2004-03)*, 2004.

[261] H. Holma, M. Kuusela, E. Malkamäki, and K. Ranta-aho, "VoIP over HSPA with 3GPP Release 7," in *IEEE Seventeenth International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2006*, Helsinki, Sep. 2006, pp. 1–5.

[262] E. Dahlman, H. Ekstrom, A. Furuskär, Y. Jading, J. Karlsson, M. Lundevall, and S. Parkvall, "The 3G long-term evolution - radio interface concepts and performance evaluation," in *Proc. IEEE Sixty-Third Vehicular Technology Conference VTC 2006-Spring*, Melbourne, Vic., May 2006, vol. 1, pp. 137–141.

HELSINKI UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF SIGNAL PROCESSING AND ACOUSTICS
REPORT SERIES

ISBN 978-952-60-3067-8 (Printed)
ISBN 978-952-60-3068-5 (Electronic)
ISSN 1797-4267