Second edition

# 3G Evolution

## HSPA and LTE for Mobile Broadband

Erik Dahlman
Stefan Parkvall
Johan Sköld
Per Beming

# 3G Evolution

## HSPA and LTE for Mobile Broadband

Second edition

Erik Dahlman, Stefan Parkvall, Johan Sköld and Per Beming

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
Linacre House, Jordan Hill, Oxford, OX2 8DP
30 Corporate Drive, Burlington, MA 01803

First edition 2007
Second edition 2008



**British Library Cataloguing in Publication Data**
3G evolution : HSPA and LTE for mobile broadband. – 2nd ed.
1. Broadband communication systems – Standards 2. Mobile communication systems – Standards 3. Cellular telephone systems – Standards
I. Dahlman, Erik
621.3′8546

**Library of Congress Control Number:** 2008931278

ISBN: 978-0-12-374538-5

For information on all Academic Press publications
visit our website at elsevierdirect.com

Typeset by Charon Tec Ltd., A Macmillan Company. (www.macmillansolutions.com)

Printed and bound in Great Britain by MPG Books Ltd, Bodmin, Cornwall

08 09 10 11 11 10 9 8 7 6 5 4 3 2 1

# Contents

# List of Figures

# List of Tables

# Preface

During the past years, there has been a quickly rising interest in radio access technologies for providing mobile as well as nomadic and fixed services for voice, video, and data. The difference in design, implementation, and use between telecom and datacom technologies is also getting more blurred. One example is cellular technologies from the telecom world being used for broadband data and wireless LAN from the datacom world being used for voice over IP.

Today, the most widespread radio access technology for mobile communication is digital cellular, with the number of users passing 3 billion by 2007, which is almost half of the world's population. It has emerged from early deployments of an expensive voice service for a few car-borne users, to today's widespread use of third generation mobile-communication devices that provide a range of mobile services and often include camera, MP3 player, and PDA functions. With this widespread use and increasing interest in 3G, a continuing evolution ahead is foreseen.

This book describes the evolution of 3G digital cellular into an advanced broadband mobile access. The focus of this book is on the evolution of the 3G mobile communication as developed in the 3GPP (*Third Generation Partnership Project*) standardization, looking at the radio access and access network evolution.

This book is divided into five parts. Part I gives the background to 3G and its evolution, looking also at the different standards bodies and organizations involved in the process of defining 3G. It is followed by a discussion of the reasons and driving forces behind the 3G evolution. Part II gives a deeper insight into some of the technologies that are included, or are expected to be included as part of the 3G evolution. Because of its generic nature, Part II can be used as a background not only for the evolution steps taken in 3GPP as described in this book, but also for readers that want to understand the technology behind other systems, such as WiMAX and CDMA2000.

Part III describes the evolution of 3G WCDMA into *High Speed Packet Access* (HSPA). It gives an overview of the key features of HSPA and its continued evolution in the context of the technologies from Part II. Following this, the different uplink and downlink components are outlined and finally more detailed descriptions of how they work together are given.

Part IV introduces the *Long Term Evolution* (LTE) and *System Architecture Evolution* (SAE). As a start, the agreed requirements and objectives for LTE are described. This is followed by an introductory technical overview of LTE, where the most important technology components are introduced, also here, based on the generic technologies given in Part II. As a second step, a more detailed description of the protocol structure is given, with further details on the uplink and downlink transmission schemes and procedures, access procedures and flexible bandwidth operation. The system architecture evolution, applicable to both LTE and HSPA, is given with details of radio access network and core network. The ongoing work on LTE-Advanced is also presented.

Finally in Part V, an assessment is made on the 3G evolution. An evaluation of the performance puts the 3G evolution tracks in relation to the targets set in 3GPP. Through an overview of similar technologies developed in other standards bodies, it will be clear that the technologies adopted for the evolution in 3GPP are implemented in many other systems as well. Finally, looking into the future, it will be seen that the 3G evolution does not stop with the HSPA Evolution and LTE.

# Acknowledgements

We thank all our colleagues at Ericsson for assisting in this project by helping with contributions to the book, giving suggestions and comments on the contents, and taking part in the huge team effort of developing HSPA and LTE.

The standardization process for 3G evolution involves people from all parts of the world, and we acknowledge the efforts of our colleagues in the wireless industry in general and in 3GPP RAN in particular. Without their work and contributions to the standardization, this book would not have been possible.

Finally, we are immensely grateful to our families for bearing with us and supporting us during the long process of writing this book.

# List of Acronyms

| | |
|---|---|
| 3GPP | Third Generation Partnership Project |
| AAS | Adaptive Antenna System |
| ACK | Acknowledgement (in ARQ protocols) |
| ACK-CH | Acknowledgement Channel (for WiMAX) |
| ACLR | Adjacent Channel Leakage Ratio |
| ACS | Adjacent Channel Selectivity |
| ACIR | Adjacent Channel Interference Ratio |
| ACTS | Advanced Communications Technology and Services |
| AM | Acknowledged Mode (RLC configuration) |
| AMC | Adaptive Modulation and Coding |
| AMPR | Additional Maximum Power Reduction |
| AMPS | Advanced Mobile Phone System |
| AMR-WB | Adaptive MultiRate-WideBand |
| AP | Access Point |
| ARIB | Association of Radio Industries and Businesses |
| ARQ | Automatic Repeat-reQuest |
| ATDMA | Advanced Time Division Mobile Access |
| ATIS | Alliance for Telecommunications Industry Solutions |
| AWGN | Additive White Gaussian Noise |
| | |
| BCCH | Broadcast Control Channel |
| BCH | Broadcast Channel |
| BE | Best Effort Service |
| BER | Bit-Error Rate |
| BLER | Block-Error Rate |
| BM-SC | Broadcast Multicast Service Center |
| BPSK | Binary Phase-Shift Keying |
| BS | Base Station |
| BSC | Base Station Controller |
| BTC | Block Turbo Code |
| BTS | Base Transceiver Station |
| | |
| CC | Convolutional Code |
| CCCH | Common Control Channel |
| CCE | Control Channel Element |
| CCSA | China Communications Standards Association |

| | |
|---|---|
| CDD | Cyclic-Delay Diversity |
| CDF | Cumulative Density Function |
| CDM | Code-Division Multiplexing |
| CDMA | Code Division Multiple Access |
| CEPT | European Conference of Postal and Telecommunications Administrations |
| CN | Core Network |
| CODIT | Code-Division Test bed |
| CP | Cyclic Prefix |
| CPC | Continuous Packet Connectivity |
| CPICH | Common Pilot Channel |
| CQI | Channel-Quality Indicator |
| CQICH | Channel Quality Indication Channel (for WiMAX) |
| CRC | Cyclic Redundancy Check |
| C-RNTI | Cell Radio-Network Temporary Identifier |
| CS | Circuit Switched |
| CTC | Convolutional Turbo Code |
| CW | Continuous Wave |
| | |
| DCCH | Dedicated Control Channel |
| DCH | Dedicated Channel |
| DCI | Downlink Control Information |
| DFE | Decision-Feedback Equalization |
| DFT | Discrete Fourier Transform |
| DFTS-OFDM | DFT-spread OFDM, see also SC-FDMA |
| DL | Downlink |
| DL-SCH | Downlink Shared Channel |
| DPCCH | Dedicated Physical Control Channel |
| DPCH | Dedicated Physical Channel |
| DPDCH | Dedicated Physical Data Channel |
| DRS | Demodulation Reference Signal |
| DRX | Discontinuous Reception |
| DTCH | Dedicated Traffic Channel |
| DTX | Discontinuous Transmission |
| D-TxAA | Dual Transmit-Diversity Adaptive Array |
| DwPTS | The downlink part of the special subframe (for TDD operation). |
| | |
| E-AGCH | E-DCH Absolute Grant Channel |
| E-DCH | Enhanced Dedicated Channel |
| EDGE | Enhanced Data rates for GSM Evolution and Enhanced Data rates for Global Evolution |

| | |
|---|---|
| E-DPCCH | E-DCH Dedicated Physical Control Channel |
| E-DPDCH | E-DCH Dedicated Physical Data Channel |
| E-HICH | E-DCH Hybrid ARQ Indicator Channel |
| eNodeB | E-UTRAN NodeB |
| EPC | Evolved Packet Core |
| E-RGCH | E-DCH Relative Grant Channel |
| ErtPS | Extended Real-Time Polling Service |
| E-TFC | E-DCH Transport Format Combination |
| E-TFCI | E-DCH Transport Format Combination Index |
| ETSI | European Telecommunications Standards Institute |
| E-UTRA | Evolved UTRA |
| E-UTRAN | Evolved UTRAN |
| EV-DO | Evolution-Data Optimized (of CDMA2000 1x) |
| EV-DV | Evolution-Data and Voice (of CDMA2000 1x) |
| EVM | Error Vector Magnitude |
| FACH | Forward Access Channel |
| FBSS | Fast Base-Station Switching |
| FCC | Federal Communications Commission |
| FCH | Frame Control Header (for WiMAX) |
| FDD | Frequency Division Duplex |
| FDM | Frequency-Division Multiplex |
| FDMA | Frequency-Division Multiple Access |
| F-DPCH | Fractional DPCH |
| FEC | Forward Error Correction |
| FFT | Fast Fourier Transform |
| FIR | Finite Impulse Response |
| F-OSICH | Forward link Other Sector Indication Channel (for IEEE 802.20) |
| FPLMTS | Future Public Land Mobile Telecommunications Systems |
| FRAMES | Future Radio Wideband Multiple Access Systems |
| FTP | File Transfer Protocol |
| FUSC | Fully Used Subcarriers (for WiMAX) |
| FSTD | Frequency Shift Transmit Diversity |
| GERAN | GSM/EDGE Radio Access Network |
| GGSN | Gateway GPRS Support Node |
| GP | Guard Period (for TDD operation) |
| GPRS | General Packet Radio Services |
| GPS | Global Positioning System |
| G-RAKE | Generalized RAKE |
| GSM | Global System for Mobile communications |

| | |
|---|---|
| HARQ | Hybrid ARQ |
| HC-SDMA | High Capacity Spatial Division Multiple Access |
| H-FDD | Half-duplex FDD |
| HHO | Hard Handover |
| HLR | Home Location Register |
| HRPD | High Rate Packet Data |
| HSDPA | High-Speed Downlink Packet Access |
| HS-DPCCH | High-Speed Dedicated Physical Control Channel |
| HS-DSCH | High-Speed Downlink Shared Channel |
| HSPA | High-Speed Packet Access |
| HS-PDSCH | High-Speed Physical Downlink Shared Channel |
| HSS | Home Subscriber Server |
| HS-SCCH | High-Speed Shared Control Channel |
| HSUPA | High-Speed Uplink Packet Access |
| ICIC | Inter-Cell Interference Coordination |
| ICS | In-Channel Selectivity |
| IDFT | Inverse DFT |
| IEEE | Institute of Electrical and Electronics Engineers |
| IFDMA | Interleaved FDMA |
| IFFT | Inverse Fast Fourier Transform |
| IMS | IP Multimedia Subsystem |
| IMT-2000 | International Mobile Telecommunications 2000 |
| IP | Internet Protocol |
| IPsec | Internet Protocol security |
| IPv4 | IP version 4 |
| IPv6 | IP version 6 |
| IR | Incremental Redundancy |
| IRC | Interference Rejection Combining |
| ISDN | Integrated Services Digital Network |
| ITU | International Telecommunications Union |
| ITU-R | International Telecommunications Union-Radiocommunications Sector |
| Iu | The interface used for communication between the RNC and the core network. |
| Iu_cs | The interface used for communication between the RNC and the GSM/WCDMA circuit switched core network. |
| Iu_ps | The interface used for communication between the RNC and the GSM/WCDMA packet switched core network. |
| Iub | The interface used for communication between the NodeB and the RNC. |
| Iur | The interface used for communication between different RNCs. |

| | |
|---|---|
| J-TACS | Japanese Total Access Communication System |
| | |
| LAN | Local Area Network |
| LCID | Logical Channel Index |
| LDPC | Low-Density Parity Check Code |
| LMMSE | Linear Minimum Mean Square Error |
| LTE | Long-Term Evolution |
| | |
| MAC | Medium Access Control |
| MAN | Metropolitan Area Network |
| MAP | Map message (for WiMAX) |
| MBFDD | Mobile Broadband FDD (for IEEE 802.20) |
| MBMS | Multimedia Broadcast/Multicast Service |
| MBS | Multicast and Broadcast Service |
| MBSFN | Multicast-Broadcast Single Frequency Network |
| MBTDD | Mobile Broadband TDD (for IEEE 802.20) |
| MBWA | Mobile Broadband Wireless Access |
| MCCH | MBMS Control Channel |
| MC | Multi-Carrier |
| MCE | MBMS Coordination Entity |
| MCH | Multicast Channel |
| MCS | Modulation and Coding Scheme |
| MDHO | Macro-Diversity Handover |
| MIB | Master Information Block |
| MICH | MBMS Indicator Channel |
| MIMO | Multiple-Input Multiple-Output |
| ML | Maximum Likelihood |
| MLD | Maximum Likelihood Detection |
| MLSE | Maximum-Likelihood Sequence Estimation |
| MME | Mobility Management Entity |
| MMS | Multimedia Messaging Service |
| MMSE | Minimum Mean Square Error |
| MPR | Maximum Power Reduction |
| MRC | Maximum Ratio Combining |
| MSC | Mobile Switching Center |
| MSCH | MBMS Scheduling Channel |
| MTCH | MBMS Traffic Channel |
| | |
| NAK | Negative Acknowledgement (in ARQ protocols) |
| NAS | Non-Access Stratum (a functional layer between the core network and the terminal that supports signaling and user data transfer) |

| | |
|---|---|
| *NMT* | *Nordisk MobilTelefon (Nordic Mobile Telephony)* |
| NodeB | NodeB, a logical node handling transmission/reception in multiple cells. Commonly, but not necessarily, corresponding to a base station. |
| nrtPS | Non-Real-Time Polling Service |
| OFDM | Orthogonal Frequency-Division Multiplexing |
| OFDMA | Orthogonal Frequency-Division Multiple Access |
| OOB | Out-Of-Band (emissions) |
| OOK | On–Off Keying |
| OVSF | Orthogonal Variable Spreading Factor |
| PAN | Personal Area Network |
| PAPR | Peak-to-Average Power Ratio |
| PAR | Peak-to-Average Ratio (same as PAPR) |
| PARC | Per-Antenna Rate Control |
| PBCH | Physical Broadcast Channel |
| PCCH | Paging Control Channel |
| PCFICH | Physical Control Format Indicator Channel |
| PCG | Project Coordination Group (in 3GPP) |
| PCH | Paging Channel |
| PCI | Pre-coding Control Indication |
| PCS | Personal Communications Systems |
| PDC | Personal Digital Cellular |
| PDCCH | Physical Downlink Control Channel |
| PDCP | Packet Data Convergence Protocol |
| PDSCH | Physical Downlink Shared Channel |
| PDSN | Packet Data Serving Node |
| PDN | Packet Data Network |
| PDU | Protocol Data Unit |
| PF | Proportional Fair (a type of scheduler) |
| PHICH | Physical Hybrid-ARQ Indicator Channel |
| PHY | Physical layer |
| PHS | Personal Handy-phone System |
| PMCH | Physical Multicast Channel |
| PMI | Precoding-Matrix Indicator |
| PoC | Push to Talk over Cellular |
| PRACH | Physical Random Access Channel |
| PRB | Physical Resource Block |
| PS | Packet Switched |
| PSK | Phase Shift Keying |

| | |
|---|---|
| PSS | Primary Synchronization Signal |
| PSTN | Public Switched Telephone Networks |
| PUCCH | Physical Uplink Control Channel |
| PUSC | Partially Used Subcarriers (for WiMAX) |
| PUSCH | Physical Uplink Shared Channel |
| | |
| QAM | Quadrature Amplitude Modulation |
| QoS | Quality-of-Service |
| QPP | Quadrature Permutation Polynomial |
| QPSK | Quadrature Phase-Shift Keying |
| | |
| RAB | Radio Access Bearer |
| RACE | Research and development in Advanced Communications in Europe |
| RACH | Random Access Channel |
| RAN | Radio Access Network |
| RA-RNTI | Random Access RNTI |
| RAT | Radio Access Technology |
| RB | Resource Block |
| RBS | Radio Base Station |
| RF | Radio Frequency |
| RI | Rank Indicator |
| RIT | Radio Interface Technology |
| RLC | Radio Link Protocol |
| RNC | Radio Network Controller |
| RNTI | Radio-Network Temporary Identifier |
| ROHC | Robust Header Compression |
| RR | Round-Robin (a type of scheduler) |
| RRC | Radio Resource Control |
| RRM | Radio Resource Management |
| RS | Reference Symbol |
| RSN | Retransmission Sequence Number |
| RSPC | IMT-2000 radio interface specifications |
| RTP | Real Time Protocol |
| rtPS | Real-Time Polling Service |
| RTWP | Received Total Wideband Power |
| RV | Redundancy Version |
| | |
| S1 | The interface between eNodeB and the Evolved Packet Core. |
| SA | System Aspects |
| SAE | System Architecture Evolution |

| | |
|---|---|
| S-CCPCH | Secondary Common Control Physical Channel |
| SC-FDMA | Single-Carrier FDMA |
| SDMA | Spatial Division Multiple Access |
| SDO | Standards Developing Organization |
| SDU | Service Data Unit |
| SEM | Spectrum Emissions Mask |
| SF | Spreading Factor |
| SFBC | Space-Frequency Block Coding |
| SFN | Single-Frequency Network or System Frame Number (in 3GPP) |
| SFTD | Space–Frequency Time Diversity |
| SGSN | Serving GPRS Support Node |
| SI | System Information message |
| SIB | System Information Block |
| SIC | Successive Interference Combining |
| SIM | Subscriber Identity Module |
| SINR | Signal-to-Interference-and-Noise Ratio |
| SIR | Signal-to-interference ratio |
| SMS | Short Message Service |
| SNR | Signal-to-noise ratio |
| SOHO | Soft Handover |
| SR | Scheduling Request |
| SRNS | Serving Radio Network Subsystem |
| SRS | Sounding Reference Signal |
| SSS | Secondary Synchronization Signal |
| STBC | Space–Time Block Coding |
| STC | Space–Time Coding |
| STTD | Space-Time Transmit Diversity |
| TACS | Total Access Communication System |
| TCP | Transmission Control Protocol |
| TC-RNTI | Temporary C-RNTI |
| TD-CDMA | Time Division-Code Division Multiple Access |
| TDD | Time Division Duplex |
| TDM | Time Division Multiplexing |
| TDMA | Time Division Multiple Access |
| TD-SCDMA | Time Division-Synchronous Code Division Multiple Access |
| TF | Transport Format |
| TFC | Transport Format Combination |
| TFCI | Transport Format Combination Index |
| TIA | Telecommunications Industry Association |
| TM | Transparent Mode (RLC configuration) |

| | |
|---|---|
| TR | Technical Report |
| TrCH | Transport Channel |
| TS | Technical Specification |
| TSG | Technical Specification Group |
| TSN | Transmission Sequence Number |
| TTA | Telecommunications Technology Association |
| TTC | Telecommunications Technology Committee |
| TTI | Transmission Time Interval |
| | |
| UCI | Uplink Control Information |
| UE | User Equipment, the 3GPP name for the mobile terminal |
| UGS | Unsolicited Grant Service |
| UL | Uplink |
| UL-SCH | Uplink Shared Channel |
| UM | Unacknowledged Mode (RLC configuration) |
| UMB | Ultra Mobile Broadband |
| UMTS | Universal Mobile Telecommunications System |
| UpPTS | The uplink part of the special subframe (for TDD operation). |
| USIM | UMTS SIM |
| US-TDMA | US Time Division Multiple Access standard |
| UTRA | Universal Terrestrial Radio Access |
| UTRAN | Universal Terrestrial Radio Access Network |
| | |
| VRB | Virtual Resource Block |
| | |
| WAN | Wide Area Network |
| WMAN | Wireless Metropolitan Area Network |
| WARC | World Administrative Radio Congress |
| WCDMA | Wideband Code Division Multiple Access |
| WG | Working Group |
| WiMAX | Worldwide Interoperability for Microwave Access |
| WLAN | Wireless Local Area Network |
| VoIP | Voice-over-IP |
| WP8F | Working Party 8F |
| WRC | World Radiocommunication Conference |
| | |
| X2 | The interface between eNodeBs. |
| | |
| ZC | Zadoff-Chu |
| ZF | Zero Forcing |
| ZTCC | Zero Tailed Convolutional Code |

# Part I
# Introduction

# 1 Background of 3G evolution

From the first experiments with radio communication by Guglielmo Marconi in the 1890s, the road to truly mobile radio communication has been quite long. To understand the complex 3G mobile-communication systems of today, it is also important to understand where they came from and how cellular systems have evolved from an expensive technology for a few selected individuals to today's global mobile-communication systems used by almost half of the world's population. Developing mobile technologies has also changed, from being a national or regional concern, to becoming a very complex task undertaken by global standards-developing organizations such as the *Third Generation Partnership Project* (3GPP) and involving thousands of people.

## 1.1 History and background of 3G

The cellular technologies specified by 3GPP are the most widely deployed in the world, with more than 2.6 billion users in 2008. The latest step being studied and developed in 3GPP is an evolution of 3G into an evolved radio access referred to as the *Long-Term Evolution* (LTE) and an evolved packet access core network in the *System Architecture Evolution* (SAE). By 2009–2010, LTE and SAE are expected to be first deployed.

Looking back to when it all it started, it begun several decades ago with early deployments of analog cellular services.

### *1.1.1 Before 3G*

The US *Federal Communications Commission* (FCC) approved the first commercial car-borne telephony service in 1946, operated by AT&T. In 1947 AT&T also introduced the cellular concept of reusing radio frequencies, which became fundamental to all subsequent mobile-communication systems. Commercial mobile telephony continued to be car-borne for many years because of bulky and power-hungry equipment. In spite of the limitations of the service, there were systems deployed in many countries during the 1950s and 1960s, but the users counted only in thousands at the most.

These first steps on the road of mobile communication were taken by the monopoly telephone administrations and wire-line operators. The big uptake of subscribers and usage came when mobile communication became an international concern and the industry was invited into the process. The first international mobile communication system was the analog NMT system (*Nordic Mobile Telephony*) which was introduced in the Nordic countries in 1981, at the same time as analog AMPS (*Advanced Mobile Phone Service*) was introduced in North America. Other analog cellular technologies deployed worldwide were TACS and J-TACS. They all had in common that equipment was still bulky, mainly car-borne, and voice quality was often inconsistent, with 'cross-talk' between users being a common problem.

With an international system such as NMT came the concept of 'roaming,' giving a service also for users traveling outside the area of their 'home' operator. This also gave a larger market for the mobile phones, attracting more companies into the mobile communication business.

The analog cellular systems supported 'plain old telephony services,' that is voice with some related supplementary services. With the advent of digital communication during the 1980s, the opportunity to develop a second generation of mobile-communication standards and systems, based on digital technology, surfaced. With digital technology came an opportunity to increase the capacity of the systems, to give a more consistent quality of the service, and to develop much more attractive truly mobile devices.

In Europe, the telecommunication administrations in CEPT[1] initiated the GSM project to develop a pan-European mobile-telephony system. The GSM activities were in 1989 continued within the newly formed *European Telecommunication Standards Institute* (ETSI). After evaluations of TDMA, CDMA, and FDMA-based proposals in the mid-1980s, the final GSM standard was built on TDMA. Development of a digital cellular standard was simultaneously done by TIA in the USA resulting in the TDMA-based IS-54 standard, later simply referred to as US-TDMA. A somewhat later development of a CDMA standard called IS-95 was completed by TIA in 1993. In Japan, a second-generation TDMA standard was also developed, usually referred to as PDC.

[1]The *European Conference of Postal and Telecommunications Administrations* (CEPT) consist of the telecom administrations from 48 countries.

All these standards were 'narrowband' in the sense that they targeted 'low-bandwidth' services such as voice. With the second-generation digital mobile communications came also the opportunity to provide data services over the mobile-communication networks. The primary data services introduced in 2G were text messaging (SMS) and circuit-switched data services enabling e-mail and other data applications. The peak data rates in 2G were initially 9.6 kbps. Higher data rates were introduced later in evolved 2G systems by assigning multiple time slots to a user and by modified coding schemes.

Packet data over cellular systems became a reality during the second half of the 1990s, with *General Packet Radio Services* (GPRS) introduced in GSM and packet data also added to other cellular technologies such as the Japanese PDC standard. These technologies are often referred to as 2.5G. The success of the wireless data service iMode in Japan gave a very clear indication of the potential for applications over packet data in mobile systems, in spite of the fairly low data rates supported at the time.

With the advent of 3G and the higher-bandwidth radio interface of UTRA (*Universal Terrestrial Radio Access*) came possibilities for a range of new services that were only hinted at with 2G and 2.5G. The 3G radio access development is today handled in 3GPP. However, the initial steps for 3G were taken in the early 1990s, long before 3GPP was formed.

What also set the stage for 3G was the internationalization of cellular standardization. GSM was a pan-European project, but quickly attracted worldwide interest when the GSM standard was deployed in a number of countries outside Europe. There are today only three countries worldwide where GSM is not deployed. A global standard gains in economy of scale, since the market for products becomes larger. This has driven a much tighter international cooperation around 3G cellular technologies than for the earlier generations.

### 1.1.2 Early 3G discussions

Work on a third-generation mobile communication started in ITU (*International Telecommunication Union*) in the 1980s. The radio communication sector ITU-R issued a first recommendation defining *Future Public Land Mobile Telecommunications Systems* (FPLMTS) in 1990, later revised in 1997 [48]. The name for 3G within ITU had by then changed from FPLMTS to IMT-2000. The World Administrative Radio Congress WARC-92 identified 230MHz of spectrum for IMT-2000 on a worldwide basis. Of these 230 MHz, 2 × 60MHz were identified as paired spectrum for FDD (*Frequency Division Duplex*) and 35 MHz as unpaired spectrum for TDD (*Time Division Duplex*), both for terrestrial use. Some spectrum was also set aside for satellite services. With that, the stage was set to specify IMT-2000.

Task Group 8/1 within ITU-R developed a range of recommendations for IMT-2000, defining a framework for services, network architectures, radio interface requirements, spectrum considerations, and evaluation methodology. Both a terrestrial and a satellite component were defined.

Task Group 8/1 defined the process for evaluating IMT-2000 technologies in ITU-R recommendation M.1225 [45]. The evaluation criteria set the target data rates for the 3G circuit-switched and packet-switched data services:

- Up to 2 Mbps in an indoor environment.
- Up to 144 kbps in a pedestrian environment.
- Up to 64 kbps in a vehicular environment.

These numbers became the benchmark that all 3G technologies were compared with. However, already today, data rates well beyond 2 Mbps can be seen in deployed 3G systems.

### 1.1.3 Research on 3G

In parallel with the widespread deployment and evolution of 2G mobile-communication systems during the 1990s, substantial efforts were put into 3G research activities. In Europe the partially EU-funded project *Research into Advanced Communications in Europe* (RACE) carried out initial 3G research in its first phase. 3G in Europe was named *Universal Mobile Telecommunications Services* (UMTS). In the second phase of RACE, the CODIT project (Code Division Test bed) and the ATDMA project (*Advanced TDMA Mobile Access*) further developed 3G concepts based on *Wideband CDMA* (WCDMA) and Wideband TDMA technologies. The next phase of related European research was *Advanced Communication Technologies and Services* (ACTS), which included the UMTS-related project *Future Radio Wideband Multiple Access System* (FRAMES). The FRAMES project resulted in a multiple access concept that included both Wideband CDMA and Wideband TDMA components.

At the same time parallel 3G activities were going on in other parts of the world. In Japan, the *Association of Radio Industries and Businesses* (ARIB) was in the process of defining a 3G wireless communication technology based on Wideband CDMA. Also in the US, a Wideband CDMA concept called WIMS was developed within the T1.P1[2] committee. Also Korea started work on Wideband CDMA at this time.

The FRAMES concept was submitted to the standardization activities for 3G in ETSI,[3] where other multiple access proposals were also introduced by the industry, including the Wideband CDMA concept from the ARIB standardization in Japan. The ETSI proposals were merged into five concept groups, which also meant that the Wideband CDMA proposals from Europe and Japan were merged.

### 1.1.4 3G standardization starts

The outcome of the ETSI process in early 1998 was the selection of *Wideband CDMA* (WCDMA) as the technology for UMTS in the paired spectrum (FDD) and TD-CDMA (*Time Division CDMA*) for the unpaired spectrum (TDD). There was also a decision to harmonize the parameters between the FDD and the TDD components.

The standardization of WCDMA went on in parallel in ETSI and ARIB until the end of 1998 when the *Third Generation Partnership Project* (3GPP) was formed by standards-developing organizations from all regions of the world. This solved the problem of trying to maintain parallel development of aligned specifications in multiple regions. The present organizational partners of 3GPP are ARIB (Japan), CCSA (China), ETSI (Europe), ATIS (USA), TTA (Korea) and TTC (Japan).

## 1.2 Standardization

### 1.2.1 The standardization process

Setting a standard for mobile communication is not a one-time job, it is an ongoing process. The standardization forums are constantly evolving their standards trying to meet new demands for services and features. The standardization process is different in the different forums, but typically includes the four phases illustrated in Figure 1.1:

1. *Requirements*, where it is decided what is to be achieved by the standard.
2. *Architecture*, where the main building blocks and interfaces are decided.

---

[2] The T1.P1 committee was part of T1 which presently has joined the ATIS standardization organization.
[3] The TDMA part of the FRAMES project was also fed into 2G standardization as the evolution of GSM into EDGE (Enhanced Data rates for GSM Evolution).

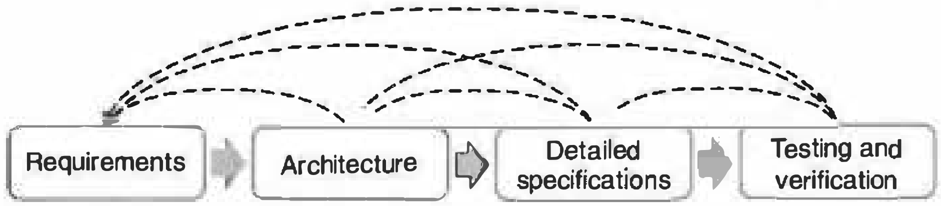


**Figure 1.1** *The standardization phases and iterative process.*

3. *Detailed specifications*, where every interface is specified in detail.
4. *Testing and verification*, where the interface specifications are proven to work with real-life equipment.

These phases are overlapping and iterative. As an example, requirements can be added, changed, or dropped during the later phases if the technical solutions call for it. Likewise, the technical solution in the detailed specifications can change due to problems found in the testing and verification phase.

Standardization starts with the *requirements* phase, where the standards body decides what should be achieved with the standard. This phase is usually relatively short.

In the *architecture* phase, the standards body decides about the architecture, i.e. the principles of how to meet the requirements. The architecture phase includes decisions about reference points and interfaces to be standardized. This phase is usually quite long and may change the requirements.

After the architecture phase, the *detailed specification* phase starts. It is in this phase the details for each of the identified interfaces are specified. During the detailed specification of the interfaces, the standards body may find that it has to change decisions done either in the architecture or even in the requirements phases.

Finally, the *testing and verification* phase starts. It is usually not a part of the actual standardization in the standards bodies, but takes place in parallel through testing by vendors and interoperability testing between vendors. This phase is the final proof of the standard. During the testing and verification phase, errors in the standard may still be found and those errors may change decisions in the detailed standard. Albeit not common, changes may need to be done also to the architecture or the requirements. To verify the standard, products are needed. Hence, the implementation of the products starts after (or during) the detailed specification phase. The testing and verification phase ends when there are

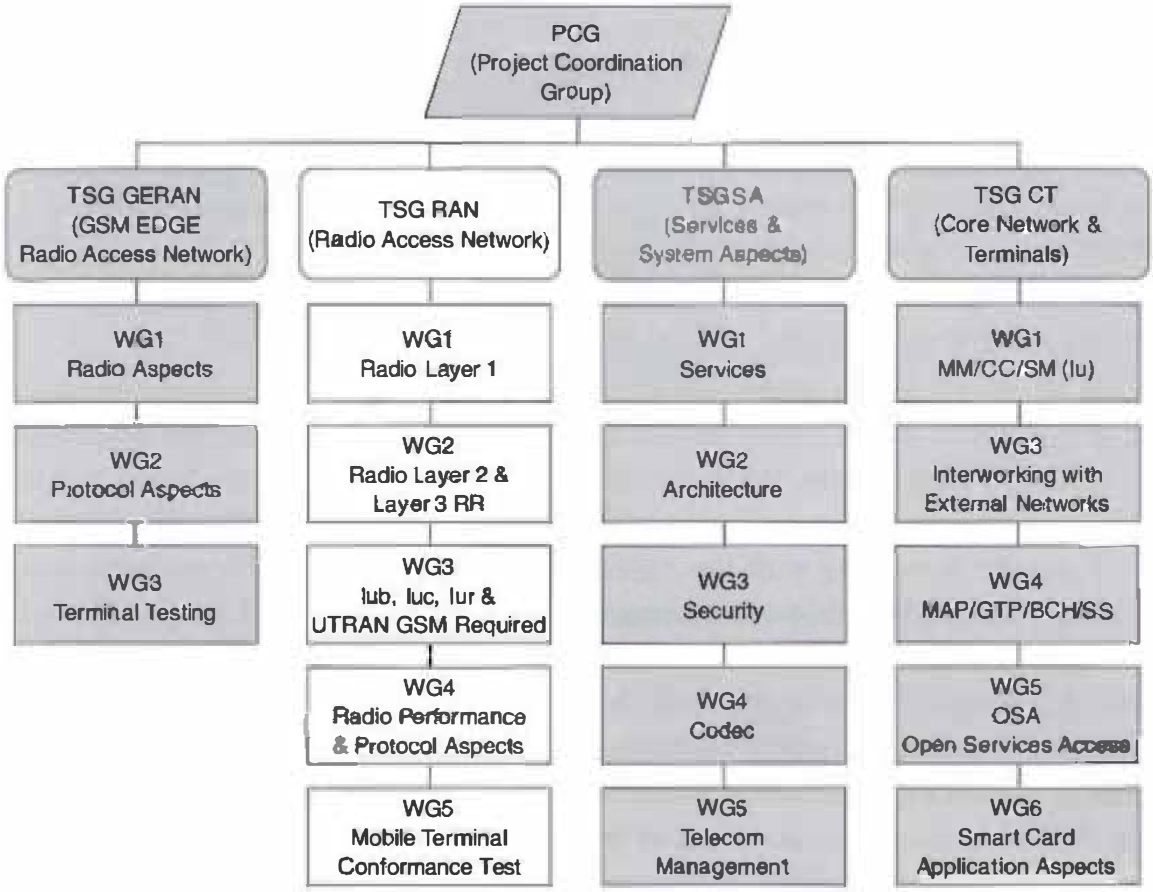


**Figure 1.2** *3GPP organization.*

stable test specifications that can be used to verify that the equipment is fulfilling the standard.

Normally, it takes one to two years from the time when the standard is completed until commercial products are out on the market. However, if the standard is built from scratch, it may take longer time since there are no stable components to build from.

## 1.2.2 3GPP

The *Third-Generation Partnership Project* (3GPP) is the standards-developing body that specifies the 3G UTRA and GSM systems. 3GPP is a partnership project formed by the standards bodies ETSI, ARIB, TTC, TTA, CCSA and ATIS. 3GPP consists of several Technical Specifications Groups (TSGs ), (see Figure 1.2).

A parallel partnership project called 3GPP2 was formed in 1999. It also develops 3G specifications, but for cdma2000, which is the 3G technology developed from the 2G CDMA-based standard IS-95. It is also a global project, and the organizational partners are ARIB, CCSA, TIA, TTA and TTC.

3GPP TSG RAN is the technical specification group that has developed WCDMA, its evolution HSPA, as well as LTE, and is in the forefront of the technology. TSG RAN consists of five working groups (WGs):

1. RAN WG1 dealing with the physical layer specifications.
2. RAN WG2 dealing with the layer 2 and layer 3 radio interface specifications.
3. RAN WG3 dealing with the fixed RAN interfaces, for example interfaces between nodes in the RAN, but also the interface between the RAN and the core network.
4. RAN WG4 dealing with the *radio frequency* (RF) and *radio resource management* (RRM) performance requirements.
5. RAN WG 5 dealing with the terminal conformance testing.

The scope of 3GPP when it was formed in 1998 was to produce global specifications for a 3G mobile system based on an evolved GSM core network, including the WCDMA-based radio access of the UTRA FDD and the TD-CDMA-based radio access of the UTRA TDD mode. The task to maintain and develop the GSM/EDGE specifications was added to 3GPP at a later stage. The UTRA (and GSM/EDGE) specifications are developed, maintained and approved in 3GPP. After approval, the organizational partners transpose them into appropriate deliverables as standards in each region.

In parallel with the initial 3GPP work, a 3G system based on TD-SCDMA was developed in China. TD-SCDMA was eventually merged into Release 4 of the 3GPP specifications as an additional TDD mode.

The work in 3GPP is carried out with relevant ITU recommendations in mind and the result of the work is also submitted to ITU. The organizational partners are obliged to identify regional requirements that may lead to options in the standard. Examples are regional frequency bands and special protection requirements local to a region. The specifications are developed with global roaming and circulation of terminals in mind. This implies that many regional requirements in essence will be global requirements for all terminals, since a roaming terminal has to meet the strictest of all regional requirements. Regional options in the specifications are thus more common for base stations than for terminals.

The specifications of all releases can be updated after each set of TSG meetings, which occur 4 times a year. The 3GPP documents are divided into releases, where

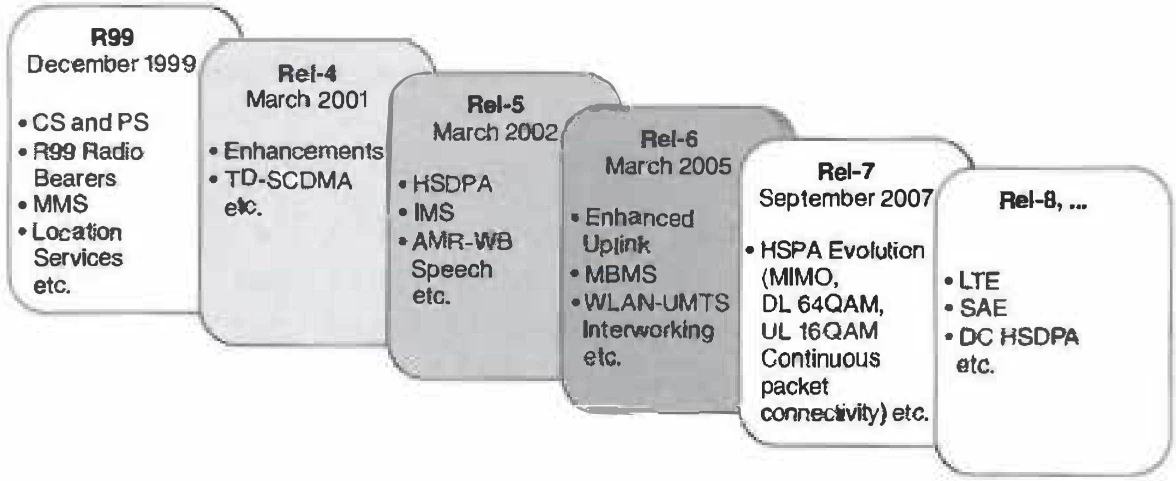


**Figure 1.3** *Releases of 3GPP specifications for UTRA.*

each release has a set of features added compared to the previous release. The features are defined in Work Items agreed and undertaken by the TSGs. The releases up to Release 8 and some main features of those are shown in Figure 1.3. The date shown for each release is the day the content of the release was frozen. For historical reasons, the first release is numbered by the year it was frozen (1999), while the following releases are numbered 4, 5, etc.

For the WCDMA Radio Access developed in TSG RAN, Release 99 contains all features needed to meet the IMT-2000 requirements as defined by ITU. There are circuit-switched voice and video services, and data services over both packet-switched and circuit-switched bearers. The first major addition of radio access features to WCDMA is Release 5 with *High Speed Downlink Packet Access* (HSDPA) and Release 6 with *Enhanced Uplink*. These two are together referred to as HSPA and are described in more detail in Part III of this book. With HSPA, UTRA goes beyond the definition of a 3G mobile system and also encompasses broadband mobile data.

With the inclusion of an Evolved UTRAN (LTE) and the related *System Architecture Evolution* (SAE) in Release 8, further steps are taken in terms of broadband capabilities. The specific solutions chosen for LTE and SAE are described in Part IV of this book.

### 1.2.3 IMT-2000 activities in ITU

The present ITU work on 3G takes place in ITU-R *Working Party 5D*[4] (WP5D), where 3G systems are referred to as IMT-2000. WP5D does not write technical

[4] The work on IMT-2000 was moved from Working Party 8F to Working Party 5D in 2008.

specifications for IMT-2000, but has kept the role of defining IMT-2000, cooperating with the regional standardization bodies and to maintain a set of recommendations for IMT-2000.

The main IMT-2000 recommendation is ITU-R M.1457 [46], which identifies the IMT-2000 *radio interface specifications* (RSPC). The recommendation contains a 'family' of radio interfaces, all included on an equal basis. The family of six terrestrial radio interfaces is illustrated in Figure 1.4, which also shows what *Standards Developing Organizations* (SDO) or Partnership Projects produce the specifications. In addition, there are several IMT-2000 satellite radio interfaces defined, not illustrated in Figure 1.4.

For each radio interface, M.1457 contains an overview of the radio interface, followed by a list of references to the detailed specifications. The actual specifications are maintained by the individual SDOs and M.1457 provides URLs locating the specifications at each SDOs web archive.

With the continuing development of the IMT-2000 radio interfaces, including the evolution of UTRA to Evolved UTRA, the ITU recommendations also need to be updated. ITU-R WP5D continuously revises recommendation M.1457 and at the time of writing it is in its seventh version. Input to the updates is provided by the SDOs and Partnership Projects writing the standards. In the latest revision of ITU-R M.1457, LTE (or E-UTRA) is included in the family through the 3GPP family members for UTRA FDD and TDD, while UMB is included through CDMA2000, as shown in the figure. WiMAX is also included as the sixth family member for IMT-2000.

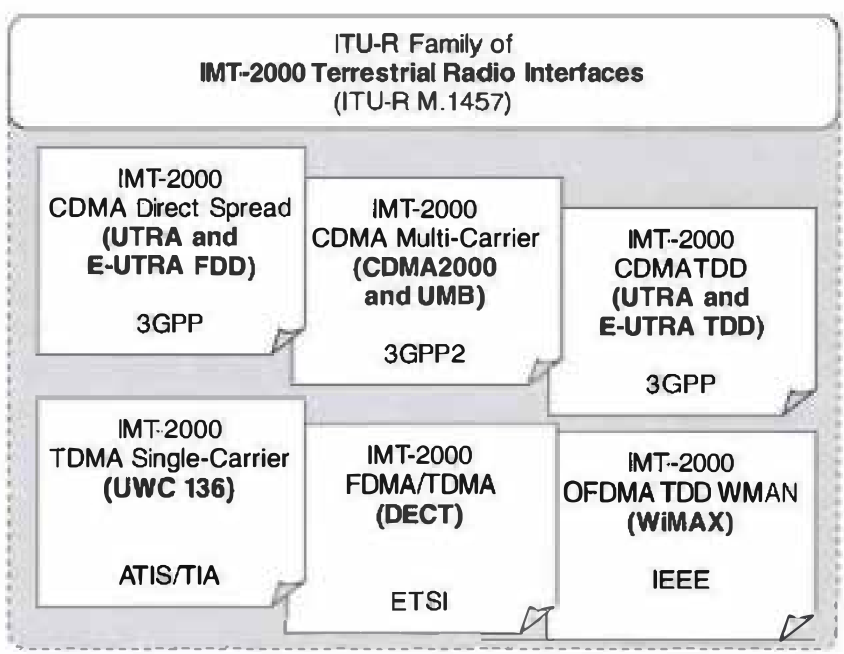


**Figure 1.4** *The definition of IMT-2000 in ITU-R.*

In addition to maintaining the IMT-2000 specifications, the main activity in ITU-R WP5D is the work on systems beyond IMT-2000, named IMT-Advanced. ITU-R has concluded studies for IMT-Advanced of services and technologies, market forecasts, principles for standardization, estimation of spectrum needs, and identification of candidate frequency bands [47]. The spectrum work has involved sharing studies between IMT-Advanced and other technologies in those bands. In March 2008, ITU-R invited the submission of candidate *Radio Interface Technologies* (RIT) in a Circular letter [139]. Submission and evaluation of RITs will be ongoing through 2009 and 2010. The target date for the final ITU-R recommendation for the IMT-Advanced radio interface specifications is February 2011.

## 1.3 Spectrum for 3G and systems beyond 3G

Spectrum for 3G was first identified at the *World Administrative Radio Congress* WARC-92. Resolution 212 [60] identified the bands 1885–2025 and 2110–2200 MHz as intended for use by national administrations that want to implement IMT–2000. Of these 230MHz of 3G spectrum, 2 × 30MHz were intended for the satellite component of IMT–2000 and the rest for the terrestrial component. Parts of the bands were during the 1990s used for deployment of 2G cellular systems, especially in the Americas. The first deployment of 3G in 2001–2002 by Japan and Europe were done in this band allocation, and it is for that reason often referred to as the IMT-2000 'core band.'

Spectrum for IMT–2000 was also identified at the World Radiocommunication Conference WRC-2000 in Resolutions 223 and 224, where it was considered that an additional need for 160 MHz of spectrum for IMT-2000 was forecasted by ITU-R. The identification includes the bands used for 2G mobile systems in 806–960MHz and 1710–1885 MHz, and 'new' 3G spectrum in the bands 2500–2690 MHz. The identification of bands assigned for 2G was also a recognition of the evolution of existing 2G mobile systems into 3G. Additional spectrum was identified at WRC'07 for IMT, encompassing both IMT-2000 and IMT-Advanced. The bands added are 450-470, 698–806, 2300–2400, and 3400–3600 MHz, but the applicability of the bands vary on a regional and national basis.

The somewhat diverging arrangement between regions of the frequency bands assigned to 3G means that there is not a single band that can be used for 3G roaming worldwide. Large efforts have however been put into defining a minimum set of bands that can be used to provide roaming. In this way, multi-band devices can provide efficient worldwide roaming for 3G. Release 8 of the 3GPP specifications includes 14 frequency bands for FDD and 8 for TDD. These are described in more detail in Chapter 20.

The worldwide frequency arrangements for IMT-2000 are outlined in ITU-R recommendation M.1036 [44]. The recommendation also identifies which parts of the spectrum that are paired and which are unpaired. For the paired spectrum, the bands for uplink (mobile transmit) and downlink (base-station transmit) are identified for *Frequency Division Duplex* (FDD) operation. The unpaired bands can for example be used for *Time Division Duplex* (TDD) operation. Note that the band that is most globally deployed for 3G is still 2 GHz.

3GPP first defined UTRA in Release 99 for the 2 GHz bands, with $2 \times 60$ MHz for UTRA FDD and $20 + 15$ MHz of unpaired spectrum for UTRA TDD. A separate definition was also made for the use of UTRA in the US PCS bands at 1900 MHz. The concept of frequency bands with separate and release-independent requirements were defined in Release 5 of the 3GPP specifications. The release-independence implies that a new frequency band added at a later release can be implemented also for earlier releases. All bands are also defined with consideration of what other bands may be deployed in the same region through special coexistence requirements for both base stations and terminals. These tailored requirements enable coexistence between 3G (and 2G) deployments in different bands in the same geographical area and even for co-location of base stations at the same sites using different bands.

# 2 The motives behind the 3G evolution

Before entering the detailed discussion on technologies being used or considered for the evolution of 3G mobile communication, it is important to understand the motivation for this evolution: that is, understanding the underlying driving forces. This chapter will try to highlight some of the driving forces giving the reader an understanding of where the technical requirements and solutions are coming from.

## 2.1 Driving forces

A key factor for success in any business is to understand the forces that will drive the business in the future. This is in particular true for the mobile-communication industry, where the rapid growth of the number of subscribers and the global presence of the technologies have attracted several new players that want to be successful. Both new operators and new vendors try to compete with the existing operators and vendors by adopting new technologies and standards to provide new and existing services better and at a lower cost than earlier systems. The existing operators and vendors will, of course, also follow or drive new technologies to stay ahead of competition. Thus, there is a key driving force in staying competitive or becoming competitive.

From the technical perspective, the development in areas like digital cameras and color displays enables new fancier services than the existing mobile-communication services. To be able to provide those services, the mobile-communication systems need to be upgraded or even replaced by new mobile-communication technologies. Similarly, the technical advancement in digital processors enables new and more powerful systems that not only can provide the new services, but also can provide the existing successful services better and to a lower cost than the dominant mobile-communication technologies of today. Thus, the key drivers are:

- staying competitive;
- services (better provisioning of old services as well as provisioning of new services);
- cost (more cost-efficient provisioning of old services as well as cost-effective provisioning of new services).

The technology advancement is necessary to provide new and more advanced services at a reasonable cost as well as to provide existing services in a better and more cost-efficient way.

### 2.1.1 Technology advancements

Technology advancements in many areas make it possible to build devices that were not possible 20, 10, or even 5 years ago. Even though Moore's law[1] is not a law of physics, it gives an indication of the rapid technology evolution for integrated circuits. This evolution enables faster processing/computing in smaller devices at lower cost. Similarly, the rapid development of color screens, small digital cameras, etc. makes it possible to envisage services to a device that were seen as utopia 10 years ago. For an example of the terminal development in the past 20 years, see Figure 2.1.

**Figure 2.1** *The terminal development has been rapid the past 20 years.*

[1] Moore's law is an empirical observation, and states that with the present rate of technological development, the complexity of an integrated circuit, with respect to minimum component cost, will double in about 18 months.

The size and weight of the mobile terminals have been reduced dramatically during the past 20 years. The standby and talk times have also been extended dramatically and the end users do not need to re-charge their devices every day. Simple black-and-white (or brown-and-gray) numerical screens have evolved into color screens capable of showing digital photos at good quality. Mega-pixel-capable digital cameras have been added making the device more attractive to use. Thus, the mobile device has become a multi-purpose device, not only a mobile phone for voice communications.

In parallel to the technical development of the mobile devices, the mobile-communication technologies are developed to meet the demands of the new services enabled, and also to enable them wireless. The development of the digital signal processors enables more advanced receivers capable of processing megabits of data in a short time, and the introduction of the optical fibers enables high-speed network connections to the base stations. In sum, this enables a fast access to information on the Internet as well as a short roundtrip time for normal communications. Thus, new and fancier services are enabled by the technical development of the devices, and new and more efficient mobile-communication systems are enabled by a similar technical development.

### 2.1.2 Services

Delivering services to the end users is the fundamental goal of any mobile-communication system. Knowing them, understanding them, managing them, and charging them properly is the key for success. It is also the most difficult task being faced by the engineers developing the mobile-communication system of the future. It is very difficult to predict what service(s) will be popular in a 5- to 10-year perspective. In fact, the engineers have to design a system that can adapt to any service that might become popular and used in the future. Unfortunately, there are also technical limitations that need to be understood, and also the technical innovations that in the future enable new services.

#### 2.1.2.1 Internet and IP technology

The success of the Internet and the IP-based services delivered over the Internet is more and more going wireless. This means that the mobile-communication systems are delivering more and more IP-based services, from the best effort-Internet data to voice-over-IP, for example in the shape of push-to-talk (PoC). Furthermore, in the wireless environment it is more natural to use, for example, location-based services and tracking services than in the fixed environment.

Thus, one can talk about mobile Internet services in addition to the traditional Internet services like browsing, e-mail, etc.[2]

Essentially, IP provides a transport mechanism that is service agnostic. Albeit there are several protocols on top of IP that are service-type specific (RTP, TCP, etc.), IP in itself is service agnostic. That enables service developers to develop services that only the imagination (and technology) sets the limit to. Thus, services will pop up, some will become popular for a while and then just fade away, whereas some others will never become a hit. Some services will become classics that will live and be used for a very long time.

#### 2.1.2.2 Traditional telephony services

Going toward IP-based services obviously does not mean that traditional services that have been provided over the circuit-switched domain, in successful mobile-communication systems like GSM, will disappear. Rather, it means that the traditional circuit-switched services will be ported over the IP networks. One particular service is the circuit-switched telephony service that will be provided as VoIP service instead. Thus, both the new advanced services that are enabled by the technology advancement of the devices and the traditional circuit-switched services will be using IP as the transporting mechanism (and are therefore called IP-based services). Hence future mobile-communication networks, including the 3G evolution, need to be optimized for IP-based applications.

#### 2.1.2.3 Wide spectrum of service needs

Trying to predict all the services that will be used over the mobile-communication systems 10 years from now is very difficult. The technology advancements in the various areas enable higher data rate connections, more memory on local devices, and more intelligent and easy to use man–machine interfaces. Furthermore, the human need of interaction and competition with other humans drives more intensive communication needs. All these combinations point toward applications and services that consume higher data rates and require lower delays compared to what today's mobile-communication systems can deliver.

However, the relative low-rate voice service will still be a very important component of the service portfolio that mobile-network operators wish to provide. In addition, services that have very relaxed delay requirements will also be there. Thus, not only high data rate services with a low-latency requirement, but also

[2]The common denominator between mobile Internet services and Internet services is the IP addressing technology with the IPv4 and IPv6 addresses identifying the end receiver. However, there is a need to handle the mobility provided by the cellular systems. Mobile IP is one possibility, but most of the cellular systems (if not all) have their own more efficient mobility mechanism.

low data rate best effort services will be provided. Furthermore, not only the data rate and delay are important to understand when talking about a service's need from a mobile-communication system, but also the setup time is very important, for example, a service can be totally useless if it takes too long to start it (for example making a phone call, downloading a web page). Thus the mobile-communication systems of the future, including the evolution of 3G mobile communication, need to be able to deliver short call setup times, low latency and a wide range of data rates.

#### 2.1.2.4 Key design services

Since it is impossible to know what services that will be popular and since service possibilities and offers will differ with time and possible also with country, the future mobile-communication systems will need to be adaptive to the changing service environment. Luckily, there are a few known key services that span the technology space. Those are:

- *Real-time-gaming applications*: These have the characteristic to require small amount of data (game update information) relatively frequent with low delay requirement.[3] Only a limited delay jitter is tolerable. A first person shooter game like Counter Strike is an example of a game that has this characteristic.
- *Voice*: This has the characteristic to require small amounts of data (voice packets) frequently with no delay jitter. The end-to-end delay has to be small enough not to be noticeable.[4]
- *Interactive file download and upload applications*: These have the characteristics of requiring low delay and high data rates.
- *Background file download and upload applications*: These have the characteristics of accepting lower bit rates and longer delays. E-mail (mostly) is an example of background file download and upload.
- *Television*: This has the characteristics of streaming downlink to many users at the same time requiring low delay jitter. The service can tolerate delays, as long as it is approximately the same delay for all users in the neighborhood. The television service has moderate data rate requirements.

A mobile-communication system designed to handle these services and the services in between will be able to facilitate most services (see Figure 2.2). Unfortunately, the upper limit of the data rate demand and the lower limit of the delay requirement are difficult to provide in a cost-efficient manner. The designers

[3]The faster the data is delivered the better. Expert Counter Strike players look for game servers with a ping time of less than 50 ms.

[4]In 3G systems the end-to-end delay requirement for circuit-switched voice is approximately 400ms. This delay is not disturbing humans in voice communications.

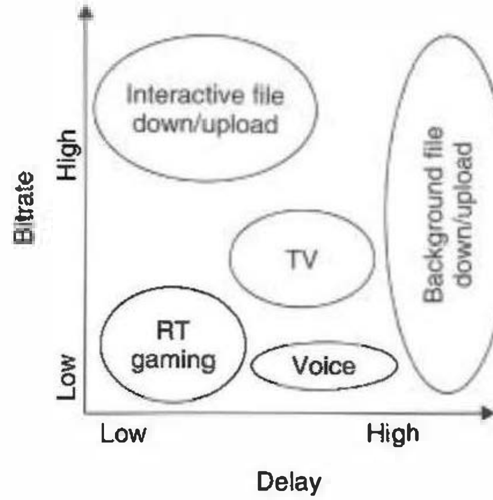


**Figure 2.2** *The bit rate – delay service space that is important to cover when designing a new cellular system.*

of the mobile-communication systems need to stop at a reasonable level, a level that the technology available at the time of standardization can provide.

### 2.1.3 Cost and performance

There is another important driving factor for future mobile-communication systems and that is the cost of the service provisioning. The technology advancement that enables new services can also be utilized to provide better mobile-communication systems using more advanced technical features. Here IP technology is not only a key enabler to allow for new services to pop up, but also a way of reducing cost of new services. The reason is that IP as a bearer of new services can be used to introduce new services as they come, not requiring an extensive special design of the system. Of course, this requires that the devices used in the mobile-communication system can be programmed by third-party providers and that the operators allow third-party service providers to use their network for communication.

Another important factor is that operators need to provide the services to all the users. Not only one user needs to get the low delay, high data rate, etc. that its service needs, but all the users with their different service needs should be served efficiently. The processing capacity evolution and Moore's law help also for this problem. New techniques are enabled by the higher processing power in the devices – techniques that delivers more bits of data per hertz. Furthermore, the coverage is increased with more advanced antennas and receivers. This enables the operators to deliver the services to more users from one base station, thus requiring fewer sites. Fewer sites imply lower operational and capitalization costs. In essence, the operators need fewer base stations and sites to provide the service.

Obviously, all services would be 'happy' if they were provided with the highest data rate, lowest delay, and lowest jitter that the system can provide. Unfortunately, this is unattainable in practice and contradictory to the operator goal of an efficient system: in other words, the more delay a service can handle the more efficient the system can be. Thus, the cost of providing lowest possible delay, jitter and call setup time is somewhat in conflict with the need of the mobile-network operator to provide it to all the users. Hence, there is a trade-off between user experience and system performance. The better the system performance is, the lower the cost of the network. However, the end users also need to get adequate performance which often is in conflict with the system performance, thus the operator cannot only optimize for system performance.

## 2.2 3G evolution: Two Radio Access Network approaches and an evolved core network

### 2.2.1 Radio Access Network evolution

TSG RAN organized a workshop on 3GPP long-term Evolution in the fall of 2004. The workshop was the starting point of the development of the Long-Term Evolution (LTE) radio interface. After the initial requirement phase in the spring of 2005, where the targets and objectives of LTE were settled, the technical specification group TSG SA launched a corresponding work on the System Architecture Evolution, since it was felt that the LTE radio interface needed a suitable evolved system architecture.

The result of the LTE workshop was that a study item in 3GPP TSG RAN was created in December 2004. The first 6 months were spent on defining the requirements, or design targets, for LTE. These were documented in a 3GPP technical report [86] and approved in June 2005. Chapter 13 will go through the requirements in more detail. Most notable are the requirements on high data rate at the cell edge and the importance of low delay, in addition to the normal capacity and peak data rate requirements. Furthermore, spectrum flexibility and maximum commonality between FDD and TDD solutions are pronounced.

During the fall 2005, 3GPP TSG RAN WG1 made extensive studies of different basic physical layer technologies and in December 2005 the TSG RAN plenary decided that the LTE radio access should be based on *OFDM* in the downlink and *single carrier FDMA* in the uplink.

TSG RAN and its working groups then worked on the LTE specifications and the specifications were approved in December 2007. However, 3GPP TSG RAN did not stop working on LTE when the first version of the specifications was completed. In fact, 3GPP will continue to evolve LTE towards LTE Advanced. Chapters 14–20 will go through the LTE radio interface in more detail.

At the same time as the LTE discussion was ongoing, 3GPP TSG RAN and its WGs continued to evolve the WCDMA system with more functionality, most notably MBMS and Enhanced Uplink. These additions were done in a backward compatible manner: that is, terminals of earlier releases can coexist on the same carrier in the same base station as terminals of the latest release. The main argument for the backward compatibility is that the installed base of equipment can be upgraded to handle the new features while still being capable of serving the old terminals. This is a cost-efficient addition of new features, albeit the new features are restricted by the solutions for the old terminals.

Naturally, HSPA[5] does not include all the technologies considered for LTE. Therefore, a study in 3GPP was initiated to see how far it is possible to take HSPA within the current spectrum allocation of 5 MHz and still respect the backward compatibility aspect. Essentially, the target with HSPA Evolution was, and still is, to reach near the characteristics of LTE when using a 5 MHz spectrum and at the same time being backward compatible. Chapters 8–12 will go through the HSPA and the HSPA Evolution in more detail.

Thus, the 3GPP 3G evolution standard has two parts: LTE and HSPA Evolution. Both parts have their merits. LTE can operate in new and more complex spectrum arrangements (although in the same spectrum bands as WCDMA and other 3G technologies) with the possibility for new designs that do not need to cater for terminals of earlier releases. HSPA Evolution can leverage on the installed base of equipment in the 5 MHz spectrum but needs to respect the backward compatibility of earlier terminals.

#### *2.2.1.1 LTE drivers and philosophy*

The 3GPP Long-Term Evolution is intended to be a mobile-communication system that can take the telecom industry into the 2020s. The philosophy behind LTE standardization is that the competence of 3GPP in specifying mobile-communication systems in general and radio interfaces in particular shall be used, but the result shall not be restricted by previous work in 3GPP. Thus, LTE does not need to be backward compatible with WCDMA and HSPA.

Leaving the legacy terminals behind, not being restricted by designs of the late 1990s, makes it possible to design a radio interface from scratch. In the LTE case,

[5] When operating with HSDPA and Enhanced UL, the system is known as HSPA.

the radio interface is purely optimized for IP transmissions not having to support ISDN traffic: that is, there is no requirement for support of GSM circuit-switch services, a requirement that WCDMA had. Furthermore, LTE also has a very large commonality of FDD and TDD operations, a situation that did not exist in 3GPP before LTE.

Instead new requirements have arisen, for example the requirement on spectrum flexibility, since the global spectrum situation becomes more and more complex. Operators get more and more scattered spectrum, spread over different bands with different contiguous bandwidths. LTE needs to be able to operate in all these bands and with the bandwidths that is available to the operator. However, in practice only a limited set of bandwidths can be used since otherwise the RF and filter design would be too costly. LTE is therefore targeted to operate in spectrum allocations from roughly 1 to 20 MHz. The spectrum flexibility support with the possibility to operate in other bandwidths than 5 MHz makes LTE very attractive for operators. The low-bandwidth operations are suitable for refarming of spectrum (for example GSM spectrum and CDMA2000 spectrum). The higher-bandwidth options are suitable for new deployments in unused spectrum, where it is more common to have larger chunks of contiguous spectrum.

Furthermore, when going to the data rates that LTE is targeting, achieving low delay and high data rates at the cell edges are more important requirements than the peak data rate. Thus, a more pronounced requirement for LTE is the low delay with high data rates at the cell edges than it was when WCDMA was designed in the late 1990s.

Although not backward compatible with WCDMA, LTE design is clearly influenced by the WCDMA and the HSPA work in 3GPP. It is the same body, the same people, and companies that are active and more importantly, WCDMA and HSPA protocols are a good foundation for the LTE design. The philosophy is to take what is good from WCDMA and HSPA, and redo those parts that have to be updated due to the new requirement situation: either there are new requirements such as the spectrum flexibility or there are requirements that no longer are valid such as the support of ISDN services. The technology advancement in the cellular area has, of course, also influenced the design choice of LTE.

#### *2.2.1.2 HSPA evolution drivers and philosophy*

WCDMA, HSDPA, and HSPA are in commercial operation throughout the world. This means that the infrastructure for HSPA is already in place with the network node sites, especially the base-station sites with their antenna arrangements and hardware. This equipment is serving millions of terminals with different characteristics and supported 3GPP releases. These terminals need to be supported by the WCDMA operators for many more years.

The philosophy of the HSPA Evolution work is to continue to add new and fancier technical features, and at the same time be able to serve the already existing terminals. This is the successful strategy of GSM that have added new features constantly since the introduction in the early 1990s. The success stems from the fact that there are millions of existing terminals at the launch time of the new features that can take the cost of the upgrade of the network for the initially few new terminals before the terminal fleet is upgraded. The time it takes to upgrade the terminal fleet is different from country to country, but a rule of thumb is that a terminal is used for 2 years before it is replaced. For HSPA Evolution that means that millions of HSPA-capable terminals need to be supported at launch. In other words, HSPA Evolution needs to be backward compatible with the previous releases in the sense that it is possible to serve terminals of earlier releases of WCDMA on the same carrier as HSPA-Evolution-capable terminals.

The backward compatibility requirement on the HSPA Evolution puts certain constraints on the technology that LTE does not need to consider, for example the physical layer fundamentals need to be the same as for WCDMA release 99. On the other hand, HSPA Evolution is built on the existing specifications and only those parts of the specifications that need to be upgraded are touched. Thus there is less standardization, implementation and testing work for HSPA Evolution than for LTE since the HSPA Evolution philosophy is to apply new more advanced technology on the existing HSPA standard. This will bring HSPA to a performance level comparable to LTE when compared on a 5MHz spectrum allocation (see Chapter 23 for a performance comparison of HSPA Evolution and LTE).

### 2.2.2 An evolved core network: system architecture evolution

Roughly at the same time as LTE and HSPA Evolution was started, 3GPP decided to make sure that an operator can coexist easily between HSPA Evolution and LTE through an evolved core network, the *Evolved Packet Core*. This work was done under the umbrella *System Architecture Evolution* study item lead by TSG SA WG2.

The *System Architecture Evolution* study focused on how the 3GPP core network will evolve into the core network of the next decades. The existing core network was designed in the 1980s for GSM, extended during the 1990s' for GPRS and WCDMA. The philosophy of the SAE is to focus on the packet-switched domain,


**Figure 2.3** *One HSPA and LTE deployment strategy: upgrade to HSPA Evolution, then deploy LTE as islands in the WCDMA/HSPA sea.*

and migrate away from the circuit-switched domain. This is done through the coming 3GPP releases ending up with the Evolved Packet Core.

Knowing that HSPA Evolution is backward compatible and knowing that the Evolved Packet Core will support both HSPA Evolution and LTE assures that LTE can be deployed in smaller islands and thus only where it is needed. A gradual deployment approach can be selected (see Figure 2.3). First the operator can upgrade its HSPA network to HSPA-Evolution-capable network, and then add LTE cells where capacity is lacking or where the operator wants to try out new services that cannot be delivered by HSPA Evolution. This approach reduces the cost of deployment since LTE do not need to be build for nationwide coverage from day one.

# 9
# High-Speed Downlink Packet Access

The introduction of *High-Speed Downlink Packet Access*, (HSDPA), implies a major extension of the WCDMA radio interface, enhancing the WCDMA downlink packet-data performance and capabilities in terms of higher peak data rate, reduced latency and increased capacity. This is achieved through the introduction of several of the techniques described in Part II, including *higher-order modulation, rate control, channel-dependent scheduling*, and *hybrid ARQ with soft combining*. The HSDPA specifications are found in [100] and the references therein.

## 9.1 Overview

### 9.1.1 Shared-channel transmission

A key characteristic of HSDPA is the use of *Shared-Channel Transmission*. Shared-channel transmission implies that a certain fraction of the total downlink radio resources available within a cell, channelization codes and transmission power in case of WCDMA, is seen as a common resource that is dynamically shared between users, primarily in the time domain. The use of shared-channel transmission, in WCDMA implemented through the *High-Speed Downlink Shared Channel* (HS-DSCH) as described below, enables the possibility to rapidly allocate a large fraction of the downlink resources for transmission of data to a specific user. This is suitable for packet-data applications which typically have bursty characteristics and thus rapidly varying resource requirements.

The basic HS-DSCH code- and time- structure is illustrated in Figure 9.1. The HS-DSCH code resource consists of a set of channelization codes of spreading factor 16, see upper part of Figure 9.1 where the number of codes available for HS-DSCH transmission is configurable between 1 and 15. Codes not reserved

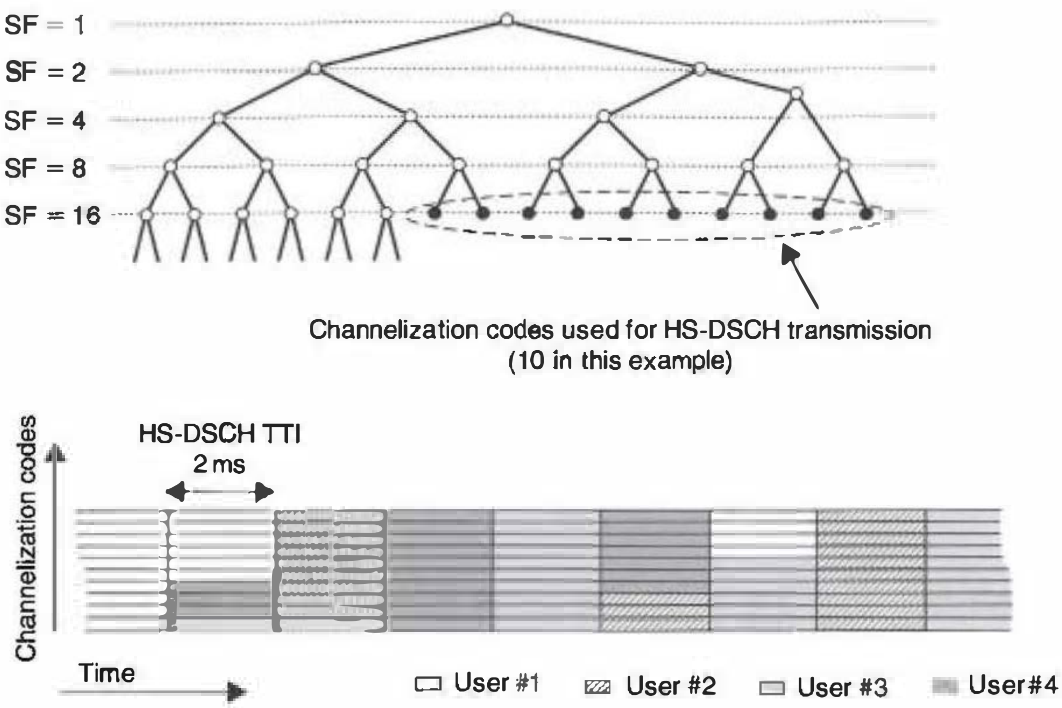


**Figure 9.1** *Time- and code-domain structure for HS-DSCH.*

for HS-DSCH transmission are used for other purposes, for example related control signaling, MBMS services, or circuit-switched services.

The dynamic allocation of the HS-DSCH code resource for transmission to a specific user is done on 2ms TTI basis, see lower part of Figure 9.1. The use of such a short TTI for HSDPA reduces the overall delay and improves the tracking of fast channel variations exploited by the rate control and the channel-dependent scheduling as discussed below.

In addition to being allocated a part of the overall code resource, a certain part of the total available cell power should also be allocated for HS-DSCH transmission. Note that the HS-DSCH is not power controlled but rate controlled as discussed below. This allows the remaining power, after serving other channels, to be used for HS-DSCH transmission and enables efficient exploitation of the overall available power resource.

### 9.1.2 Channel-dependent scheduling

*Scheduling* controls to which user the shared-channel transmission is directed at a given time instant. The scheduler is a key element and to a large extent determines the overall system performance, especially in a highly loaded network. In each TTI, the scheduler decides to which user(s) the HS-DSCH should be transmitted and, in close cooperation with the rate-control mechanism, at what data rate.

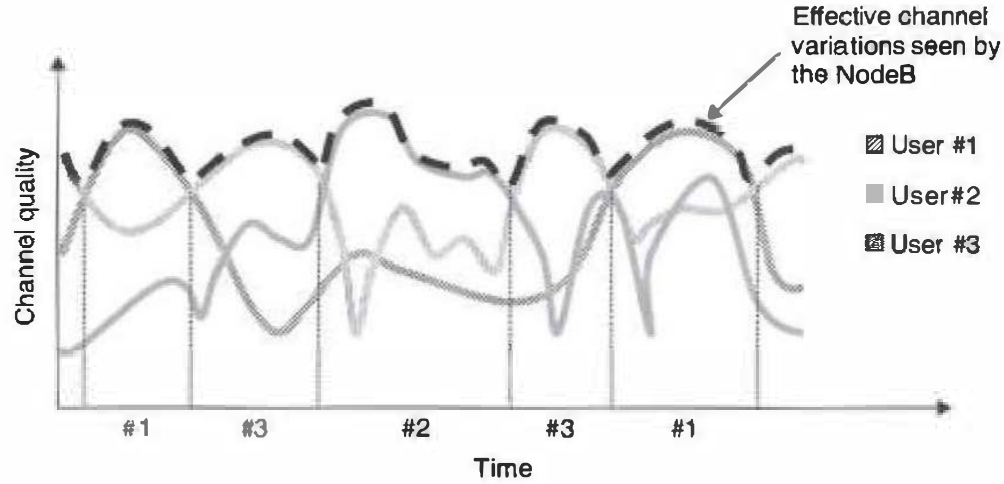


**Figure 9.2** *Channel-dependent scheduling for HSDPA.*

In Chapter 7, it was discussed how a significant increase in capacity can be obtained if the radio-channel conditions are taken into account in the scheduling decision, so-called *channel-dependent scheduling*. Since the radio conditions for the radio links to different UEs within a cell typically vary independently, at each point in time there is almost always a radio link whose channel quality is near its peak, see Figure 9.2. As this radio link is likely to have good channel quality, a high data rate can be used for this radio link. This translates into a high system capacity. The gain obtained by transmitting to users with favorable radio-link conditions is commonly known as *multi-user diversity* and the gains are larger, the larger the channel variations and the larger the number of users in a cell. Thus, in contrast to the traditional view that fast fading is an undesirable effect that has to be combated, with the possibility for channel-dependent scheduling fading is potentially beneficial and should be exploited.

Several different scheduling strategies were discussed in Chapter 7. As discussed, a practical scheduler strategy exploits the short-term variations, for example due to multi-path fading and fast interference variations, while maintaining some degree of long-term fairness between the users. In principle, the larger the long-term unfairness, the higher the cell capacity. A trade-off between fairness and capacity is therefore required.

In addition to the channel conditions, traffic conditions are also taken into account by the scheduler. For example, there is obviously no purpose in scheduling a user with no data awaiting transmission, regardless of whether the channel conditions are beneficial or not. Furthermore, some services should preferably be given higher priority. As an example, streaming services should be ensured a relatively constant long-term data rate while background services such as file download have less stringent requirements on a constant long-term data rate.

### 9.1.3 Rate control and higher-order modulation

In Chapter 7, rate control was discussed and, for packet-data services, shown to be a more efficient tool for link adaptation, compared to the fast power control typically used in CDMA-based systems, especially when used together with channel-dependent scheduling.

For HSDPA, rate control is implemented by dynamically adjusting the channel-coding rate as well as dynamically selecting between QPSK and 16QAM modulation. *Higher-order modulation* such as 16QAM allows for higher bandwidth utilization than QPSK, but requires bigher received $E_b/N_0$ as described in Chapter 3. Consequently, 16QAM is mainly useful in advantageous channel conditions. The data rate is selected independently for each 2 ms TTI by the NodeB and the rate control mechanism can therefore track rapid channel variations.

### 9.1.4 Hybrid ARQ with soft combining

Fast *hybrid ARQ with soft combining* allows the terminal to request retransmission of erroneously received transport blocks, effectively fine-tuning the effective code rate and compensating for errors made by the link-adaptation mechanism. The terminal attempts to decode each transport block it receives and reports to the NodeB its success or failure 5 ms after the reception of the transport block. This allows for rapid retransmissions of unsuccessfully received data and significantly reduces the delays associated with retransmissions compared to Release 99.

*Soft combining* implies that the terminal does not discard soft information in case it cannot decode a transport block as in traditional hybrid-ARQ protocols, but combines soft information from previous transmission attempts with the current retransmission to increase the probability of successful decoding. Incremental redundancy, IR, is used as the basis for soft combining in HSDPA, that is the retransmissions may contain parity bits not included in the original transmission. From Chapter 7, it is known that IR can provide significant gains when the code rate for the initial transmission attempts is high as the additional parity bits in the retransmission results in a lower overall code rate. Thus, IR is mainly useful in bandwidth-limited situations, for example, when the terminal is close to the base station and the amount of channelization codes, and not the transmission power, limits the achievable data rate. The set of coded bits to use for the retransmission is controlled by the NodeB, taking the available UE memory into account.

### 9.1.5 Architecture

From the previous discussion it is clear that the basic HSDPA techniques rely on fast adaptation to rapid variations in the radio conditions. Therefore, these techniques need to be placed close to the radio interface on the network side, that is in the NodeB. At the same time, an important design objective of HSDPA was to retain the Release 99 functional split between layers and nodes as far as possible. Minimization of the architectural changes is desirable as it simplifies introduction of HSDPA in already deployed networks and also secures operation in environments where not all cells have been upgraded with HSDPA functionality. Therefore, HSDPA introduces a new MAC sub-layer in the NodeB, the *MAC-hs,* responsible for scheduling, rate control and hybrid-ARQ protocol operation. Hence, apart from the necessary enhancements to the RNC such as admission control of HSDPA users, the introduction of HSDPA mainly affects the NodeB (Figure 9.3).

Each UE using HSDPA will receive HS-DSCH transmission from one cell, the *serving cell.* The serving cell is responsible for scheduling, rate control, hybrid ARQ, and all other MAC-hs functions used by HSDPA. Uplink soft handover

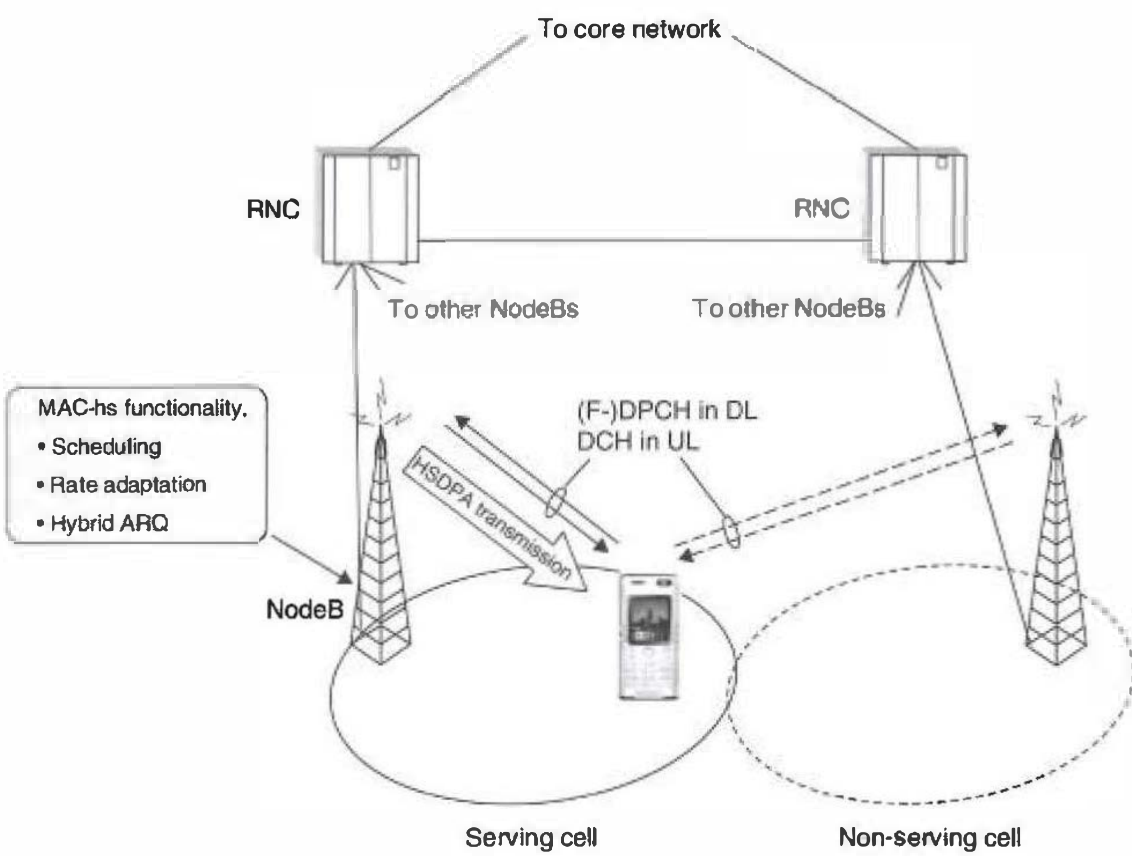


**Figure 9.3** *Illustration of the HSDPA architecture.*

is supported, in which case the uplink data transmission will be received in multiple cells and the UE will receive power control commands from multiple cells.

Mobility from a cell supporting HSDPA to a cell that is not supporting HSDPA is easily handled. Uninterrupted service to the user can be provided, albeit at a lower data rate, by using channel switching in the RNC and switch the user to a dedicated channel in the non-HSDPA cell. Similarly, a user equipped with an HSDPA-capable terminal may be switched from a dedicated channel to HSDPA when the user enters a cell with HSDPA support.

## 9.2 Details of HSDPA

### 9.2.1 HS-DSCH: Inclusion of features in WCDMA Release 5

The High-Speed Downlink Shared Channel (HS-DSCH), is the transport channel used to support shared-channel transmission and the other basic technologies in HSDPA, namely channel-dependent scheduling, rate control (including higher-order modulation), and hybrid ARQ with soft combining. As discussed in the introduction and illustrated in Figure 9.1, the HS-DSCH corresponds to a set of channelization codes, each with spreading factor 16. Each such channelization code is also known as an HS-PDSCH – *High-Speed Physical Downlink Shared Channel.*

In addition to HS-DSCH, there is a need for other channels as well, for example for circuit-switched services and for control signaling. To allow for a trade-off between the amount of code resources set aside for HS-DSCH and the amount of code resource used for other purposes, the number of channelization codes available for HS-DSCH can be configured, ranging from 1 to 15 codes. Codes not reserved for HS-DSCH transmission are used for other purposes, for example related control signaling and circuit-switched services. The first node in the code tree can never be used for HS-DSCH transmission as this node includes mandatory physical channels such as the common pilot.

Sharing of the HS-DSCH code resource should primarily take place in the time domain. The reason is to fully exploit the advantages of channel-dependent scheduling and rate control, since the quality at the terminal varies in the time domain, but is (almost) independent of the set of codes (physical channels) used for transmission. However, sharing of the HS-DSCH code resource in the code domain is also supported as illustrated in Figure 9.1. With code-domain sharing, two or more UEs are scheduled simultaneously by using different parts of the common code resource (different sets of physical channels). The reasons for code-domain sharing are twofold: support of terminals that are, for complexity reasons, not able to despread the full set of codes, and efficient support of small payloads when the transmitted data does not require the full set of allocated HS-DSCH codes. In either of these cases, it is obviously a waste of resources to assign the full code resource to a single terminal.

In addition to being allocated a part of the overall code resource, a certain part of the total available cell power should also be used for HS-DSCH transmission. To maximize the utilization of the power resource in the base station, the remaining power after serving other, power-controlled channels, should preferably be used for HS-DSCH transmission as illustrated in Figure 9.4. In principle, this results in a (more or less) constant transmission power in a cell. Since the HS-DSCH is rate controlled as discussed below, the HS-DSCH data rate can be selected to match the radio conditions and the amount of power instantaneously available for HS-DSCH transmission.

To obtain rapid allocation of the shared resources, and to obtain a small end-user delay, the TTI should be selected as small as possible. At the same time, a too small TTI would result in excessive overhead as control signaling is required for each transmission. For HSDPA, this trade-off resulted in the selection of a 2ms TTI.

Downlink control signaling is necessary for the operation of HS-DSCH in each TTI. Obviously, the identity of the UE(s) currently being scheduled must be signaled as well as the physical resource (the channelization codes) used for transmission to this UE. The UE also needs to be informed about the transport format used for the transmission as well as hybrid-ARQ-related information. The resource and transport-format information consists of the part of the code

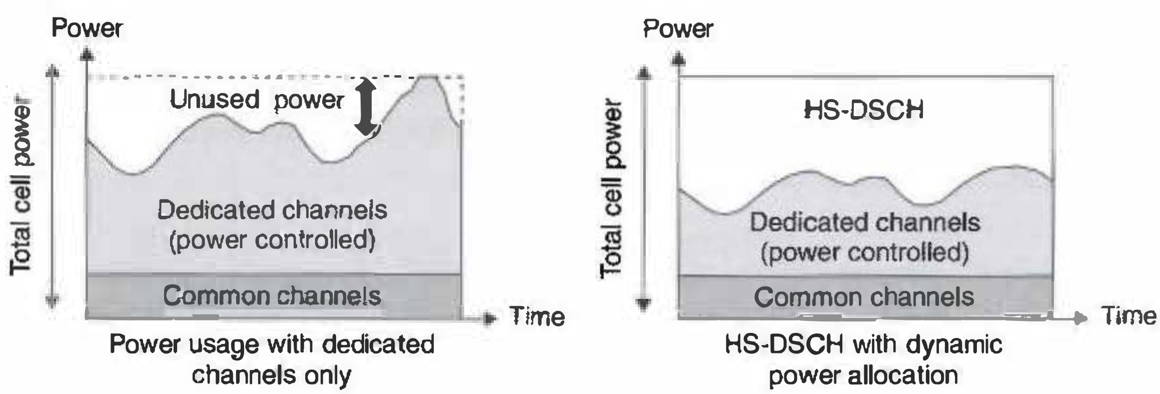


**Figure 9.4** *Dynamic power usage with HS-DSCH.*

tree used for data transmission, the modulation scheme used, and the transport-block size. The downlink control signaling is carried on the *High-Speed Shared Control Channel* (HS-SCCH), transmitted in parallel to the HS-DSCH using a separate channelization code. The HS-SCCH is a shared channel, received by all UEs for which an HS-DSCH is configured to find out whether the UE has been scheduled or not.

Several HS-SCCHs can be configured in a cell, but as the HS-DSCH is shared mainly in the time domain and only the currently scheduled terminal needs to receive the HS-SCCH, there is typically only one or, if code-domain sharing is supported in the cell, a few HS-SCCHs configured in each cell. However, each HS-DSCH-capable terminal is required to be able to monitor up to four HS-SCCHs. Four HS-SCCH has been found to provide sufficient flexibility in the scheduling of multiple UEs; if the number was significantly smaller the scheduler would have been restricted in which UEs to schedule simultaneously in case of code-domain sharing.

HSDPA transmission also requires uplink control signaling as the hybrid-ARQ mechanism must be able to inform the NodeB whether the downlink transmission was successfully received or not. For each downlink TTI in which the UE has been scheduled, an ACK or NAK will be sent on the uplink to indicate the result of the HS-DSCH decoding. This information is carried on the uplink *High-Speed Dedicated Physical Control Channel* (HS-DPCCH). One HS-DPCCH is set up for each UE with an HS-DSCH configured. In addition, the NodeB needs information about the instantaneous downlink channel conditions at the UE for the purpose of channel-dependent scheduling and rate control. Therefore, each UE also measures the instantaneous downlink channel conditions and transmits a *Channel-Quality Indicator* (CQI), on the HS-DPCCH.

In addition to HS-DSCH and HS-SCCH, an HSDPA terminal need to receive power control commands for support of fast closed-loop power control of the uplink in the same way as any WCDMA terminal. This can be achieved by a downlink dedicated physical channel, DPCH, for each UE. In addition to power control commands, this channel can also be used for user data not carried on the HS-DSCH, for example circuit-switched services.

In Release 6, support for *fractional DPCH,* F-DPCH, is added to reduce the consumption of downlink channelization codes. In principle, the only use for a dedicated channel in the downlink is to carry power control commands to the UE in order to adjust the uplink transmission. If all data transmissions, including higher-layer signaling radio bearers, are mapped to the HS-DSCH, it is a waste

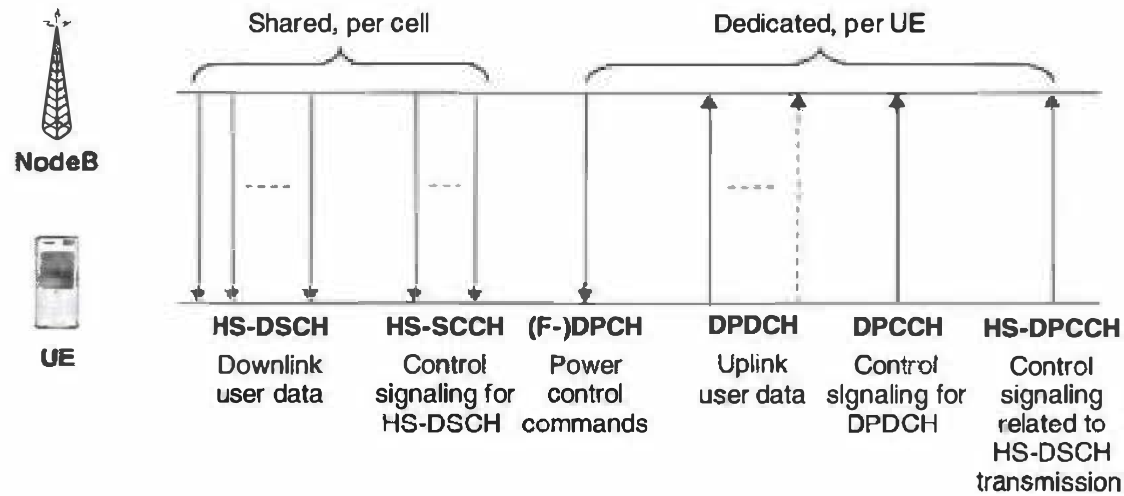


**Figure 9.5** *Channel structure with HSDPA.*

of scarce code resources to use a dedicated channel with spreading factor 256 per UE for power control only. The F-DPCH resolves this by allowing multiple UEs to share a single downlink channelization code.

To summarize, the overall channel structure with HSDPA is illustrated in Figure 9.5.

Neither the HS-PDSCH, nor the HS-SCCH, is subject to downlink macro-diversity or soft handover. The basic reason is the location of the HS-DSCH scheduling in the NodeB. Hence, it is not possible to simultaneously transmit the HS-DSCH to a single UE from multiple NodeBs, which prohibits the use of inter-NodeB soft handover. Furthermore, it should be noted that within each cell, multi-user diversity is exploited by the channel-dependent scheduler. Basically, the scheduler only transmits to a user when the instantaneous radio conditions are favorable and thus the additional gain from macro-diversity is reduced.

However, the uplink channels, as well as any dedicated downlink channels, can be in soft handover. As these channels are not subject to channel-dependent scheduling, macro-diversity provides a direct coverage benefit.

### 9.2.2 MAC-hs and physical-layer processing

As mentioned in the introduction, the MAC-hs is a new sub-layer located in the NodeB and responsible for the HS-DSCH scheduling, rate control and hybrid-ARQ protocol operation. To support these features, the physical layer has also been enhanced with the appropriate functionality, for example support for soft

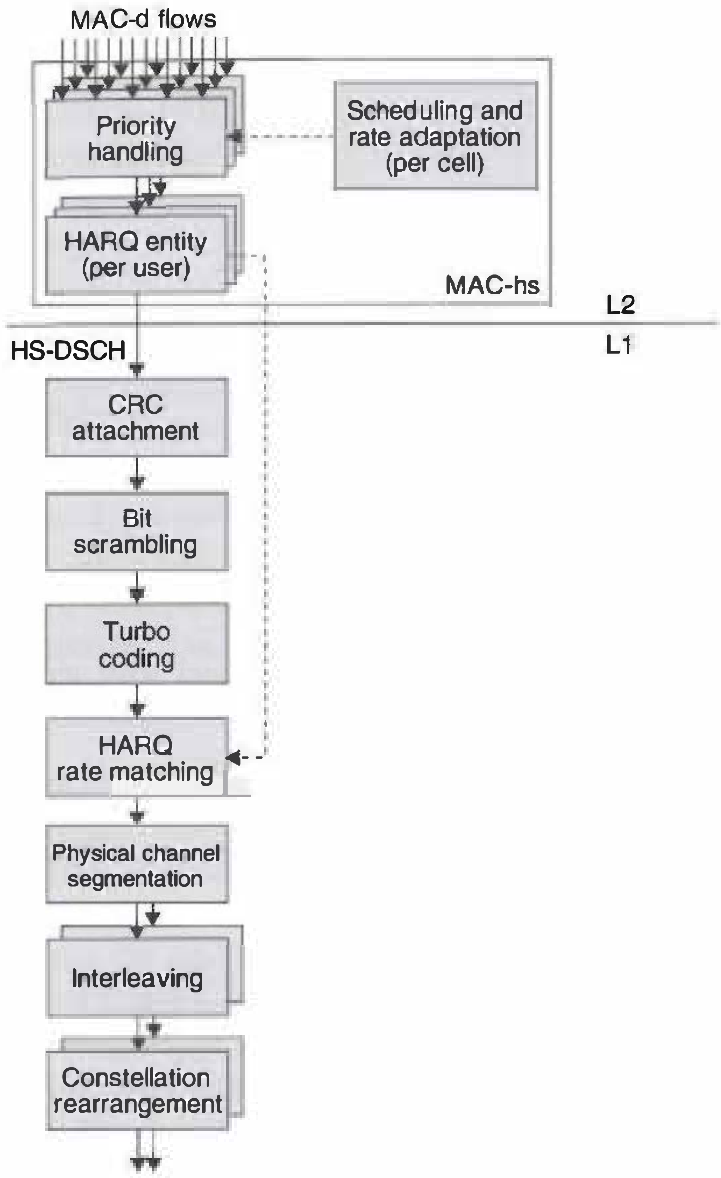


**Figure 9.6** *MAC-hs and physical-layer processing.*

combining in the hybrid ARQ. In Figure 9.6, the MAC-hs and physical-layer processing is illustrated.

The MAC-hs consists of scheduling, priority handling, transport-format selection (rate control), and the protocol parts of the hybrid-ARQ mechanism. Data, in the form of a single transport block with dynamic size, passes from the MAC-hs via the HS-DSCH transport channel to the HS-DSCH physical-layer processing.

The HS-DSCH physical-layer processing is straightforward. A 24-bit CRC is attached to each transport block. The CRC is used by the UE to detect errors in the received transport block.

Demodulation of 16QAM, which is one of the modulation schemes supported by the HS-DSCH, requires amplitude knowledge at the receiver in order to correctly form the soft values prior to Turbo decoding. This is different from QPSK, where no such knowledge is required as all information is contained in the phase of the received signal. To ease the estimation of the amplitude reference in the receiver, the bits after CRC attachment are scrambled. This results in a sufficiently random sequence out from the Turbo coder to cause both inner and outer signal points in the 16QAM constellation to be used, thereby aiding the UE in the estimation of the amplitude reference. Note that bit scrambling is done regardless of the modulation scheme used, even if it is strictly speaking only needed in case of 16QAM modulation.

The fundamental coding scheme in HSDPA is rate-1/3 Turbo coding. To obtain the code rate selected by the rate-control mechanism in the MAC-hs, rate matching, that is, puncturing or repetition, is used to match the number of coded bits to the number of physical-channel bits available. The rate-matching mechanism is also part of the physical-layer hybrid-ARQ and is used to generate different redundancy versions for incremental redundancy. This is done through the use of different puncturing patterns; different bits are punctured for initial transmissions and retransmission.

Physical-channel segmentation distributes the bits to the channelization codes used for transmission, followed by channel interleaving.

Constellation rearrangement is used only for 16QAM. If Chase combining is used in combination with 16QAM, a gain in performance can be obtained if the signal constellation is changed between retransmissions. This is further elaborated upon below.

### 9.2.3 Scheduling

One of the basic principles for HSDPA is the use of channel-dependent scheduling. The scheduler in the MAC-hs controls what part of the shared code and power resource is assigned to which user in a certain TTI. It is a key component and to a large extent determines the overall HSDPA system performance, especially in a loaded network. At lower loads, only one or a few users are available for scheduling and the differences between different scheduling strategies are less pronounced.

Although the scheduler is implementation specific and not specified by 3GPP, the overall goal of most schedulers is to take advantage of the channel variations between users and preferably schedule transmissions to a user when the channel conditions are advantageous. As discussed in Chapter 7, several scheduling strategies are possible. However, efficient scheduling strategies require at least:

- information about the instantaneous channel conditions at the UE,
- information about the buffer status and priorities of the data flows.

Information about the instantaneous channel quality at the UE is typically obtained through a 5-bit Channel-Quality Indicator (CQI), which each UE feed back to the NodeB at regular intervals. The CQI is calculated at the UE based on the signal-to-noise ratio of the received common pilot. Instead of expressing the CQI as a received signal quality, the CQI is expressed as a recommended transport-block size, taking into account also the receiver performance. This is appropriate as the quantity of relevance is the instantaneous data rate a terminal can support rather than the channel quality alone. Hence, a terminal with a more advanced receiver, being able to receive data at a higher rate at the same channel quality, will report a larger CQI than a terminal with a less advanced receiver, all other conditions being identical.

In addition to the instantaneous channel quality, the scheduler should typically also take buffer status and priority levels into account. Obviously UEs for which there is no data awaiting transmission should not be scheduled. There could also be data that is important to transmit within a certain maximum delay, regardless of the channel conditions. One important example hereof is RRC signaling, for example, related to cell change in order to support mobility, which should be delivered to the UE as soon as possible. Another example, although not as time critical as RRC signaling, is streaming services, which has an upper limit on the acceptable delay of a packet to ensure a constant average data rate. To support priority handling in the scheduling decision, a set of priority queues is defined into which the data is inserted according to the priority of the data as illustrated in Figure 9.7. The scheduler selects data from these priority queues for transmission based on the channel conditions, the priority of the queue, and any other relevant information. To efficiently support streaming applications, which require a minimum average data rate, there is a possibility for the RNC to 'guarantee' this data rate by providing information about the average data rate to the scheduler in the NodeB. The scheduler can take this constraint into account in the scheduling process.

### 9.2.4 Rate control

As described in Chapter 7, rate control denotes the process of adjusting the data rate to match the instantaneous radio conditions. The data rate is adjusted by

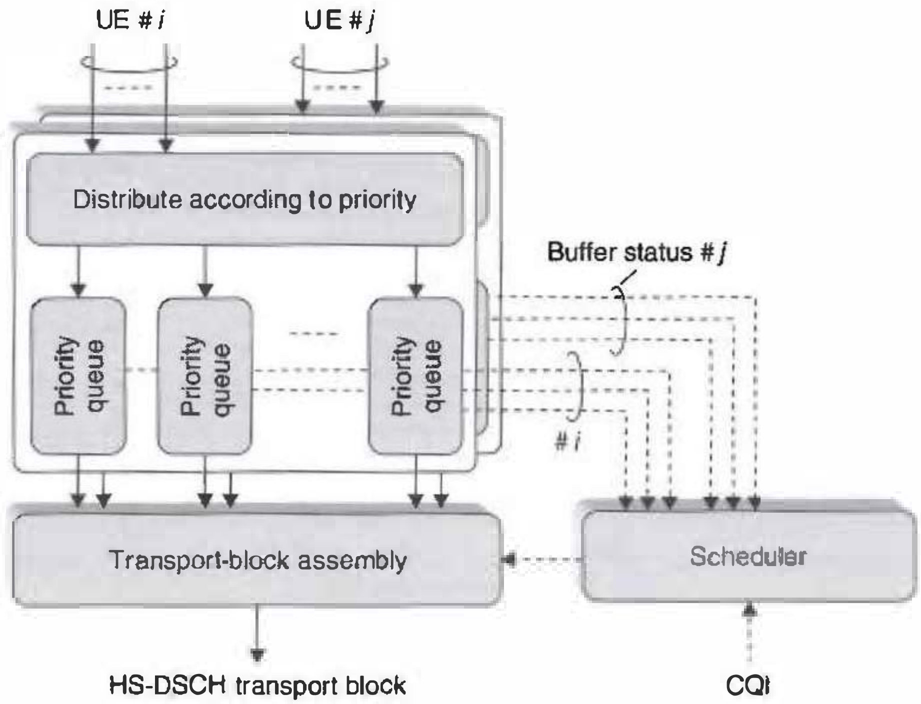


**Figure 9.7** *Priority handling in the scheduler.*

changing the modulation scheme and the channel-coding rate. For each TTI, the rate-control mechanism in the scheduler selects, for the scheduled user(s), the transport format(s) and channelization-code resources to use. The transport format consists of the modulation scheme (QPSK or 16QAM) and the transport-block size.

The resulting code rate after Turbo coding and rate matching is given implicitly by the modulation scheme, the transport-block size, and the channelization-code set allocated to the UE for the given TTI. The number of coded bits after rate matching is given by the modulation scheme and the number of channelization codes, while the number of information bits prior coding is given by the transport-block size. Hence, by adjusting some or all of these parameters, the overall code rate can be adjusted.

Rate control is implemented by allowing the MAC-hs to set the transport format independently for each 2 ms HS-DSCH TTI. Hence, both the modulation scheme and the instantaneous code rate can be adjusted to obtain a data rate suitable for the current radio conditions. The relatively short TTI of 2ms allows the rate control to track reasonable rapid variations in the instantaneous channel quality.

The HS-DSCH transport-block size can take one of 254 values. These values, illustrated in Figure 9.8, are listed in the specifications and therefore known

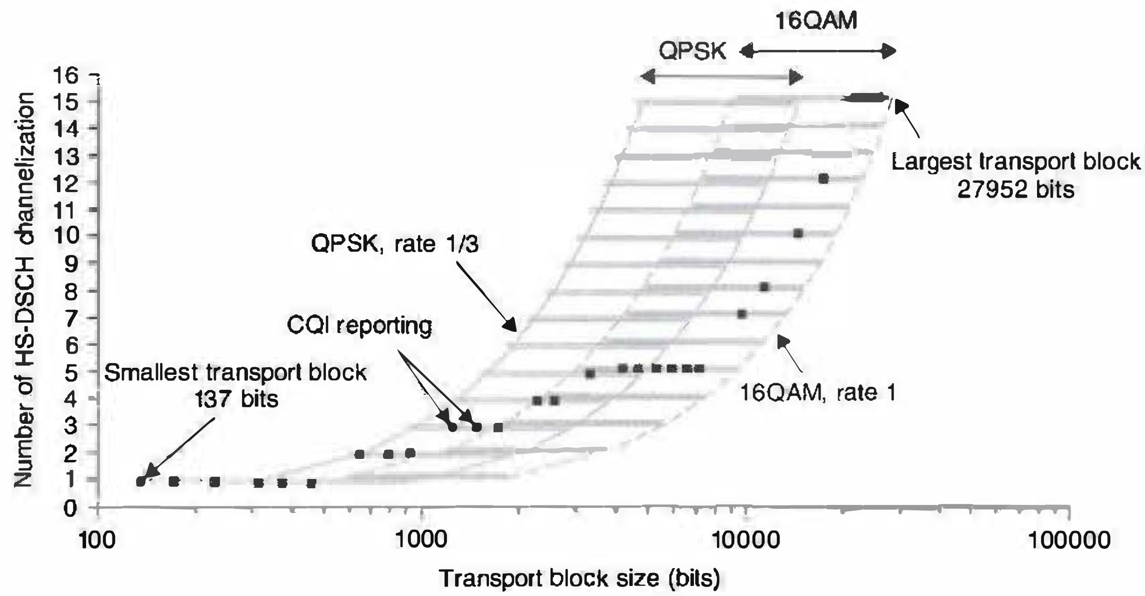


**Figure 9.8** *Transport-block sizes vs. the number of channelization codes for QPSK and 16QAM modulation. The transport-block sizes used for CQI reporting are also illustrated.*

to both the UE and the NodeB. Thus, there is no need for (re)configuration of transport-block sizes at channel setup or when switching serving cell, which reduces the amount of overhead associated with mobility. Each combination of HS-DSCH channelization codes and modulation scheme defines a subset containing 63 out of the 254 different transport-block sizes and the 6-bit 'HS-DSCH transport-block size information' indicates one out of the 63 transport-block sizes possible for this subset. With this scheme, transport-block sizes in the range of 137–27 952 bits can be signaled, with channel-coding rates ranging from 1/3 up to 1.

For retransmissions, there is a possibility for a code rate >1. This is achieved by exploiting the fact that the transport-block size cannot change between transmission and retransmission. Hence, instead of signaling the transport-block size for the retransmission, a reserved value can be used, indicating that no transport-block-size information is provided by the HS-SCCH and the value from the original transmission should be used. This is useful for additional scheduling flexibility, for example to retransmit only a small amount of parity bits in case the latest CQI report indicates that the UE was 'almost' able to decode the original transmission.

As stated in the introduction, the primary way of adapting to rapid variations in the instantaneous channel quality is rate control as no fast closed-loop power control is specified for HS-DSCH. This does not imply that the HS-DSCH transmission power cannot change for other reasons, for example due to variations in the power required by other downlink channels. In Figure 9.4 on page 147, an example of a dynamic HS-DSCH power allocation scheme is illustrated, where the HS-DSCH uses all power not used by other channels. Of course, the overall interference created in the cell must be taken into account when allocating the amount of HS-DSCH power. This is the responsibility of the radio-resource control in the RNC, which can set an upper limit on the power used by the NodeB for the HS-DSCHs and all HS-SCCHs.[1] As long as the NodeB stays within this limit, the power allocation for HSDPA is up to the NodeB implementation. Corresponding measurements, used by the NodeB to report the current power usage to the RNC are also defined. Knowledge about the amount of power used for non-HSDPA channels is useful to the admission control functionality in the RNC. Without this knowledge, the RNC would not be able to determine whether there are resources left for non-HSDPA users trying to enter the cell.

Unlike QPSK, the demodulation of 16QAM requires an amplitude reference at the UE. How this is achieved is implementation specific. One possibility is to use a channel estimate formed from the common pilot and obtain the ratio between the HS-DSCH and common pilot received powers through averaging over 2ms. The instantaneous amplitude estimate necessary for 16QAM demodulation can then be obtained from the common pilot and the estimated offset. This is the reason for the bit level scrambling prior Turbo coding in Figure 9.6; with scrambling both the inner and outer signal points in the 16QAM constellation will be used with a high probability and an accurate estimate of the received HS-DSCH power can be formed.

What criteria to use for the rate control, that is, the transport-format selection process in the MAC-hs, are implementation specific and not defined in the standard. Principally, the target for the rate control is to select a transport format resulting in transmitting an as large transport block as possible with a reasonable error probability, given the instantaneous channel conditions. Naturally, selecting a transport-block size larger than the amount of data to be transmitted in a TTI is not useful, regardless of whether the instantaneous radio conditions allows for a larger transport block to be transmitted. Hence, the transport-format selection does not only depend on the instantaneous radio conditions, but also on the instantaneous source traffic situation.

Since the rate control typically depends on the instantaneous channel conditions, rate control relies on the same estimate of the instantaneous radio quality

[1] If the cell is configured to support E-DCH as well, this limit also covers the power used for the related E-DCH downlink control signaling. See Chapter 10.

at the UE as the scheduler. As discussed above, this knowledge is typically obtained from the CQI although other quantities may also be useful. This is further elaborated upon in Section 9.3.6.

### 9.2.5 Hybrid ARQ with soft combining

The hybrid-ARQ functionality spans both the MAC-hs and the physical layer. As the MAC-hs is located in the NodeB, erroneous transport blocks can be rapidly retransmitted. Hybrid-ARQ retransmissions are therefore significantly less costly in terms of delay compared to RLC-based retransmissions. There are two fundamental reasons for this difference:

- There is no need for signaling between the NodeB and the RNC for the hybrid-ARQ retransmission. Consequently, any Iub/Iur delays are avoided for retransmissions. Handling retransmission in the NodeB is also beneficial from a pure Iub/Iur capacity perspective; hybrid-ARQ retransmissions come at no cost in terms of transport-network capacity.
- The RLC protocol is typically configured with relatively infrequent status reports of erroneous data blocks (once per several TTIs) to reduce the signaling load, while the HSDPA hybrid-ARQ protocol allows for frequent status reports (once per TTI), thus reducing the roundtrip time.

In HSDPA, the hybrid ARQ operates per transport block or, equivalently, per TTI that is, whenever the HS-DSCH CRC indicates an error, a retransmission representing the same information as the original transport block is requested. As there is a single transport block per TTI, the content of the whole TTI is retransmitted in case of an error. This reduces the amount of uplink signaling as a single ACK/NAK bit per TTI is sufficient. Furthermore, studies during the HSDPA design phase indicated that the benefits of having multiple transport blocks per TTI with the possibility for individual retransmissions were quite small. A major source of transmission errors are sudden interference variations on the channel and errors in the link-adaptation mechanism. Thanks to the short TTI, the channel is relatively static during the transmission of a transport block and in most cases errors are evenly distributed over the TTI. This limits the potential benefits of individual retransmissions.

Incremental redundancy is the basic scheme for soft combining, that is, retransmissions may consist of a different set of coded bits than the original transmission. Different redundancy versions, that is, different sets of coded bits, are generated as part of the rate-matching mechanism. The rate matcher uses puncturing (or repetition) to match the number of code bits to the number

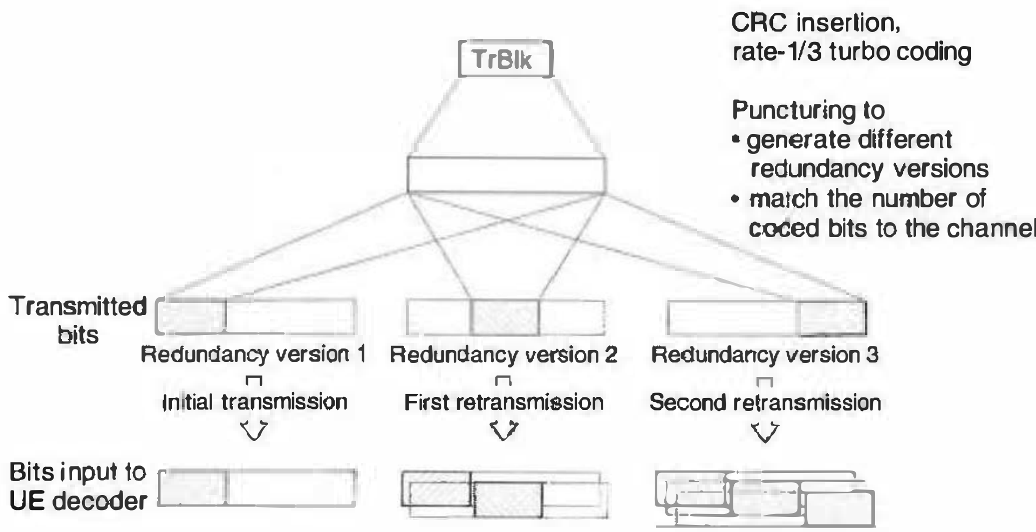


**Figure 9.9** *Generation of redundancy versions.*

of physical channel bits available. By using different puncturing patterns, different sets of coded bits, that is different redundancy versions, result. This is illustrated in Figure 9.9. Note that Chase combining is a special case of incremental redundancy; the NodeB decides whether to use incremental redundancy or Chase combining by selecting the appropriate puncturing pattern for the retransmission.

The UE receives the coded bits and attempts to decode them. In case the decoding attempts fails, the UE buffers the received soft bits and requests a retransmission by sending a NAK. Once the retransmission occurs, the UE combines the buffered soft bits with the received soft bits from the retransmission and tries to decode the combination.

For soft combining to operate properly, the UE need to know whether the transmission is a retransmission of previously transmitted data or whether it is transmission of new data. For this purpose, the downlink control signaling includes a new-data indicator, used by the UE to control whether the soft buffer should be cleared (the current transmission is new data) or whether soft combining of the soft buffer and the received soft bits should take place (retransmission).

To minimize the delay associated with a retransmission, the outcome of the decoding in the UE should be reported to the NodeB as soon as possible. At the same time, the amount of overhead from the feedback signaling should be minimized. This lead to the choice of a stop-and-wait structure for HSDPA, where a single bit is transmitted from the UE to the NodeB a well-specified time,

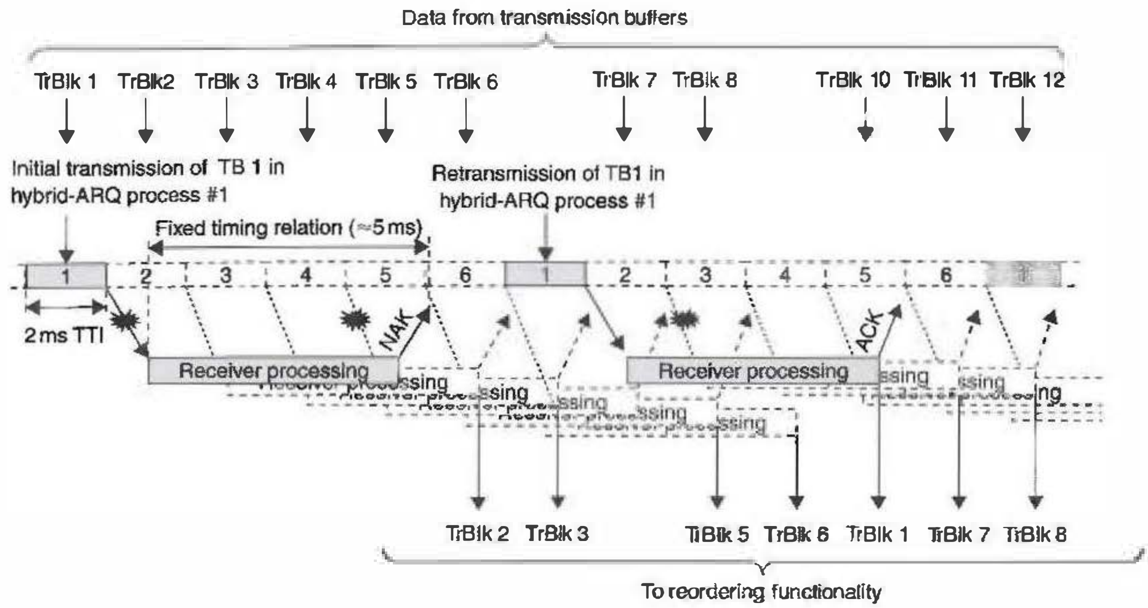


**Figure 9.10** *Multiple hybrid-ARQ process (six in this example).*

approximately 5 ms, after the reception of a transport block. To allow for continuous transmission to a single UE, multiple stop-and-wait structures, or *hybrid-ARQ processes*, are operated in parallel as illustrated in Figure 9.10. Hence, for each user there is one *hybrid-ARQ entity*, each consisting of *multiple* hybrid-ARQ processes.

The number of hybrid-ARQ processes should match the roundtrip time between the UE and NodeB, including their respective processing time, to allow for continuous transmission to a UE. Using a larger number of processes than motivated by the roundtrip time does not provide any gains but introduces unnecessary delays between retransmissions.

Since the NodeB processing time may differ between different implementations, the number of hybrid-ARQ processes is configurable. Up to eight processes can be set up for a user, although a typical number of processes is six. This provides approximately 2.8 ms of processing time in the NodeB from the reception of the ACK/NAK until the NodeB can schedule a (re)transmission to the UE in the same hybrid-ARQ process.

Downlink control signaling is used to inform the UE which of the hybrid-ARQ processes that is used for the current TTI. This is important information to the UE as it is needed to do soft combining with the correct soft buffer; each hybrid-ARQ process has its own soft buffer.

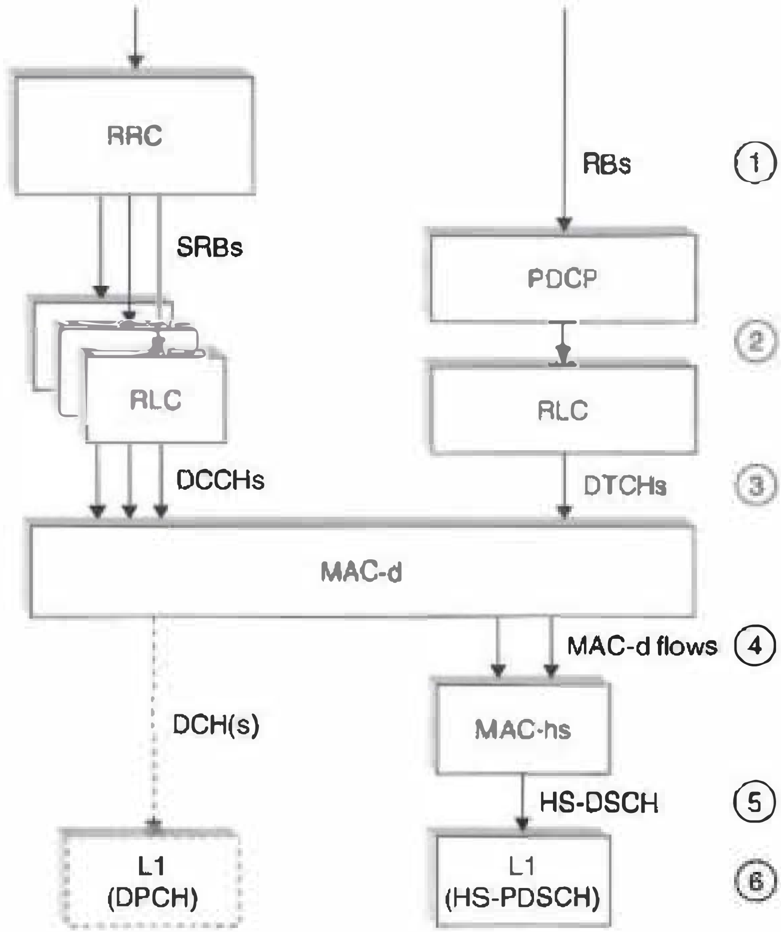


**Figure 9.11** *Protocol configuration when HS-DSCH is assigned. The numbers in the rightmost part of the figure corresponds to the numbers to the right in Figure 9.12.*

One result of having multiple independent hybrid-ARQ processes operated in parallel is that decoded transport blocks may appear out-of-sequence. For example, a retransmission may be needed in hybrid-ARQ process number one, while process number two did successfully receive the data after the first transmission attempts. Therefore, the transport block transmitted in process number two will be available for forwarding to higher layers at the receiver side before the transport block transmitted in process number one, although the transport blocks were originally transmitted in a different order. This is illustrated in Figure 9.10. As the RLC protocol assumes data to appear in the correct order, a reordering mechanism is used between the outputs from the multiple hybrid-ARQ processes and the RLC protocol. The reordering mechanism is described in more detail in Section 9.3.4.

## 9.2.6 Data flow

To illustrate the flow of user data through the different layers, an example radio-interface protocol configuration is shown in Figure 9.11. For the UE in this example, an IP-based service is assumed, where the user data is mapped to the HS-DSCH.

For signaling purposes in the radio network, several signaling radio bearers are configured in the control plane. In Release 5, signaling radio bearers cannot be mapped to the HS-DSCH, and consequently dedicated transport channels must be used, while this restriction is removed in Release 6 to allow for operation completely without dedicated transport channels in the downlink.

Figure 9.12 illustrates the data flow at the reference points shown in Figure 9.11. In this example an IP-based service is assumed. The PDCP performs (optional) IP header compression. The output from the PDCP is fed to the RLC protocol entity. After possible concatenation, the RLC SDUs are segmented into smaller blocks of typically 40 bytes. An RLC PDU is comprised of a data segment and the RLC header. If logical-channel multiplexing is performed in MAC-d, a 4-bit header is added to form a MAC-d PDU. In MAC-hs, a number of MAC-d PDUs, possibly of variable size, are assembled and a MAC-hs header is attached

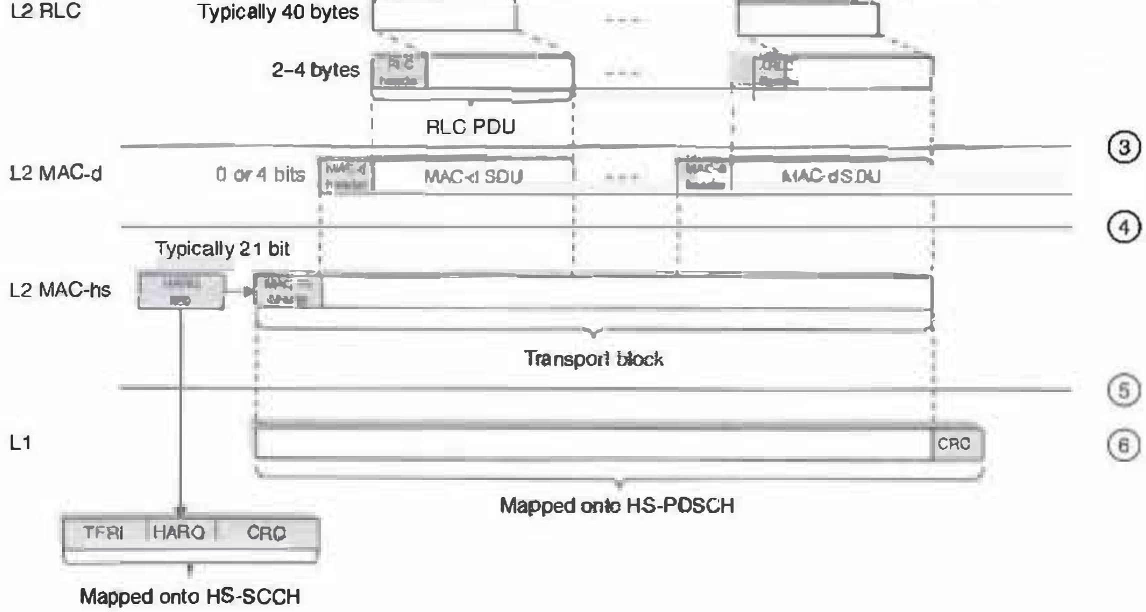


**Figure 9.12** *Data flow at UTRAN side.*

to form one transport block, subsequently coded and transmitted by the physical layer.

### 9.2.7 Resource control for HS-DSCH

With the introduction of HSDPA, parts of the radio resource management are handled by the NodeB instead of the RNC. This is a result of introducing channel-dependent scheduling and rate control in the NodeB in order to exploit rapid channel variations. However, the RNC still has the overall responsibility for radio-resource management, including admission control and handling of inter-cell interference. Therefore, new measurement reports from the NodeB to the RNC have been introduced, as well as mechanisms for the RNC to set the limits within which the NodeB are allowed to handle the HSDPA resources[2] in the cell.

To limit the transmission power used for HSDPA, the RNC can set the maximum amount of power the NodeB is allowed to use for HSDPA-related downlink transmissions. This ensures that the RNC has control of the maximum amount of interference a cell may generate to neighboring cells. Within the limitation set by the RNC, the NodeB is free to manage the power spent on the HSDPA downlink channels. If the quantity is absent (or larger than the tot al NodeB power), the NodeB may use all available power for downlink transmissions on the HS-DSCH and HS-SCCH.

Admission control in the RNC needs to take into account the amount of power available in the NodeB. Only if there is a sufficient amount of transmission power available in the NodeB can a new user be admitted into the cell. The *Transmitted carrier power* measurement is available for this purpose. However, with the introduction of HSDPA, the NodeB can transmit at full power, even with a single user in the cell, to maximize the data rates. To the admission control in the RNC, it would appear as the cell is fully loaded and no more users would be admitted. Therefore, a new measurement, *Transmitted carrier power of all codes not used for HS-PDSCH or HS-SCCH,* is introduced, which can be used in admission control to determine whether new users can be admitted into the cell or not (Figure 9.13).

In addition to the power-related signaling discussed above, there is also signaling useful to support streaming services. To efficiently support streaming, where

[2]Note that many of these measurements were extended in Rel6 to include Enhanced Uplink Downlink control channels in addition to the HSDPA-related channels.

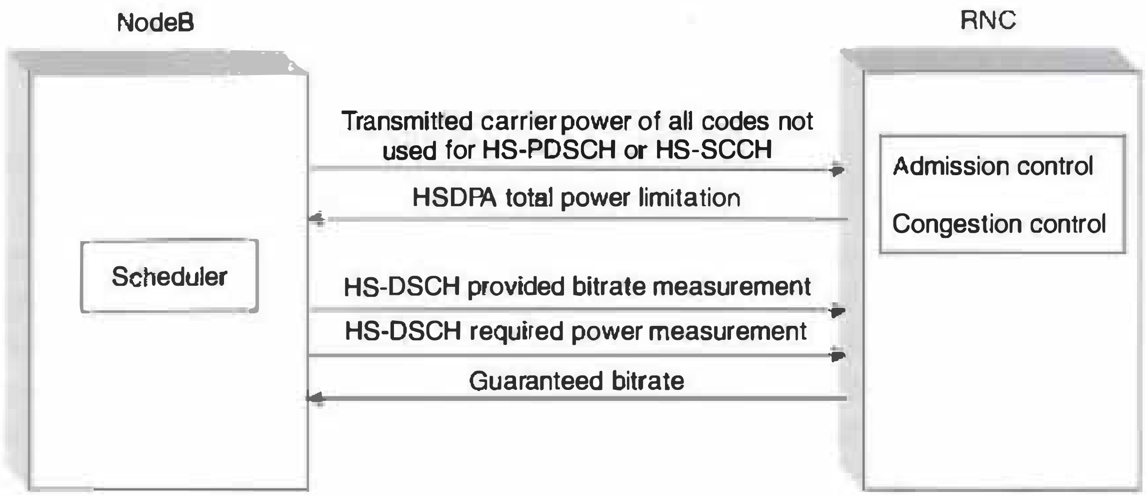


**Figure 9.13** *Measurements and resource limitations for HSDPA.*

a certain minimum data rate needs to be provided on average, the RNC can signal the *MAC-hs Guaranteed Bit Rate*. The scheduler can use this information to ensure that, averaged over a longer period of time, a sufficiently high data rate is provided for a certain MAC-d priority queue. To monitor the fulfillment of this, and to be able to observe the load in the cell due to these restrictions, the NodeB can report the required transmission power for each priority class configured by the RNC in order to identify 'costly' UEs. The NodeB can also report the data rate, averaged over 100 ms, it actually provides for each priority class.

## 9.2.8 Mobility

Mobility for HSDPA, that is, change of serving cell, is handled through RRC signaling using similar procedures as for dedicated channels. The basics for mobility are network-controlled handover and UE measurement reporting. Measurements are reported from the UE to the RNC, which, based on the measurements, reconfigures the UE and involved NodeBs, resulting in a change of serving cell.

Several measurement mechanisms are specified already in the first release of WCDMA and used for, for example, active set update, hard handover, and intra-frequency measurements. One example is Measurement Event 1D, 'change of best cell,' which is reported by the UE whenever the common-pilot strength from a neighboring cell (taking any measurement offsets into account) becomes stronger than for the current best cell. This can be used to determine when to switch the HS-DSCH serving cell as illustrated in Figure 9.14. Updates of the active set are not included in this example; it is assumed that both the source serving cell and the target serving cell are part of the active set.

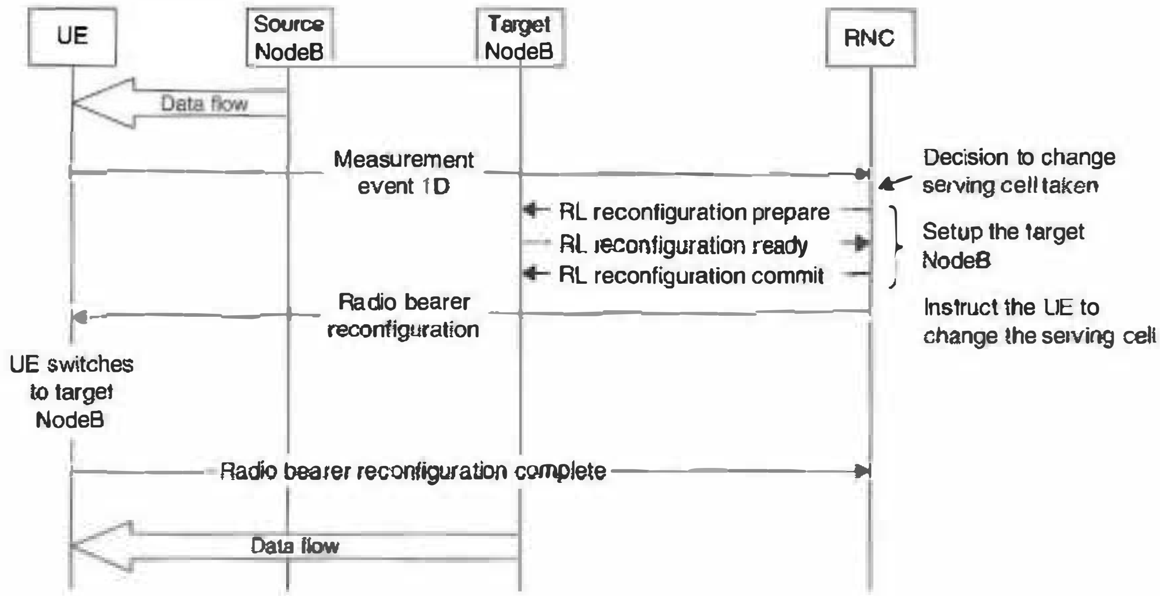


**Figure 9.14** *Change of serving cell for HSPA. It is assumed that both the source and target NodeB are part of the active set.*

The reconfiguration of the UE and involved NodeBs can be either synchronous or asynchronous. With synchronous reconfiguration, an activation time is defined in the reconfiguration message, ensuring that all involved parties change their reconfiguration at the same time. Due to unknown delays between the NodeB and the RNC, as well as processing and protocol delays, a suitable margin may need to be taken into account in the choice of activation time. Asynchronous reconfiguration implies that the involved nodes obey the reconfiguration message as soon it is received. However, in this case, data transmission from the new cell may start before the UE has been switched from the old cell, which would result in some data loss that has to be retransmitted by the RLC protocol. Hence, synchronous reconfigurations are typically used for HS-DSCH serving cell change. The MAC-hs protocol is reset when moving from one NodeB to another. Thus the hybrid-ARQ protocol state is not transferred between the two NodeBs. Any packet losses at the time of cell change are instead handled by the RLC protocol.

Related to mobility is the flow control between the NodeB and the RNC, used to control the amount of data buffered in the MAC-hs in the NodeB and avoid overflow in the buffers. The requirements on the flow control are, to some extent, conflicting as it shall ensure that MAC-hs buffers should be large enough to contain a sufficient amount of data to fully utilize the physical channel resources (in case of advantageous channel conditions), while at the same time MAC-hs buffers should be kept as small as possible to minimize the amounts of packets that need to be resent to a new NodeB in case of inter-NodeB handover.

### 9.2.9 UE categories

To allow for a range of UE implementations, different UE capabilities are specified. The UE capabilities are divided into a number of parameters, which are sent from the UE at the establishment of a connection and if/when the UE capabilities are changed during an ongoing connection. The UE capabilities may then be used by the network to select a configuration that is supported by the UE. Several of the UE capabilities applicable to other channels are valid for HS-DSCH as well, but there are also some HS-DSCH-specific capabilities.

Basically, the physical-layer UE capabilities are used to limit the requirements for three different UE resources: the despreading resource, the soft-buffer memory used by the hybrid-ARQ functionality, and the Turbo decoder. The despreading resource is limited in terms of the maximum number of HS-PDSCH codes the UE simultaneously needs to despread. Three different capabilities exist in terms of de-spreading resources, corresponding to the capability to despread a maximum of 5, 10, or 15 physical channels (HS-PDSCH).

The amount of soft-buffer memory is in the range of 14 400--172 800 soft bits, depending on the UE category. Note that this is the total amount of buffer memory for all hybrid-ARQ processes, not the value per process. The memory is divided among the multiple hybrid-ARQ processes, typically with an equal amount of memory per process although non-equal allocation is also possible.

The requirements on the Turbo-decoding resource are defined through two parameters: the maximum number of transport-channel bits that can be received within an HS-DSCH TTI and the minimum inter-TTI interval, that is the distance in time between subsequent transport blocks. The decoding time in a Turbo decoder is roughly proportional to the number of information bits which thus provides a limit on the required processing speed. In addition, for low-end UEs, there is a possibility to avoid continuous data transmission by specifying an inter-TTI interval larger than one.

In order to limit the number of possible combinations of UE capabilities and to avoid parameter combinations that do not make sense, the UE capability parameters relevant for the physical layer are lumped into 12 different *categories* as illustrated in Table 9.1.

## 9.3 Finer details of HSDPA

### 9.3.1 Hybrid ARQ revisited: Physical-layer processing

Hybrid ARQ with soft combining has been described above, although some details of the physical-layer and protocol operation were omitted in order to

**Table 9.1** *HSDPA UE categories [99].*

| HS-DSCH category | Maximum number of HS-DSCH codes received | Minimum inter-TTI interval | Maximum transport-block size soft bits | | Maximum number of schemes | Supported modulation |
|---|---|---|---|---|---|---|
| 1 | 5 | 3 | 7298 | (3.6 Mbit/s) | 19200 | 16QAM, QPSK |
| 2 | 5 | 3 | 7298 | (3.6 Mbit/s) | 28800 | 16QAM, QPSK |
| 3 | 5 | 2 | 7298 | (3.6 Mbit/s) | 28800 | 16QAM, QPSK |
| 4 | 5 | 2 | 7298 | (3.6 Mbit/s) | 38400 | 16QAM, QPSK |
| 5 | 5 | 1 | 7298 | (3.6 Mbit/s) | 57600 | 16QAM, QPSK |
| 6 | 5 | 1 | 7298 | (3.6 Mbit/s) | 67200 | 16QAM, QPSK |
| 7 | 10 | 1 | 14411 | (7.2 Mbit/s) | 115200 | 16QAM, QPSK |
| 8 | 10 | 1 | 14411 | (7.2 Mbit/s) | 134400 | 16QAM, QPSK |
| 9 | 15 | 1 | 20251 | (10.1 Mbit/s) | 172800 | 16QAM, QPSK |
| 10 | 15 | 1 | 27952 | (14 Mbit/s) | 172800 | 16QAM,QPSK |
| 11 | 5 | 2 | 3630 | (1.8 Mbit/s) | 14400 | QPSK |
| 12 | 5 | 1 | 3630 | (1.8 Mbit/s) | 28800 | QPSK |

simplify the description. This section provides a more detailed description of the processing, aiming at filling the missing gaps.

As already mentioned, the hybrid ARQ operates on a single transport block, that is, whenever the HS-DSCH CRC indicates an error, a retransmission representing the same information as the original transport block is requested. Since there is a single transport block per TTI, this implies that it is not possible to mix transmissions and retransmissions within the same TTI.

Since incremental redundancy is the basic hybrid-ARQ soft-combining scheme, retransmissions generally consist of a different set of coded bits. Furthermore, the modulation scheme, the channelization-code set, and the transmission power can be different compared to the original transmission. Incremental redundancy generally has better performance, especially for high initial code rates, but poses higher requirements on the soft buffering in the UE since soft bits from all transmission attempts must be buffered prior to decoding. Therefore, the NodeB needs to have knowledge about the soft-buffer size in the UE (for each active hybrid-ARQ process). Coded bits that do not fit within the soft buffer shall not be transmitted. For HSDPA, this problem is solved through the use of *two-stage rate matching*. The first rate-matching stage limits the number of coded bits to what is possible to fit in the soft buffer, while the second rate-matching stage generates the different redundancy versions.

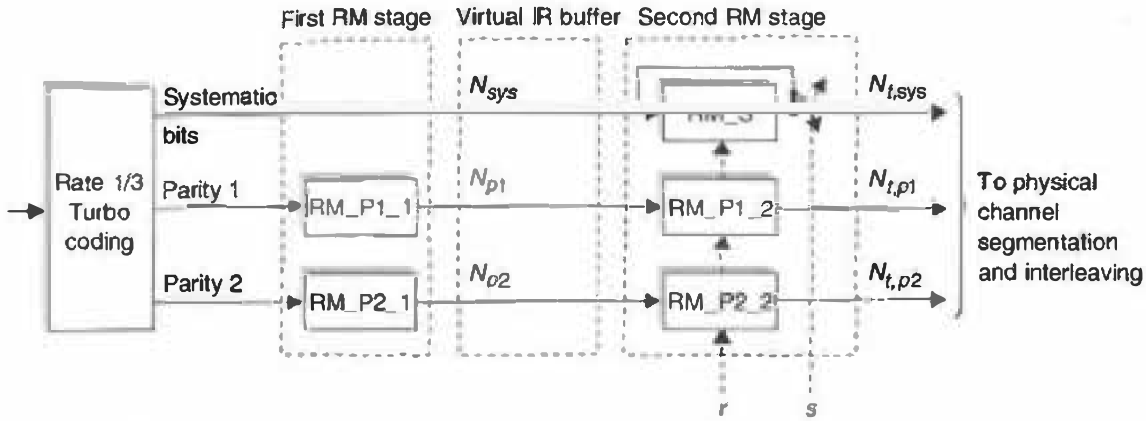


**Figure 9.15** *The principle of two-stage rate matching.*

Each rate-matching stage uses several identical rate-matching blocks, denoted RM in Figure 9.15. An RM block can be configured to puncture or repeat every *n*th bit.

The first rate-matching stage is used to limit the number of coded bits to the available UE soft buffer for the hybrid-ARQ process currently being addressed. A sufficient number of coded bits are punctured to ensure that all coded bits at the output of the first rate-matching stage will fit in the soft buffer (known as *virtual IR buffer* at the transmitter side). Hence, depending on the soft-buffer size in the UE, the lowest code rate may be higher than the rate-1/3 mother code rate in the Turbo coder. Note that, if the number of bits from the channel coding does not exceed the UE soft-buffering capability, the first rate-matching stage is transparent and no puncturing is performed.

The second rate-matching stage serves two purposes:

- Matching the number of bits in the virtual IR buffer to the number of available channel bits. The number of available channel bits is given by the size of the channelization-code set and the modulation scheme selected for the TTI.
- Generating different sets of coded bits as controlled by the two redundancy-version parameters $r$ and $s$, described below.

Equal repetition for all three streams is applied if the number of available channel bits is larger than the number of bits in the virtual IR buffer, otherwise puncturing is applied.

To support full incremental redundancy, that is, to have the possibility to transmit only/mainly parity bits in a retransmission, puncturing of systematic bits is possible as controlled by the parameter $s$. Setting $s = 1$ implies that the systematic bits are prioritized and puncturing is primarily applied with an equal amount to the two parity-bit streams. On the other hand, for a transmission prioritizing the parity bits, $s = 0$ and primarily the systematic bits are punctured. If, for a transmission prioritizing the systematic bits, the number of coded bits is larger than the number of physical channel bits, despite all the parity bits have been punctured, further puncturing is applied to the systematic bits. Similarly, if puncturing the systematic bits is not sufficient for a transmission prioritizing the parity bits, puncturing is applied to the parity bits as well.

For good performance, all systematic bits should be transmitted in the initial transmission, corresponding to $s = 1$, and the code rate should be set to less than one. For the retransmission (assuming the initial transmission did not succeed), different strategies can be applied. If the NodeB received neither ACK, nor NAK, in response to the initial transmission attempt, the UE may have missed the initial transmission. Setting $s = 1$ also for the retransmission is therefore appropriate. This is also the case if NAK is received and Chase combining is used for retransmissions. However, if a NAK is received and incremental redundancy is used, that is, the parity bits should be prioritized, setting $s = 0$ is appropriate.

The parameter $r$ controls the puncturing pattern in each rate-matching block in Figure 9.15 and determines which bits to puncture. Typically, $r = 0$ is used for the initial transmission attempt. For retransmissions, the value of $r$ is typically increased, effectively leading to a different puncturing pattern. Thus, by varying $r$, multiple, possibly partially overlapping, sets of coded bits representing the same set of information bits can be generated. It should be noted that changing the number of channel bits by changing the modulation scheme or the number of channelization codes also affects which coded bits that are transmitted even if the $r$ and $s$ parameters are unchanged between the transmission attempts.

With the two-stage rate-matching scheme, both incremental redundancy and Chase combining can easily be supported. By setting $s = 1$ and $r = 0$ for all transmission attempts, the same set of bits is used for the retransmissions as for the original transmission, that is Chase combining. Incremental redundancy is easily achieved by setting $s = 1$ and using $r = 0$ for the initial transmission, while retransmissions use $s = 0$ and $r > 0$. Partial IR, that is, incremental redundancy with the systematic bits included in each transmission, results if $s = 1$ for all the retransmissions as well as the initial transmission.

In Figure 9.16, a simple numerical example is shown to further illustrate the operation of the physical-layer hybrid-ARQ processing of data. Assume that,

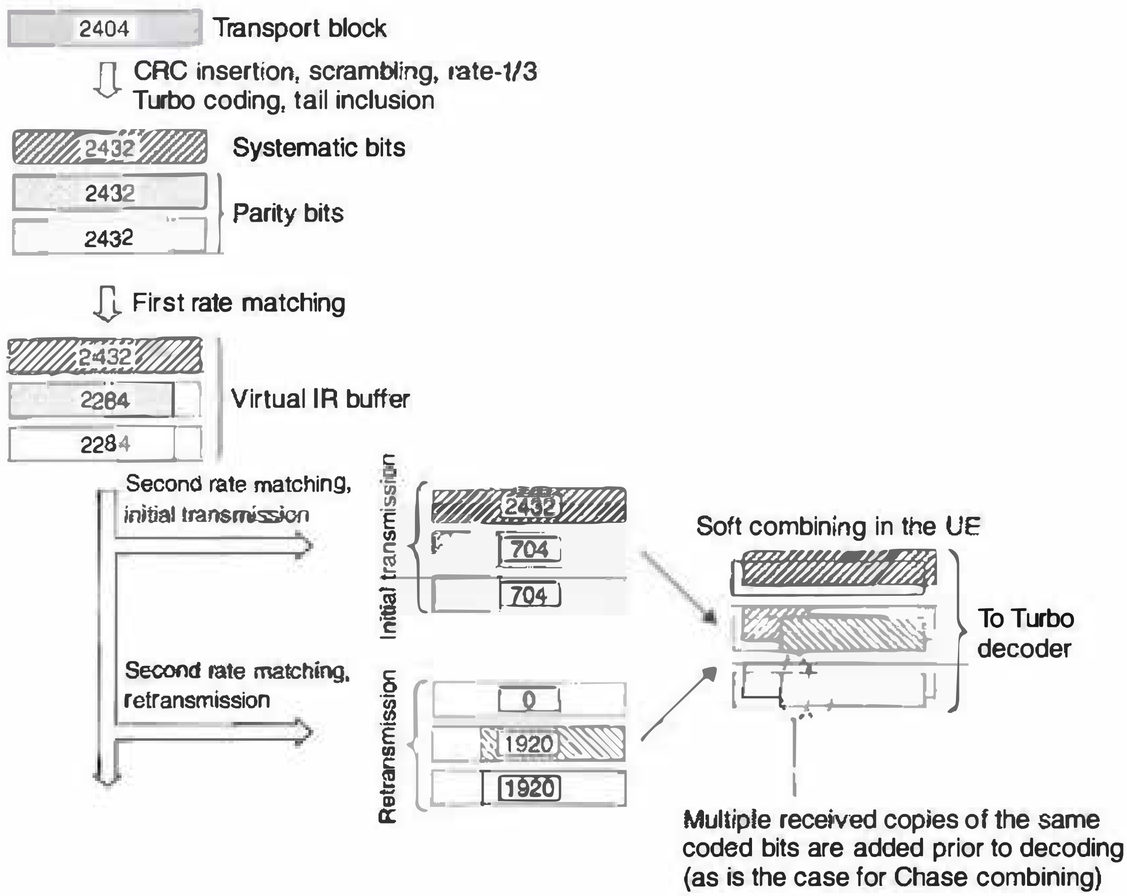


**Figure 9.16** *An example of the generation of different redundancy versions in the case of IR. The numbers indicate the number of bits after the different stages using the example case in the text.*

as an example, a transport block of 2404 bits is to be transmitted using one of the hybrid-ARQ processes. Furthermore, assume the hybrid-ARQ process in question is capable of buffering at most 7000 soft values due to memory limitations in the UE and the soft memory configuration set by higher layers. Finally, the channel can carry 3840 coded bits in this example (QPSK modulation, 4 channelization codes).

A 24-bit CRC is appended to the transport block, rate-1/3 Turbo coding is applied and a 12-bit tail appended, resulting in 7296 coded bits. The coded bits are fed to the first stage rate matching, which punctures parity bits such that 2432 systematic bits and $2 \times 2284$ parity bits, in total 7000 bits, are fed to the second-stage rate-matching block. Since at most 7000 coded bits can be transmitted, the lowest possible code rate is $2432/7000 = 0.35$, which is slightly higher than the mother code's rate of 1/3 due to the limited soft buffer in the UE.

For the initial transmission, the second-stage rate matching matches the 7000 coded bits to the 3840 channel bits by puncturing the parity bits only. This is achieved by using $r = 0$ and $s = 1$, that is, a self-decodable transmission, and the resulting code rate is 2432/3840 = 0.63.

For retransmissions, either Chase combining or incremental redundancy can be used, as chosen by the NodeB. If Chase combining is used by setting $s = 1$ and $r = 0$, the same 3840 bits as used for the initial transmission are retransmitted (assuming unchanged modulation scheme and channelization-code set). The resulting effective code rate remains 0.63 as no additional parity has been transmitted, but an energy gain has been obtained as, in total, twice the amount of energy has been transmitted for each bit. Note that this example assumed identical transport formats for the initial transmission and the retransmission.

If incremental redundancy is used for the retransmission, for example, by using $s = 0$ and $r = 1$, the systematic bits are punctured and only parity bits are retransmitted, of which 3840 (out of 4568 parity bits available after the first stage rate matching) fit into the physical channel. Note that some of these parity bits were included already in the original transmission as the number of unique parity bits is not large enough to fill both the original transmission and the retransmissions. After the retransmission, the resulting code rate is 2432/7000 = 0.35. Hence, contrary to Chase combining, there is a coding gain in addition to the energy gain.

### *9.3.2 Interleaving and constellation rearrangement*

For 16QAM, two of the four bits carried by each modulation symbol will be more reliable at the receiver due to the difference in the number of nearest neighbors in the constellation. This is in contrast to QPSK, where both bits are of equal reliability. Furthermore, for Turbo codes, systematic bits are of greater importance in the decoding process, compared to parity bits. Hence, it is desirable to map as many of the systematic bits as possible to the more reliable positions in a 16QAM symbol. A dual interleaver scheme, illustrated in Figure 9.17, has been adopted for HS-DSCH in order to control the mapping of systematic and parity bits onto the 16QAM modulation symbols.

For QPSK, only the upper interleaver in Figure 9.17 is used, while for 16QAM, two identical interleavers are used in parallel. Systematic bits are primarily fed into the upper interleaver, whereas parity bits are primarily fed into the lower interleaver. The 16QAM constellation is defined such that the output from the upper interleaver is mapped onto the reliable bit positions and the output from the lower interleaver onto the less reliable positions.

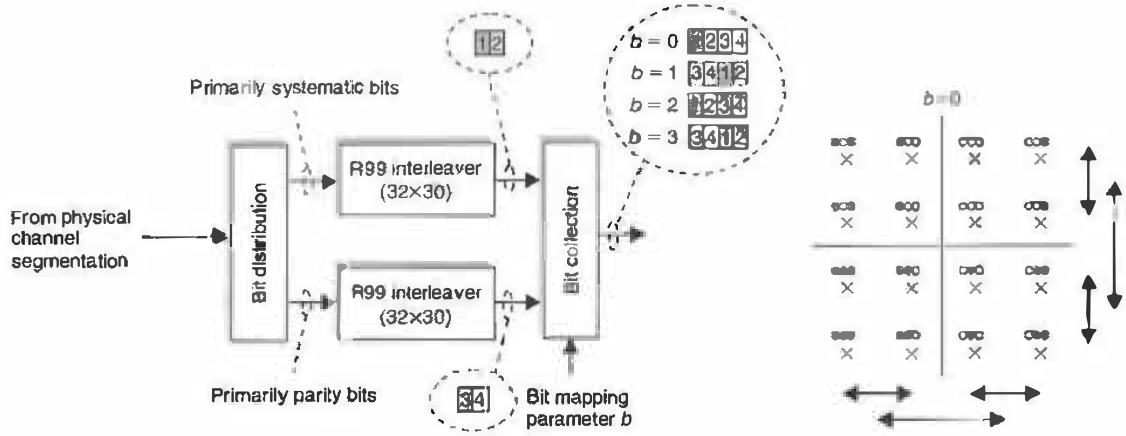


**Figure 9.17** *The channel interleaver for the HS-DSCH. The shaded parts are only used for 16QAM. Colors illustrate the mapping order for a sequence of 4 bits, where a bar on top of the figure denotes bit inversion.*

If 16QAM is used in conjunction with hybrid ARQ using Chase combining, there is a performance gain by rearranging the 16QAM symbol constellations between multiple transmission attempts as this provides an averaging effect among the reliability of the bits. However, note that this gain is only available for retransmissions and not for the initial transmission. Furthermore, the gains with constellation rearrangement in combination with incremental redundancy are minor. Hence, its use is mainly applicable when Chase combining is used.

Constellation rearrangement is obtained through bit manipulations in the bit collector block and is controlled by a four-state bit mapping parameter, controlling two independent operations. First, the output of the two interleavers can be swapped. Second, the output of the lower interleaver (or the upper interleaver if swapping is used) can be inverted. In essence, this results in the selection of one out of four different signal constellations for 16QAM.

## 9.3.3 Hybrid ARQ revisited: Protocol operation

As stated earlier, each hybrid-ARQ entity is capable of supporting multiple (up to eight) stop-and-wait hybrid-ARQ processes. The motivation behind this is to allow for continuous transmission to a single UE, which cannot be achieved by a single stop-and-wait scheme. The number of hybrid-ARQ processes is configurable by higher-layer signaling. Preferably, the number of hybrid-ARQ processes is chosen to match the roundtrip time, consisting of the TTI itself, any radio-interface delay in downlink and uplink, the processing time in the UE, and the processing time in the NodeB.

The protocol design assumes a well-defined time between the end of the received transport block and the transmission of the ACK/NAK as discussed in

Section 9.2.5. In essence, this time is the time the UE has available for decoding of the received data. From a delay perspective, this time should be as small as possible, but a too small value would put unrealistic requirements on the UE processing speed. Although in principle the time could be made a UE capability, this was not felt necessary and a value of 5 ms was agreed as a good trade-off between performance and complexity. This value affects the number of hybrid-ARQ processes necessary. Typically, a total of six processes are configured, which leaves around 2.8 ms for processing of retransmissions in the NodeB.

Which of the hybrid-ARQ processes that is used for the current transmission is controlled by the scheduler and explicitly signaled to the UE. Note that the hybrid-ARQ processes can be addressed in any order. The amount of soft-buffering memory available in the UE is semi-statically split between the different hybrid-ARQ processes. Thus the larger the number of hybrid-ARQ processes is, the smaller the amount of soft-buffer memory available to a hybrid-ARQ process for incremental redundancy. The split of the total soft-buffer memory between the hybrid-ARQ processes is controlled by the RNC and does not necessarily have to be such that the soft-buffer memory per hybrid-ARQ process is the same. Some hybrid-ARQ processes can be configured to use more soft-buffer memory than others, although the typical case is to split the available memory equally among the processes.

Whenever the current transmission is not a retransmission, the NodeB MAC-hs increments the single-bit new-data indicator. Hence, for each new transport block, the bit is toggled. The indicator is used by the UE to clear the soft buffer for initial transmissions since, by definition, no soft combining should be done for an initial transmission. The indicator is also used to detect error cases in the status signaling, for example, if the 'new-data' indicator is not toggled despite the fact that the previous data for the hybrid-ARQ process in question was correctly decoded and acknowledged, an error in the uplink signaling has most likely occurred. Similarly, if the indicator is toggled but the previous data for the hybrid-ARQ process was not correctly decoded, the UE will replace the data previously in the soft buffers with the new received data.

Errors in the status (ACK/NAK) signaling will impact the overall performance. If an ACK is misinterpreted as a NAK, an unnecessary hybrid-ARQ retransmission will take place, leading to a (small) reduction in the throughput. On the other hand, misinterpreting a NAK as an ACK will lead to loss of data as the NodeB will not perform a hybrid-ARQ retransmission despite the UE was not able to successfully decode the data. Instead, the missing data has to be retransmitted by the RLC protocol, a more time-consuming procedure than hybrid-ARQ retransmissions. Therefore, the requirements on the ACK/NAK errors are typically asymmetric with Pr{NAK|ACK} = $10^{-2}$ and Pr{ACK|NAK} = $10^{-3}$ (or $10^{-4}$) as typical values. With these error probabilities, the impact on the end-user TCP performance due to hybrid-ARQ signaling errors is small [75].

### 9.3.4 In-sequence delivery

The multiple hybrid-ARQ processes cannot themselves ensure in-sequence delivery as there is no interaction between the processes. Hence, in-sequence delivery must be implemented on top of the hybrid-ARQ processes and a reordering queue in the UE MAC-hs is used for this purpose. Related to the reordering queues in the UE are the priority queues in the NodeB, used for handling priorities in the scheduling process.

The NodeB MAC-hs receives MAC-d PDUs in one or several MAC-d flows. Each such MAC-d PDU has a priority assigned to it and MAC-d PDUs with different priorities can be mixed in the same MAC-d flow. The MAC-d flows are split if necessary and the MAC-d PDUs are sorted into priority queues as illustrated in Figure 9.18. Each priority queue corresponds to a certain MAC-d flow and a certain MAC-d priority, where RRC signaling is used to set up the mapping between the priority queues and the MAC-d flows. Hence, the scheduler in the MAC-hs can, if desired, take the priorities into account when making

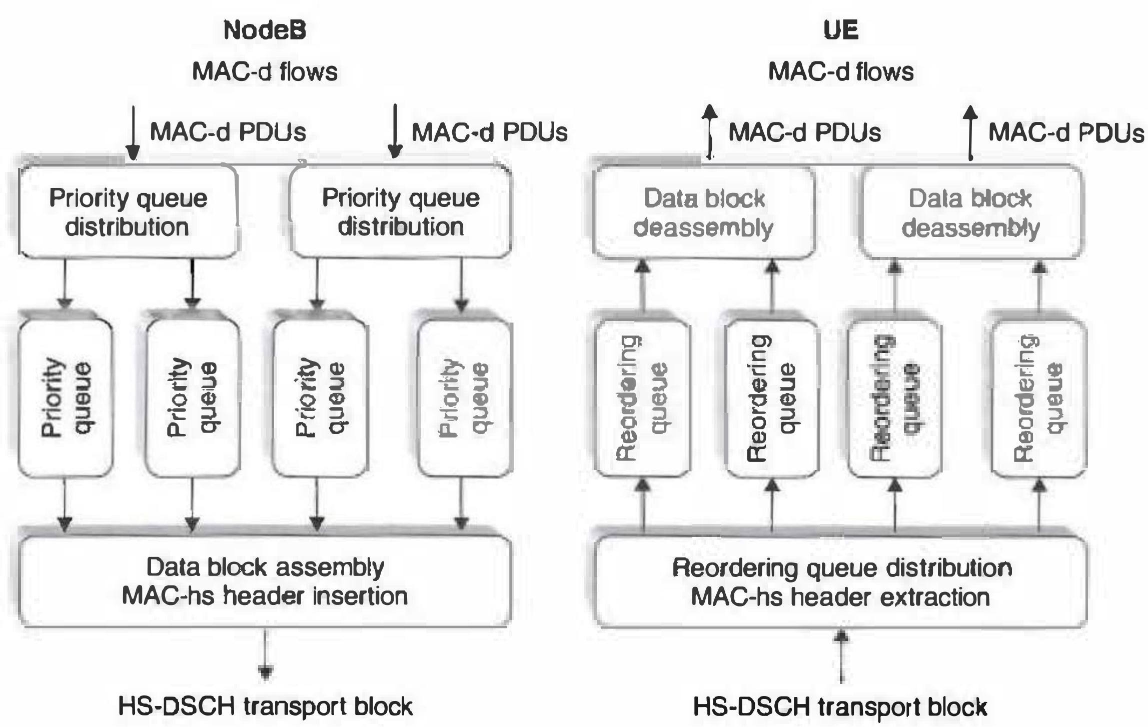


**Figure 9.18** *The priority queues in the NodeB MAC-hs (left) and the reordering queues in the UE MAC-hs (right).*

the scheduling decision. One or several MAC-d PDUs from one of the priority queues are assembled into a data block, where the number of MAC-d PDUs and the priority queue selection is controlled by the scheduler. A MAC-hs header containing, among others, queue identity and a transmission sequence number, is added to form a transport block. The transport block is forwarded to the physical layer for further processing. As there is only a single transmission sequence number and queue identity in the transport block, all MAC-d PDUs within the same transport block come from the same priority queue. Thus, mixing MAC-d PDUs from different priority queues within the same TTI is not possible.

In the UE, the reordering-queue identity is used to place the received data block, containing received MAC-d PDUs, into the correct reordering queue as illustrated in Figure 9.18. Each reordering queue corresponds to a priority queue in the NodeB, although the priority queues buffer MAC-d PDUs, while the reordering queues buffer data blocks. Within each reordering queue, the transmission sequence number, sent in the MAC-hs header is used to ensure in-sequence delivery of the MAC-d PDUs. The transmission sequence number is unique within the reordering queue, but not between different reordering queues.

The basic idea behind reordering, illustrated in Figure 9.19, is to store data blocks in the reordering queue until all data blocks with lower sequence numbers have been delivered. As an example, at time $t_0$ in Figure 9.19, the NodeB has transmitted data blocks with sequence numbers 0 through 3. However, the data block with sequence number 1 has not yet reached the MAC-hs reordering queue in the UE, possibly due to hybrid-ARQ retransmissions or errors in the hybrid-ARQ uplink signaling. Data block 0 has been disassembled into MAC-d

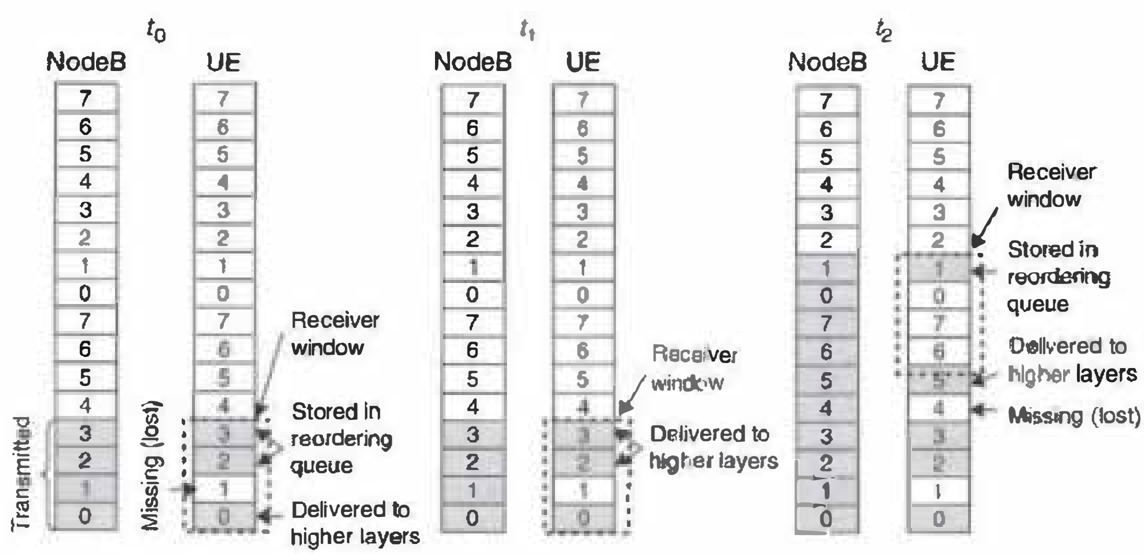


**Figure 9.19** *Illustration of the principles behind reordering queues.*

PDUs and delivered to upper layers by the UE MAC-hs, while data blocks 2 and 3 are buffered in the reordering queue since data block 1 is missing.

Evidently, there is a risk of stalling the reordering queue if missing data blocks (data block 1 in this example) are not successfully received within a finite time. Therefore, a timer-based stall avoidance mechanism is defined for the MAC-hs. Whenever a data block is successfully received but cannot be delivered to higher layers, a timer is started. In Figure 9.19, this occurs when data block 2 is received since data block 1 is missing in the reordering buffer. Note that there is at maximum one stall avoidance timer active. Therefore no timer is started upon reception of data block 3 as there is already one active timer started for data block 2. Upon expiration of the timer, which occurs at time $t_1$ in Figure 9.19, data block 1 is considered to be lost. Any subsequent data blocks up to the first missing data block are to be disassembled into MAC-d PDUs and delivered to higher layers. In Figure 9.19, data blocks 2 and 3 are delivered to higher layers.

Relying on the timer-based mechanism alone would limit the possible values of the timer and limit the performance if the sequence numbers are to be kept unique. Hence, a window-based stall avoidance mechanism is defined in addition to the timer-based mechanism to ensure a consistent UE behavior. If a data block with a sequence number higher than the end of the window is received by the reordering function, the data block is inserted into the reordering buffer at the position indicated by the sequence number. The receiver window is advanced such that the received data block forms the last data block within the window. Any data blocks not within the window after the window advancement are delivered to higher layers. In the example in Figure 9.19, the window size of 4 is used, but the MAC-hs window size is configurable by RRC. In Figure 9.19, a data block with sequence number 1 is received at time $t_2$, which causes the receiver window to be advanced to cover sequence numbers 6 through 1. Data block 4 is considered to be lost, since it is now outside the window whereas data block 5 is disassembled and delivered to higher layers. In order for the reordering functionality in the UE to operate properly, the NodeB should not retransmit MAC-hs PDUs with sequence numbers lower than the highest transmitted sequence number minus the UE receiver window size.

### 9.3.5 MAC-hs header

To support reordering and de-multiplexing of MAC-d PDUs in the UE as discussed above, the necessary information needs to be signaled to the UE. As this information is required only after successful decoding of a transport block, in-band signaling in the form of a MAC-hs header can be used.

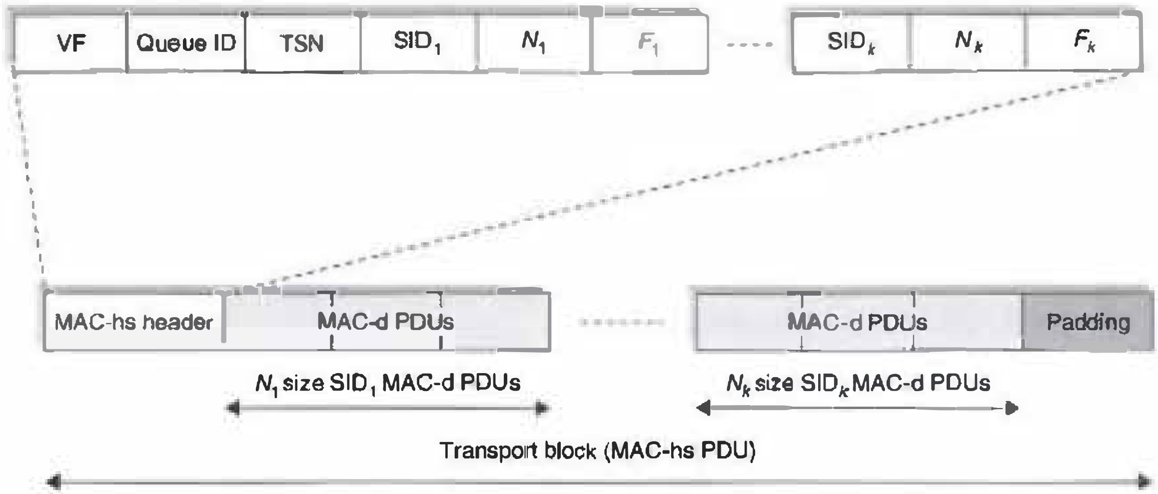


**Figure 9.20** *The structure of the MAC-hs header.*

The MAC-hs header contains

- reordering-queue identity,
- Transmission Sequence Number (TSN),
- number and size of the MAC-d PDUs.

The structure of the MAC-hs header is illustrated in Figure 9.20. The Version Flag (VF) is identical to zero and reserved for future extensions of the MAC-hs header. The 3-bit Queue ID identifies the reordering queue to be used in the receiver. All MAC-d PDUs in one MAC-hs PDU belong to the same reordering queue. The 6-bit TSN field identifies the transmission sequence number of the MAC-hs data block. The TSN is unique within a reordering buffer but not between different reordering buffers. Together with the Queue ID, the TSN provides support for in-sequence delivery as described in the previous section.

The MAC-hs payload consists of one or several MAC-d PDUs. The 3-bit *SID*, size index identifier, provides the MAC-d PDU size and the 7-bit *N* field identifies the number of MAC-d PDUs. The flag *F* is used to indicate the end of the MAC-hs header. One set of *SID*, *N*, and *F* is used for each set of consecutive MAC-d PDUs and multiple MAC-d PDU sizes are supported by forming groups of MAC-d PDUs of equal size. Note that all the MAC-d PDUs within a data block must be in consecutive order since the sequence numbering is per data block. Hence, if a sequence of MAC-d PDUs with sizes given by $SID_1$, $SID_2$, $SID_1$ is to be transmitted, three groups has to be formed despite that there are only two MAC-d PDU sizes. Finally, the MAC-hs PDU is padded (if necessary) such that the MAC-hs PDU size equals a suitable block size. It should be noted that, in most cases, there is only a single MAC-d PDU size and, consequently, only a single set of *SID*, *N*, and *F*.

### 9.3.6 CQI and other means to assess the downlink quality

Obviously, some of the key HSDPA functions, primarily scheduling and rate control, rely on rapid adaptation of the transmission parameters to the instantaneous channel conditions as experienced by the UE. The NodeB is free to form an estimate of the channel conditions using any available information, but, as already discussed, uplink control signaling from the UEs in the form of a Channel-Quality Indicator (CQI), is typically used.

The CQI does not explicitly indicate the channel quality, but rather the data rate supported by the UE given the current channel conditions. More specifically, the CQI is a recommended transport-block size (which is equivalent to a recommended data rate).

The reason for not reporting an explicit channel-quality measure is that different UEs might support different data rates in identical environments, depending on the exact receiver implementation. By reporting the data rate rather than an explicit channel-quality measure, the fact that a UE has a relatively better receiver can be utilized to provide better service (higher data rates) to such a UE. It is interesting to note that this provides a benefit with advanced receiver structures for the end user. For a power-controlled channel, the gain from an advanced receiver is seen as a lower transmit power at the NodeB, thus providing a benefit for the network but not the end user. This is in contrast to the HS-DSCH using rate control, where a UE with an advanced receiver can receive the HS-DSCH with higher data rate compared to a standard receiver.

Each 5-bit CQI value corresponds to a given transport-block size, modulation scheme, and number of channelization codes. These values are shown in Figure 9.8 on page 153 (assuming a high-end terminal, capable of receiving 15 codes). Different tables are used for different UE categories as a UE shall not report a CQI exceeding its capabilities. For example, a UE only supporting 5 codes shall not report a CQI corresponding to 15 codes, while a 15-code UE may do so. Therefore, power offsets are used for channel qualities exceeding the UE capabilities. A power offset of $x$ dB indicates that the UE can receive a certain transport-block size, but at $x$ dB lower transmission power than the CQI report was based upon. This is illustrated in Table 9.2 for some different UE categories. UEs belonging to category 1–6 can only receive up to 5 HS-DSCH channelization codes and therefore must use a power offset for the highest CQI values, while category 10 UEs are able to receive up to 15 codes.

The CQI values listed are sorted in ascending order and the UE shall report the highest CQI for which transmission with parameters corresponding to the

**Table 9.2** *Example of CQI reporting for two different UE categories [97].*

| CQI value | Transport block size | | Modulation scheme | Number of HS-DSCH channelization codes | | Power offset (dB) | |
|---|---|---|---|---|---|---|---|
| | Category 1–6 | Category 10 | | Category 1–6 | Category 10 | Category 1–6 | Category 10 |
| 0 | N/A | | Out of range | | | | |
| 1 | 137 | | QPSK | 1 | | 0 | |
| 2 | 173 | | QPSK | 1 | | 0 | |
| 3 | 233 | | QPSK | 1 | | 0 | |
| 4 | 317 | | QPSK | 1 | | 0 | |
| 5 | 377 | | QPSK | 1 | | 0 | |
| 6 | 461 | | QPSK | 1 | | 0 | |
| 7 | 650 | | QPSK | 2 | | 0 | |
| 8 | 792 | | QPSK | 2 | | 0 | |
| 9 | 931 | | QPSK | 2 | | 0 | |
| 10 | 1262 | | QPSK | 3 | | 0 | |
| 11 | 1483 | | QPSK | 3 | | 0 | |
| 12 | 1742 | | QPSK | 3 | | 0 | |
| 13 | 2279 | | QPSK | 4 | | 0 | |
| 14 | 2583 | | QPSK | 4 | | 0 | |
| 15 | 3319 | | QPSK | 5 | | 0 | |
| 16 | 3565 | | 16QAM | 5 | | 0 | |
| 17 | 4189 | | 16QAM | 5 | | 0 | |
| 18 | 4664 | | 16QAM | 5 | | 0 | |
| 19 | 5287 | | 16QAM | 5 | | 0 | |
| 20 | 5887 | | 16QAM | 5 | | 0 | |
| 21 | 6554 | | 16QAM | 5 | | 0 | |
| 22 | 7168 | | 16QAM | 5 | | 0 | |
| 23 | 7168 | 9719 | 16QAM | 5 | 7 | −1 | 0 |
| 24 | 7168 | 11418 | 16QAM | 5 | 8 | −2 | 0 |
| 25 | 7168 | 14411 | 16QAM | 5 | 10 | −3 | 0 |
| 26 | 7168 | 17237 | 16QAM | 5 | 12 | −4 | 0 |
| 27 | 7168 | 21754 | 16QAM | 5 | 15 | −5 | 0 |
| 28 | 7168 | 23370 | 16QAM | 5 | 15 | −6 | 0 |
| 29 | 7168 | 24222 | 16QAM | 5 | 15 | −7 | 0 |
| 30 | 7168 | 25558 | 16QAM | 5 | 15 | −8 | 0 |

CQI result in a block error probability not exceeding 10%. The CQI values are chosen such that an increase in CQI by one step corresponds to approximately 1 dB increase in the instantaneous carrier-to-interference ratio on an AWGN channel. Measurements on the common pilot form the basis for the CQI. The CQI represents the instantaneous channel conditions in a predefined 3-slot interval ending one slot prior to the CQI transmission. Specifying which

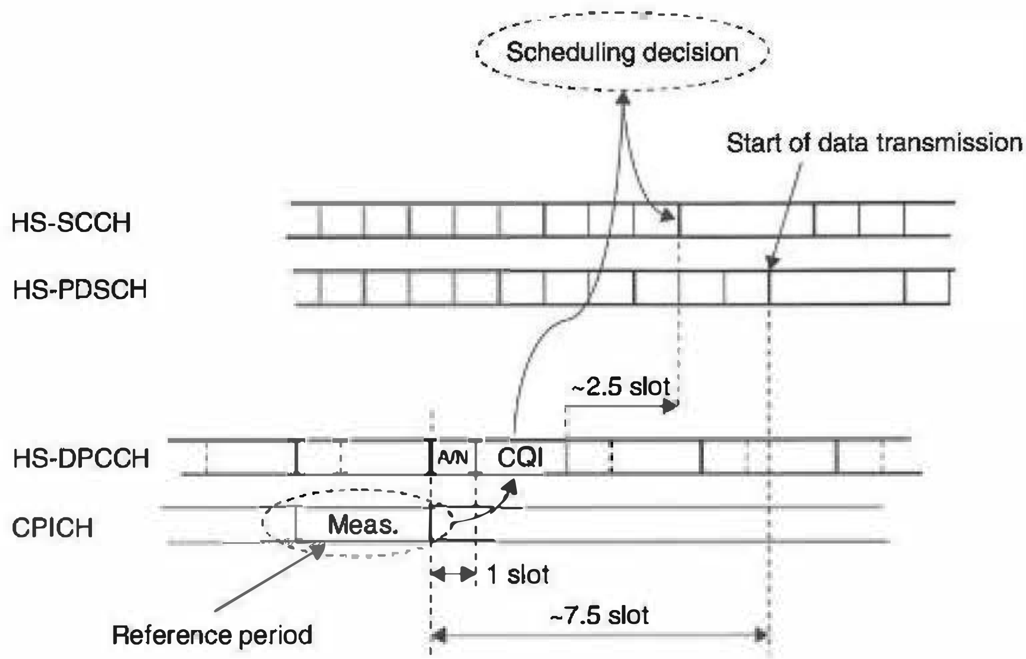


**Figure 9.21** *Timing relation for the CQI reports.*

interval the CQI relates to allows the NodeB to track changes in the channel quality between the CQI reports by using the power control commands for the associated downlink (F-) DPCH as described below. The timing of the CQI reports and the earliest possible time the report can be used for scheduling purposes is illustrated in Figure 9.21.

The rate of the channel-quality reporting is configurable in the range of one report per 2–160ms. The CQI reporting can also be switched off completely.

As the scheduling and rate-adaptation algorithms are vendor specific, it is possible to perform rate control based on other criteria than the UE reports as well, either alone or in combination. Using the transmit power level of the associated DPCH is one such possibility, where a high transmit power indicates unfavorable channel conditions and a low DPCH transmit power indicates favorable conditions. Since the power level is a relative measure of the channel quality and not reflects an absolute subjective channel quality, this technique is advantageously combined with infrequent UE quality reports. The UE reports provide an absolute quality and the transmission power of the power-controlled DPCH can be used to update this quality report between the reporting instances. This combined scheme works quite well and can significantly reduce the frequency of the UE CQI reports as long as the DPCH is not in soft handover. In soft handover the transmit power of the different radio links involved in the soft handover are power controlled such that the combined received signal is of sufficient quality. Consequently, the DPCH transmit power at the serving

HS-DSCH cell does not necessarily reflect the perceived UE channel quality. Hence, more frequent UE quality reports are typically required in soft handover scenarios.

### 9.3.7 Downlink control signaling: HS-SCCH

The HS-SCCH, sometimes referred to as the *shared control channel*, is a shared downlink physical channel that carries control signaling information needed for a UE to be able to properly despread, demodulate and decode the HS-DSCH.

In each 2ms interval corresponding to one HS-DSCH TTI, one HS-SCCH carries physical-layer signaling to a single UE. As HSDPA supports HS-DSCH transmission to multiple users in parallel by means of code multiplexing, see Section 9.1.1, multiple HS-SCCH may be needed in a cell. According to the specification, a UE should be able to decode four HS-SCCHs in parallel. However, more than four HS-SCCHs can be configured within a cell, although the need for this is rare.

HS-SCCH uses a spreading factor of 128 and has a time structure based on a subframe of length 2ms which is the same length as the HS-DSCH TTI. The following information is carried on the HS-SCCH:

- The HS-DSCH transport format, consisting of:
  - HS-DSCH channelization-code set [7 bits]
  - HS-DSCH modulation scheme, QPSK /16QAM [1 bit]
  - HS-DSCH transport-block size information [6 bits].
- Hybrid-ARQ-related information, consisting of:
  - hybrid-ARQ process number [3 bits]
  - redundancy version [3 bits]
  - new-data indicator [1 bit].
- A UE ID that identifies the UE for which the HS-SCCH information is intended [16 bits]. As will be described below, the UE ID is not explicitly transmitted but implicitly included in the CRC calculation and HS-SCCH channel coding.

As described in Section 9.2.4, the HS-DSCH transport block can take 1 out of 254 different sizes. Each combination of channelization-code-set size and modulation scheme corresponds to a subset of these transport-block sizes, where each subset consists of 63 possible transport-block sizes. The 6-bit 'HS-DSCH transport-block size information' indicates which out of the 63 possible transport-block sizes is actually used for the HS-DSCH transmission in the corresponding TTI. The transport-block sizes have been defined to make full use of code rates ranging from 1/3 to 1 for initial transmissions. For retransmissions, instantaneous code rates larger than one can be achieved by indicating 'the transport-block size is identical to the previous transmission in this hybrid-ARQ process.' This is indicated setting the 'HS-DSCH transport-block size information' field to 111111. This is useful for additional scheduling flexibility, for example to retransmit only a small amount of parity bit in case the latest CQI report indicates the UE 'almost' was able to decode the original transmission.

Requirements on when different parts of the HS-SCCH information need to be available to the UE has affected the detailed structure of the HS-SCCH channel coding and physical-channel mapping. For UE complexity reasons, it is beneficial if the channelization-code set is known to the UE prior to the start of the HS-DSCH transmission. Otherwise, the UE would have to buffer the received signal on a sub-chip level prior to despreading or, alternatively, despread all potential HS-DSCH codes up to the maximum of 15 codes. Knowing the modulation scheme prior to the HS-DSCH subframe is also preferred as it allows for 'on-the-fly' demodulation. On the other hand, the transport-block size and the hybrid-ARQ-related information are only needed at HS-DSCH decoding/soft combining, which can anyway not start until the end of the HS-DSCH TTI. Thus, the HS-SCCH information is split into two parts:

- Part 1 consisting of channelization-code set and modulation scheme [total of 8 bits].
- Part 2 consisting of transport-block size and hybrid-ARQ-related parameters [total of 13 bits].

The HS-SCCH coding, physical-channel mapping and timing relation to the HS-DSCH transmission is illustrated in Figure 9.22. The HS-DSCH channel coding is based on rate-1/3 convolutional coding, carried out separately for part 1 and part 2. Part 1 is coded and rate matched to 40 bits to fit into the first slot of the HS-SCCH subframe. Before mapping to the physical channel, the coded part 1 is scrambled by a 40 bits UE-specific bit sequence. The sequence is derived from the 16 bits UE ID using rate-1/2 convolutional coding followed by puncturing. With the scheme of Figure 9.22, the part 1 information can be decoded after one slot of the HS-SCCH subframe. Furthermore, in case of more than one HS-SCCH, the UE can find the correct HS-SCCH from the soft metric of the channel decoder already after the first slot. One possible way for the UE to utilize the soft metric for determining which (if any) of the multiple HS-SCCHs that carries control information for the UE is to form the log-likelihood ratio between the most likely code word and the second most likely code word for each HS-SCCH. The HS-SCCH with the largest ratio is likely to be intended for the UE and can be selected for further decoding of the part-2 information.

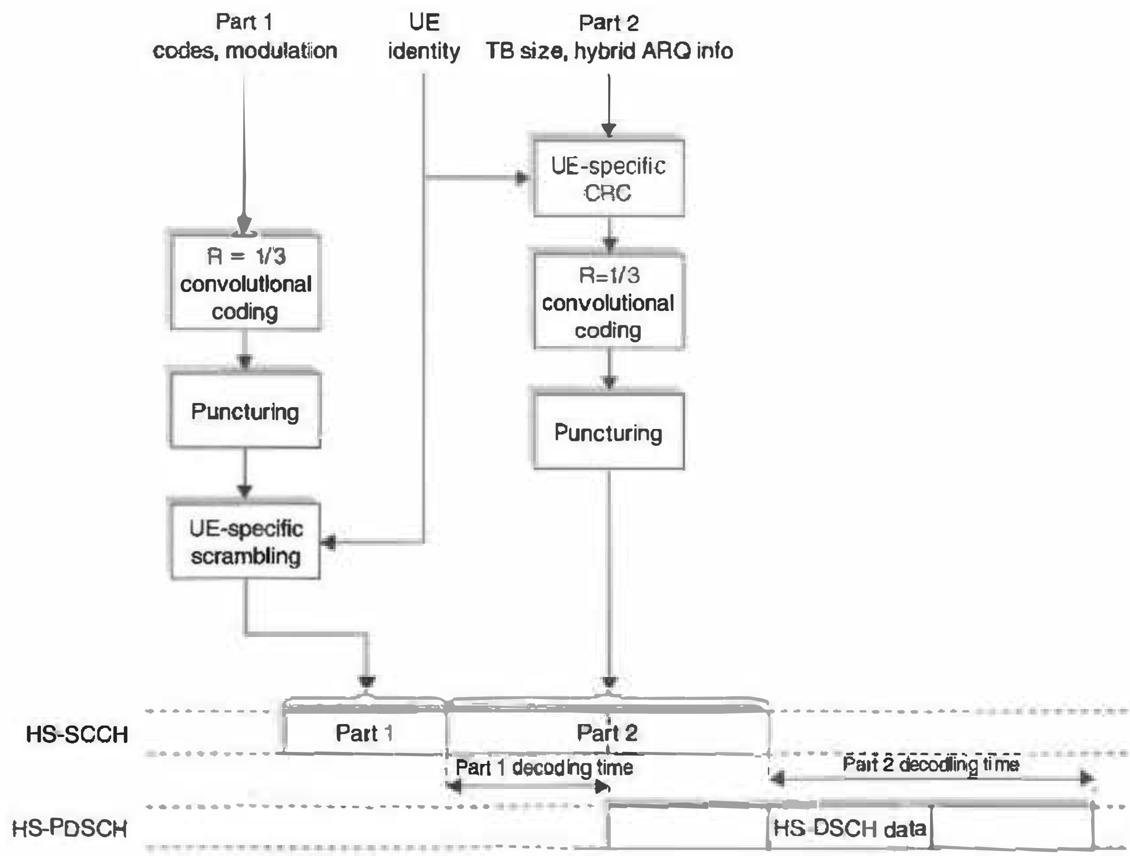


**Figure 9.22** *HS-SCCH channel coding.*

Part 2 is coded and rate matched to 80 bits to fit into the second and third slot of the HS-SCCH. Part 2 includes a UE-specific CRC for error detection. The CRC is calculated over all the information bits, both part 1 and part 2, as well as the UE identity. The identity is not explicitly transmitted, but by including its ID when calculating the CRC at the receiver, the UE can decide whether it was the intended recipient or not. If the transmission is intended for another UE, the CRC will not check.

In case of HS-DSCH transmission to a single UE in consecutive TTIs, the UE must despread the HS-SCCH in parallel to the HS-DSCH channelization codes. To reduce the number of required despreaders, the same HS-SCCH shall be used when HS-DSCH transmission is carried out in consecutive TTI. This implies that, when simultaneously receiving HS-DSCH, the UE only needs to despread a single HS-SCCH.

In order to avoid waste of capacity, the HS-SCCH transmit power should be adjusted to what is needed to reach the intended UE. Similar information used for rate control of the HS-DSCH, for example the CQI reports, can be used to power control the HS-SCCH.

### 9.3.8 Downlink control signaling: F-DPCH

As described in Section 9.2.1, for each UE for which HS-DSCH can be transmitted, there is also an associated downlink DPCH. In principle, if all data transmission, including RRC signaling, is mapped to the HS-DSCH, there is no need to carry any user data on the DPCH. Consequently, there is no need for downlink Transport-Format Combination Indicator (TFCI) or dedicated pilots on such a DPCH. In this case, the only use for the downlink DPCH in case of HS-DSCH transmission is to carry power control commands to the UE in order to adjust the uplink transmission power. This fact is exploited by the F-DPCH or *fractional DPCH*, introduced in Release 6 as a means to reduce the amount of downlink channelization codes used for dedicated channels. Instead of allocating one DPCH with spreading factor 256 for the sole purpose of transmitting one power control command per slot, the F-DPCH allows up to ten UEs to share a single channelization code for this purpose. In essence, the F-DPCH is a slot format supporting TPC bits only. Two TPC bits (one QPSK symbol) is transmitted in one tenth of a slot, using a spreading factor 256, and the rest of the slot remains unused. By setting the downlink timing of multiple UEs appropriately, as illustrated in Figure 9.23, up to ten UEs can then share a single channelization code. This can also be seen as time-multiplexing power control commands to several users on one channelization code.

### 9.3.9 Uplink control signaling: HS-DPCCH

For operation of the hybrid-ARQ protocol and to provide the NodeB with knowledge about the instantaneous downlink channel conditions, uplink control signaling is required. This signaling is carried on an additional new uplink physical channel, the HS-DPCCH, using a channelization code separate from the conventional uplink DPCCH. The use of a separate channelization code for the HS-DPCCH makes the HS-DPCCH 'invisible' to non-HSDPA-capable base stations and allows for the uplink being in soft handover even if not all NodeBs in the active set support HSDPA.

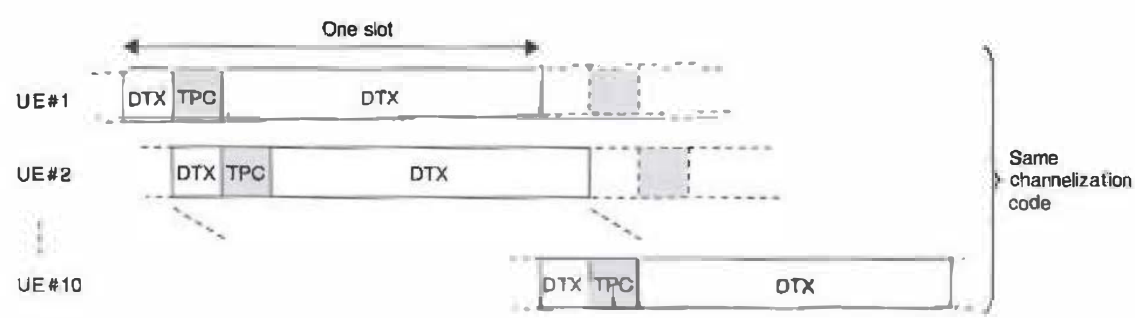


**Figure 9.23** *Fractional DPCH (F-DPCH), introduced in Release 6.*

The HS-DPCCH uses a spreading factor of 256 and is transmitted in parallel with the other uplink channels as illustrated Figure 9.24. To reduce the uplink peak-to-average ratio, the channelization code used for HS-DPCCH and if the HS-DPCCH is mapped to the I or Q branch of this code depends on the maximum number of DPDCHs used by the transport-format combination set configured for the UE.

As the HS-DPCCH spreading factor is 256, the HS-DPCCH allows for a total of 30 channel bits per 2 ms subframe (3 slots). The HS-DPCCH information is divided in such a way that the hybrid-ARQ acknowledgement is transmitted in the first slot of the subframe while the channel-quality indication is transmitted in the second and third slot, see Figure 9.24.

In order to minimize the hybrid-ARQ roundtrip time, the HS-DPCCH transmission timing is not slot aligned to the other uplink channels. Instead, the HS-DPCCH timing is defined relative to the end of the subframe carrying the corresponding HS-DSCH data as illustrated in Figure 9.24. The timing is such that there are approximately 7.5 slots (19 200 chips) of UE processing time available, from the end of the HS-DSCH TTI to the transmission of the corresponding uplink hybrid-ARQ acknowledgement. If the HS-DPCCH had been slot aligned to the uplink DPCCH, there would have been an uncertainty of one slot in the HS-DSCH /HS-DPCCH timing. This uncertainty would have reduced the processing time available for the UE /NodeB by one slot.

Due to the alignment between the uplink HS-DPCCH and the downlink HS-DSCH, the HS-DPCCH will not necessarily be slot aligned with the uplink DPDCH/DPCCH. However, note that the HS-DPCCH is always aligned to the uplink DPCCH/DPDCH on a 256-chip basis in order to keep uplink orthogonality. As a consequence, the HS-DPCCH cannot have a completely fixed transmit timing relative to the received HS-DSCH. Instead the HS-DPCCH transmit timing

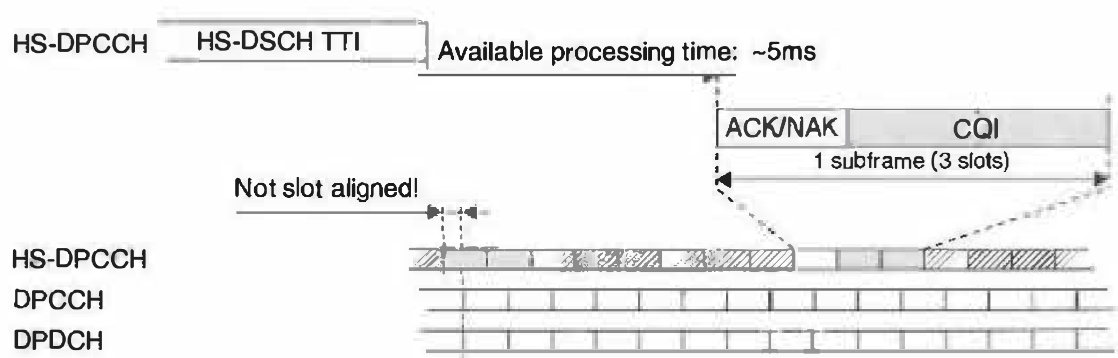


**Figure 9.24** *Basic structure of uplink signaling with IQ/code-multiplexed HS-DPCCH.*

varies in an interval 19 200 chips to 19 200 + 255 chips. Note that CQI and ACK/NAK are transmitted independently of each other. In subframes where no ACK/NAK or CQI is to be transmitted, nothing is transmitted in the corresponding HS-DPCCH field.

The hybrid-ARQ acknowledgement consists of a single information bit, ACK or NAK, indicating whether the HS-DSCH was correctly decoded (the CRC checked) or not. ACK or NAK is only transmitted in case the UE correctly received the HS-SCCH control signaling. If no HS-SCCH control signaling intended for the UE was detected, nothing is transmitted in the ACK/NAK field (DTX). This reduces the uplink load as only the UEs to which HS-DSCH data was actually sent in a TTI transmit an ACK/NAK on the uplink. The single-bit ACK is repetition coded into 10 bits to fit into the first slot of a HS-DPCCH subframe.

Reliable reception of the uplink ACK/NAK requires a sufficient amount of energy. In some situations where the UE is power limited, it may not be possible to collect enough energy by transmitting the ACK/NAK over a single slot. Therefore, there is a possibility to configure the UE to repeat the ACK/NAK in $N$ subsequent ACK/NAK slots. Naturally, when the UE is configured to transmit repeated acknowledgements, it cannot receive HS-DSCH data in consecutive TTIs, as the UE would then not be able to acknowledge all HS-DSCH data. Instead there must be at least $N-1$ idle 2 ms subframes between each HS-DSCH TTI in which data is to be received. Examples when repetition of the acknowledgements can be useful are very large cells, or in some soft handover situations. In soft handover, the uplink can be power controlled by multiple NodeBs. If any of the non-serving NodeBs has the best uplink, the received HS-DPCCH quality at the serving NodeB may not be sufficient and repetition may therefore be necessary.

As mentioned earlier, the impact of ACK-to-NAK and NAK-to-ACK errors are different, leading to different requirements. In addition, the DTX-to-ACK error case also has to be handled. If the UE misses the scheduling information and the NodeB misinterprets the DTX as ACK, data loss in the hybrid ARQ will occur. An asymmetric decision threshold in the ACK/NAK detector should therefore preferably be used as illustrated in Figure 9.25. Based on the noise variance at the ACK/NAK detector, the threshold can be computed to meet a certain DTX-to-ACK error probability, for example, $10^{-2}$, after which the transmission power of the ACK and NAK can be set to meet the remaining error requirements (ACK-to-NAK and NAK-to-ACK).

In Release 6 of the WCDMA specifications, an enhancement to the ACK/NAK signaling has been introduced. In addition to the ACK and NAK, the UE may also

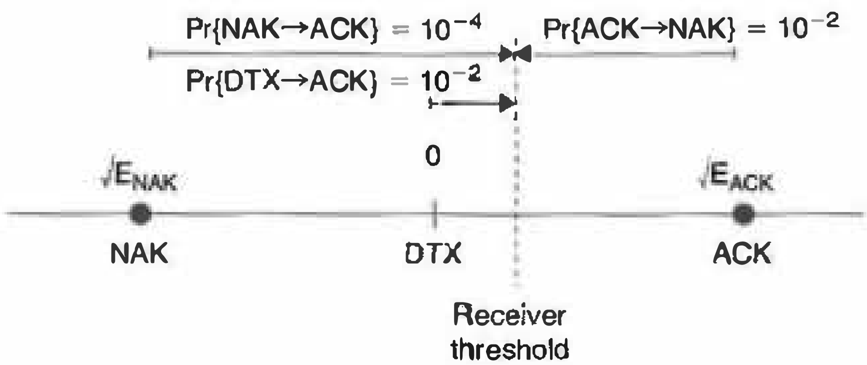


**Figure 9.25** *Detection threshold for the ACK/NAK field of HS-DPCCH.*

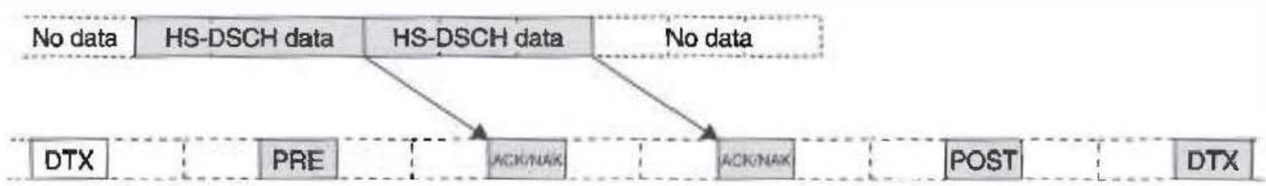


**Figure 9.26** *Enhanced ACK/NAK using PRE and POST.*

transmit two additional code words, PRE and POST, on the HS-DPCCH. A UE configured to use the enhancement will transmit PRE and POST in the subframes preceding and succeeding, respectively, the ACK/NAK (unless these subframes were used by the ACK/NAK for other transport blocks). Thus, an ACK will cause a transmission spanning multiple subframes and the power can therefore be reduced while maintaining the same ACK-to-NAK error rate (Figure 9.26).

The CQI consists of five information bits. A (20,5) block code is used to code this information to 20 bits, which corresponds to two slots on the HS-DPCCH. Similarly to the ACK/NAK, repetition of the CQI field over multiple 2 ms subframes is possible and can be used to provide improved coverage.

# 10
# Enhanced Uplink

*Enhanced Uplink,* also known as *High-Speed Uplink Packet Access* (HSUPA), has been introduced in WCDMA Release 6. It provides improvements in WCDMA uplink capabilities and performance in terms of higher data rates, reduced latency, and improved system capacity, and is therefore a natural complement to HSDPA. Together, the two are commonly referred to as High-Speed Packet Access (HSPA). The specifications of Enhanced Uplink can be found in [101] and the references therein.

## 10.1 Overview

At the core of Enhanced Uplink are two basic technologies used also for HSDPA – fast scheduling and fast hybrid ARQ with soft combining. For similar reasons as for HSDPA, Enhanced Uplink also introduces a short 2 ms uplink TTI. These enhancements are implemented in WCDMA through a new transport channel, the *Enhanced Dedicated Channel* (E-DCH).

Although the same technologies are used both for HSDPA and Enhanced Uplink, there are fundamental differences between them, which have affected the detailed implementation of the features:

- In the downlink, the shared resource is transmission power and the code space, both of which are located in *one* central node, the NodeB. In the uplink, the shared resource is the amount of allowed uplink interference, which depends on the transmission power of *multiple distributed* nodes, the UEs.
- The scheduler and the transmission buffers are located in the same node in the downlink, while in the uplink the scheduler is located in the NodeB while the data buffers are distributed in the UEs. Hence, the UEs need to signal buffer status information to the scheduler.
- The WCDMA uplink, also with Enhanced Uplink, is inherently non-orthogonal, and subject to interference between uplink transmissions within the same

cell. This is in contrast to the downlink, where different transmitted channels are *orthogonal*. Fast power control is therefore essential for the uplink to handle the near-far problem.[1] The E-DCH is transmitted with a power offset relative to the power-controlled uplink control channel and by adjusting the maximum allowed power offset, the scheduler can control the E-DCH data rate. This is in contrast to HSDPA, where a (more or less) constant transmission power with rate adaptation is used.

- Soft handover is supported by the E-DCH. *Receiving* data from a terminal in multiple cells is fundamentally beneficial as it provides diversity, while transmission from multiple cells in case of HSDPA is cumbersome and with questionable benefits as discussed in the previous chapter. Soft handover also implies *power control by multiple cells*, which is necessary to limit the amount of interference generated in neighboring cells and to maintain backward compatibility and coexistence with UE not using the E-DCH for data transmission.
- In the downlink, higher-order modulation, which trades power efficiency for bandwidth efficiency, is useful to provide high data rates in some situations, for example when the scheduler has assigned a small number of channelization codes for a transmission but the amount of available transmission power is relatively high. The situation in the uplink is different; there is no need to share channelization codes between users and the channel coding rates are therefore typically lower than for the downlink. Hence, unlike the downlink, higher-order modulation is less useful in the uplink macro-cells and therefore not part of the first release of enhanced uplink.[2]

With these differences in mind, the basic principles behind Enhanced Uplink can be discussed.

### 10.1.1 Scheduling

For Enhanced Uplink, the scheduler is a key element, controlling *when* and *at what data rate* the UE is allowed to transmit. The higher the data rate a terminal is using, the higher the terminal's received power at the NodeB must be to maintain the $E_b/N_0$ required for successful demodulation. By increasing the transmission power, the UE can transmit at a higher data rate. However, due to the non-orthogonal uplink, the received power from one UE represents interference for other terminals. Hence, the shared resource for Enhanced Uplink is the

[1] The near-far problem describes the problem of detecting a weak user, located far from the transmitter, when a user close to the transmitter is active. Power control ensured the signals are received at a similar strength, therefore, enabling detection of both users' transmissions.

[2] Uplink higher-order modulation is introduced in Release 7; see Chapter 12 for further details.

amount of tolerable interference in the cell. If the interference level is too high, some transmissions in the cell, control channels and non-scheduled uplink transmissions, may not be received properly. On the other hand, a too low interference level may indicate that UEs are artificially throttled and the full system capacity not exploited. Therefore, Enhanced Uplink relies on a scheduler to give users with data to transmit permission to use an as high data rate as possible without exceeding the maximum tolerable interference level in the cell.

Unlike HSDPA, where the scheduler and the transmission buffers both are located in the NodeB, the data to be transmitted resides in the UEs for the uplink case. At the same time, the scheduler is located in the NodeB to coordinate different UEs transmission activities in the cell. Hence, a mechanism for communicating the scheduling decisions to the UEs and to provide buffer information from the UEs to the scheduler is required. The scheduling framework for Enhanced Uplink is based on *scheduling grants* sent by the NodeB scheduler to control the UE transmission activity and *scheduling requests* sent by the UEs to request resources. The scheduling grants control the maximum allowed E-DCH-to-pilot power ratio the terminal may use; a larger grant implies the terminal may use a higher data rate but also contributes more to the interference level in the cell. Based on measurements of the (instantaneous) interference level, the scheduler controls the scheduling grant in each terminal to maintain the interference level in the cell at a desired target (Figure 10.1).

In HSDPA, typically a single user is addressed in each TTI. For Enhanced Uplink, the implementation-specific uplink scheduling strategy in most cases schedules multiple users in parallel. The reason is the significantly smaller transmit power of a terminal compared to a NodeB: a single terminal typically cannot utilize the full cell capacity on its own.

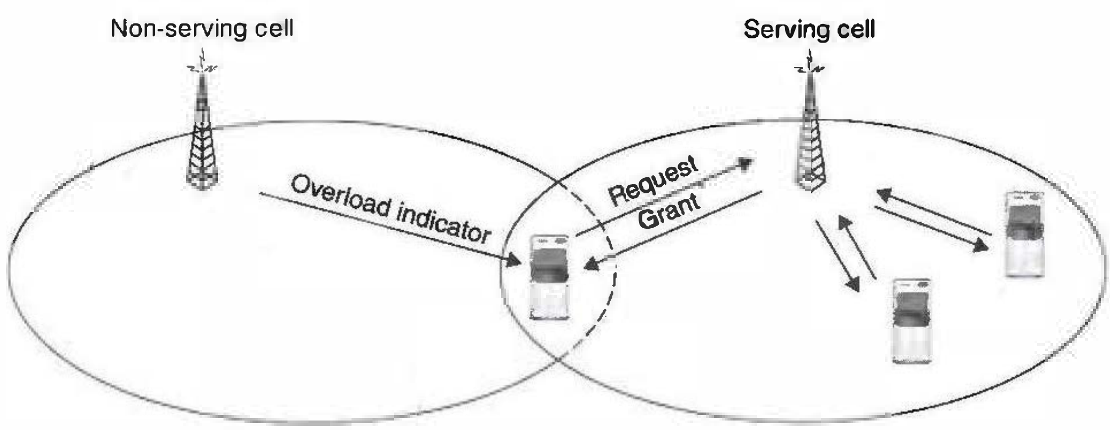


**Figure 10.1** *Enhanced Uplink scheduling framework.*

Inter-cell interference also needs to be controlled. Even if the scheduler has allowed a UE to transmit at a high data rate based on an acceptable intra-cell interference level, this may cause non-acceptable interference to neighboring cells. Therefore, in soft handover, the *serving cell* has the main responsibility for the scheduling operation, but the UE monitors scheduling information from all cells with which the UE is in soft handover. The non-serving cells can request all its non-served users to lower their E-DCH data rate by transmitting an *overload indicator* in the downlink. This mechanism ensures a stable network operation.

Fast scheduling allows for a more relaxed connection admission strategy. A larger number of bursty high-rate packet-data users can be admitted to the system as the scheduling mechanism can handle the situation when multiple users need to transmit in parallel. If this creates an unacceptably high interference level in the system, the scheduler can rapidly react and restrict the data rates they may use. Without fast scheduling, the admission control would have to be more conservative and reserve a margin in the system in case of multiple users transmitting simultaneously.

### 10.1.2 Hybrid ARQ with soft combining

Fast hybrid ARQ with soft combining is used by Enhanced Uplink for basically the same reason as for HSDPA – to provide robustness against occasional transmission errors. A similar scheme as for HSDPA is used. For each transport block received in the uplink, a single bit is transmitted from the NodeB to the UE to indicate successful decoding (ACK) or to request a retransmission of the erroneously received transport block (NAK).

One main difference compared to HSDPA stems from the use of soft handover in the uplink. When the UE is in soft handover, this implies that the hybrid ARQ protocol is *terminated in multiple cells*. Consequently, in many cases, the transmitted data may be successfully received in some NodeBs but not in others. From a UE perspective, it is sufficient if at least one NodeB successfully receives the data. Therefore, in soft handover, all involved NodeBs attempt to decode the data and transmits an ACK or a NAK. If the UE receives an ACK from at least one of the NodeBs, the UE considers the data to be successfully received.

Hybrid ARQ with soft combining can be exploited not only to provide robustness against unpredictable interference, but also to improve the link efficiency to increase capacity and/or coverage. One possibility to provide a data rate of $x$ Mbit/s is to transmit at $x$ Mbit/s and set the transmission power to target a low error probability (in the order of a few percent) in the first transmission attempt.

Alternatively, the same resulting data rate can be provided by transmitting using $n$ times higher data rate at an unchanged transmission power and use multiple hybrid ARQ retransmissions. From the discussion in Chapter 7, this approach on average results in a lower cost per bit (a lower $E_b/N_0$) than the first approach. The reason is that, on average, less than $n$ transmissions will be used. This is sometimes known as *early termination gain* and can be seen as implicit rate adaptation. Additional coded bits are only transmitted when necessary. Thus, the code rate after retransmissions is determined by what was needed by the instantaneous channel conditions. This is exactly what rate adaptation also tries to achieve, the main difference being that rate adaptation tries to find the correct code rate prior to transmission. The same principle of implicit rate adaptation can also be used for HS-DSCH in the downlink to improve the link efficiency.

### 10.1.3 Architecture

For efficient operation, the scheduler should be able to exploit rapid variations in the interference level and the channel conditions. Hybrid ARQ with soft combining also benefits from rapid retransmissions as this reduces the cost of retransmissions. These two functions should therefore reside close to the radio-interface. As a result, and for similar reasons as for HSDPA, the scheduling and hybrid ARQ functionalities of Enhanced Uplink are located in the NodeB. Furthermore, also similar to the HSDPA design, it is preferable to keep all radio-interface layers above MAC intact. Hence, ciphering, admission control, etc., is still under the control of the RNC. This also allows for a smooth introduction of Enhanced Uplink in selected areas; in cells not supporting E-DCH transmissions, channel switching can be used to map the user's data flow onto the DCH instead.

Following the HSDPA design philosophy, a new MAC entity, the *MAC-e*, is introduced in the UE and NodeB. In the NodeB, the MAC-e is responsible for support of fast hybrid ARQ retransmissions and scheduling, while in the UE, the MAC-e is responsible for selecting the data rate within the limits set by the scheduler in the NodeB MAC-e.

When the UE is in soft handover with multiple NodeBs, different transport blocks may be successfully decoded in different NodeBs. Consequently, one transport block may be successfully received in one NodeB while another NodeB is still involved in retransmissions of an earlier transport block. Therefore, to ensure in-sequence delivery of data blocks to the RLC protocol, a reordering functionality is required in the RNC in the form of a new MAC entity, the MAC-es. In soft handover, multiple MAC-e entities are used per UE as the data is received in multiple cells. However, the MAC-e in the serving cell

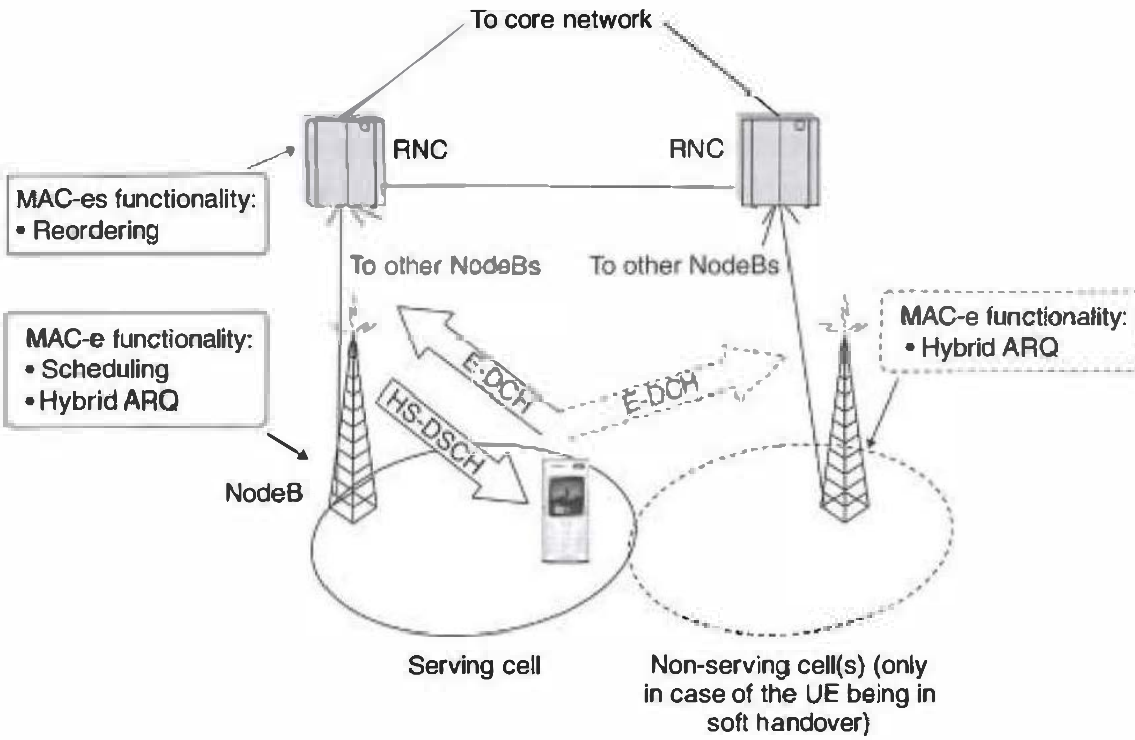


**Figure 10.2** *The architecture with E-DCH (and HS-DSCH) configured.*

has the main responsibility for the scheduling; the MAC-e in a non-serving cell is mainly handling the hybrid ARQ protocol (Figure 10.2).

## 10.2 Details of Enhanced Uplink

To support uplink scheduling and hybrid ARQ with soft combining in WCDMA, a new transport-channel type, the Enhanced Dedicated Channel (E-DCH) has been introduced in Release 6. The E-DCH can be configured simultaneously with one or several DCHs. Thus, high-speed packet-data transmission on the E-DCH can occur at the same time as services using the DCH from the same UE.

A low delay is one of the key characteristics of Enhanced Uplink and required for efficient packet-data support. Therefore, a short TTI of 2ms is supported by the E-DCH to allow for rapid adaptation of transmission parameters and reduction of the end-user delays associated with packet-data transmission. Not only does this reduce the cost of a retransmission, the transmission time for the initial transmission is also reduced. Physical-layer processing delay is typically proportional to the amount of data to process and the shorter the TTI, the smaller the amount of data to process in each TTI for a given data rate. At the same time, in deployments with relatively modest data rates, for example in large cells, a longer TTI may be beneficial as the payload in a 2 ms TTI can become unnecessarily small

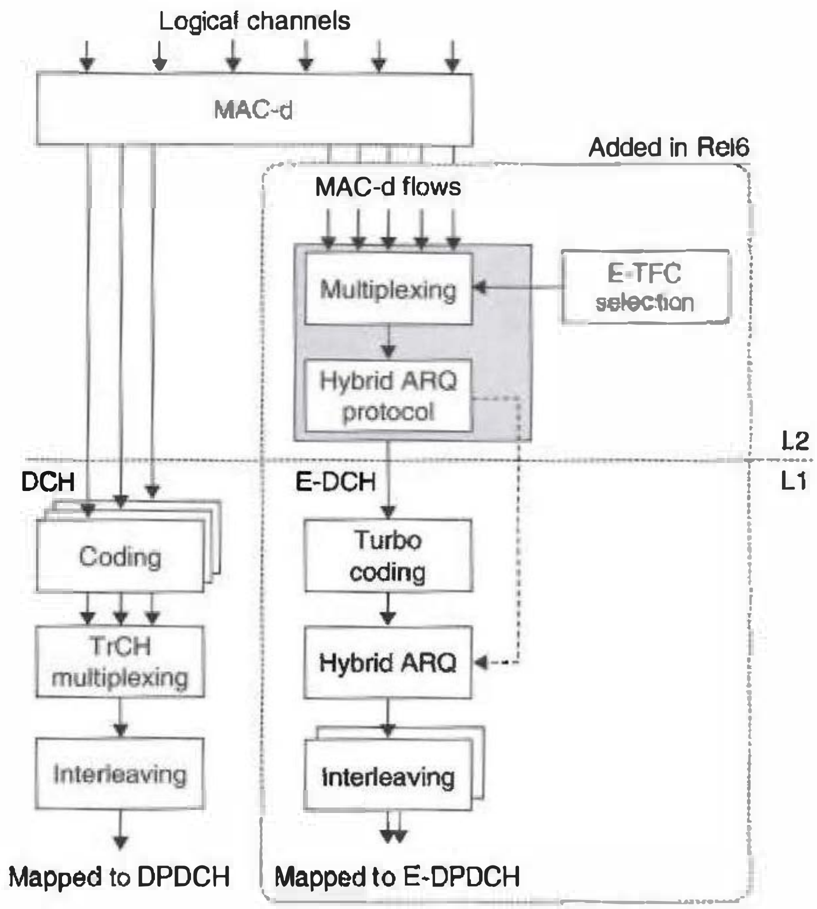


**Figure 10.3** *Separate processing of E-DCH and DCH.*

and the associated relative overhead too large. Hence, the E-DCH supports two TTI lengths, 2 and 10 ms, and the network can configure the appropriate value. In principle, different UEs can be configured with different TTIs.

The E-DCH is mapped to a set of uplink channelization codes known as *E-DCH Dedicated Physical Data Channels* (E-DPDCHs). Depending on the instantaneous data rate, the number of E-DPDCHs and their spreading factors are both varied.

Simultaneous transmission of E-DCH and DCH is possible as discussed above. Backward compatibility requires the E-DCH processing to be invisible to a NodeB not supporting Enhanced Uplink. This has been solved by separate processing of the DCH and E-DCH and mapping to different channelization code sets as illustrated in Figure 10.3. If the UE is in soft handover with multiple cells, of which some does not support Enhanced Uplink, the E-DCH transmission is invisible to these cells. This allows for a gradual upgrade of an existing network. An additional benefit with the structure is that it simplifies the introduction of the 2 ms TTI and also provides greater freedom in the selection of hybrid ARQ processing.

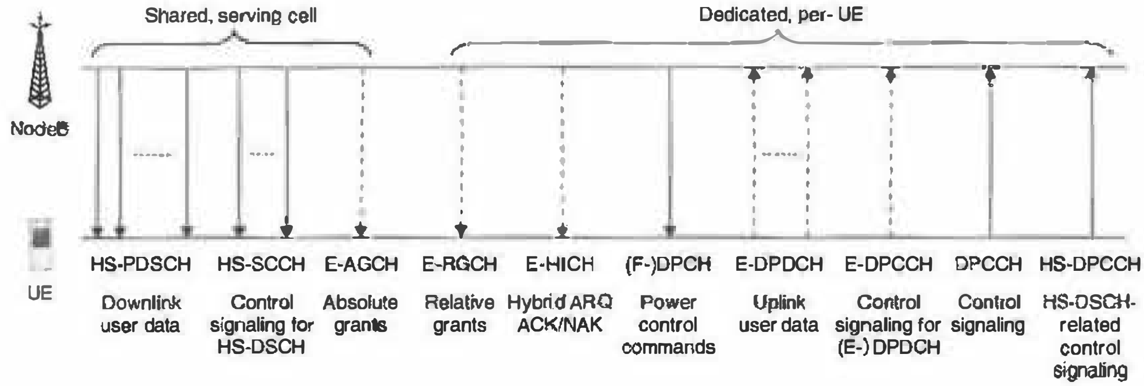


**Figure 10.4** *Overall channel structure with HSDPA and Enhanced Uplink. The new channels introduced as part of Enhanced Uplink are shown with dashed lines.*

Downlink control signaling is necessary for the operation of the E-DCH. The downlink, as well as uplink, control channels used for E-DCH support are illustrated in Figure 10.4, together with the channels used for HSDPA.

Obviously, the NodeB needs to be able to request retransmissions from the UE as part of the hybrid ARQ mechanism. This information, the ACK/NAK, is sent on a new downlink dedicated physical channel, the *E-DCH Hybrid ARQ Indicator Channel* (E-HICH). Each UE with E-DCH configured receives one E-HICH of its own from each of the cells which the UE is in soft handover with.

Scheduling grants, sent from the scheduler to the UE to control when and at what data rate the UE is transmitting, can be sent to the UE using the shared *E-DCH Absolute Grant Channel* (E-AGCH). The E-AGCH is sent from the serving cell only as this is the cell having the main responsibility for the scheduling operation and is received by all UEs with an E-DCH configured. In addition, scheduling grant information can also be conveyed to the UE through an *E-DCH Relative Grant Channel* (E-RGCH). The E-AGCH is typically used for large changes in the data rate, while the E-RGCH is used for smaller adjustments during an ongoing data transmission. This is further elaborated upon in the discussion on scheduling operation below.

Since the uplink by design is non-orthogonal, fast closed-loop power control is necessary to address the near-far problem. The E-DCH is no different from any other uplink channel and is therefore power controlled in the same way as other uplink channels. The NodeB measures the received signal-to-interference ratio and sends power control commands in the downlink to the UE to adjust the transmission power. Power control commands can be transmitted using DPCH or, to save channelization codes, the fractional DPCH, F-DPCH.

In the uplink, control signaling is required to provide the NodeB with the necessary information to be able to demodulate and decode the data transmission. Even though, in principle, the serving cell could have this knowledge as it has issued the scheduling grants, the non-serving cells in soft handover clearly do not have this information. Furthermore, as discussed below, the E-DCH also supports non-scheduled transmissions. Hence, there is a need for out-band control signaling in the uplink, and the *E-DCH Dedicated Physical Control Channel* (E-DPCCH) is used for this purpose.

### 10.2.1 MAC-e and physical layer processing

Similar to HSDPA, short delays and rapid adaptation are important aspects of the Enhanced Uplink. This is implemented by introducing the MAC-e, a new entity in the NodeB responsible for scheduling and hybrid ARQ protocol operation. The physical layer is also enhanced to provide the necessary support for a short TTI and for soft combining in the hybrid ARQ mechanism.

In soft handover, uplink data can be received in multiple NodeBs. Consequently, there is a need for a MAC-e entity in each of the involved NodeBs to handle the hybrid ARQ protocol. The MAC-e in the serving cell is, in addition, responsible for handling the scheduling operation.

To handle the Enhanced Uplink processing in the terminal, there is also a MAC-e entity in the UE. This can be seen in Figure 10.5, where the Enhanced Uplink processing in the UE is illustrated. The MAC-e in the UE consists of MAC-e multiplexing, transport format selection, and the protocol parts of the hybrid ARQ mechanism.

Mixed services, for example simultaneous file upload and VoIP, are supported. Hence, as there is only a single E-DCH transport channel, data from multiple MAC-d flows can be multiplexed through MAC-e multiplexing. The different services are in this case typically transmitted on different MAC-d flows as they may have different quality-of-service requirements.

Only the UE has accurate knowledge about the buffer situation and power situation in the UE at the time of transmission of a transport block in the uplink. Hence, the UE is allowed to autonomously select the data rate or, strictly speaking, the *E-DCH Transport Format Combination* (E-TFC). Naturally, the UE needs to take the scheduling decisions into account in the transport format selection; the scheduling decision represents an upper limit of the data rate the UE is not allowed to exceed. However, it may well use a lower data rate, for example if the

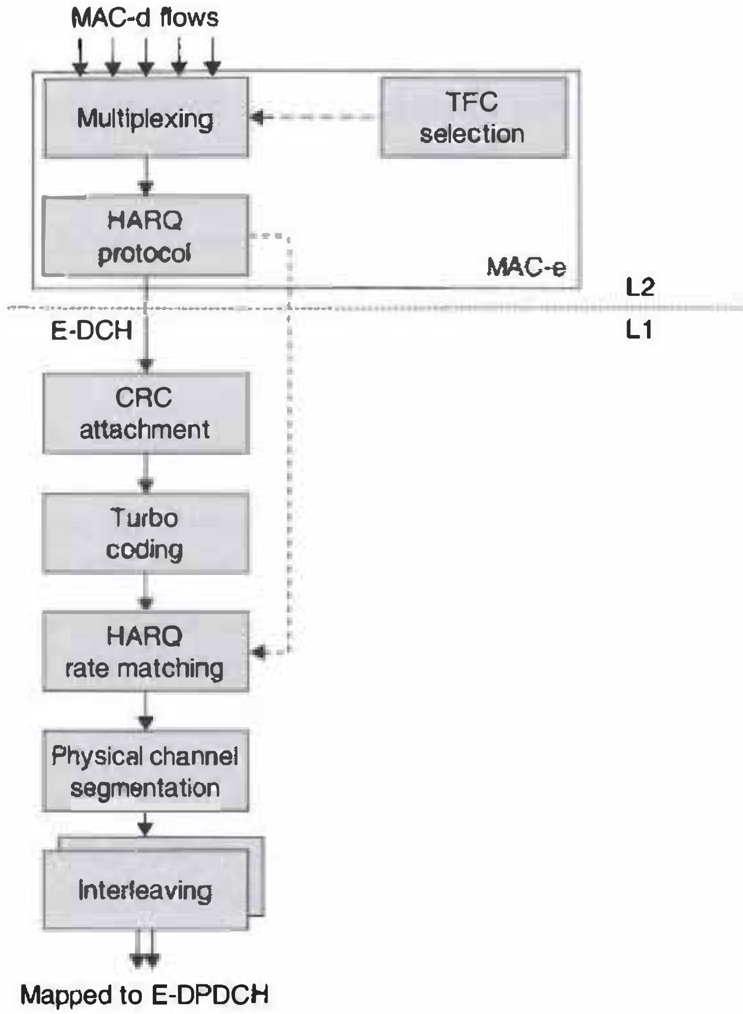


**Figure 10.5** *MAC-e and physical-layer processing.*

transmit power does not support the scheduled data rate. E-TFC selection, including MAC-e multiplexing, is discussed further in conjunction with scheduling.

The hybrid ARQ protocol is similar to the one used for HSDPA, that is multiple stop-and-wait hybrid ARQ processes operated in parallel. There is one major difference though – when the terminal is in soft handover with several NodeBs, the hybrid ARQ protocol is terminated in multiple nodes.

Physical layer processing is straightforward and has several similarities with the HS-DSCH physical layer processing. From the MAC-e in the UE, data is passed to the physical layer in the form of one transport block per TTI on the E-DCH. Compared to the DCH coding and multiplexing chain, the overall structure of the E-DCH physical layer processing is simpler as there is only a single E-DCH and hence no transport channel multiplexing.

A 24-bit CRC is attached to the single E-DCH transport block to allow the hybrid ARQ mechanism in the NodeB to detect any errors in the received transport block. Coding is done using the same rate 1/3 Turbo coder as used for HSDPA.

The physical layer hybrid ARQ functionality is implemented in a similar way as for HSDPA. Repetition or puncturing of the bits from the Turbo coder is used to adjust the number of coded bits to the number of channel bits. By adjusting the puncturing pattern, different redundancy versions can be generated.

Physical channel segmentation distributes the coded bits to the different channelization codes used, followed by interleaving and modulation.

### 10.2.2 Scheduling

Scheduling is one of the fundamental technologies behind Enhanced Uplink. In principle, scheduling is possible already in the first version of WCDMA, but Enhanced Uplink supports a significantly faster scheduling operation thanks to the location of the scheduler in the NodeB.

The responsibility of the scheduler is to control *when* and *at what data rate* a UE is allowed to transmit, thereby controlling the amount of interference affecting other users at the NodeB. This can be seen as controlling each UE's consumption of common resources, which in case of Enhanced Uplink is the amount of tolerable interference, that is the total received power at the base station. The amount of common uplink resources a terminal is using depends on the data rate used. Generally, the higher the data rate, the larger the required transmission power and thus the higher the resource consumption.

The term *noise rise* or *rise-over-thermal* is often used when discussing uplink operation. Noise rise, defined as $(I_0 + N_0)/N_0$ where $N_0$ and $I_0$ are the noise and interference power spectral densities, respectively, is a measure of the increase in interference in the cell due to the transmission activity. For example, 0dB noise rise indicates an unloaded system and 3dB noise rise implies a power spectral density due to uplink transmission equal to the noise spectral density. Although noise rise as such is not of major interest, it has a close relation to coverage and uplink load. A too large noise rise would result in loss of coverage for some channels – a terminal may not have sufficient transmission power available to reach the required $E_b/N_0$ at the base station. Hence, the uplink scheduler must keep the noise rise within acceptable limits.

*Channel-dependent scheduling*, which typically is used in HSDPA, is possible for the uplink as well, but it should be noted that the benefits are different. As fast power control is used for the uplink, a terminal transmitting when the channel conditions are favorable will generate the same amount of interference in the cell as a terminal transmitting in unfavorable channel conditions, given the same data rate for the two. This is in contrast to HSDPA, where in principle a constant transmission power is used and the data rates are adapted to the channel conditions, resulting in a higher data rate for users with favorable radio conditions. However, for the uplink the transmission power will be different for the two terminals. Hence, the amount of interference generated in *neighboring* cells will differ. Channel-dependent scheduling will therefore result in a lower noise rise in the system, thereby improving capacity and/or coverage.

In practical cases, the transmission power a UE is limited by several factors, both regulatory restrictions and power amplifier implementation restrictions. For WCDMA, different power classes are specified limiting the maximum power the UE can use to, where 21 dBm is a common value of the maximum power. This affects the discussion on uplink scheduling, making channel-dependent scheduling beneficial also from an intra-cell perspective. A UE scheduled when the channel conditions are beneficial encounters a reduced risk of hitting its transmission power limitation. This implies that the UE is likely to be able to transmit at a higher data rate if scheduled to transmit at favorable channel conditions. Therefore, taking channel conditions into account in the uplink scheduling decisions will improve the capacity, although the difference between non-channel-dependent and channel-dependent scheduling in most cases not are as large as in the downlink case.

Round-robin scheduling is one simple example of an uplink scheduling strategy, where terminals take turn in transmitting in the uplink. Similar to round-robin scheduling in HSDPA, this results in TDMA-like operation and avoids intra-cell interference due to the non-orthogonal uplink. However, as the maximum transmission power of the terminals is limited, a single terminal may not be able to fully utilize the uplink capacity when transmitting and thus reducing the uplink capacity in the cell. The larger the cells, the higher the probability that the UE does not have sufficient transmit power available.

To overcome this, an alternative is to assign the same data rate to all users having data to transmit and to select this data rate such that the maximum cell load is respected. This results in maximum fairness in terms of the same data rate for all users, but does not maximize the capacity of the cell. One of the benefits though is the simple scheduling operation – there is no need to estimate the uplink channel quality and the transmission power status for each UE. Only the buffer status of each UE and the total interference level in the cell is required.

With *greedy filling*, the terminal with the best radio conditions is assigned as high data rate as possible. If the interference level at the receiver is smaller than the maximum tolerable level, the terminal with the second best channel conditions is allowed to transmit as well, continuing with more and more terminals until the maximum tolerable interference level at the receiver is reached. This strategy maximizes the radio-interface utilization but is achieved at the cost of potentially large differences in data rates between users. In the extreme case, a user at the cell border with poor channel conditions may not be allowed to transmit at all.

Strategies between these two can also be considered such as different proportional fair strategies. This can be achieved by including a weighting factor for each user, proportional to the ratio between the instantaneous and average data rates, into the greedy filling algorithm. In a practical scenario, it is also necessary to take the transport network capacity and the processing resources in the base station into account in the scheduling decision, as well as the priorities for different data flows.

The above discussion of different scheduling strategies assumed all UEs having an infinite amount of data to transmit (full buffers). Similarly as the discussion for HSDPA, the traffic behavior is important to take into account when comparing different scheduling strategies. Packet-data applications are typically bursty in nature with large and rapid variations in their resource requirements. Hence, the overall target of the scheduler is to allocate a large fraction of the shared resource to users momentarily requiring high data rates, while at the same time ensuring stable system operation by keeping the noise rise within limits.

A particular benefit of fast scheduling is the fact that it allows for a more relaxed connection admission strategy. For the DCH, admission control typically has to reserve resources relative to the peak data rate as there are limited means to recover from an event when many or all users transmit simultaneously with their maximum rate. Admission relative to the peak rate results in a rather conservative admission strategy for bursty packet-data applications. With fast scheduling, a larger number of packet-data users can be admitted since fast scheduling provides means to control the load in case many users request for transmission simultaneously.

#### *10.2.2.1 Scheduling framework for Enhanced Uplink*

The scheduling framework for Enhanced Uplink is generic in the sense the control signaling allows for several different scheduling implementations. One major difference between uplink and downlink scheduling is the location of the scheduler and the information necessary for the scheduling decisions.

In HSDPA, the scheduler and the buffer status are located at the same node, the NodeB. Hence, the scheduling strategy is completely implementation dependent and there is no need to standardize any buffer status signaling to support the scheduling decisions.

In Enhanced Uplink, the scheduler is still located in the NodeB to control the transmission activity of different UEs, while the buffer status information is distributed among the UEs. In addition to the buffer status, the scheduler also needs information about the available transmission power in the UE: if the UE is close to its maximum transmission power there is no use in scheduling a (significantly) higher data rate. Hence, there is a need to specify signaling to convey buffer status and power availability information from the UE to the NodeB.

The basis for the scheduling framework is *scheduling grants* sent by the NodeB to the UE and limiting the E-DCH data rate and *scheduling requests* sent from the UE to the NodeB to request permission to transmit (at a higher rate than currently allowed). Scheduling decisions are taken by the *serving cell*, which has the main responsibility for scheduling as illustrated in Figure 10.6 (in case of simultaneous HSDPA and Enhanced Uplink, the same cell is the serving cell for both). However, when in soft handover, the non-serving cells have a possibility to influence the UE behavior to control the inter-cell interference.

Providing the scheduler with the necessary information about the UE situation, taking the scheduling decision based on this information, and communicating

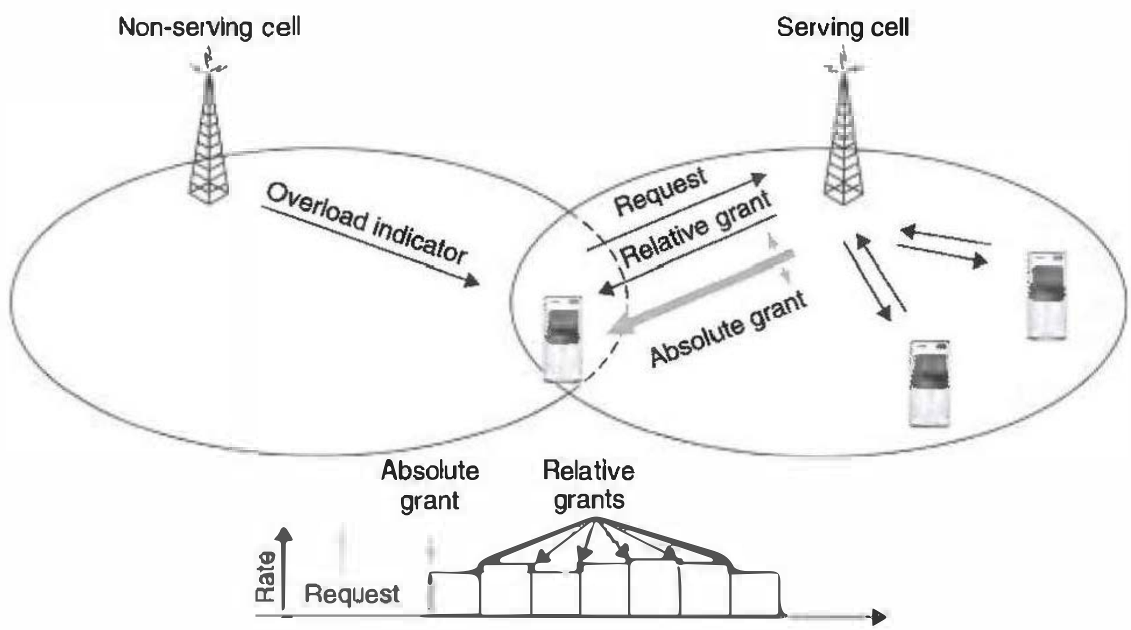


**Figure 10.6** *Overview of the scheduling operation.*

the decision back to the UE takes a non-zero amount of time. The situation at the UE in terms of buffer status and power availability may therefore be different at the time of transmission compared to the time of providing the information to the NodeB UE buffer situation. For example, the UE may have less data to transmit than assumed by the scheduler, high-priority data may have entered the transmission buffer or the channel conditions may have worsened such that the UE has less power available for data transmission. To handle such situations and to exploit any interference reductions due to a lower data rate, the scheduling grant does not set the E-DCH data rate, but rather an *upper limit* of the resource usage. The UE select the data rate or, more precisely, the *E-DCH Transport Format Combination* (E-TFC) within the restrictions set by the scheduler.

The *serving grant* is an internal variable in each UE, used to track the maximum amount of resource the UE is allowed to use. It is expressed as a maximum E-DPDCH-to-DPCCH power ratio and the UE is allowed to transmit from any MAC-d flow and using any transport block size as long as it does not exceed the serving grant. Hence, the scheduler is responsible for scheduling between UEs, while the UEs themselves are responsible to schedule between MAC-d flows according to rules in the specifications. Basically, a high-priority flow should be served before a low-priority flow.

Expressing the serving grant as a maximum power ratio is motivated by the fact that the fundamental quantity the scheduler is trying to control is uplink interference, which is directly proportional to transmission power. The E-DPDCH transmission power is defined relative to the DPCCH to ensure the E-DPDCH is affected by the power control commands. As the E-DPDCH transmission power typically is significantly larger than the DPCCH transmission power, the E-DPDCH-to-DPCCH power ratio is roughly proportional to the total transmission power, $(P_{\text{E-2DPDCH}} + P_{\text{DPCCH}})/P_{\text{DPCCH}} \approx P_{\text{E-DPDCH}}/P_{\text{DPCCH}}$, and thus setting a limit on the maximum E-DPCCH-to-DPCCH power ration corresponds to control of the maximum transmission power of the UE.

The NodeB can update the serving grant in the UE by sending an *absolute grant* or a *relative grant* to the UE (Figure 10.7). Absolute grants are transmitted on the shared E-AGCH and are used for absolute changes of the serving grant. Typically, these changes are relatively large, for example to assign the UE a high data rate for an upcoming packet transmission.

Relative grants are transmitted on the E-RGCH and are used for relative changes of the serving grant. Unlike the absolute grants, these changes are small; the change in transmission power due to a relative grant is typically in the order of

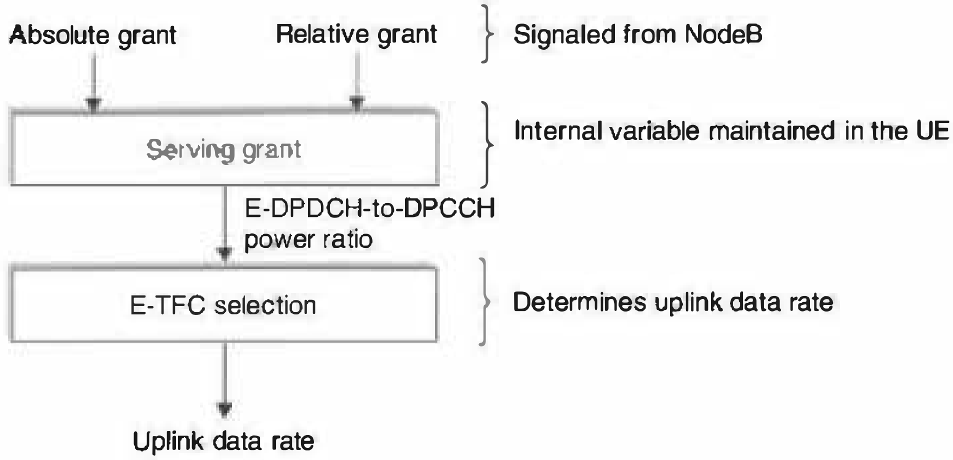


**Figure 10.7** *The relation between absolute grant, relative grant and serving grant.*

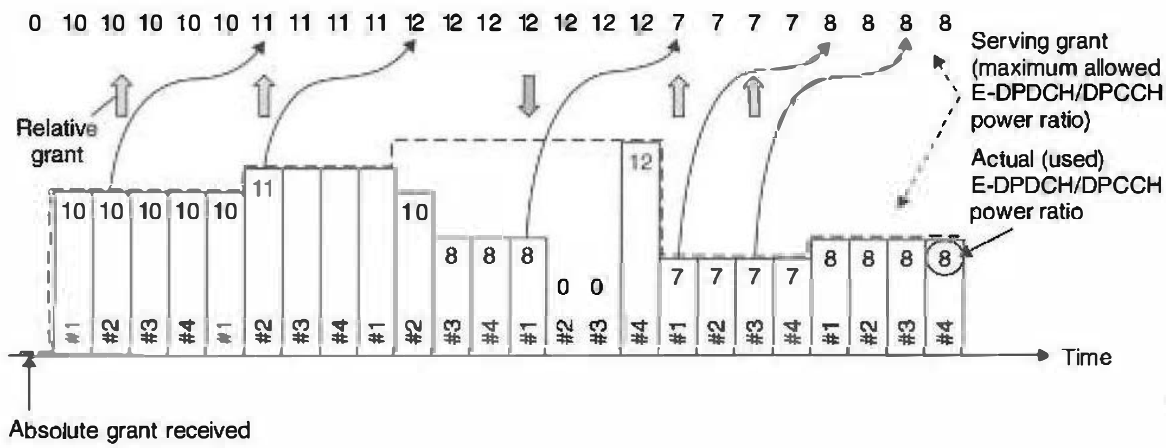


**Figure 10.8** *Illustration of relative grant usage.*

1 dB. Relative grants can be sent from both serving and, in case of the UE being in soft handover, also from the non-serving cells. However, there is a significant difference between the two and the two cases deserve to be treated separately.

Relative grants from the serving cell are dedicated for a single UE, that is each UE receives its own relative grant to allow for individual adjustments of the serving grants in different UEs. The relative grant is typically used for small, possibly frequent, updates of the data rate during an ongoing packet transmission. A relative grant from the serving cell can take one of the three values: 'UP',' 'HOLD,' or 'DOWN.' The 'up' ('down') command instructs the UE to increase (decrease) the serving grant, that is to increase (decrease) the maximum allowed E-DPDCH-to-DPCCH power ratio compared to the last used power ratio, where the last used power ratio refers to the previous TTI in the same hybrid ARQ process. The 'hold' command instructs the UE not to change the serving grant. An illustration of the operation is found in Figure 10.8.

Relative grants from non-serving cells are used to control inter-cell interference. The scheduler in the serving cell has no knowledge about the interference to neighboring cells due to the scheduling decisions taken. For example, the load in the serving cell may be low and from that perspective, it may be perfectly fine to schedule a high-rate transmission. However, the neighboring cell may not be able to cope with the additional interference caused by the high-rate transmission. Hence, there must be a possibility for the non-serving cell to influence the data rates used. In essence, this can be seen as an 'emergency break' or an 'overload indicator,' commanding non-served UEs to lower their data rate.

Although the name 'relative grant' is used also for the overload indicator, the operation is quite different from the relative grant from the serving cell. First, the overload indicator is a common signal received by all UEs. Since the non-serving cell only is concerned about the total interference level from the neighboring cell, and not which UE that is causing the interference, a common signal is sufficient. Furthermore, as the non-serving cell is not aware of the traffic priorities, etc., of the UEs it is not serving, there would be no use in having dedicated signaling from the non-serving cell.

Second, the overload indicator only takes two, not three, values – 'DTX' and 'down,' where the former does not affect the UE operation. All UEs receiving 'down' from any of the non-serving cells decrease their respective serving grant relative to the previous TTI in the same hybrid ARQ process.

#### *10.2.2.2 Scheduling information*

For efficient scheduling, the scheduler obviously needs information about the UE situation, both in terms of buffer status and in terms of the available transmission power. Naturally, the more detailed the information is, the better the possibilities for the scheduler to take accurate and efficient decisions. However, at the same time, the amount of information sent in the uplink should be kept low not to consume excessive uplink capacity. These requirements are, to some extent, contradicting and are in Enhanced Uplink addressed by providing two mechanism complementing each other: the out-band 'happy bit' transmitted on the E-DPCCH and in-band scheduling information transmitted on the E-DCH.

Out-band signaling is done through a single bit on the E-DPCCH, the 'happy bit.' Whenever the UE has available power for the E-DCH to transmit at a higher data rate compared to what is allowed by the serving grant, and the number of bits in the buffer would require more than a certain number of TTIs, the UE shall set the bit to 'not happy' to indicate that it would benefit from a higher serving grant. Otherwise, the UE shall declare 'happy.' Note that the happy bit is only transmitted in conjunction with an ongoing data transmission as the E-DPCCH is only transmitted together with the E-DPDCH.

In-band scheduling information provides detailed information about the buffer occupancy, including priority information, and the transmission power available for the E-DCH. The in-band information is transmitted in the same way as user data, either alone or as part of a user data transmission. Consequently, this information benefits from hybrid ARQ with soft combining. As in-band scheduling information is the only mechanism for the unscheduled UE to request resources, the scheduling information can be sent non-scheduled and can therefore be transmitted regardless of the serving grant. Non-scheduled transmissions are not restricted to scheduling information only; the network can configure non-scheduled transmissions also for other data.

### 10.2.3 E-TFC selection

The E-TFC selection is responsible for selecting the transport format of the E-DCH, thereby determining the data rate to be used for uplink transmission, and to control MAC-e multiplexing. Clearly, the selection needs to take the scheduling decisions taken by the NodeB into account, which is done through the serving grant as previously discussed. MAC-e multiplexing, on the other hand, is handled autonomously by the UE. Hence, while the scheduler handles resource allocation *between* UEs, the E-TFC selection controls resource allocation between flows *within* the UE. The rules for multiplexing of the flows are given by the specification; in principle, high-priority data shall be transmitted before any data of lower priority.

Introduction of the E-DCH needs to take coexistence with DCHs into account. If this is not done, services mapped onto DCHs could be affected. This would be a non-desirable situation as it may require reconfiguration of parameters set for DCH transmission. Therefore, a basic requirement is to serve DCH traffic first and only spend otherwise unused power resources on the E-DCH. Comparisons can be made with HSDPA, where any dedicated channels are served first and the HS-DSCH may use the otherwise unused transmission power. Therefore, TFC selection is performed in two steps. First, the normal DCH TFC selection is performed as in previous releases. The UE then estimates the remaining power and a second TFC selection step is performed where E-DCH can use the remaining power. The overall E-TFC selection procedure is illustrated in Figure 10.9.

Each E-TFC has an associated E-DPDCH-to-DPCCH power offset. Clearly, the higher the data rate, the higher the power offset. When the required transmitter

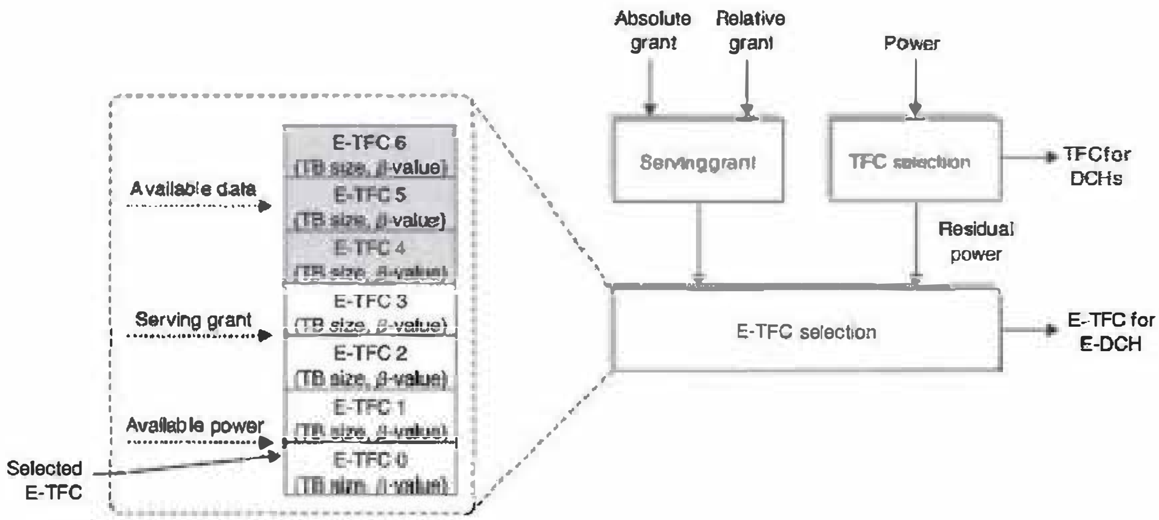


**Figure 10.9** *Illustration of the E-TFC selection process.*

power for different E-TFCs has been calculated, the UE can calculate the possible E-TFCs to use from a power perspective. The UE then selects the E-TFC by maximizing the amount of data that can be transmitted given the power constraint and the scheduling grant.

The possible transport block sizes being part of the E-TFCs are predefined in the specifications, similar to HS-DSCH. This reduces the amount of signaling, for example at handover between cells, as there is no need to configure a new set of E-TFCs at each cell change. Generally, conformance tests to ensure the UE obeys the specifications are also simpler the smaller the amount of configurability in the terminal.

To allow for some flexibility in the transport block sizes, there are four tables of E-TFCs specified; for each of the two TTIs specified there is one table optimized for common RLC PDU sizes and one general table with constant maximum relative overhead. Which one of the predefined tables that the UE shall use is determined by the TTI and RRC signaling.

### 10.2.4 Hybrid ARQ with soft combining

Hybrid ARQ with soft combining for Enhanced Uplink serves a similar purpose as the hybrid ARQ mechanism for HSDPA – to provide robustness against transmission errors. However, hybrid ARQ with soft combining is not only a tool for providing robustness against occasional errors; it can also be used for enhanced capacity as discussed in the introduction. As hybrid ARQ retransmissions are fast, many services allow for a retransmission or two. Combined with incremental redundancy, this forms an implicit rate control mechanism. Thus, hybrid ARQ with soft combining can be used in several (related) ways:

- To provide robustness against variations in the received signal quality.
- To increase the link efficiency by targeting multiple transmission attempts, for example by imposing a maximum number of transmission attempts and operating the outer loop power control on the residual error event after soft combining.

To a large extent, the requirements on hybrid ARQ are similar to HSDPA and, consequently, the hybrid ARQ design for Enhanced Uplink is fairly similar to the design used for HSDPA, although there are some differences as well, mainly originating from the support of soft handover in the uplink.

Similar to HSDPA, Enhanced Uplink hybrid ARQ spans both the MAC layer and the physical layer. The use of multiple parallel stop-and-wait processes for the hybrid ARQ protocol has proven efficient for HSDPA and is used for Enhanced Uplink for the same reasons – fast retransmission and high throughput combined with low overhead of the ACK/NAK signaling. Upon reception of the single transport block transmitted in a certain TTI and intended for a certain hybrid ARQ process, the NodeB attempts to decode the set of bits and the outcome of the decoding attempt, ACK or NAK, is signaled to the UE. To minimize the cost of the ACK/NAK, a single bit is used. Clearly, the UE must know which hybrid ARQ process a received ACK/NAK bit is associated with. Again, this is solved using the same approach as in HSDPA, that is the timing of the ACK/NAK is used to associate the ACK/NAK with a certain hybrid ARQ process. A well-defined time after reception of the uplink transport block on the E-DCH, the NodeB generates an ACK/NAK. Upon reception of a NAK, the UE performs a retransmission and the NodeB performs soft combining using incremental redundancy.

The handling of retransmissions, more specifically when to perform a retransmission, is one of the major differences between the hybrid ARQ operation in the uplink and the downlink (Figure 10.10). For HSDPA, retransmissions are scheduled as any other data and the NodeB is free to schedule the retransmission to the UE at any time instant and using a redundancy version of its choice. It may also address the hybrid ARQ processes in any order, that is it may decide to perform retransmissions for one hybrid ARQ process, but not for another process in the same UE. This type of operation is often referred to as adaptive asynchronous hybrid ARQ. Adaptive since the NodeB may change the transmission format and asynchronous since retransmissions may occur at any time after receiving the ACK/NAK.

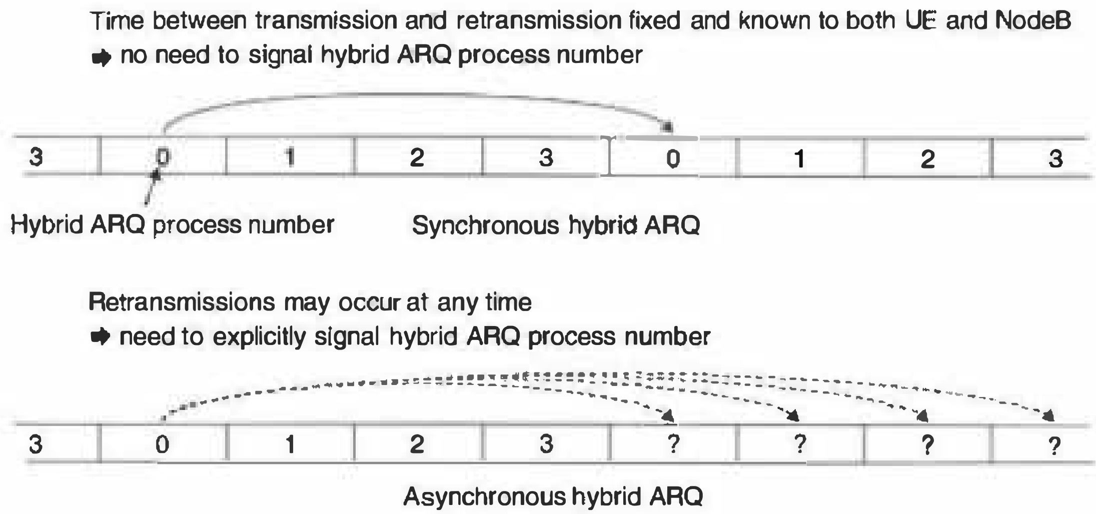


**Figure 10.10** *Synchronous vs. asynchronous hybrid ARQ.*

For the uplink, on the other hand, a synchronous, non-adaptive hybrid ARQ operation is used. Hence, thanks to the synchronous operation, retransmissions occur a predefined time after the initial transmission, that is they are not explicitly scheduled. Likewise, the non-adaptive operation implies the transport format and redundancy version to be used for each of the retransmissions is also known from the time of the original transmission. Therefore, neither is there a need to explicitly scheduling the retransmissions nor is there a need for signaling the redundancy version the UE shall use. This is the main benefit of synchronous operation of the hybrid ARQ – minimizing the control signaling overhead. Naturally, the possibility to adapt the transmission format of the retransmissions to any changes in the channel conditions are lost, but as the uplink scheduler in the NodeB has less knowledge of the transmitter status – this information is located in the UE and provided to the NodeB using in-band signaling not available until the hybrid ARQ has successfully decoded the received data – than the downlink scheduler, this loss is by far outweighed by the gain in reduced control signaling overhead.

Apart from the synchronous vs. asynchronous operation of the hybrid ARQ protocol, the other main difference between uplink and downlink hybrid ARQ is the use of soft handover in the former case. In soft handover between different NodeBs, the hybrid ARQ protocol is terminated in multiple nodes, namely all the involved NodeBs. For HSDPA, on the other hand, there is only a single termination point for the hybrid ARQ protocol – the UE. In Enhanced Uplink, the UE therefore needs to receive ACK/NAK from all involved NodeBs. As it, from the UE perspective, is sufficient if at least one of the involved NodeBs receive the transmitted transport block correctly, it considers the data to be successfully delivered to the network if at least one of the NodeBs signals an ACK. This rule is sometimes called 'or-of-the-ACKs.' A retransmission occurs only if all involved NodeBs signal a NAK, indicating that none of them has been able to decode the transmitted data.

As known from the HSDPA description, the use of multiple parallel hybrid ARQ processes cannot itself provide in-sequence delivery and a reordering mechanism is required (Figure 10.11). For HSDPA, reordering is obviously located in the UE. The same aspect with out-of-sequence delivery is valid also for the uplink, which calls for a reordering mechanism also in this case. However, due to the support of soft handover, reordering cannot be located in the NodeB. Data transmitted in one hybrid ARQ process may be successfully decoded in one NodeB, while data transmitted in the next hybrid ARQ process may happen to be correctly decoded in another NodeB. Furthermore, in some situations, several involved NodeBs may succeed in decoding the same transport block. For these reasons, the reordering mechanism needs to have access to the transport blocks delivered from all involved NodeBs and therefore the reordering is located in the RNC. Reordering also removes any duplicates of transport blocks detected in multiple NodeBs.

The presence of soft handover in the uplink has also impacted the design of the control signaling. Similar to HSDPA, there is a need to indicate to the receiving end whether the soft buffer should be cleared, that is the transmission is an initial transmission, or if soft combining with the soft information stored from previous transmissions in this hybrid ARQ process should take place. HSDPA relies on a single-bit new-data indicator, for this purpose. If the NodeB misinterpreted an uplink NAK as an ACK and continues with the next packet, the UE can capture this error event by observing the single-bit 'new-data indicator' which is incremented for each new packet transmission. If the new-data indicator is incremented, the UE will clear the soft buffer, despite its contents were not successfully decoded, and try to decode the new transmission. Although a

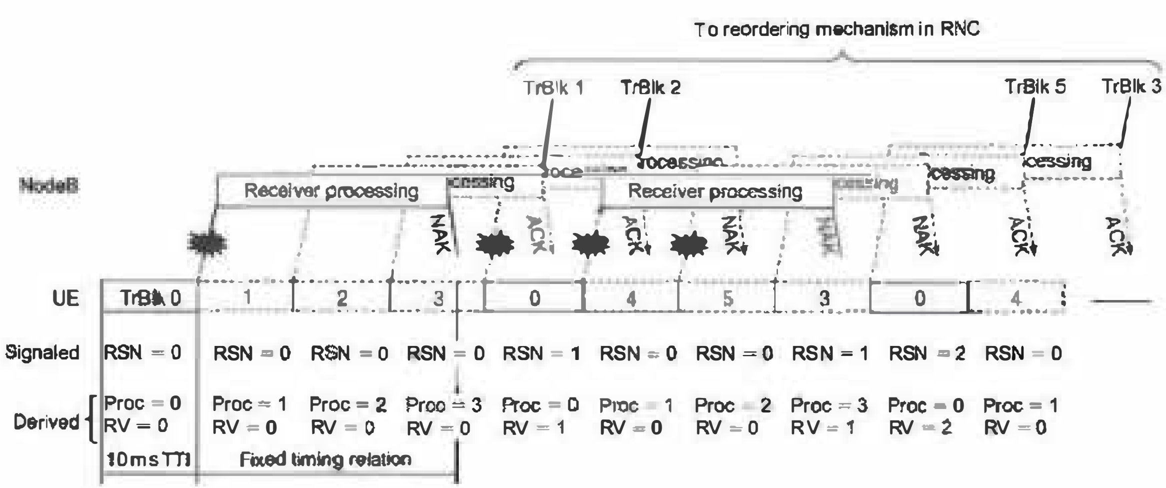


**Figure 10.11** *Multiple hybrid ARQ processes for Enhanced Uplink.*

transport block is lost and has to be retransmitted by the RLC protocol, the UE does not attempt to soft combine coded bits originating from different transport blocks and therefore the soft buffer is not corrupted. Only if the uplink NAK *and* the downlink new-data indicator are *both* misinterpreted, which is a rare event, will the soft buffer be corrupted.

For Enhanced Uplink, a single-bit new-data indicator would work in absence of soft handover. Only if the downlink NAK *and* the uplink control signaling both are misinterpreted will the NodeB soft buffer be corrupted. However, in presence of soft handover, this simple method is not sufficient. Instead, a 2-bit *Retransmission Sequence Number* (RSN) is used for Enhanced Uplink. The initial transmission sets RSN to zero and for each subsequent transmission the RSN is incremented by one. Even if the RSN only can take values in the range of 0 to 3, any number of retransmissions is possible; the RSN simply remains at 3 for the third and later retransmissions. Together with the synchronous protocol operation, the NodeB knows when a retransmission is supposed to occur and with what RSN. The simple example in Figure 10.12 illustrates the operation. As the first NodeB acknowledged packet A, the UE continues with packet B, despite that the second NodeB did not correctly decode the packet. At the point of transmission of packet B, the second NodeB expects a retransmission of packet A, but due to the uplink channel conditions at this point in time, the second NodeB does not even detect the presence of a transmission. Again, the first NodeB acknowledge the transmission and the UE continues with packet C. This time, the second

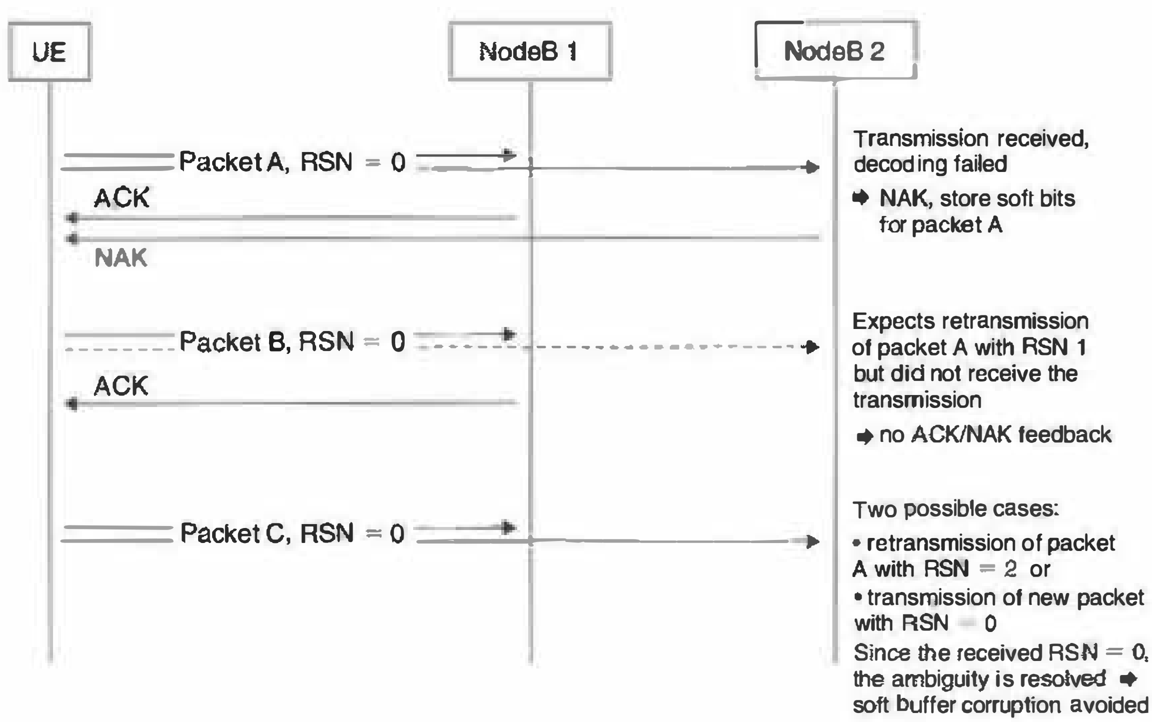


**Figure 10.12** *Retransmissions in soft handover.*

NodeB does receive the transmissions and, thanks to the synchronous hybrid ARQ operation, can immediately conclude that it must be a transmission of a new packet. If it were a retransmission of packet A, the RSN would have been equal to two. This example illustrated the improved robustness from a 2-bit RSN together with a synchronous hybrid ARQ operation. A scheme with a single-bit 'new-data indicator,' which can be seen as a 1-bit RSN, would not have been able to handle the fairly common case of a missed transmission in the second NodeB. The new-data indicator would in this case be equal to zero, both in the case of a retransmission of packet A and in the case of an initial transmission of packet C, thereby leading to soft buffer corruption.

Soft combining in the hybrid ARQ mechanism for Enhanced Uplink is based on incremental redundancy. Generation of redundancy versions is done in a similar way as for HSDPA by using different puncturing patterns for the different redundancy versions. The redundancy version is controlled by the RSN according to a rule in the specifications, see further Section 10.3.2.

For Turbo codes, the systematic bits are of higher importance than the parity bits as discussed in Chapter 7. Therefore, the systematic bits should be included in the initial transmission to allow for decodability already after the first transmission attempts. Furthermore, for the best gain with incremental redundancy, the retransmissions should contain additional parity. This leads to a design where the initial transmission is self-decodable and includes all systematic bits as well as some parity bits, while the retransmission mainly contains additional parity bits not previously transmitted.

However, in soft handover, not all involved NodeBs may have received all transmissions. There is a risk that a NodeB did not receive the first transmission with the systematic bits, but only the parity bits in the retransmission. As this would lead to degraded performance, it is preferable if all redundancy versions used when in soft handover are self-decodable and contains the systematic bits. The above-mentioned rule used to map RSN into redundancy versions does this by making all redundancy versions self-decodable for lower data rates, which typically is used in the soft handover region at the cell edge, while using full incremental redundancy for the highest data rates, unlikely to be used in soft handover.

### *10.2.5 Physical channel allocation*

The mapping of the coded E-DCH onto the physical channels is straightforward. As illustrated in Figure 10.13, the E-DCH is mapped to one or several E-DPDCHs, separate from the DPDCH. Depending on the E-TFC selected a different number

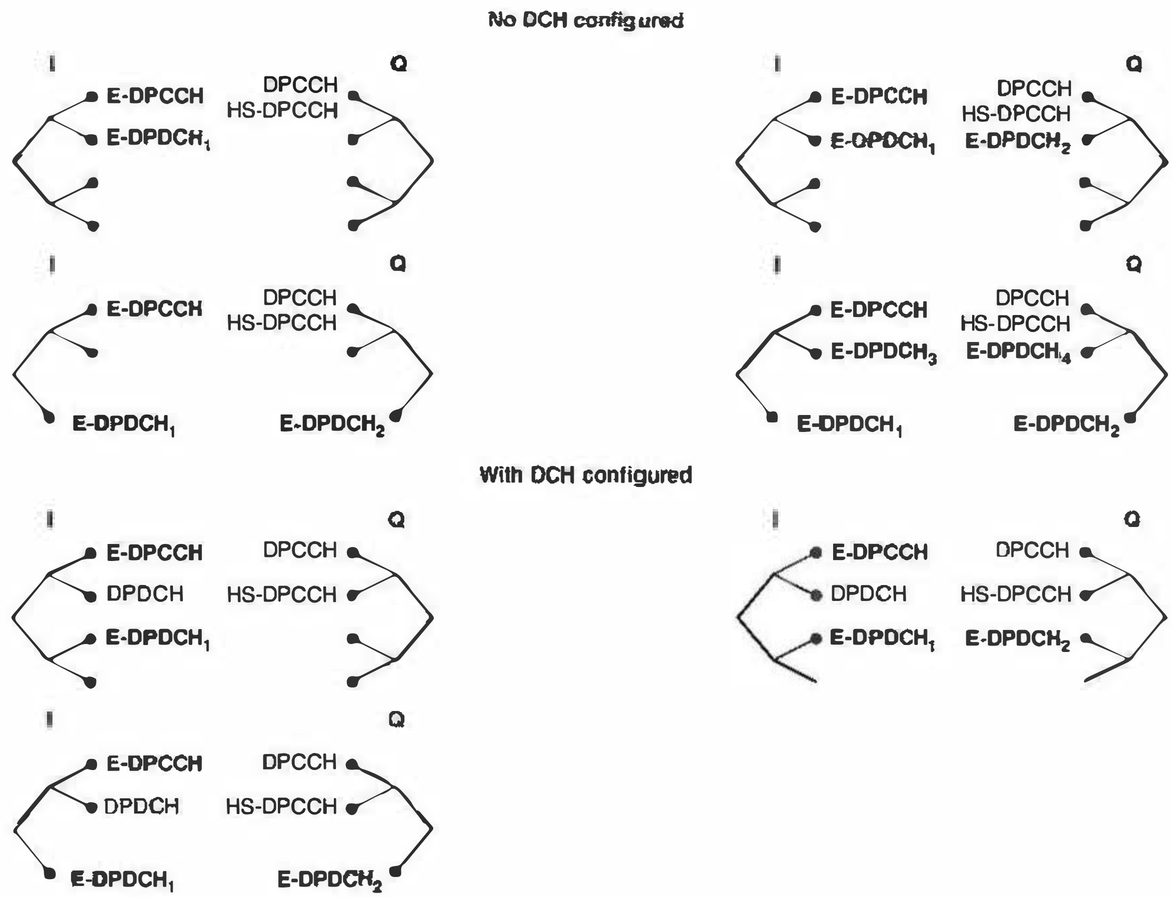


**Figure 10.13** *Code allocation in case of simultaneous E-DCH and HS-DSCH operation (note that the code allocation is slightly different when no HS-DPCCH is configured). Channels with SF > 4 are shown on the corresponding SF4 branch for illustrative purposes.*

of E-DPDCHs is used. For the lowest data rates, a single E-DPDCH with a spreading factor inversely proportional to the data rate is sufficient.

To maintain backward compatibility, the mapping of the DPCCH, DPDCH, and HS-DPCCH remains unchanged compared to previous releases.

The order in which the E-DPDCHs are allocated is chosen to minimize the peak-to-average power ratio (PAR) in the UE, and it also depends on whether the HS-DPCCH and the DPDCH are present or not. The higher the PAR, the larger the back-off required in the UE power amplifier, which impact the uplink coverage. Hence, a low PAR is a highly desirable property. PAR is also the reason why SF2 is introduced as it can be shown that two codes of SF2 have a lower PAR than four codes of SF4. For the highest data rates, a mixture of spreading factors, $2 \times SF2 + 2 \times SF4$, is used. The physical channel configurations possible are listed in Table 10.1, and in Figure 10.13 the physical channel allocation with a simultaneous HS-DPCCH is illustrated.

**Table 10.1** *Possible physical channel configurations. The E-DPDCH data rates are raw data rate, the maximum E-DCH data rate will be lower due to coding and limitations set by the UE categories.*

| #DPCCH | #DPDCH | #HS-DPCCH | #E-DPCCH | #E-DPDCH | Comment |
|---|---|---|---|---|---|
| 1 | 1–6 | 0 or 1 | – | – | Rel 5 configurations |
| 1 | 0 or 1 | 0 or 1 | 1 | $1 \times SF \geq 4$ | 0.96 Mbit/s E-DPDCH raw data rate |
| 1 | 0 or 1 | 0 or 1 | 1 | $2 \times SF4$ | 1.92 Mbit/s E-DPDCH raw data rate |
| 1 | 0 or 1 | 0 or 1 | 1 | $2 \times SF2$ | 3.84 Mbit/s E-DPDCH raw data rate |
| 1 | 0 | 0 or 1 | 1 | $2 \times SF2 + 2 \times SF4$ | 5.76 Mbit/s E-DPDCH raw data rate |

### 10.2.6 Power control

The E-DCH power control works in a similar manner as for the DCH and there is no change in the overall power control architecture with the introduction of the E-DCH. A single inner power control loop adjusts the transmission power of the DPCCH. The E-DPDCH transmission power is set by the E-TFC selection relative to the DPCCH power in a similar way as the DPDCH transmission power is set by the TFC selection. The inner loop power control located in the NodeB, bases its decision on the SIR target set by the outer loop power control located in the RNC.

The outer loop in earlier releases is primarily driven by the DCH BLER visible to the RNC. If a DCH is configured, the outer loop, which is an implementation-specific algorithm, may continue to operate on the DCH only. This approach works well as long as there are sufficiently frequent transmissions on the DCH, but the performance is degraded if DCH transmissions are infrequent.

If no DCH is configured, and possibly also if only infrequent transmissions occur at the DCH, information on the E-DCH transmissions need to be taken into account. However, due to the introduction of hybrid ARQ for the E-DCH, the residual E-DCH BLER may not be an adequate input for the outer loop power control. In most cases, the residual E-DCH BLER visible to the RNC is close to zero, which would cause the outer loop to lower the SIR target and potentially

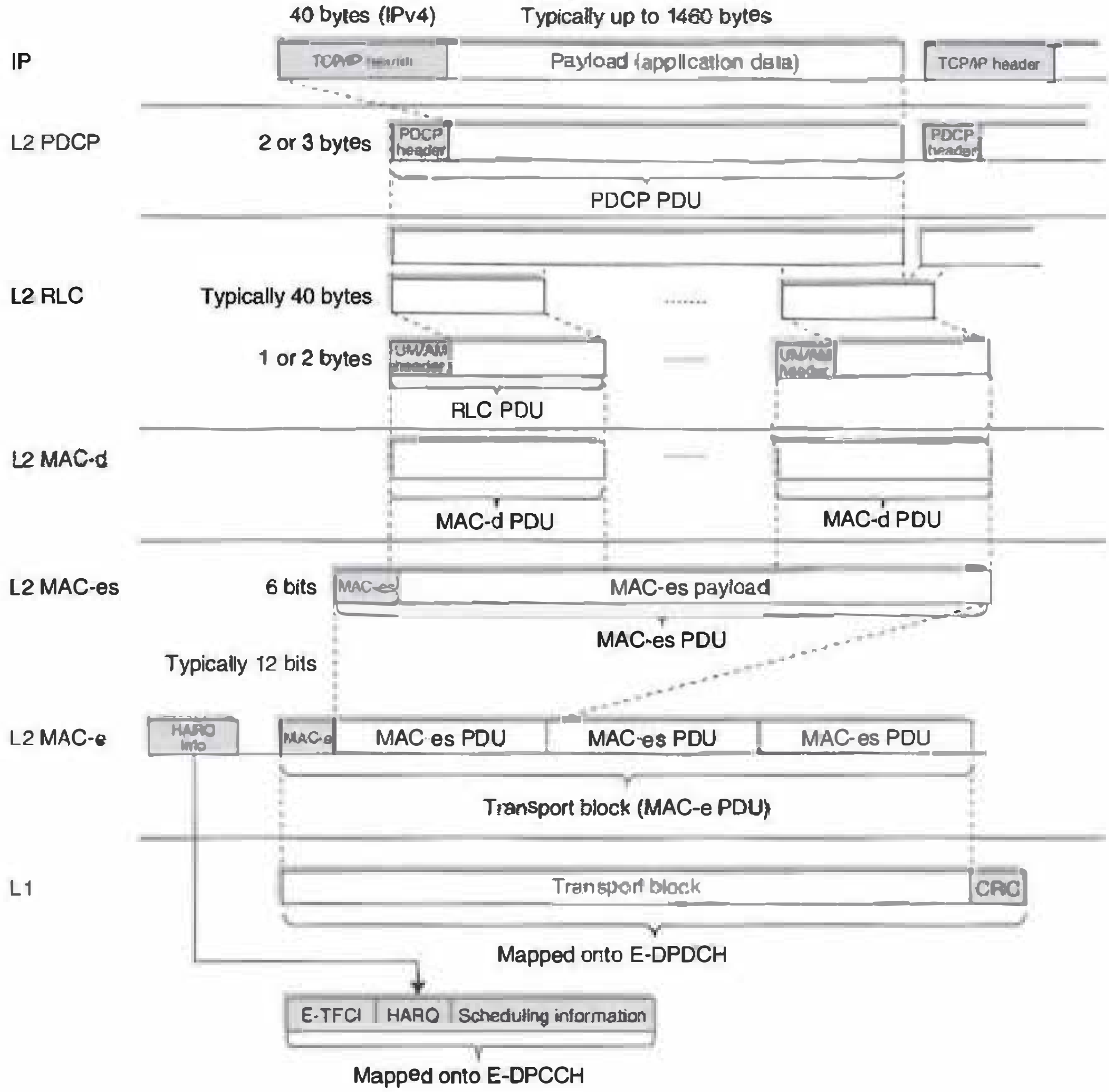


**Figure 10.14** *Data flow.*

cause a loss of the uplink DPCCH if the residual E-DCH BLER alone is used as input to the outer loop mechanism. Therefore, to assist the outer loop power control, the number of retransmissions actually used for transmission of a transport block is signaled from the NodeB to the RNC. The RNC can use this information as part of the outer loop to set the SIR target in the inner loop.

### 10.2.7 Data flow

In Figure 10.14, the data flow from the application through all the protocol layers is illustrated in a similar way as for HSDPA. In this example, an IP service is assumed. The PDCP optionally performs IP header compression. The output from the PDCP is fed to the RLC. After possible concatenation, the RLC SDUs are segmented into smaller blocks of typically 40 bytes and an RLC header is attached. The RLC PDU is passed via the MAC-d, which is transparent for Enhanced Uplink, to the MAC-e. The MAC-e concatenates one or several MAC-d PDUs from one or several MAC-d flows and inserts MAC-es and MAC-e headers to form a transport block, which is forwarded on the E-DCH to the physical layer for further processing and transmission.

### 10.2.8 Resource control for E-DCH

Similar to HSDPA, the introduction of Enhanced Uplink implies that a part of the radio resource management is handled by the NodeB instead of the RNC. However, the RNC still has the overall responsibility for radio resource management, including admission control and handling of inter-cell interference. Thus, there is a need to monitor and control the resource usage of E-DCH channels to achieve a good balance between E-DCH and non-E-DCH users. This is illustrated in Figure 10.15.

For admission control purposes, the RNC relies on the *Received Total Wideband Power* (RTWP) measurement, which indicates the total uplink resource usage in the cell. Admission control may also exploit the *E-DCH provided bit rate*, which is a NodeB measurement reporting the aggregated served E-DCH bit rate per priority class. Together with the RTWP measurement, it is possible to design an admission algorithm evaluating the E-DCH scheduler headroom for a particular priority class.

To control the load in the cell, the RNC may signal an RTWP target to the NodeB in which case the NodeB should schedule E-DCH transmissions such that the RTWP is within this limit. The RNC may also signal a reference RTWP, which the NodeB may use to improve its estimate of the uplink load in the cell. Note that whether the scheduler uses an absolute measure, such as the RTWP, or

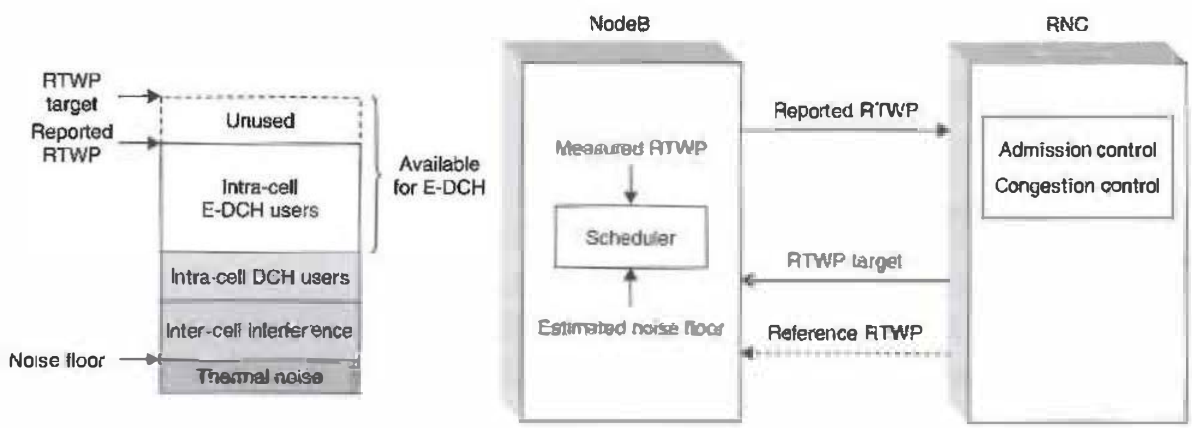


**Figure 10.15** *Illustration of the resource sharing between E-DCH and DCH channels.*

a relative measure such as noise rise is not specified. Internally, the NodeB performs any measurements useful to a particular scheduler design.

To provide the RNC with a possibility to control the ratio between inter-cell and intra-cell interference, the RNC may signal a *Target Non-serving E-DCH to Total E-DCH Power Ratio* to the NodeB. The scheduler must obey this limitation when setting the overload indicator and is not allowed to suppress non-serving E-DCH UEs unless the target is exceeded. This prevents a cell to starve users in neighboring cells. If this was not the case, a scheduler could in principle permanently set the overload indicator to 'steal' resources from neighboring cells: a situation which definitely not is desirable.

Finally, the measurement *Transmitted carrier power of all codes not used for HS-PDSCH, HS-SCCH, E-AGCH, E-RGCH, or E-HICH transmission* also includes the E-DCH-related downlink control signaling.

### 10.2.9 Mobility

Active set management for the E-DCH uses the same mechanisms as for Release 99 DCH, that is the UE measures the signal quality from neighboring cells and informs the RNC. The RNC may then take a decision to update the active set. Note that the E-DCH active set is a subset of the DCH active set. In most cases, the two sets are identical, but in situations where only part of the network support E-DCH, the E-DCH active set may be smaller than the DCH active set as the former only includes cells capable of E-DCH reception.

Changing serving cell is performed in the same way as for HSDPA (see Chapter 9) as the same cell is the serving cell for both E-DCH and HS-DSCH.

### 10.2.10 UE categories

Similar to HSDPA, the physical layer UE capabilities have been grouped into six categories. Fundamentally, two major physical layer aspects, the number of supported channelization codes and the supported TTI values, are determined by the category number. The E-DCH UE categories are listed in Table 10.2. Support for 10ms E-DCH TTI is mandatory for all UE categories, while only a subset of the categories support a 2 ms TTI. Furthermore, note that the highest data rate supported with 10ms TTI is 2 Mbit/s. The reason for this is to limit the amount of buffer memory for soft combining in the NodeB; a larger transport block size translates into a larger soft buffer memory in case of retransmissions. A UE supporting E-DCH must also be able to support HS-DSCH.

**Table 10.2** *E-DCH UE categories [99].*

| E-DCH category | Max # E-DPDCHs, min SF | Supports 2ms TTI | Max transport block size | |
|---|---|---|---|---|
| | | | 10ms TTI | 2ms TTI |
| 1 | 1 × SF4 | – | 7110 (0.7Mbit/s) | – |
| 2 | 2 × SF4 | Y | 14 484 (1.4Mbit/s) | 2798 (1.4Mbit/s) |
| 3 | 2 × SF4 | – | 14 484 (1.4Mbit/s) | – |
| 4 | 2 × SF2 | Y | 20 000 (2Mbit/s) | 5772 (2.9Mbit/s) |
| 5 | 2 × SF2 | – | 20 000 (2Mbit/s) | – |
| 6 | 2 × SF4 + 2 × SF2 | Y | 20 000 (2Mbit/s) | 11 484(5.74Mbit/s) |

## 10.3 Finer details of Enhanced Uplink

### 10.3.1 Scheduling – the small print

The use of a serving grant as a means to control the E-TFC selection has already been discussed, as has the use of absolute and relative grants to update the serving grant. Absolute grants are transmitted on the shared E-AGCH physical and are used for absolute changes of the serving grant as already stated. In addition to conveying the maximum E-DPDCH-to-DPCCH power ratio, the E-AGCH also contains an activation flag, whose usage will be discussed below. Obviously, the E-AGCH is also carrying the identity of the UE for which the E-AGCH information is intended. However, although the UE receives only *one* E-AGCH, it is assigned *two* identities, one primary and one secondary. The primary identity is UE specific and unique for each UE in the cell, while the secondary identity is a group identity shared by a group of UEs. The reason for having two identities is to allow for scheduling strategies based on both common, or group-wise, scheduling, where multiple terminals are addressed with a single identity and individual per-UE scheduling (Figure 10.16).

Common scheduling means that multiple terminals are assigned the same identity; the secondary identity is common to multiple UEs. A grant sent with the secondary identity is therefore valid for multiple UEs and each of these UEs may transmit up to the limitation set by the grant. Hence, this approach is suitable for scheduling strategies not taking the uplink radio conditions into account, for example CDMA-like strategies where scheduler mainly strives to control the total cell interference level. A low-load condition is one such example. At low cell load, there is no need to optimize for capacity. Optimization can instead focus on minimizing the delays by assigning the same grant level to multiple UEs using the secondary identity. As soon as a UE has data to transmit, the

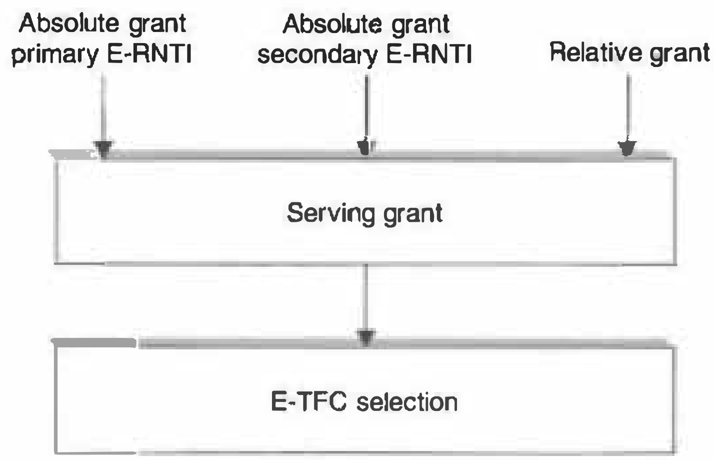


**Figure 10.16** *The relation between absolute grant, relative grant and serving grant.*

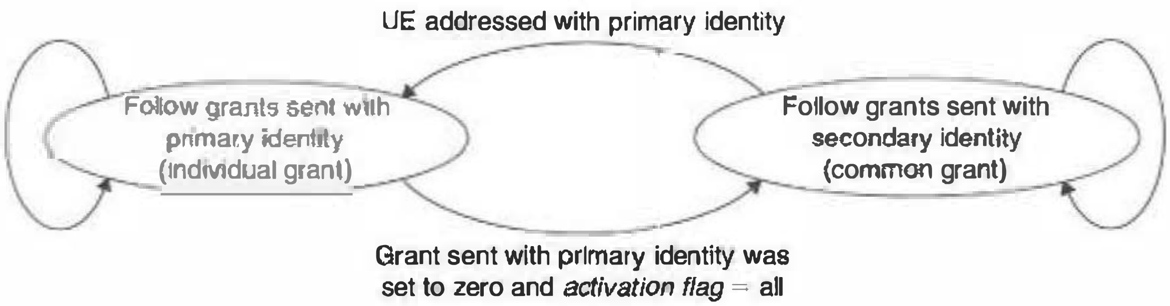


**Figure 10.17** *Illustration of UE monitoring of the two identities.*

UE can directly transmit up to the common grant level. There is no need to go through a request phase first, as the UE already has been assigned a non-zero serving grant. Note that multiple UEs may, in this case, transmit simultaneously, which must be taken into account when setting the serving grant level.

Individual per-UE scheduling provides tighter control of the uplink load and is useful to maximize the capacity at high loads. The scheduler determines which user that is allowed to transmit and set the serving grant of the intended user by using the primary identity, unique for a specific UE. In this case, the UEs resource utilization is individually controlled, for example to exploit advantageous uplink channel conditions. The greedy filling strategy discussed earlier is one example of a strategy requiring individual grants.

Which of the two identities, the primary or the secondary, a UE is obeying can be described by a state diagram, illustrated in Figure 10.17. Depending on the state the UE is in, it follows either grants sent with the primary or the secondary identity. Addressing the UE with its unique primary identity causes the UE to stop obeying grants sent using the secondary common identity. There is also a mechanism to force the UE back to follow the secondary, common grant level. The usefulness of this is best illustrated with the example below.

Consider the example in Figure 10.18, illustrating the usage of common and dedicated scheduling. The UEs are all initialized to follow the secondary identity and a suitable common grant level is set using the secondary identity. Any UE that has been assigned a grant level using the secondary identity may transmit using a data rate up to the common grant level; a level that is adjusted as the load in the system varies, for example due to non-scheduled transmissions. As time evolves, UE #1 is in need of a high data rate to upload a huge file. Note that UE #1 may start the upload using the common grant level while waiting for the scheduler to grant a higher level. The scheduler decides to lower the common grant level using the secondary, common identity to reduce the load from other UEs. A large grant is sent to UE #1 using UE #1's primary and unique identity to grant UE #1 a high data rate (or, more accurately, a higher E-DPDCH-to-DPCCH power ratio). This operation also causes UE #1 to enter the 'primary' state in Figure 10.17. At a later point in time, the scheduler decides send a zero grant to UE #1 with activation flag set to *all*, which forces UE #1 back to follow the secondary identity (back to common scheduling).

From this example, it is seen that the two identities each UE is assigned – one primary, UE-specific identity; and one secondary, common identity – facilitates a flexible scheduling strategy.

#### *10.3.1.1 Relative grants*

Relative grants from the serving cell can take one of the values 'up,' 'down,' and 'no change.' This is used to fine-tune an individual UE's resource utilization as already discussed. To implement the increase (decrease) of the serving grant, the UE maintains a table of possible E-DPDCH-to-DPCCH power ratios as

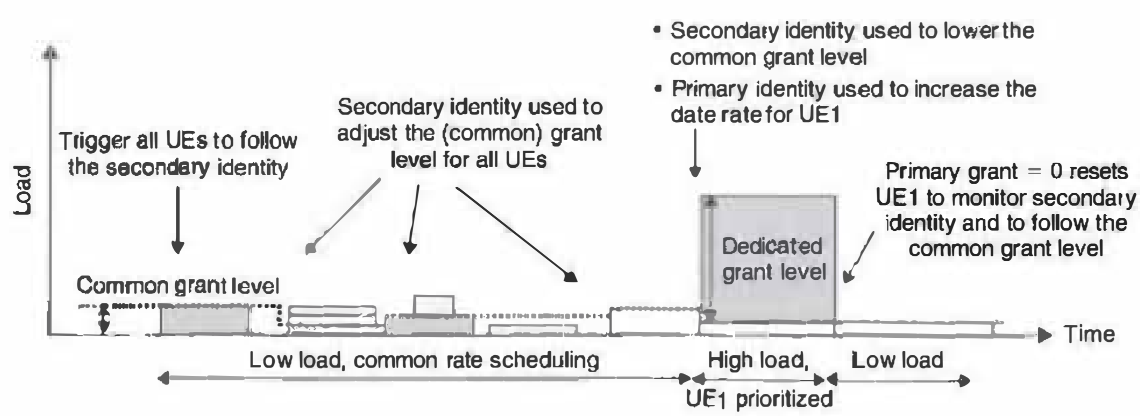


**Figure 10.18** *Example of common and dedicated scheduling.*

illustrated in Figure 10.19. The up/down commands corresponds to an increase/decrease of power ratio in the table by one step compared to the power ratio used in the previous TTI in the same hybrid ARQ process. There is also a possibility to have a larger increase (but not decrease) for small values of the serving grant. This is achieved by (through RRC signaling) configuring two thresholds in the E-DPDCH-to-DPCCH power ratio table, below which the UE may increase the serving grant by three and two steps, respectively, instead of only a single step. The use of the table and the two thresholds allow the network to increase the serving grant efficiently without extensive repetition of relative grants for small data rates (small serving grants) and at the same time avoiding large changes in the power offset for large serving grants.

The 'overload indicator' (relative grant from non-serving cells) is used to control the inter-cell interference (in contrast to the grants from the serving cell which provide the possibility to control the intra-cell interference). As previously described, the overload indicator can take two values: 'down' or 'DTX,' where the latter does not affect the UE operation. If the UE receives 'down' from any of the non-serving cells, the serving grant is decreased relative to the previous TTI in the same hybrid ARQ process.

'Ping-pong effects' describe a situation where the serving grant level in a UE starts to oscillate. One example when this could happen is if a non-serving cell

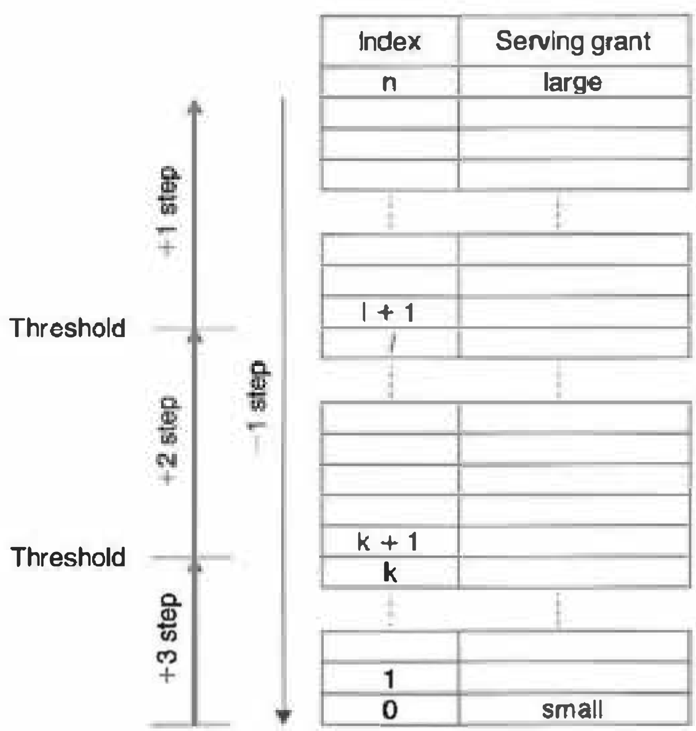


**Figure 10.19** *Grant table.*

requests the UEs to lower their transmission power (and hence data rate) due to a too high interference level in the non-served cell. When the UE has reacted to this, the serving cell will experience a lower interference level and may decide to increase the grant level in the UE. The UE utilizes this increased grant level to transmit at a higher power, which again triggers the overload indicator from the non-serving cell and the process repeats.

To avoid 'ping-pong effects,' the UE ignores any 'up' commands from the serving cell for one hybrid ARQ roundtrip time after receiving an 'overload indicator.' During this time, the UE shall not allow the serving grant to increase beyond the limit resulting from the 'overload indicator.' This avoids situations where the non-serving cell reduces the data rate to avoid an overload situation in the non-serving cell, followed by the serving cell increasing the data rate to utilize the interference headroom suddenly available, thus causing the overload situation to reappear. The serving cell may also want to be careful with immediately increasing the serving grant to is previous level as the exact serving grant in the UE is not known to the serving cell in this case (although it may partly derive it by observing the happy bit). Furthermore, to reduce the impact from erroneous relative grants, the UE shall ignore relative grants from the serving cell during one hybrid ARQ roundtrip time after having received an absolute grant with the primary identity.

### *10.3.1.2 Per-process scheduling*

Individual hybrid ARQ processes can be (de)activated, implying that not all processes in a UE are available for data transmission as illustrated in Figure 10.20. Process (de)activation is only possible for 2 ms TTI, for 10 ms TTI all processes are permanently enabled. The reason for process deactivation is mainly to be able to limit the data rate in case of 2ms E-DCH TTI (with 320-bit RLC PDUs and 2ms TTI, the minimum data rate is 160kbit/s unless certain processes are disabled), but it can also be used to enable TDMA-like operation in case of all UEs uplink transmissions being synchronized. Activation of individual processes can either be done through RRC signaling or by using the activation flag being part of

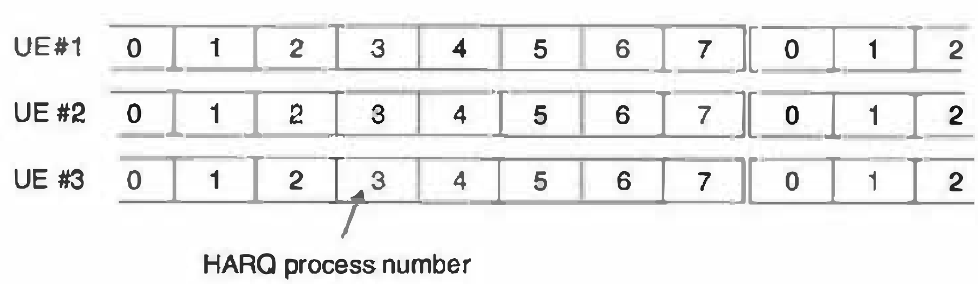


**Figure 10.20** *Example of activation of individual hybrid ARQ processes.*

the absolute grant. The activation flag indicates whether only the current process is activated or whether all processes not disabled by RRC signaling are activated.

Non-scheduled transmission can be restricted to certain hybrid ARQ processes. The decision is taken by the serving NodeB and informed to the other NodeBs in the active set through the RNC. Normally, the scheduler needs to operate with a certain margin to be able to handle any non-scheduled transmissions that may occur and restricting non-scheduled transmissions to certain processes can therefore allow the scheduler to operate with smaller margins in the remaining processes.

#### *10.3.1.3 Scheduling requests*

The scheduler needs information about the UE status and, as already discussed, two mechanisms are defined in Enhanced Uplink to provide this information: the in-band *scheduling information* and the out-band *happy bit*.

In-band scheduling information can be transmitted either alone or in conjunction with uplink data transmission. From a baseband perspective, scheduling information is no different from uplink user data. Hence, the same baseband processing and hybrid ARQ operation is used.

In case of a standalone scheduling information, the E-DPDCH-to-DPCCH power offset to be used is configured by RRC signaling. To ensure that the scheduling information reaches the scheduler, the transmission is repeated until an ACK is received from the *serving* cell (or the maximum number of transmission attempts is reached). This is different from a normal data transmission, where an ACK from *any* cell in the active set is sufficient.

In case of simultaneous data transmission, the scheduling information is transmitted using the same hybrid ARQ profile as the highest-priority MAC-d flow in the transmission (see Section 10.3.1.5 for a discussion on hybrid ARQ profiles). In this case, periodic triggering will be relied upon for reliability. Scheduling information can be transmitted using any hybrid ARQ process, including processes deactivated for data transmission. This is useful to minimize the delay in the scheduling mechanism.

The in-band scheduling information consists of 18 bits, containing information about:

- Identity of the highest-priority logical channel with data awaiting transmission, 4 bits.

- Buffer occupancy, 5 bits indicating the total number of bytes in the buffer and 4 bits indicating the fraction of the buffer occupied with data from the highest-priority logical channel.
- Available transmission power relative to DPCCH, 5 bits.

Since the scheduling information contains information about both the total number of bits and the number of bits in the highest-priority buffer, the scheduler can ensure that UEs with high-priority data is served before UEs with low-priority data, a useful feature at high loads. The network can configure for which flows scheduling information should be transmitted.

Several rules when to transmit scheduling information are defined. These are:

- If padding allows transmission of scheduling information. Clearly, it makes sense to fill up the transport block with useful scheduling information rather than dummy bits.
- If the serving grant is zero or all hybrid ARQ processes are deactivated and data enters the UE buffer. Obviously, if data enters the UE but the UE has no valid grant for data transmission, a grant should be requested.
- If the UE does have a grant, but incoming data has higher priority than the data currently in the buffer. The presence of higher-priority data should be conveyed to the NodeB as it may affect its decision to scheduler the UE in question.
- Periodically as configured by RRC signaling (although scheduling information is not transmitted if the UE buffer size equals zero).
- At cell change to provide the new cell with information about the UE status.

#### *10.3.1.4 NodeB hardware resource handling in soft handover*

From a NodeB internal hardware allocation point of view, there is a significant difference between the serving cell and the non-serving cells in soft handover: the serving cell has information about the scheduling grant sent to the UE and, therefore, knowledge about the maximum amount of hardware resources needed for processing transmissions from this particular UE, information that is missing in the non-serving cells. Internal resource management in the non-serving cells therefore requires some attention when designing the scheduler. One possibility is to allocate sufficient resources for the highest possible data rate the UE is capable of. Obviously, this does not imply any restrictions to the data rates the serving cell may schedule, but may, depending on the implementation, come at a cost of less efficient usage of processing resources in the non-serving NodeBs. To reduce this cost, the highest data rates the scheduler is allowed to assign could be limited by the scheduler design. Alternatively, the non-serving NodeB may under-allocate processing resources, knowing that it may not be able to decode the first few TTIs of a UE transmission. Once the UE starts to transmit at a high data rate, the non-serving NodeB can reallocate resources to this UE, assuming that it will continue to transmit for some time. Non-serving cells may also try to listen to the scheduling requests from the UE to the serving cell to get some information about the amount of resources the UE may need.

#### 10.3.1.5 Quality-of-service support

The scheduler operates per UE, that is it determines which UE that is allowed to transmit and at what maximum resource consumption. However, each UE may have several different flows of different priority. For example, VoIP and RRC signaling typically have a higher priority than a background file upload. Since the scheduler operates per UE, the control of different flows within a UE is *not* directly controlled by the scheduler. In principle, this could be possible, but it would increase the amount of downlink control signaling. For Enhanced Uplink, an E-TFC-based mechanism for quality-of-service support has been selected. Hence, as described earlier, the scheduler handles resource allocation *between UEs,* while the E-TFC selection controls resource allocation between flows *within the UE.*

The basis for QoS support is so-called hybrid ARQ profiles, one per MAC-d flow in the UE. A hybrid ARQ profile consists of a power offset attribute and a maximum number of transmissions allowed for a MAC-d flow.

The power offset value is used to determine the hybrid ARQ operating point, which is directly related to the number of retransmissions. In many cases, several retransmissions may fit within the allowed delay budget. Exploiting multiple transmission attempts together with soft combining is useful to reduce the cost of transmitting at a certain data rate as discussed in conjunction with hybrid ARQ.

However, for certain high-priority MAC-d flows, the delays associated with multiple hybrid ARQ retransmissions may not be acceptable. This could, for example, be the case for RRC signaling such as handover messages for mobility. Therefore, for these flows, it is desirable to increase the E-DPDCH transmission power, thereby increasing the probability for the data to be correctly received at the first transmission attempt. This is achieved by configuring a higher power offset for hybrid ARQ profiles associated with high-priority flows. Of course, the transmission must be within the limits set by the serving grant. Therefore, the payload is smaller when transmitting high-priority data with a larger power offset than when transmitting low-priority data with a smaller power offset.

The power needed for the transmission of an E-DCH transport block is calculated including the power offset obtained from the hybrid ARQ profile for flow

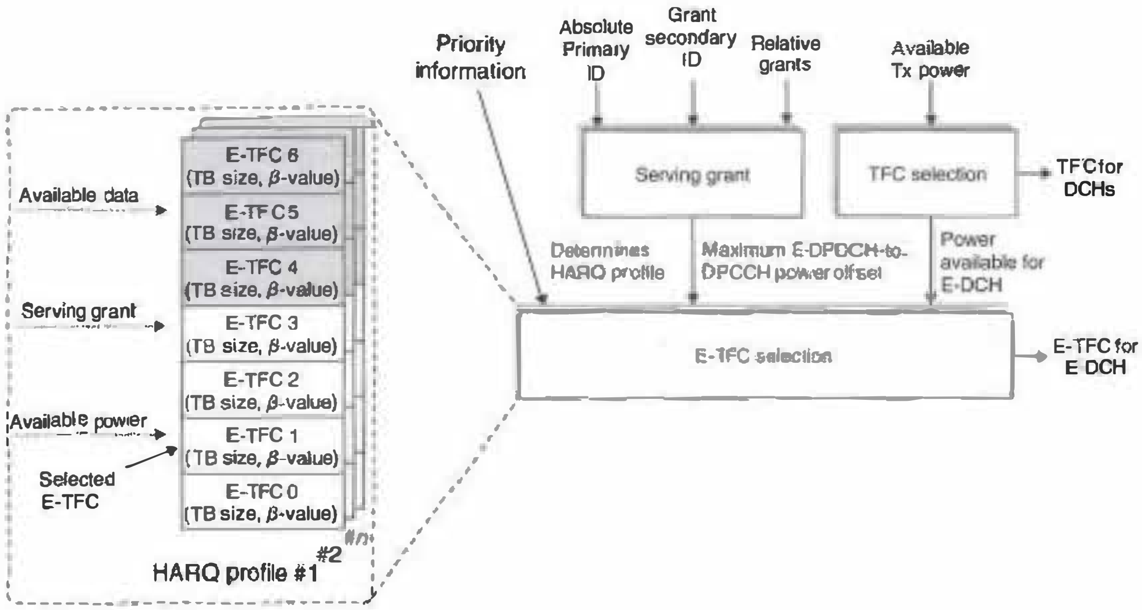


**Figure 10.21** *E-TFC selection and hybrid ARQ profiles.*

to be transmitted. The required transmit power for each possible transport block size can then be calculated by adding (in dB) the E-DPDCH-to-DPCCH power offset given by the transport block size and the power offset associated with the hybrid ARQ profile. The UE then selects the largest possible payload, taking these power offsets into account, which can be transmitted within the power available for the E-DCH (Figure 10.21).

Absolute priorities for logical channels are used, that is the UE maximizes the data rate for high-priority data and only transmits data from a low priority in a TTI if all data with higher priority has been transmitted. This ensures that any high-priority data in the UE is served before any low-priority data.

If data from more than one MAC-d flow is included in a TTI, the power offset associated with the MAC-d flow with the logical channel with the highest priority shall be used in the calculation. Therefore, if multiple MAC-d flows are multiplexed within a given transport block, the low-priority flows will get a 'free ride' in this TTI when multiplexed with high-priority data.

There are two ways of supporting guaranteed bit rate services: scheduled and non-scheduled transmissions. With scheduled transmission, the NodeB schedules the UE sufficiently frequent and with sufficiently high bit rate to support the guaranteed bit rate. With non-scheduled transmission, a flow using non-scheduled transmission is defined by the RNC and configured in the UE through RRC signaling. The UE can transmit data belonging to such a flow without first receiving any scheduling grant. An advantage with the scheduled approach is that the network has more control of the interference situation and the power required for downlink ACK/NAK signaling and may, for example, allocate a high bit rate during a fraction of the time while still maintaining the guaranteed bit rate in average. Non-scheduled transmissions, on the other hand, are clearly needed at least for transmitting the scheduling information in case the UE does not have a valid scheduling grant.

### 10.3.2 Further details on hybrid ARQ operation

Hybrid ARQ for Enhanced Uplink serves a similar purpose as for the HSDPA – to provide robustness against occasional transmissions errors. It can also, as already discussed, be used to increase the link efficiency by targeting multiple hybrid ARQ retransmission attempts.

The hybrid ARQ for the E-DCH operates on a single transport block, that is whenever the E-DCH CRC indicates an error, the MAC-e in the NodeB can request a retransmission representing the same information as the original transport block. Note that there is a single transport block per TTI. Thus it is not possible to mix initial transmission and retransmissions within the same TTI.

Incremental redundancy is used as the basic soft combining mechanism, that is retransmissions typically consists of a different set of coded bits than the initial transmission. Note that, per definition, the set of information bits must be identical between the initial transmission and the retransmissions. For Enhanced Uplink, this implies that the E-DCH transport format, which is defined by the transport block size and includes the number of physical channels and their spreading factors, remains unchanged between transmission and retransmission. Thus, the number of channel bits is identical for the initial transmission and the retransmissions. However, the rate matching pattern will change in order to implement incremental redundancy. The transmission power may also be different for different transmission attempts, for example, due to DCH activity.

The physical layer processing supporting the hybrid ARQ operation is similar to the one used for HS-DSCH, although only a single rate matching stage is used. The reason for two-stage rate matching for HS-DSCH was to handle memory limitations in the UE, but for the E-DCH, any NodeB memory limitations can be handled by proper network configuration. For example, the network could restrict the number of E-TFCs in the UE such that the UE cannot transmit more bits than the NodeB can buffer.

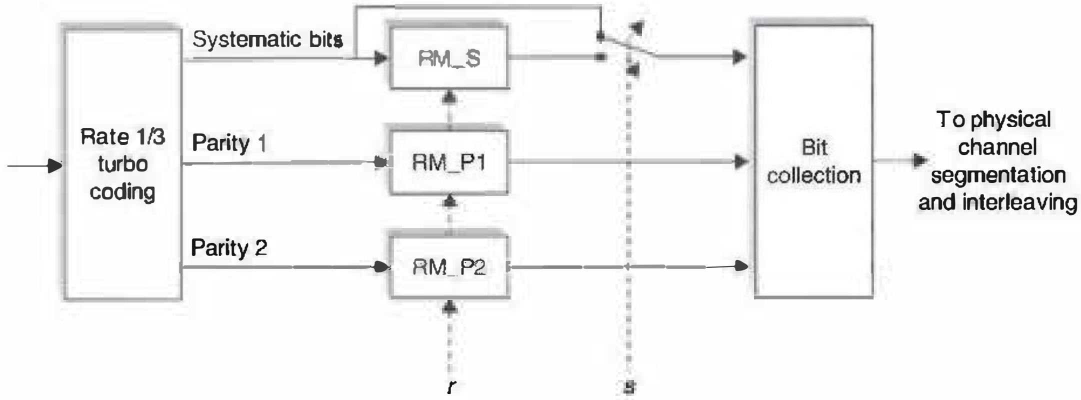


**Figure 10.22** *E-DCH rate matching and the r and s parameters. The bit collection procedure is identical to the QPSK bit collection for HS-DSCH.*

The purpose of the E-DCH rate matching, illustrated in Figure 10.22, is twofold:

- To match the number of coded bits to the number of physical channel bits on the E-DPDCH available for the selected E-DCH transport format.
- To generate different sets of coded bits for incremental redundancy as controlled by the two parameters $r$ and $s$ as described below.

The number of physical channel bits depends on the spreading factor and the number of E-DPDCHs allocated for a particular E-DCH transport format. In other words, part of the E-TFC selection is to determine the number of E-DPDCHs and their respective spreading factors. From a performance perspective, coding is always better than spreading and, preferably, the number of channelization codes should be as high as possible and their spreading factor as small as possible. This would avoid puncturing and result in full utilization of the rate 1/3 mother Turbo code. At the same time, there is no point in using a lower spreading factor than necessary to reach rate 1/3 as this only would lead to excessive repetition in the rate matching block. Furthermore, from an implementation perspective, the number of E-DPDCHs should be kept as low as possible to minimize the processing cost in the NodeB receiver as each E-DPDCH requires one set of de-spreaders.

To fulfill these, partially contradicting requirements, the concept of *Puncturing Limit* (PL) is used to control the maximum amount of puncturing the UE is allowed to perform. The UE will select an as small number of channelization codes and as high spreading factor as possible without exceeding the puncturing limits, that is not puncture more than a fraction of $(1 - \mathrm{PL})$ of the coded bits. This is illustrated in Figure 10.23, where it is seen that puncturing is allowed up to a limit before additional E-DPDCHs are used. Two puncturing limits, $\mathrm{PL}_{\mathrm{max}}$ and $\mathrm{PL}_{\mathrm{non\text{-}max}}$, are defined. The limit $\mathrm{PL}_{\mathrm{max}}$ is determined by the UE category and is used if the number of E-DPDCHs and their spreading factor is equal to the UE capability and the UE therefore cannot increase the number of E-DPDCHs.

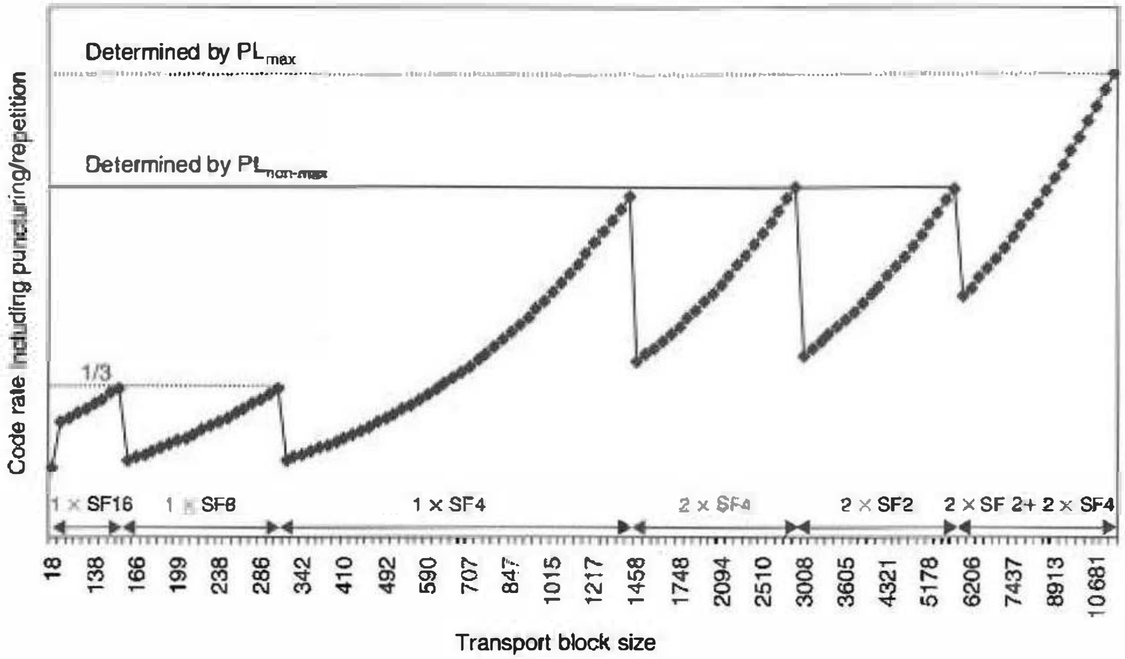


**Figure 10.23** *Amount of puncturing as a function of the transport block size.*

Otherwise, $PL_{non\text{-}max}$, which is signaled to the UE at the setup of the connection, is used. The use of two different puncturing limits, instead of a single one as for the DCH, allows for a higher maximum data rate as more puncturing can be applied for the highest data rates. Typically, additional E-DPDCHs are used when the code rate is larger than approximately 0.5. For the highest data rates, on the other hand, a significantly larger amount of puncturing is necessary as it is not possible to further increase the number of codes.

The puncturing (or repetition) is controlled by the two parameters $r$ and $s$ in the same way as for the second HS-DSCH rate matching stage (Figure 10.22). If $s = 1$, systematic bits are prioritized and an equal amount of puncturing is applied to the two streams of parity bits, while if $s = 0$, puncturing is primarily applied to the systematic bits. The puncturing pattern is controlled by the parameter $r$. For the initial transmission attempt $r$ is set to zero and is increased for the retransmissions. Thus, by varying $r$, multiple, possibly partially overlapping, sets of coded bits representing the same set of information bits can be generated. Note that a change in $r$ also affects the puncturing pattern, even if $r$ is unchanged, as different amounts of systematic and parity bits will be punctured for the two possible values of $s$.

Equal repetition for all three streams is applied if the number of available channel bits is larger than the number of bits from the Turbo coder, otherwise puncturing is applied. Unlike the DCH, but in line with the HS-DSCH, the E-DCH rate matching may puncture the systematic bits as well and not only the parity bits. This is used for incremental redundancy, where some retransmissions contain mainly parity bits.

The values of $s$ and $r$ are determined from the Redundancy Version (RV), which in turn is linked to the Retransmission Sequence Number (RSN). The RSN is set to zero for the initial transmission and incremented by one for each retransmission as described earlier.

Compared to the HS-DSCH, one major difference is the support for soft handover on the E-DCH. As briefly mentioned above, not all involved cells may receive all transmission attempts in soft handover. Self-decodable transmissions, $s = 1$, are therefore beneficial in these situations as the systematic bits are more important than the parity bits for successful decoding. If full incremental redundancy is used in soft handover, there is a possibility that the first transmission attempt, containing the systematic bits ($s = 1$), is not reliably received in a cell, while the second transmission attempt, containing mostly parity bits ($s = 0$), is received. This could result in degraded performance. However, the data rates in soft handover are typically somewhat lower (the code rate is lower) as the UE in most cases are far from the base station when entering soft handover. Therefore, the redundancy versions are defined such that all transmissions are self-decodable ($s = 1$) for transport formats where the initial code rate is less than 0.5, while the remaining transport formats include retransmissions that are not self-decodable. Thus, thanks to this design, self-decodability 'comes for free' when in soft handover. The design is also well matched to the fact that incremental redundancy (i.e., $s = 0$ for some of the retransmissions) provides most of the gain when the initial code rate is high.

The mapping from RSN via RV to the $r$ and $s$ parameters, illustrated in Figure 10.24, is mandated in the specification and is not configurable with the exception that higher layer signaling can be used to mandate the UE to always use RV = 0,

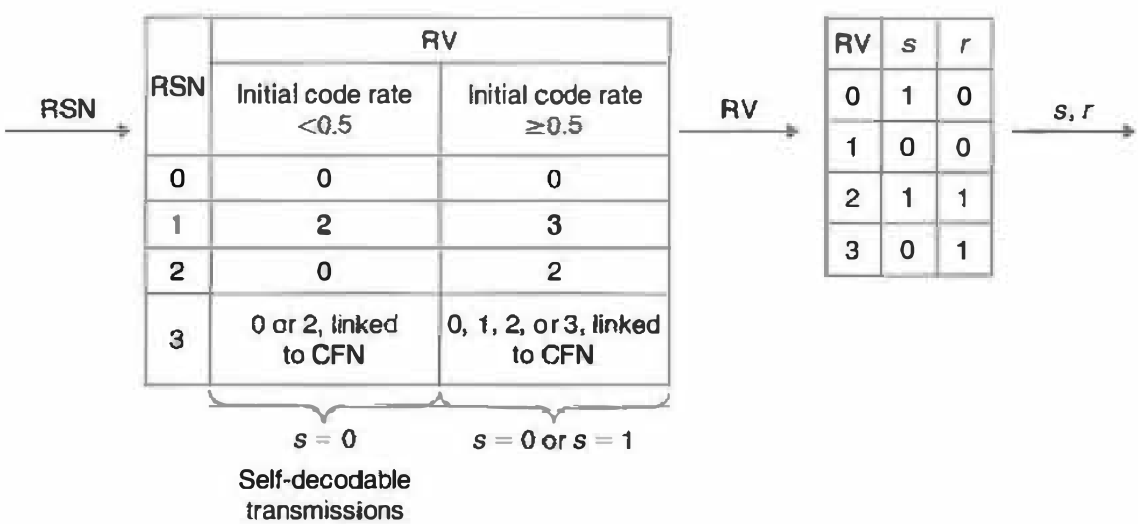


**Figure 10.24** *Mapping from RSN via RV to s and r.*

regardless of the RSN. This implies that the retransmission consists of exactly the same coded bits as the initial transmission (Chase combining), and can be used if the memory capabilities of the NodeB are limited. Note that, for RSN = 3, the RV is linked to the (sub)frame number. The reason is to allow for variations in the puncturing pattern even for situations when more than three retransmissions are used.

#### *10.3.2.1 Protocol operation*

The hybrid ARQ protocol uses multiple stop-and-wait hybrid ARQ processes similar to HS-DSCH. The motivation is to allow continuous transmission, which cannot be achieved with a single stop-and-wait scheme, while at the same time having some of the simplicity of a stop-and-wait protocol.

As already touched upon several times, the support for soft handover is one of the major differences between the uplink and the downlink. This has also impacted the number of hybrid ARQ processes. For HSDPA, this number is configurable to allow for different NodeB implementations. Although the same approach could be taken for Enhanced Uplink, soft handover between two NodeBs from different vendors would be complicated. In soft handover, all involved NodeBs need to use the same number of hybrid ARQ processes, which partially removes the flexibility of having a configurable number as all NodeBs must support at least one common configuration. To simplify the overall structure, a fixed number of hybrid ARQ processes to be used by all NodeBs are defined. The number of processes is strongly related to the timing of the ACK/NAK transmission in the downlink (see the discussion on control signaling timing for details). For the two TTIs of 10 and 2 ms, the number of processes, $N_{\mathrm{HARQ}}$, is 4 and 8, respectively. This results in a total hybrid ARQ roundtrip time of $4.10 = 40$ and $8.2 = 16$ ms, respectively.

The use of a synchronous hybrid ARQ protocol is a distinguishing feature compared to HSDPA. In a synchronous scheme, the hybrid ARQ process number is derived from the (sub)frame number and is not explicitly signaled. This implies that the transmissions in a given hybrid ARQ process can only be made once every $N_{\mathrm{HARQ}}$ TTI. This also implies that a retransmission (when necessary) always occur $N_{\mathrm{HARQ}}$ TTIs after the previous (re)transmission. Note that this does not affect the delay until a first transmission can be made since a data transmission can be started in any available process. Once the transmission of data in a process has started, retransmissions will be made until either an ACK is received or the maximum number of retransmissions has been reached (the maximum number of retransmissions is configurable by the RNC via RRC signaling). The retransmissions are done without the need for scheduling grants; only the initial transmission needs to be scheduled. As the scheduler in the NodeB is aware of whether a retransmission is expected or not, the interference from the (non-scheduled) retransmissions can be taken into account when forming the scheduling decision for other users.

#### 10.3.2.2 In-sequence delivery

Similar to the case for HS-DSCH, the multiple hybrid ARQ processes of E-DCH cannot, in themselves, ensure in-sequence delivery, as there is no interaction between the processes. Also, in soft handover situations, data is received independently in several NodeBs and can therefore be received in the RNC in a different order than transmitted. In addition, differences in Iub/Iur transport delay can cause out-of sequence delivery to RLC. Hence, in-sequence delivery must be implemented on top of the MAC-e entity and a reordering entity in the RNC has been defined for this purpose in a separate MAC entity, the MAC-es. In E-DCH, the reordering is always performed per logical channel such that all data for a logical channel is delivered in-sequence to the corresponding RLC entity. This can be compared to HS-DSCH where the reordering is performed in configurable reordering queues.

The actual mechanism to perform reordering in the RNC is implementation specific and not standardized, but typically similar principles as specified for the HS-DSCH are used. Therefore, each MAC-es PDU transmitted from the UE includes a *Transmission Sequence Number* (TSN), which is incremented for each transmission on a logical channel. By ordering the MAC-es PDUs based on TSN, in-sequence delivery to the RLC entities is possible.

To illustrate the reordering mechanism consider the situation shown in Figure 10.25. The MAC-es PDUs 0, 2, 3, and 5 have been received in the RNC while MAC-es PDUs 1 and 4 have not yet been received. The RNC can in this situation not know why PDUs 1 and 4 are missing and needs to store PDUs 2, 3, and 5 in the reordering buffer. As soon as PDU 1 arrives, PDU 1, 2, and 3 can be delivered to RLC.

The reordering mechanism also needs to handle the situation where PDUs are permanently lost due to, for example, loss over Iub, errors in the hybrid ARQ

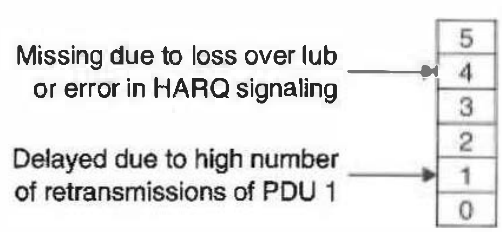


**Figure 10.25** *Reordering mechanism.*

signaling, or in case the maximum number of retransmissions was reached without successful decoding. In those situations a stall avoidance mechanism is needed, that is a mechanism to prevent that the reordering scheme waits for PDUs that never will arrive. Otherwise, PDU 5 in Figure 10.25 would never be forwarded to RLC.

Stall avoidance can be achieved with a timer similar to what is specified for the UE in HS-DSCH. The stall avoidance timer delivers packets to the RLC entity if a PDU has been missing for a certain time. If the stall avoidance mechanism delivers PDUs to the RLC entity too early, it may result in unnecessary RLC retransmissions when the PDU is only delayed, for example, due to too many hybrid ARQ retransmissions. If, on the other hand, the PDUs are kept too long in the reordering buffer, it will also degrade the performance since the delay will increase.

To improve the stall avoidance mechanism, the NodeB signals the time (frame and subframe number) when each PDU was correctly decoded to the RNC, as well as how many retransmissions were needed before the PDU was successfully received. The RNC can use this information to optimize the reordering functionality. Consider the example in Figure 10.25. If PDU 5 in the example above needed 4 retransmissions and the maximum number of retransmission attempts configured equals 5, the RNC knows that if PDU 4 has not arrived within one hybrid ARQ roundtrip time (plus some margin to consider variations in Iub delay) after PDU 5, it is permanently lost. In this case, the RNC only have to wait one roundtrip time before delivering PDU 5 to RLC.

#### *10.3.2.3 MAC-e and MAC-es headers*

To support reordering and de-multiplexing of the PDUs from different MAC-d flows, the appropriate information is signaled in-band in the form of MAC-es and MAC-e headers. The structures of the MAC-e/es headers are illustrated in Figure 10.26.

Several MAC-d PDUs of the same size and from the same logical channel are concatenated. The *Data Description Indicator* (DDI) provides information about the logical channel from which the MAC-d PDUs belong, as well as the size of the MAC-d PDUs. The number of MAC-d PDUs is indicated by *N*. The Transmission Sequence Number (TSN), used to support reordering as described in the previous section, is also attached to the set of MAC-d PDUs.

The MAC-e header consists of a number of *DDI* and *N* pairs. A mapping is provided by RRC from the *DDI* field to a MAC-d PDU size, logical channel ID and MAC-d flow ID. The logical channel also uniquely identifies the reordering queue since reordering in E-DCH is performed per logical channel.

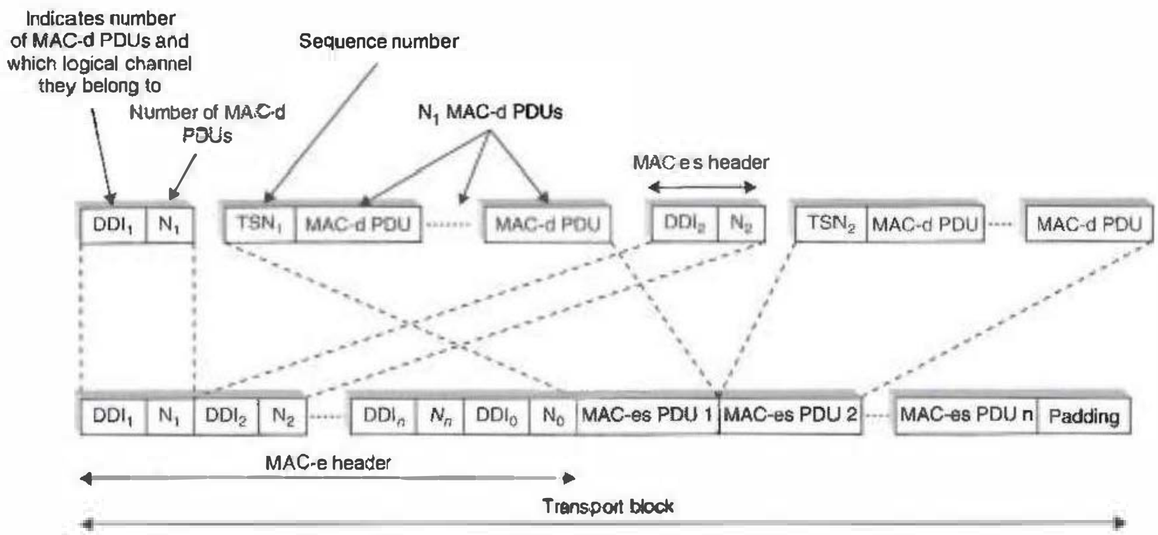


**Figure 10.26** *Structure and format of the MAC-e/es PDU.*

The sequence of *DDI* and *N* fields is completed with a predefined value of *DDI* to indicate the end of the MAC-e header. After the MAC-e header follows a number of MAC-es PDUs, where the number of MAC-es PDUs is the same as the number of *DDI* and *N* pairs in the MAC-e header (not counting the predefined DDI value indicating the end of the MAC-e header). After the last MAC-es PDU, the MAC-e PDU may contain padding to fit the current transport block size.

When appropriate, the MAC-e header also includes 18 bits of scheduling information using a special DDI value.

### 10.3.3 Control signaling

To support E-DCH transmissions in the uplink, three downlink channels carrying out-band control signaling are defined:

- The E-HICH is a dedicated physical channel transmitted from each cell in the active set and used to carry the hybrid ARQ acknowledgments.
- The E-AGCH is a shared physical channel transmitted from the serving cell only and used to carry the absolute grants.
- The E-RGCH carries relative grants. From the serving cell, the E-RGCH is a dedicated physical channel, carrying the relative grants. From non-serving cells, the E-RGCH is a common physical channel, carrying the overload indicator.

Thus, a single UE will receive multiple downlink physical control channels. From the serving cell, the UE receives the E-HICH, E-AGCH, and E-RGCH. From each of the non-serving cells, the UE receives the E-HICH and the E-RGCH.

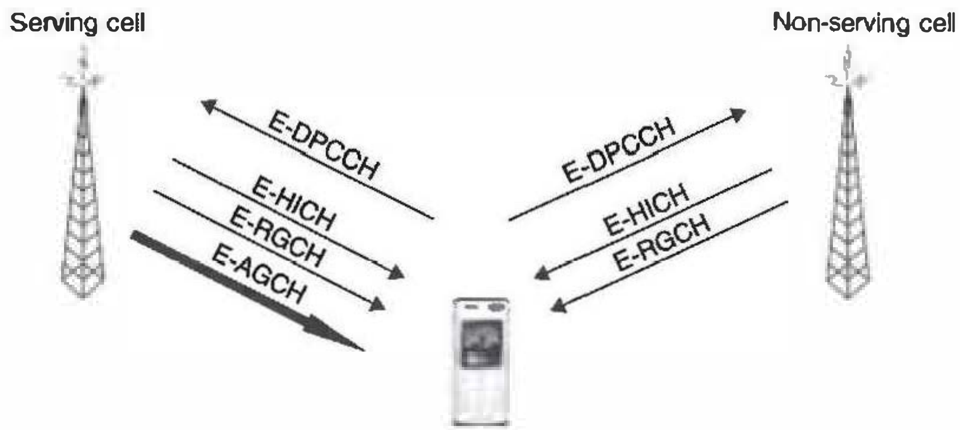


**Figure 10.27** *E-DCH-related out-band control signaling.*

Out-band uplink control signaling is also required to indicate the E-TFC the UE selected, the RSN, and the happy bit. This information is carried on the uplink E-DPCCH.

In addition to the E-DCH-related out-band control signaling, downlink control signaling for transmission of power control bits is required. This is no different from WCDMA in general and is carried on the (F-)DPCH. Similarly, the DPCCH is present in the uplink to provide a phase reference for coherent demodulation as well. The overall E-DCH-related out-band control signaling is illustrated in Figure 10.27.

#### *10.3.3.1 E-HICH*

The E-HICH is a downlink dedicated physical channel, carrying the binary hybrid ARQ acknowledgments to inform the UE about the outcome of the E-DCH detection at the NodeB. The NodeB transmits either ACK or NAK, depending on whether the decoding of the corresponding E-DCH transport block was successful or a retransmission is requested. To not unnecessarily waste downlink transmission power, nothing is transmitted on the E-HICH if the NodeB did not detect a transmission attempt, that is no energy was detected on the E-DPCCH or the E-DPDCH.

Despite the fact that the ACK/NAK is a single bit of information, the ACK/NAK is transmitted with a duration of 2 or 8 ms, depending on the TTI configured.[3] This ensures that a sufficient amount of energy can be obtained to satisfy the relatively stringent error requirements of the ACK/NAK signaling, without requiring a too high peak power for the E-HICH.

[3]The reason for 8ms and not 10ms is to provide some additional processing time in the NodeB. See the timing discussion for further details.

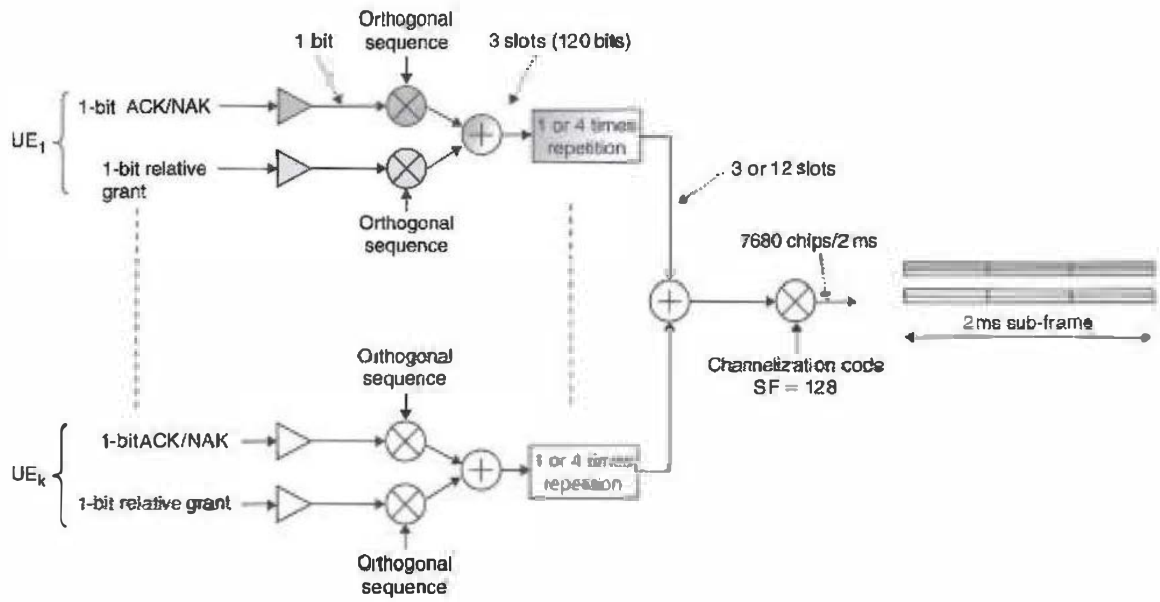


**Figure 10.28** *E-HICH and E-RGCH structures (from the serving cell).*

To save channelization codes in the downlink, multiple ACK/NAKs are transmitted on a single channelization code of spreading factor 128. Each user is assigned an orthogonal signature sequence to generate a signal spanning 2 or 8 ms as illustrated in Figure 10.28. The single-bit ACK/NAK is multiplied with a signature sequence of length 40 bits,[4] which equals one slot of bits at the specified spreading factor of 128. The same procedure is used for 3 or 12 slots, depending on the E-DCH TTI, to obtain the desired signaling interval of 2 or 8 ms. This allows multiple UEs to share a single channelization code and significantly reduces amount of channelization codes that needs to be assigned for E-HICH.

As the mutual correlation between different signature sequences varies with the sequence index, signature sequence hopping is used to average out these differences. With hopping, the signature sequence of a certain UE changes from slot to slot using a hopping pattern[5] as illustrated in Figure 10.29.

Both the E-HICH and the E-RGCH use the same structure and to simplify the UE implementation, the E-RGCH and E-HICH for a certain UE shall be allocated the same channelization code and scrambling code. Thus, with length 40 signature sequences, 20 users, each with 1 E-RGCH and 1 E-HICH, can share a single channelization code. Note that the power for different users' E-HICH and E-RGCH can be set differently, despite the fact that they share the same code.

[4]In essence, this is identical to defining a spreading factor of $40 \cdot 128$ is = 5120.

[5]The use of hopping could also be expressed as a corresponding three-slot-long signature sequence.

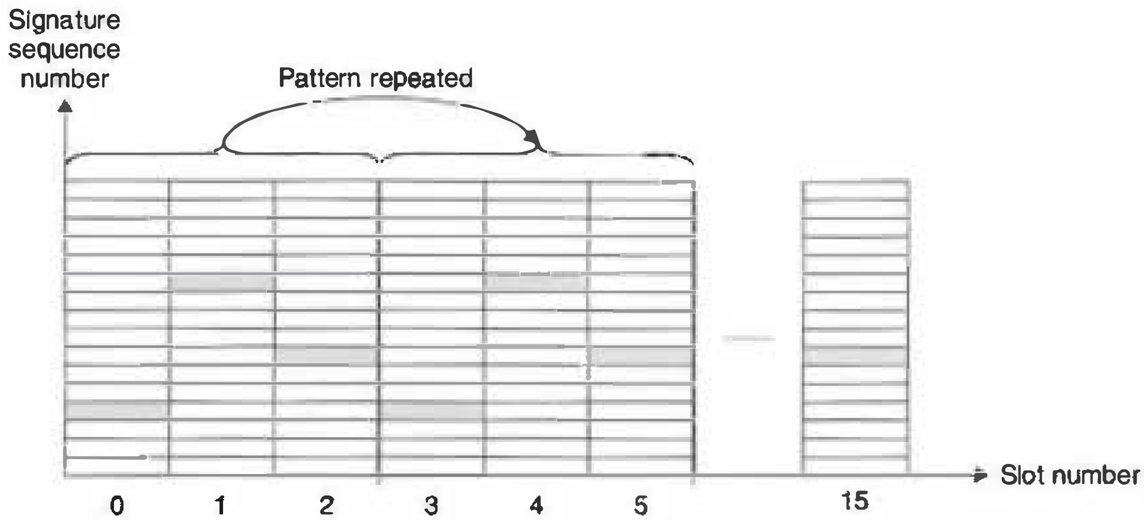


**Figure 10.29** *Illustration of signature sequence hopping.*

When a single NodeB is handling multiple cells and a UE is connected to several of those cells, that is, the UE is in softer handover between these cells, it is reasonable to assume that the NodeB will transmit the same ACK/NAK information to the UE in all these cells. Hence, the UE shall perform soft combining of the E-HICH in this case and the received signal on each of the E-HICH-es being received from the same NodeB shall be coherently added prior to decoding. This is the same approach as used for combining of power control bits already from the first release of WCDMA.

The modulation scheme used for the E-HICH is different for the serving and the non-serving cells. In the serving radio link set, BPSK is used, while for non-serving radio link sets, On-Off Keying (OOK) is used such that NAK is mapped to DTX (no energy transmitted). The reason for having different mappings is to minimize downlink power consumption. Generally, BPSK is preferable if ACK is transmitted for most of the transmissions, while the average power consumption is lower for OOK when NAK is transmitted more than 75% of the time as no energy is transmitted for the NAK. When the UE is not in soft handover, there is only the serving cell in the active set and this cell will detect the presence of an uplink transmission most of the time. Thus, BPSK is preferred for the serving cells. In soft handover, on the other hand, at most one cell is typically able to decode the transmission, implying that most of the cells will transmit a NAK, making OOK attractive. However, note that the NodeB will only transmit an ACK or a NAK in case it detected the presence of an uplink transmission attempts. If not even the presence of a data transmission is detected in the NodeB, nothing will be transmitted as described above. Hence, the E-HICH receiver in the UE must be able to handle the DTX case as well, although from a protocol point of view only the values ACK and NAK exist.

#### 10.3.3.2 E-AGCH

The E-AGCH is a shared channel, carrying absolute scheduling grants consisting of:

- The maximum E-DPDCH/DPCCH power ratio the UE is allowed to use for the E-DCH (5 bits).
- An activation flag (1 bit), used for (de)activating individual hybrid ARQ processes.
- An identity that identifies the UE (or group of UEs) for which the E-AGCH information is intended (16 bits). The identity is not explicitly transmitted but implicitly included in the CRC calculation.

Rate 1/3 convolutional coding is used for the E-AGCH and the coded bits are rate matched to 60 bits, corresponding to 2ms duration at the E-AGCH spreading factor of 256 (Figure 10.30). In case of a 10ms E-DCH TTI, the 2ms structure is repeated 5 times. Note that a single channelization code can handle a cell with both TTIs and therefore it is not necessary to reserve two channelization codes in a cell with mixed TTIs. UEs with 2ms TTI will attempt to decode each subframe of a 10ms E-AGCH without finding its identity. Similarly, a 10ms UE will combine five subframes before decoding and the CRC check will fail unless the grant was 10ms long. For group-wise scheduling, it is unlikely that both 2 and 10ms UEs will be given the same grant (although the above behavior might be exploited) and the absolute grants for these two groups of UEs can be sent separated in time on the same channelization code.

Each E-DCH-enabled UE receives one E-AGCH (although there may be one or several E-AGCH configured in a cell) from the serving cell. Although the UE is required to monitor the E-AGCH for valid information every TTI, a typical scheduling algorithm may only address a particular UE using the E-AGCH occasionally. The UE can discover whether the information is valid or not by checking the ID-specific CRC.

#### 10.3.3.3 E-RGCH

Relative grants are transmitted on the E-RGCH and the transmission structure used for the E-RGCH is identical to that of the E-HICH. The UE is expected to receive one relative grant per TTI from each of the cells in its active set. Thus, relative grants can be transmitted from both the serving and the non-serving cells.

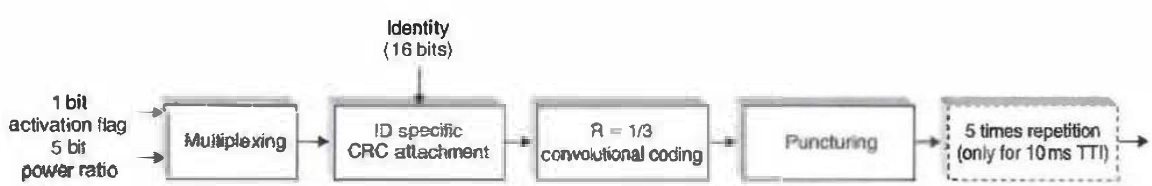


**Figure 10.30** *E-AGCH coding structure.*

From the serving cell, the E-RGCH is a dedicated physical channel and the signaled value can be one of +1, DTX, and −1, corresponding to UP, HOLD, and DOWN, respectively. Similar to the E-HICH, the duration of the E-RGCH equals 2 or 8 ms, depending on the E-DCH TTI configured.

From the non-serving cells, the E-RGCH is a common physical channel, in essence a common 'overload indicator' used to limit the amount of inter-cell interference. The value on the E-RGCH from the non-serving cells (overload indicator) can only take the values DTX and −1, corresponding to 'no overload' and DOWN, respectively. E-RGCH from the non-serving cells span 10 ms, regardless of the E-DCH TTI configured. Note that Figure 10.28 is representative for the serving cell as each UE is assigned a separate relative grant (from the non-serving cell the E-RGCH is common to multiple UEs).

#### *10.3.3.4 Timing*

The timing structure for the E-DCH downlink control channels (E-AGCH, E-RGCH, E-HICH) is designed to fulfill a number of requirements. Additional timing bases in the UE are not desirable from a complexity perspective, and hence the timing relation should either be based on the common pilot or the downlink DPCH as the timing of those channels anyway needs to be handled by the UE.

Common channels, the E-RGCH from the non-serving cell and the E-AGCH, are monitored by multiple UEs and must have a common timing. Therefore, the timing relation of these channels is defined as an offset relative to the common pilot. The duration of the E-AGCH is equal to the E-DCH TTI for which the UE is configured. For the E-RGCH from the non-serving cell, the duration is always 10 ms, regardless of the TTI. This simplifies mixing UEs with different TTIs in a single cell while providing sufficiently rapid inter-cell interference control.

Dedicated channels, the E-RGCH from the serving cell and the E-HICH, are unique for each UE. To maintain a similar processing delay in the UE and NodeB, regardless of the UE timing offset to the common pilot, their timing is defined relative to the downlink DPCH.

The structure of the E-HICH, where multiple E-HICHs share a common channelization code, has influenced the design of the timing relations. To preserve orthogonality between users sharing the same channelization code, the (sub)frame structure of the E-HICHs must be aligned. Therefore, the E-HICH timing is derived from the downlink DPCH timing, adjusted to the closest 2 ms subframe not violating the smallest UE processing requirement.

The number of hybrid ARQ processes directly affects the delay budget in the UE and NodeB. The smaller the number of hybrid ARQ processes, the better from

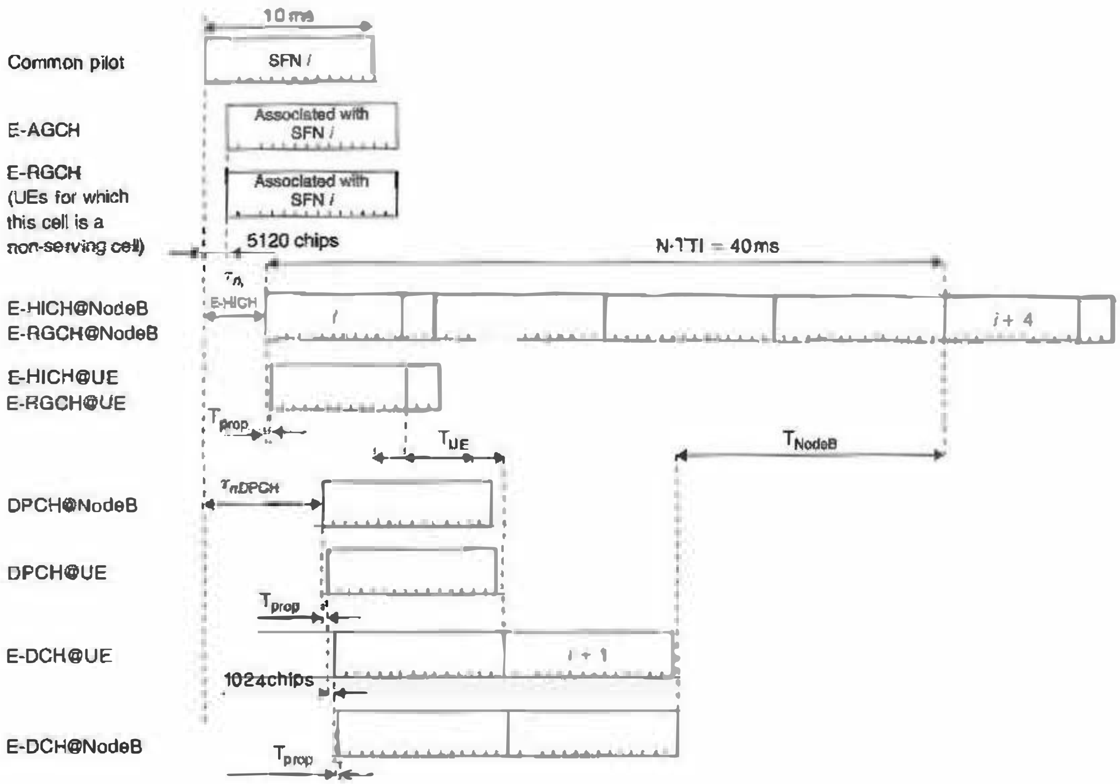


**Figure 10.31** *Timing relation for downlink control channels, 10ms TTI.*

a roundtrip time perspective but also the tighter the implementation requirements. The number of hybrid ARQ processes for E-DCH is fixed to four processes in case of a 10 ms TTI and ten processes in case of a 2 ms TTI. The total delay budget is split between the UE and the NodeB as given by the expressions relating the downlink DPCH timing to the corresponding E-HICH subframe. To allow for 2 ms extra NodeB processing delays, without tightening the UE requirements, the E-HICH duration is 8 ms, rather than 10 ms in case of a 10 ms E-DCH TTI. Note that the acceptable UE and NodeB processing delays vary in a 2 ms interval depending on the downlink DPCH timing configuration. For the UE, this effect is hard to exploit as it has no control over the network configuration and the UE design therefore must account for the worst case. The NodeB, on the other hand, may, at least in principle, exploit this fact if the network is configured to obtain the maximum NodeB processing time.

For simplicity, the timing of the E-RGCH from the serving cell is identical to that of the E-HICH. This also matches the interpretation of the relative grant in the UE as it is specified relative to the previous TTI in the same hybrid ARQ process, that is the same relation that is valid for the ACK/NAK.

The downlink timing relations are illustrated in Figure 10.31 for 10 ms E-DCH TTI and in Figure 10.32 for 2 ms TTI. An overview of the approximate processing delays in the UE and NodeB can be found in Table 10.3.

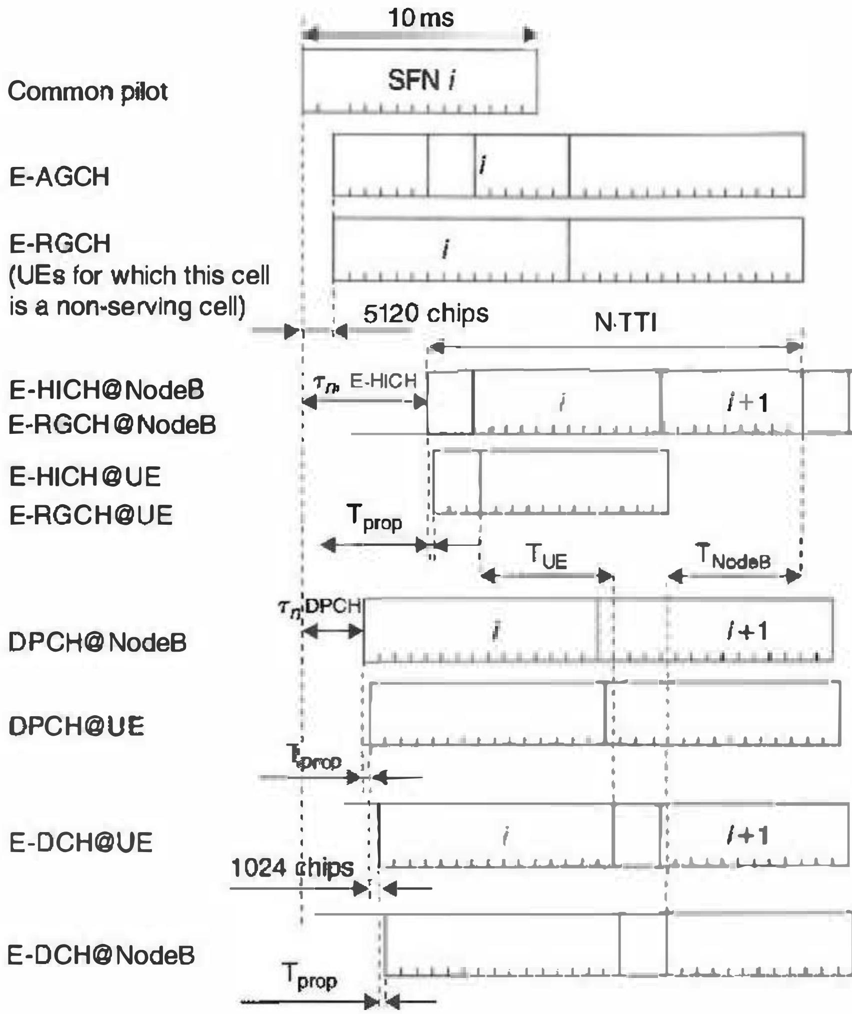


**Figure 10.32** *Timing relation for downlink control channels, 2 ms TTI.*

**Table 10.3** *Minimum UE and NodeB processing time. Note that the propagation delay has to be included in the NodeB timing budget.*

| | 10ms E-DCH TTI | 2ms E-DCH TTI |
|---|---|---|
| Number of hybrid ARQ processes | 4 | 8 |
| Minimum UE processing time | 5.56ms | 3.56ms |
| Minimum NodeB processing time | 14.1 ms | 6.1 ms |

### *10.3.3.5 Uplink control signaling: E-DPCCH*

The uplink E-DCH-related out-band control signaling, transmitted on the E-DPCCH physical channel, consist of:

- 2-bit RSN,
- 7-bit E-TFCI,
- 1-bit rate request ('happy bit').

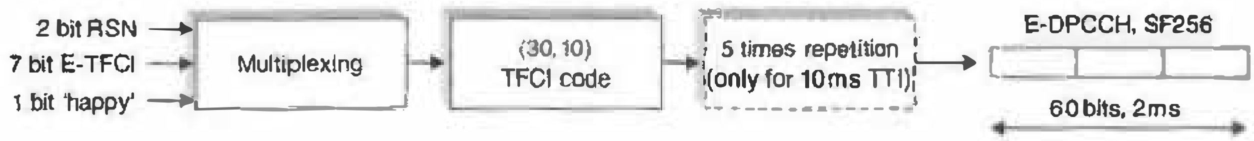


**Figure 10.33** *E-DPCCH coding.*

The E-DPCCH is transmitted in parallel to the uplink DPCCH on a separate channelization code with spreading factor 256. In this way, backward compatibility is ensured in the sense that the uplink DPCCH has retained exactly the same structure as in earlier WCDMA releases. An additional benefit of transmitting the DPCCH and the E-DPCCH in parallel, instead of time multiplexed on the same channelization code, is that it allows for independent power-level setting for the two channels. This is useful as the NodeB performance may differ between implementations.

The complete set of 10 E-DPCCH information bits are encoded into 30 bits using a second-order Reed-Müller code (the same block code as used for coding of control information on the DPCCH). The 30 bits are transmitted over three E-DPCCH slots for the case of 2 ms E-DCH TTI (Figure 10.33). In case of 10 ms E-DCH TTI, the 2 ms structure is repeated 5 times. The E-DPCCH timing is aligned with the DPCCH (and consequently the DPDCH and the E-DPDCH).

To minimize the interference generated in the cell, the E-DPCCH is only transmitted when the E-DPDCH is transmitted. Consequently, the NodeB has to detect whether the E-DPCCH is present or not in a certain subframe (DTX detection) and, if present, decode the E-DPCCH information. Several algorithms are possible for DTX detection, for example, comparing the E-DPCCH energy against a threshold depending on the noise variance.

# 20
# Flexible bandwidth in LTE

Spectrum flexibility is a key feature of the LTE radio access and is set out in the LTE requirements [86]. It consists of several components, including deployment in different-sized spectrum allocations and deployment in diverse frequency ranges, both in paired and unpaired frequency bands.

There are a number of frequency bands identified for mobile use and for IMT-2000 today. Most of these bands are already defined for operation with WCDMA/HSPA, and LTE is the next step in the 3G evolution to be deployed in those bands. Both paired and unpaired bands are included in the LTE specifications. The additional challenge with LTE operation in some bands is the possibility of using channel bandwidths up to 20MHz.

The use of OFDM in LTE gives flexibility both in terms of the size of the spectrum allocation needed and in the instantaneous transmission bandwidth used. The OFDM physical layer also enables frequency-domain scheduling. Beyond the physical layer implications described in Chapters 16 and 17, these properties also impact the RF implementation in terms of filters, amplifiers, and all other RF components that are used to transmit and receive the signal. This means that the RF requirements for the receiver and transmitter will have to be expressed with the flexibility in mind.

## 20.1 Spectrum for LTE

LTE can be deployed both in existing IMT bands and in future bands that may be identified. The possibility to operate a radio-access technology in different frequency bands is, in itself, nothing new. For example, quad-band GSM terminals are common, capable of operating in the 850, 900, 1800, and 1900 MHz bands. From a radio-access functionality perspective, this has no or limited impact and the LTE physical-layer specifications [106–109] do not assume any specific frequency band. What may differ, in terms of specification, between different bands

**Table 20.1** *Paired frequency bands defined by 3GPP for LTE.*

| Band | Uplink range (MHz) | Downlink range (MHz) | Main region(s) |
|---|---|---|---|
| 1 | 1920–1980 | 2110–2170 | Europe, Asia |
| 2 | 1850–1910 | 1930–1990 | Americas (Asia) |
| 3 | 1710–1785 | 1805–1880 | Europe, Asia (Americas) |
| 4 | 1710–1755 | 2110–2155 | Americas |
| 5 | 824–849 | 869–894 | Americas |
| 6 | 830–840 | 875–885 | Japan |
| 7 | 2500–2570 | 2620–2690 | Europe, Asia |
| 8 | 880–915 | 925–960 | Europe, Asia |
| 9 | 1749.9–1784.9 | 1844.9–1879.9 | Japan |
| 10 | 1710–1770 | 2110–2170 | Americas |
| 11 | 1427.9–1452.9 | 1475.9–1500.9 | Japan |
| 12 | 698–716 | 728–746 | Americas |
| 13 | 777–787 | 746–756 | Americas |
| 14 | 788–798 | 758–768 | Americas |

are mainly the more specific RF requirements such as the allowed maximum transmit power, requirements/limits on out-of-band (OOB) emission, and so on. One reason for this is that external constraints, imposed by regulatory bodies, may differ between different frequency bands.

### 20.1.1 Frequency bands for LTE

The frequency bands where LTE will operate will be in both paired and unpaired spectrum, requiring flexibility in the duplex arrangement. For this reason, LTE supports both FDD and TDD as discussed in the previous chapters.

Release 8 of the 3GPP specifications for LTE includes fourteen frequency bands for FDD and eight for TDD. The paired bands for FDD operation are numbered from 1 to 14 [126] as shown in Table 20.1. The unpaired bands for TDD operation are numbered from 33 to 40 as shown in Table 20.2. Note that the frequency bands for UTRA FDD use the same numbers as the paired LTE bands, but are labeled with Roman numerals from I to XIV. All bands for LTE are summarized in Figure 20.1 and Figure 20.2, which also show the corresponding frequency allocation defined by the ITU.

Some of the frequency bands are partly or fully overlapping. This is in most cases explained by regional differences in how the bands defined by the ITU are implemented. At the same time, a high degree of commonality between the

**Table 20.2** *Unpaired frequency bands defined by 3GPP for LTE.*

| Band | Frequency range (MHz) | Main region(s) |
|---|---|---|
| 33 | 1900–1920 | Europe, Asia (not Japan) |
| 34 | 2010–2025 | Europe, Asia |
| 35 | 1850–1910 | – |
| 36 | 1930–1990 | – |
| 37 | 1910–1930 | – |
| 38 | 2570–2620 | Europe |
| 39 | 1880–1920 | China |
| 40 | 2300–2400 | Europe, Asia |

bands is desired to enable global roaming. The set of bands have evolved over time as bands for UTRA, with each band originating in global, regional, and local spectrum developments. The complete set of UTRA bands was then transferred to the LTE specifications.

*Bands 1, 33, and 34* are the same paired and unpaired bands that were defined first for UTRA in Release 99 of the 3GPPP specifications. *Band 2* was added later for operation in the US PCS1900 band and *Band 3* for 3G operation in the GSM1800 band. The unpaired *Bands 35, 36, and 37* are also defined for the PCS1900 frequency ranges, but are not deployed anywhere today.

*Band 4* was introduced as a new band for the Americas following the addition of the 3G bands at WRC-2000. Its downlink overlaps completely with the downlink of Band 1, which facilitates roaming and eases the design of dual Band 1 + 4 terminals. *Band 10* is an extension of Band 4 from 2 × 45 to 2 × 60 MHz.

*Band 9* overlaps with Band 3, but is also intended only for Japan. The specifications are drafted in such a way that implementation of roaming dual Band 3 + 9 terminals is possible. The 1500MHz frequency band is also identified in 3GPP for Japan as *Band 11*. It is allocated globally to mobile service on a co-primary basis and was previously used for 2G in Japan.

With WRC-2000, the band 2500–2690 MHz was identified for IMT-2000 and it is identified as *Band 7* in 3GPP for FDD and *Band 38* for TDD operation in the 'center gap' of the FDD allocation. *Band 39* is an extension of the unpaired Band 33 from 20 to 40MHz for use in China.

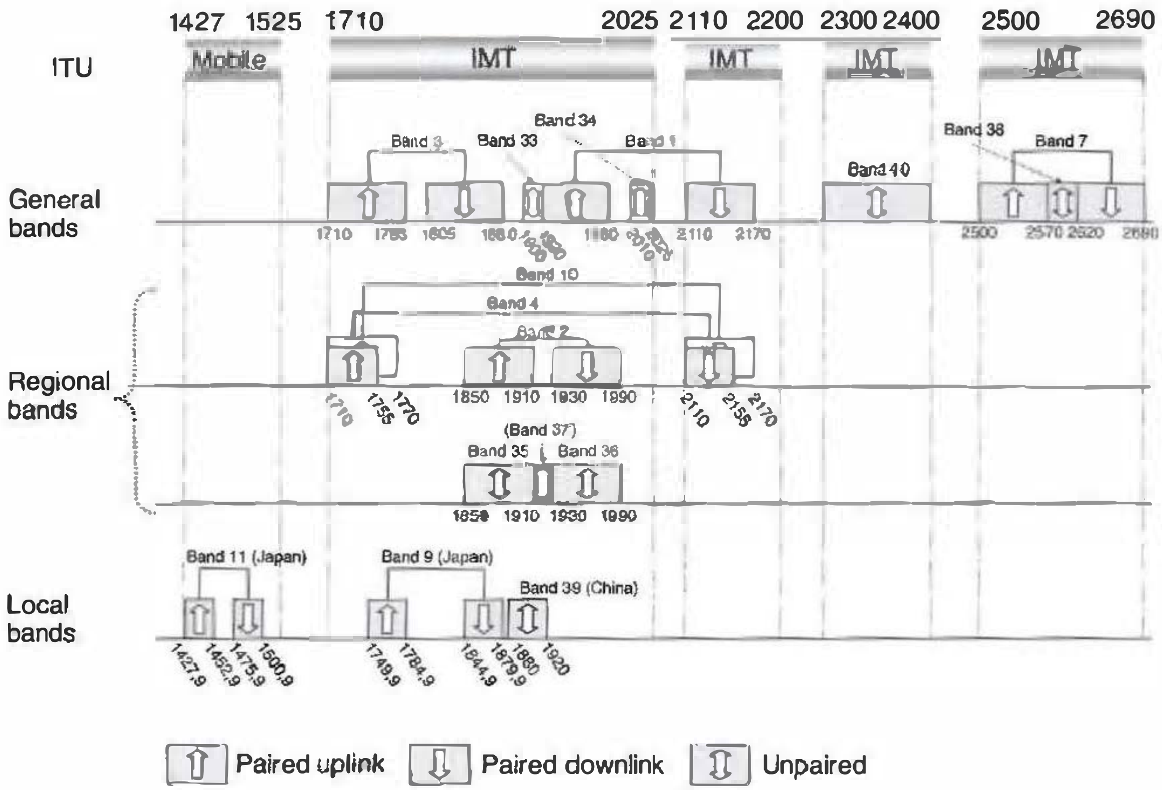


**Figure 20.1** *Operating bands specified in 3GPP above 1 GHz and the corresponding ITU allocation.*

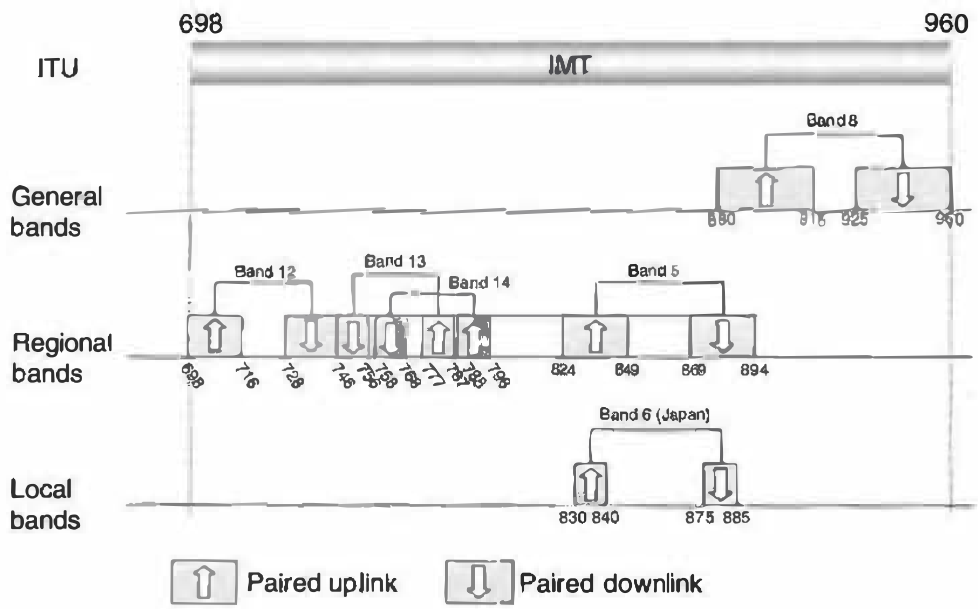


**Figure 20.2** *Operating bands specified in 3GPP below 1 GHz and the corresponding ITU allocation.*

WRC-2000 also identified the frequency ranges 806–960 MHz for IMT-2000. As shown in Figure 20.2, *Bands 5, 6, and 8* are defined for FDD operation in this range. Band 8 uses the same band plan as GSM900. Bands 5 and 6 overlap, but are intended for different regions. Band 5 is based on the US cellular band, while Band 6 is restricted to Japan in the specifications. 2G systems in Japan had a very specific band plan and Band 6 is a way of aligning the Japanese spectrum plan in the 810–960 MHz range to that of other parts of the world.

*Bands 12, 13, and 14* is the first set of bands defined for what is called the *digital dividend*, that is for spectrum previously used for broadcasting. This spectrum is partly migrated to be used by other wireless technologies, since TV broadcasting is migrated from analog to more spectrum efficient digital technologies.

### 20.1.2 New frequency bands

Additional frequency bands are continuously specified for UTRA and LTE. WRC-07 identified additional frequency bands for IMT, which encompasses both IMT-2000 and IMT-Advanced. Several bands were defined by WRC-07 that will be available partly or fully for deployment on a global basis:

- *450–470 MHz* was identified for IMT globally. It is already today allocated to mobile service globally, but it is a not a very large band.
- *698–806 MHz* was allocated to mobile service and identified to IMT to some extent in all regions. Together with the band 806–960 MHz identified at WRC-2000, it forms a wide frequency range from 698 to 960MHz that is partly identified to IMT in all regions, with some variations.
- *2300–2400 MHz* was identified for IMT on a worldwide basis in all three regions.
- *3400–3600 MHz* was allocated to the mobile service on a primary basis in Europe and Asia and partly in some countries in the Americas. There is also satellite use in the bands today.

For the frequency ranges below 1 GHz identified at WRC-07, 3GPP has already specified several operating bands as shown in Figure 20.2. In addition to Bands 5, 6, and 8 described above, *Bands 12, 13, and 14* are defined for operation mainly for US allocations. Note that Band 14 has a special configuration, since the upper part of this band is intended for a public safety network that is to be operated in a private/public partnership by a commercial operator. Work is also ongoing in Europe within Electronic Communications Committee (ECC) Task Group 4 [138] on the technical feasibility for harmonized European spectrum allocations for fixed and mobile applications in the digital dividend (below 862MHz).

*Band 40* is an unpaired band specified for the new frequency range 2300–2400MHz identified for IMT.

Work in 3GPP is initiated also for the frequency band 3.4–3.8 GHz [125]. In Europe, a majority of countries already license the band 3.4–3.6GHz for both Fixed Wireless Access and mobile use. Licensing of 3.6–3.8GHz for Wireless Access is more limited. There is a European spectrum decision for 3.4–3.8 GHz with 'flexible usage modes' for deployment of fixed, nomadic, and mobile networks. Frequency arrangements considered in the decision include FDD use with 100MHz block offset between paired blocks and/or TDD use. In Japan, not only 3.4–3.6GHz but also 3.6–4.2GHz will be available to terrestrial mobile services such as IMT to use after 2010. The band 3.4–3.6 GHz has been licensed for wireless access also in Latin America.

## 20.2 Flexible spectrum use

Many of the frequency bands identified above for deployment of LTE are existing IMT-2000 bands and some also have other systems deployed in those bands, including WCDMA/HSPA and GSM. Bands are also in some regions defined in a 'technology neutral' manner, which means that coexistence between different technologies is a necessity.

The fundamental LTE requirement to operate in different frequency bands [85] does not, in itself, impose any specific requirements on the radio interface design. There are however implications for the RF requirements and how those are defined, in order to support the following:

- *Coexistence between operators in the same geographical area in the band*: These other operators may deploy LTE or other IMT-2000 technologies, such as UMTS/HSPA and GSM/EDGE. There may also be non-IMT-2000 technologies. Such coexistence requirements are to a large extent developed within 3GPP, but there may also be regional requirements defined by regulatory bodies in some frequency bands.
- *Co-location of BS equipment between operators*: There are in many cases limitations to where BS equipment can be deployed. Often sites must be shared between operators or an operator will deploy multiple technologies in one site. This puts additional requirement on both BS receivers and transmitters.
- *Coexistence with services in adjacent frequency bands and across country borders*: The use of the RF spectrum is regulated through complex international

agreements, involving many interests. There will therefore be requirements for coordination between operators in different countries and for coexistence with services in adjacent frequency bands. Most of these are defined in different regulatory bodies. Sometimes the regulators request that 3GPP includes such coexistence limits in the 3GPP specifications.

- *Release independent frequency band principles*: Frequency bands are defined regionally and new bands are added continuously. This means that every new release of 3GPP specifications will have new bands added. Through the 'release independence' principle, it is possible to design terminals based on an early release of 3GPP specifications that support a frequency band added in a later release.

## 20.3 Flexible channel bandwidth operation

The frequency allocations in Figures 20.1 and 20.2 are up to 2 × 75 MHz, but the spectrum available for a single operator may be from 2 × 20 MHz down to 2 × 5 MHz for FDD and down to 1 × 5 MHz for TDD. Furthermore, the migration to LTE in frequency bands currently used for other radio-access technologies must often take place gradually to ensure that sufficient amount of spectrum remains to support the existing users. Thus the amount of spectrum that can initially be migrated to LTE may be relatively small, but may then gradually increase, as shown in Figure 20.3. The variation of possible spectrum scenarios

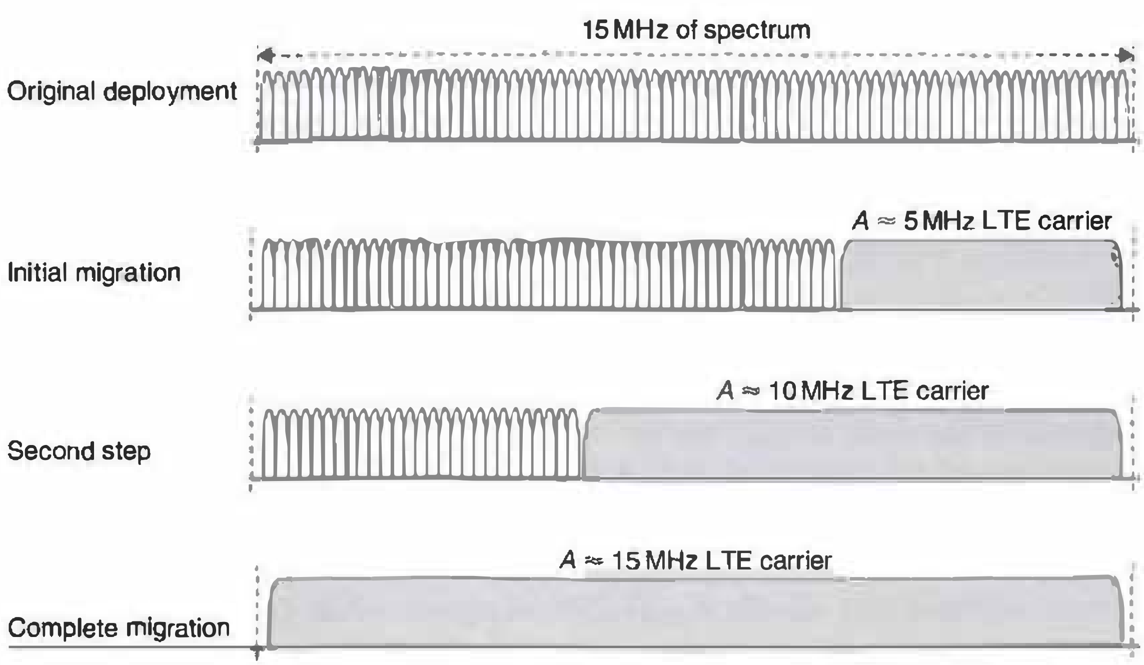


**Figure 20.3** *Example of how LTE can be migrated step-by-step into a spectrum allocation with an original GSM deployment.*

will imply a requirement for spectrum flexibility for LTE in terms of the transmission bandwidths supported.

The spectrum flexibility requirement points out the need for LTE to be scalable in the frequency domain. This flexibility requirement is in [86] stated as a list of LTE spectrum allocations from 1.25 to 20MHz. Note that the final channel bandwidths selected differ slightly from this initial assumption.

As shown in Chapters 16 and 17, the frequency-domain structure of LTE is based on resource blocks consisting of 12 subcarriers with a total bandwidth of $12 \times 15$kHz = 180kHz. The basic radio-access specification including the physical-layer and protocol specifications enable *transmission bandwidth configurations* from 6 up to 110 resource blocks on one LTE RF carrier. This allows for channel bandwidths ranging from 1.4MHz up to beyond 20MHz in steps of 180kHz and is fundamental to providing the required spectrum flexibility.

In order to limit implementation complexity, only a limited set of bandwidths are defined in the RF specifications. Based on the frequency bands available for LTE deployment today and in the future as described above and considering the known migration and deployment scenarios in those bands, a limited set of six channel bandwidths are specified. The RF requirements for the BS and terminal are defined only for those six channel bandwidths. The channel bandwidths range from 1.4 to 20MHz as shown in Table 20.3. The lower bandwidths 1.4 and 3MHz are chosen specifically to ease migration to LTE in spectrum where CDMA2000 is operated, and also to facilitate migration of GSM and TD-SCDMA to LTE. The specified bandwidths target relevant scenarios in different frequency bands. For this reason, the set of bandwidths available for a specific band is not necessarily the same as in other bands. At a later stage, if new frequency bands are made available that have other spectrum scenarios requiring additional channel bandwidths, the corresponding RF parameters and requirements can be added in the RF specifications, without actually having to update the physical-layer specifications. The process of adding new channel bandwidths is in this way similar to adding new frequency bands.

Figure 20.4 illustrates in principle the relationship between the channel bandwidth and the number of resource blocks for one RF carrier. Note that for all channel bandwidths except 1.4MHz, the resource blocks in the transmission bandwidth configuration fill up 90% of the channel bandwidth. The spectrum emissions shown in Figure 20.4 are for a pure OFDM signal, while the actual transmitted emissions will depend also on the transmitter RF chain and other components. The emissions outside the channel bandwidth are called *unwanted emissions* and the requirements for those are discussed further below.

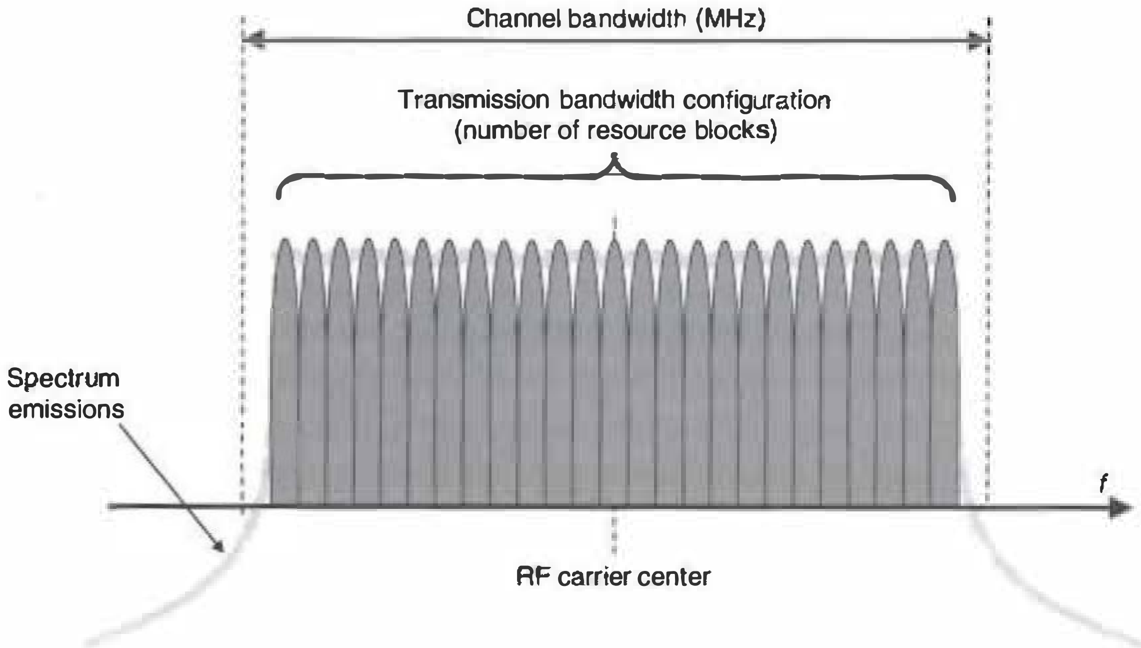


**Figure 20.4** *The channel bandwidth for one RF carrier and the corresponding transmission bandwidth configuration.*

**Table 20.3** *Channel bandwidths specified in LTE.*

| Channel bandwidth (MHz) | Number of resource blocks |
|---|---|
| 1.4 | 6 |
| 3 | 15 |
| 5 | 25 |
| 10 | 50 |
| 15 | 75 |
| 20 | 100 |

## 20.4 Requirements to support flexible bandwidth

### *20.4.1 RF requirements for LTE*

The RF requirements define the receiver and transmitter RF characteristics of a BS or UE. The BS is the physical node that transmits and receives RF signals on one or more antenna connectors to cover one cell. UE is the 3GPP term for the terminal.

The set of RF requirements defined for LTE is fundamentally the same as those defined for UTRA or any other radio system. Some requirements are also based on regulatory requirements and are more related to the frequency band of operation and/or the place where the system is deployed, than it is related to the type of system.

What is particular to LTE is the flexible bandwidth and the related multiple channel bandwidths of the system, which makes some requirements more difficult to define. It has special implications for the transmitter requirements on unwanted emissions, where the definition of the limits in international regulation depends on the channel bandwidth, which becomes difficult for a system where the BS may operate with multiple channel bandwidths and the terminal may vary its channel bandwidth of operation. The properties of the flexible OFDM-based Layer 1 also have implications for specifying the transmitter modulation quality and how to define the receiver selectivity and blocking requirements.

There are also some differences in how the requirements for the terminal and BS requirements respectively are defined. For this reason, they are treated separately in this chapter. The detailed background of the RF requirements for LTE is described in [130] and [132]. The RF requirements for the BS are specified in [127] and for the terminal (UE) in [126]. The RF requirements are divided into transmitter and receiver characteristics. There are in addition 'performance characteristics' which define the receiver baseband performance and are thus not strictly RF requirements, though the performance will also depends on the RF to some extent.

*Transmitter characteristics* are maximum output power, output power dynamics, transmitted signal quality (mainly frequency error and Error Vector Magnitude, (EVM), unwanted emissions, and transmitter intermodulation.

*Receiver characteristics* are reference sensitivity level, receiver dynamic range, Adjacent Channel Selectivity (ACS), receiver blocking (including spurious response for the terminal), receiver intermodulation, and receiver spurious emissions.

Each RF requirement has a corresponding test defined in the LTE test specifications for the BS [128] and the terminal [131]. These specifications define the test setup, test procedure, test signals, tolerances, etc. needed to show compliance with the RF and performance requirements.

The discussion below will focus on requirements where the flexible bandwidth properties of LTE have particular implications.

### 20.4.2 Regional requirements

There are a number of regional variations to the RF requirements and their application. The variations originate in different regional and local regulation of spectrum and its use. The most obvious regional variation is the different frequency bands and their use as discussed above. Many of the regional RF requirements are also tied to specific frequency bands.

When there is a regional requirement on for example spurious emissions, this requirement should be reflected in the 3GPP specifications. For the BS it is entered as an optional requirement and is marked as 'regional.' For the terminal, the same procedure is not possible, since a terminal may roam between different regions and will therefore have to fulfill all regional requirements that are tied to an operating band in the regions where the band is used. For LTE, this becomes more complex than for UTRA, since there is an additional variation in the transmitter (and receiver) bandwidth used, making some regional requirements difficult to meet as a mandatory requirement. The concept of *network signaling* of RF requirements is therefore introduced for LTE, where a terminal can be informed at call setup of whether some specific RF requirements apply when the terminal is connected to a network.

Examples of regional requirements are:

- *Spurious emissions*: Different 'categories' of emission levels are defined by ITU-R [134] and applied in different regions. These are called categories A and B.
- *Coexistence with other systems in the same geographical area*: Since the type of system to coexist with varies between regions, this is often a regional requirement. In each region, the requirement is however usually mandatory. For terminals, it will normally be mandatory for any roaming device.
- *Co-location with other BS*: The type of BS to be potentially co-located with also varies between regions. Co-location requirements are however usually not mandatory from a regulatory point of view.

The way regional regulation is set also varies considerably. In Europe, most requirements are developed in cooperation between the standards body ETSI and the ECC, who work under mandate from the European Commission. The regulation for the US operating bands is developed by the FCC. Also Japan has a local radio regulation that is reflected in the 3GPP specifications.

### 20.4.3 BS transmitter requirements

Unwanted emissions from the transmitter are divided into *OOB emission* and *spurious emissions* in ITU-R recommendations [134]. OOB emissions are defined as emission on a frequency close to the RF carrier, which results from the modulation process. Spurious emissions are emissions outside the RF carrier that may be reduced without affecting the corresponding transmission of information, but excluding OOB emissions. Examples of spurious emissions are harmonic emissions, intermodulation products, and frequency conversion products. The frequency range where OOB emissions are normally defined is called the *OOB domain* whereas spurious emission limits are normally defined in the *spurious domain*.

ITU-R also defines the limit between the OOB and spurious domains at a frequency separation from the carrier center of 2.5 times the *necessary bandwidth*, which equals 2.5 times the Channel bandwidth for E-UTRA. This division of the requirements is applied for UTRA which has a fixed channel bandwidth, but becomes more difficult for LTE, which is a flexible bandwidth system implying that the frequency range where requirements apply would then vary with the channel bandwidth.

As shown in Chapter 4 the spectrum of an OFDM signal decays rather slowly outside of the transmission bandwidth configuration. Since the transmitted signal occupies 90% of the channel bandwidth, it is not possible to directly meet the unwanted emission limits with a 'pure' OFDM signal. The techniques used for achieving the transmitter requirements are however not specified or mandated in LTE specifications. Time-domain windowing is one method commonly used in OFDM-based transmission systems to control spectrum emissions. Filtering is always used, both time-domain digital filtering of the baseband signal and analog filtering of the RF signal. Since the RF signal in the downlink needs to be amplified with a power amplifier that has nonlinear characteristics, linearization schemes are also an essential part of controlling spectrum emissions.

The discussion below is related to the unwanted emission requirements. Those limits are to be fulfilled over a specified dynamic range of the transmitter, both in terms of variations of the total transmitted power and of the power per resource element in the OFDM signal. There are in addition requirements for the modulation quality in terms of *frequency error* and *EVM*, both of which define the difference between an ideal OFDM signal at the assigned channel frequency and the actual transmitted RF signal.

#### *20.4.3.1 Operating band unwanted emissions*

For the reasons above, a unified concept of *operating band unwanted emissions* is used for the LTE BS instead of the usual spectrum mask defined for OOB emissions. This requirement applies over the whole BS transmitter operating band, plus an additional 10 MHz on each side as shown in Figure 20.5. All requirements outside of that range are set by the 'regular' spurious emission

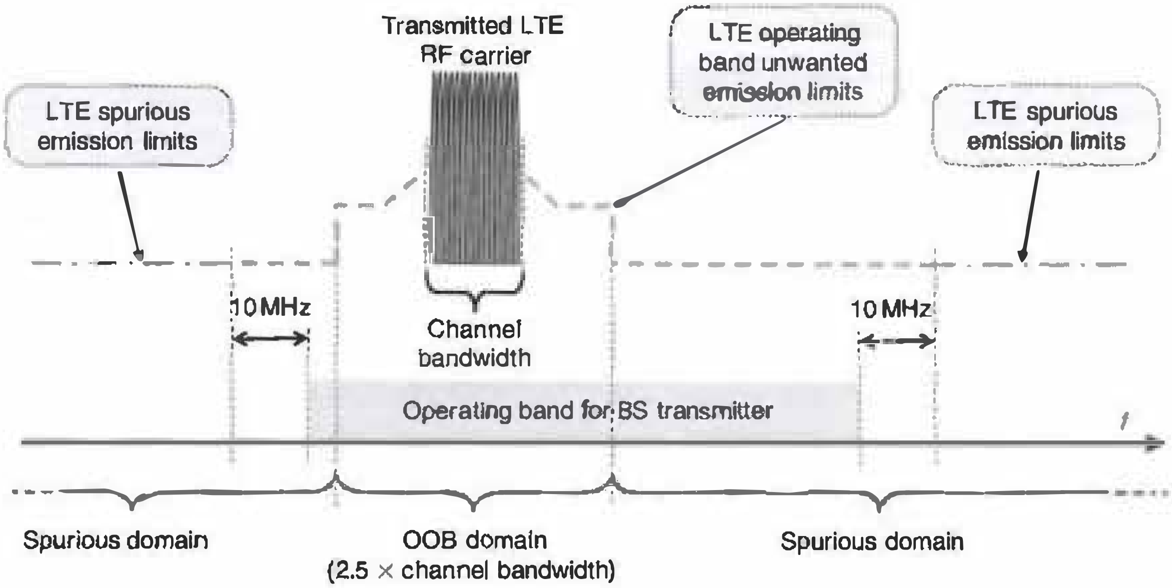


**Figure 20.5** *Defined frequency ranges for spurious emissions and operating band unwanted emissions.*

limits, based on the regulatory limits. Since the operating band unwanted emissions are defined over a frequency range that for the wider channel bandwidths are completely in the OOB domain, while it for the smaller channel bandwidths can be both in spurious and OOB domain, the limits are for all cases set in a way that complies with the ITU-R recommendations for spurious emissions [134]. The operating band unwanted emissions are defined with a 100 kHz measurement bandwidth.

There are special limits defined by FCC regulation [143] for the operating bands used in the US. Those are specified as separate limits in addition to the operating band unwanted emission limits.

#### *20.4.3.2 Adjacent Channel Leakage Ratio*

In addition, the OOB emissions are defined by an *Adjacent Channel Leakage Ratio* (ACLR) requirement. The ACLR concept is very useful for analysis of coexistence between two systems that operate on adjacent frequencies. ACLR defines the ratio of the power transmitted within the assigned channel bandwidth, to the power of the unwanted emissions transmitted on an adjacent channel. There is a corresponding receiver requirement called *Adjacent Channel Selectivity* (ACS), which defines a receiver's ability to suppress a signal on an adjacent channel.

The definitions of ACLR and ACS are illustrated in Figure 20.6 for a wanted and an interfering signal received in adjacent channels. The interfering signal's leakage of unwanted emissions at the wanted signal receiver is given by the ACLR and the ability of the receiver of the wanted signal to suppress the interfering

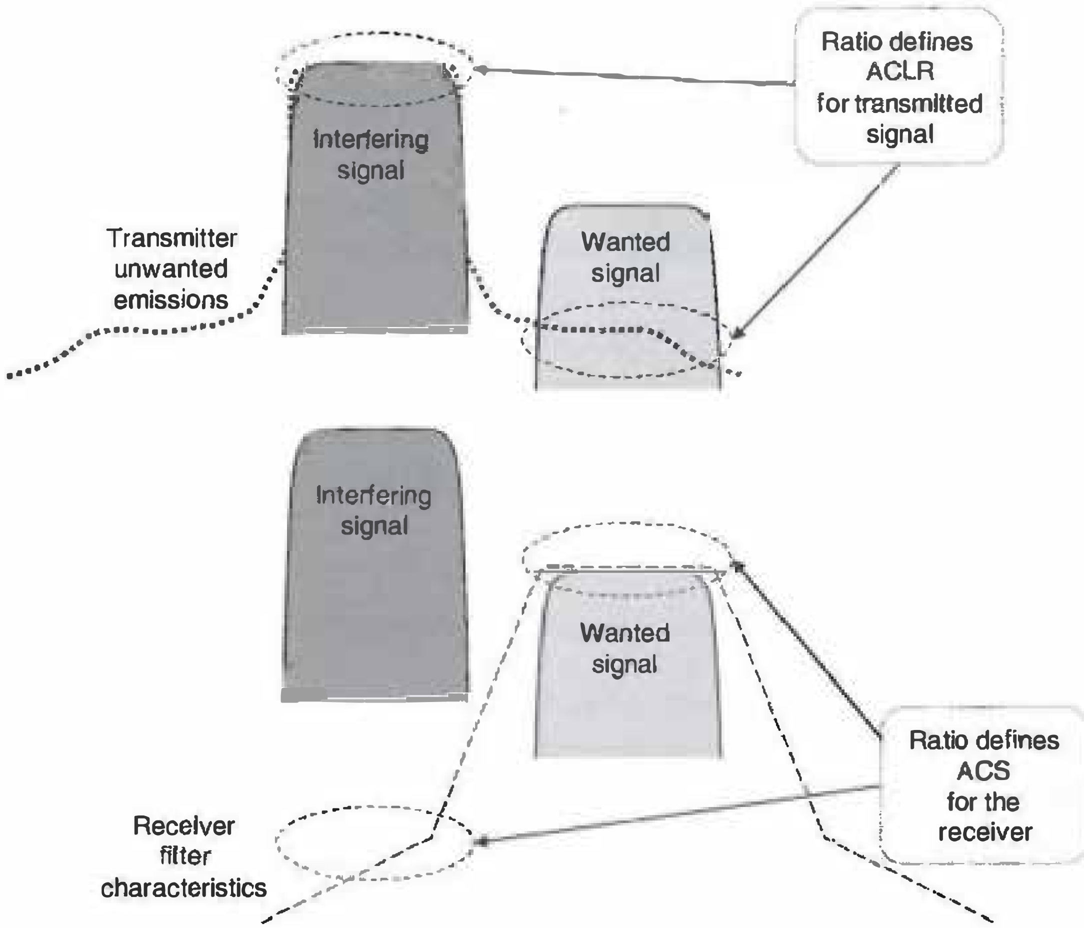


**Figure 20.6** *Definitions of ACLR and ACS, using example characteristics of an 'aggressor' interfering and a 'victim' wanted signal.*

signal in the adjacent channel is defined by the ACS. The two parameters when combined define the total leakage between two transmissions on adjacent channels. That ratio is called *Adjacent Channel Interference Ratio* (ACIR) and is defined as the ratio of the power transmitted on one channel to the total interference received by a receiver on the adjacent channel, due to both transmitter (ACLR) and receiver (ACS) imperfections.

This relation between the adjacent channel parameters is [135]

$$\mathrm{ACIR} = \frac{1}{\dfrac{1}{\mathrm{ACLR}} + \dfrac{1}{\mathrm{ACS}}} \tag{20.1}$$

Both ACLR and ACS can be defined with different channel bandwidths for the two adjacent channels, which is the case for some requirements set for LTE due to the bandwidth flexibility. The equation above will also apply for different channel bandwidths, but only if the same two channel bandwidths are used for defining all three parameters ACIR, ACLR, and ACS used in the equation.

The ACLR limits for LTE are derived based on extensive coexistence analysis [133] between LTE and potential LTE or other systems on adjacent carriers. Requirements on ACLR and operating band unwanted emissions both cover the OOB domain, but the operating band unwanted emission limits are set slightly more relaxed compared to the ACLR, since they are defined in a much narrower measurement bandwidth of 100 kHz. This allows for some variations in the unwanted emissions due to intermodulation products from varying power allocation between resource blocks within the channel.

For an LTE BS, there are ACLR requirements both for an adjacent channel with a UTRA receiver and with an LTE receiver of the same channel bandwidth.

#### 20.4.3.3 Spurious emissions

The limits for spurious emissions are taken from international recommendations [134], but are only defined in the region outside the frequency range of operating band unwanted emissions limits as described above, that is at frequencies that are separated from the BS transmitter operating band with at least 10 MHz. There are also the additional regional or optional limits for protection of other systems that LTE may coexist with or even be co-located with. Examples of other systems considered in those additional spurious emissions requirements are GSM, UTRA FDD/TDD, CDMA, and PHS.

### 20.4.4 BS receiver requirements

The set of receiver requirements for LTE is quite similar to what is defined for UTRA, but many of them need to be defined differently, due to the flexible bandwidth properties. The receiver characteristics are fundamentally specified in three parts:

- Requirements for receiving the wanted signal in itself, including reference sensitivity and dynamic range.
- Requirements for the receiver's susceptibility to different types of interfering signals.
- Requirements on unwanted emissions from the receiver.

#### 20.4.4.1 Reference sensitivity and receiver dynamic range

The primary purpose of the *reference sensitivity requirement* is to verify the receiver *Noise Figure*, which is a measure of how much the receiver's RF signal chain degrades the SNR of the received signal. For this reason, a low SNR transmission scheme using QPSK is chosen as reference channel for the reference sensitivity test. The reference sensitivity is defined at a receiver input level where the throughput is 95% of the maximum throughput for the reference channel.

A terminal in LTE may be assigned only a small part of the uplink channel bandwidth, implying that the sensitivity should be defined for smaller bandwidths, ideally per resource block. For complexity reasons, a maximum granularity of 25 resource block has been chosen, which means that for channel bandwidths larger than 5 MHz, sensitivity is verified over multiple adjacent 5 MHz blocks, while it is only defined over the full channel for smaller channel bandwidths.

The intention of the *dynamic range requirement* is to ensure that the BS can receive with high throughput also in the presence of increased interference and corresponding higher wanted signal levels, thereby testing the effects of different receiver impairments. In order to stress the receiver a higher SNR transmission scheme using 16QAM is applied for the test. In order to further stress the receiver to higher signal levels, an interfering AWGN signal at a level 20 dB above the assumed noise floor is added to the received signal.

#### *20.4.4.2 Receiver susceptibility to interfering signals*

There is a set of requirements for defining the BS ability to receive a wanted signal in the presence of an interfering signal. The reason for the multiple requirements is that depending on the frequency offset of the interferer from the wanted signal, the interference scenario may look very different and different types of receiver impairments will impact the performance. The intention of the different combinations of interfering signals is to model as far as possible the range of possible scenarios with interfering signals of different bandwidths that may be encountered inside and outside the BS receiver operating band.

The following requirements are defined, starting from interferers with large frequency separation and going close-in (see also Figure 20.7). In all cases where the interfering signal is an LTE signal, it has the same bandwidth as the wanted signal, but at the most 5 MHz.

- *Blocking*: Corresponds to the scenario with strong interfering signals received outside the operating band (out-of-band) or inside the operating band, but not adjacent to the wanted signal (in-band, including the first 20 MHz outside the band). The scenarios are modeled with a *Continuous Wave* (CW) signal for the out-of-band case and an LTE signal for the in-band case. There are additional

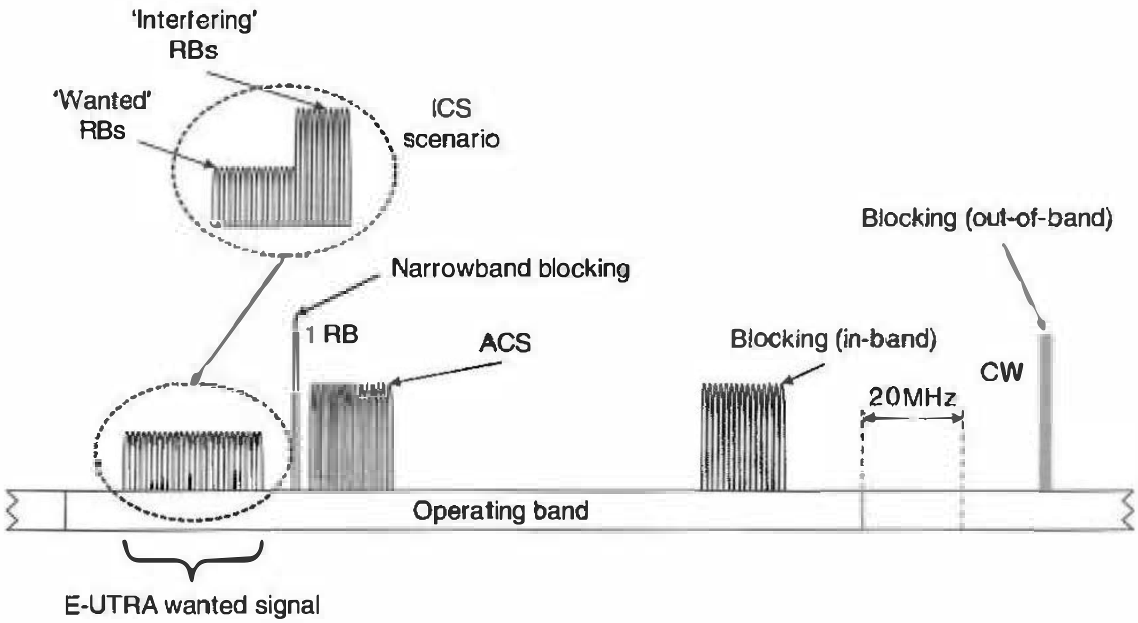


**Figure 20.7** *Requirements for receiver susceptibility to interfering signals in terms of blocking, ACS, narrowband blocking, and in-channel selectivity (ICS).*

(optional) blocking requirements for the scenario when the BS is co-sited with another BS in a different operating band.

- *Adjacent Channel Selectivity*: The ACS scenario is a strong signal in the channel adjacent to the wanted signal and is closely related to the corresponding ACLR requirement for the terminal (see also the discussion in Section 20.4.3.2). The adjacent interferer is an LTE signal.
- *Narrowband blocking*: The scenario is an adjacent strong narrowband interferer, which in the requirement is modeled as a single resource block LTE signal.
- *In-channel selectivity* (ICS): The scenario is multiple received signals of different received power levels inside the channel bandwidth, where the performance of the weaker 'wanted' signal is verified in presence of the stronger 'interfering' signal.
- *Receiver intermodulation*: The scenario is *two* interfering signals near adjacent to the wanted signal, where the interferers are one CW and one LTE signal (not shown in Figure 20.7). The interferers are placed in frequency in such a way that the main intermodulation product falls inside the wanted signal's channel bandwidth. There is also a *narrowband intermodulation* requirement where the CW signal is very close to the wanted signal and the LTE interferer is a single RB signal.

For all requirements except in-channel selectivity, the wanted signal uses the same reference channel as in the reference sensitivity requirement. With the interference added, the same 95% relative throughput is met for the reference channel, but at a 'de-sensitized' higher wanted signal level, for most requirements at 6 dB above the reference sensitivity level.

### 20.4.5 Terminal transmitter requirements

The type of transmitter requirements defined for the terminal is very similar to what is defined for the BS as explained above and the definitions of the requirements are often similar. The output power levels are however considerably lower for a terminal, while the restrictions on the terminal implementation are much higher. There is a tight pressure on cost and complexity for all telecommunications equipment, but this is much more pronounced for terminals, due to the scale of the total market which is approximately one *billion* devices per year.

In the following, the focus will be on requirements where the terminal requirements differ in definition from the corresponding ones for the BS and where there are particular implications from the flexible bandwidth in LTE.

#### 20.4.5.1 Channel bandwidths supported

For the terminal, the channel bandwidths supported are not only a function of the E-UTRA band, but also have a relation to the transmitter and receiver RF requirements. For some of the higher channel bandwidths supported in paired frequency bands where the duplex band gap between uplink and downlink is small, there are certain relaxations of the terminal performance. These relaxations and limitations may consist of a reduced lower power level or an allowed receiver sensitivity reduction (for the highest transmission bandwidths) due to duplex filter constrains.

#### 20.4.5.2 Terminal power level

The terminal output power level is defined in three steps:

- *Terminal power class* defines a *nominal* maximum output power for QPSK modulation. It may be different in different operating bands, but the main terminal power class is today set at 23 dBm.
- *Maximum Power Reduction (MPR)* defines an allowed reduction of maximum power level for certain combinations of modulation used (QPSK or 16QAM) and number of resource blocks that are assigned.
- *Additional Maximum Power Reduction (A-MPR)* may be applied in some regions and is usually connected to specific transmitter requirements such as regional emission limits. For each such set of requirement, there is an associated network signaling value that identifies the allowed A-MPR and the associated conditions.

#### 20.4.5.3 Unwanted emission limits

The unwanted emissions are defined in a slightly different way for the LTE terminal than for the BS. The limits are divided into three parts:

- *In-band emissions* are emissions within the occupied bandwidth (channel bandwidth). The requirement limits how much a terminal can transmit into other resource blocks within the channel bandwidth. Unlike the OOB emissions, the in-band emissions are measured after cyclic prefix removal and FFT, since this is how a terminal transmitter affects a real eNodeB receiver. In-band emissions are in the specification defined as a part of the transmit signal quality requirements.
- *OOB emissions* are defined in terms of a *Spectrum Emissions Mask* (SEM) and an ACLR requirement.
  - The *SEM* is defined as a general mask and a set of additional masks that can be applied to reflect different regional requirements. Each additional mask has an associated network signaling value.
  - *ACLR* limits are set both with assumed UTRA and LTE receivers on the adjacent channel. As for the BS, the limit is also set stricter than the corresponding SEM, thereby accounting for variations in the spectrum emissions resulting from variations in resource block allocations. The ACLR limits are set based on extensive coexistence analysis [133].
- *Spurious emission* limits are defined for all frequency ranges outside the frequency range covered by the SEM. The limits are in general based on international regulations [134], but there are also additional requirements for coexistence with other bands when the mobile is roaming. The additional spurious emission limits can have an associated network signaling value.

The limit between the frequency ranges for OOB limits and spurious limits do not follow the same principle as for the BS. For 5 MHz channel bandwidth, it is set at 250% of the necessary bandwidth as recommended by ITU-R, but for higher channel bandwidths it is set closer than 250%.

### 20.4.6 Terminal receiver requirements

Also the set of terminal transmitter requirements is similar to what is defined for the BS. The requirements are defined for the full channel bandwidth signals and with all resource blocks allocated for the wanted signal. All receiver requirements assume that the receiver is equipped with two Rx ports using antenna diversity, which does not preclude that a single port terminal can meet a specific requirement.

#### 20.4.6.1 Reference sensitivity

The reference sensitivity is defined using a low SNR reference channel with QPSK modulation in order to verify the terminal noise figure. For the higher channel bandwidths (≥5 MHz) in some operating bands, the nominal reference sensitivity needs to be met with a minimum number of allocated resource blocks. For larger allocation, a certain relaxation is allowed.

#### 20.4.6.2 Receiver susceptibility to interfering signals

The set of requirements that defines the terminal's ability to receive a wanted signal in the presence of an interfering signal is very similar to the corresponding BS requirements, as illustrated in Figure 20.7. The requirement levels are different for the terminal, since the interference scenarios for the BS and terminal are very different. There is also no terminal requirement corresponding to the BS in-channel selectivity requirement.

The following requirements are defined:

- *Blocking*: There are both out-of-band blocking and in-band blocking requirements, where in-band includes the first 15 MHz outside the operating band and excludes frequencies adjacent to the carriers. Limits are defined with CW interferers for out-of-band and LTE signals for in-band blocking. A fixed number of *exceptions* are allowed from the terminal out-of-band blocking requirement, for each assigned frequency channel and at the respective *spurious response frequencies*. At those frequencies, the terminal must comply with the more relaxed spurious response requirement.
- *Adjacent Channel Selectivity*: The ACS is specified for two cases with a lower and a higher signal level. The adjacent signal is an LTE signal.
- *Narrowband blocking*: The narrowband blocking requirement is defined with an adjacent CW interfering signal.
- *Receiver intermodulation*: The requirement is defined with one CW and one LTE signal, placed in such a way that the main intermodulation product falls inside the wanted signal. There is also in addition a narrowband intermodulation requirement.

[89] '3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Network Architecture', 3GPP, 3GPP TS 23.002.

[90] '3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS); Service Description; Stage 2', 3GPP, 3GPP TS 23.060.

[91] '3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3GPP System Architecture Evolution: GPRS Enhancements for LTE Access; Release 8', 3GPP, 3GPP TS 23.401.

[92] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; User Equipment (UE) Radio Transmission and Reception (FDD)', 3GPP, 3GPP TS 25.101.

[93] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; User Equipment (UE) Radio Transmission and Reception (TDD)', 3GPP., 3GPP TS 25.102.

[94] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Physical Channels and Mapping of Transport Channels onto Physical Channels (FDD)', 3GPP., 3GP TS 25.211.

[95] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Multiplexing and Channel Coding (FDD)', 3GPP, 3GP TS 25.212.

[96] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Spreading and Modulation (FDD)', 3GPP., 3GP TS 25.213.

[97] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Physical Layer Procedures (FDD)', 3GPP, 3GP TS 25.214.

[98] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Radio Interface Protocol Architecture', 3GPP, 3GPP TS 25.301.

[99] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; UE Radio Access Capabilities', 3GPP, 3GPP TS 25.306.

[100] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; High Speed Downlink Packet Access (HSDPA); Overall Description; Stage 2', 3GPP, 3GPP TS 25.308.

[101] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; FDD Enhanced Uplink; Overall Description; Stage 2', 3GPP, 3GPP TS 25.309.

[102] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Introduction of the Multimedia Broadcast Multicast Service (MBMS) in the Radio Access Network (RAN); Stage 2 (Release 6)', 3GPP, 3GPP TS 25.346.

[103] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; UTRAN Overall Description', 3GPP, 3GPP TS 25.401.

[104] '3rd Generation Partnership Project; Technical Specification Group Radio Access Network; UTRAN Iu Interface RANAP Signalling (Release 7)', 3GPP., 3GPP TS 25.413.