TECHNISCHE UNIVERSITÄT WIEN

VIENNA UNIVERSITY OF TECHNOLOGY

DISSERTATION

Link Error Analysis and Modeling for Cross-Layer Design in UMTS Mobile Communication Networks

Conducted for the purpose of receiving the academic title
'Doktor der technischen Wissenschaften'

Submitted at Vienna University of Technology
Faculty of Electrical Engineering and Information Technology

by

Dipl.-Ing. Wolfgang Karner
9725076
Hangstraße 2, A-4600 Thalheim

Vienna, December 2007 _____

Supervisor:

Univ.Prof. Dr.-Ing. Markus Rupp
Institut für Nachrichtentechnik und Hochfrequenztechnik
Technische Universität Wien, Österreich

Examiner:

Prof. Dr. Roberto Verdone
Dipartimento di Elettronica, Informatica e Sistemistica
Università di Bologna, Italia

Abstract

Particularly in wireless mobile communications, link errors severely affect the quality of the services due to the high error probability and the specific error characteristics (burst errors) in the radio access part of the network.

In this thesis it is shown that a thorough analysis and the appropriate modeling of the radio-link error behaviour is essential not only to evaluate and optimize the higher layer protocols and services. It is also the basis for finding network-aware cross-layer processing algorithms which are capable of exploiting the specific properties of the link error statistics (e.g. the predictability).

This thesis presents the analysis of the radio link errors based on measurements in live UMTS (Universal Mobile Telecommunication System) radio access networks. It is shown that due to the link error characteristics basically two scenarios have to be distinguished: static and dynamic (regardless of which kind of mobility). Furthermore, the analysis shows that apart from fading effects, also the quality-based power control mechanism induces correlation to the link error characteristics, resulting in error bursts, error clusters and the predictability of the link errors.

It is presented that 'well-known' classical error models are not capable of describing the measured specific error characteristics of the UMTS DCH (Dedicated Channel) properly. Especially the error predictability cannot be represented at all due to the geometric distributed output of the models. Thus, a new modeling approach based on semi-Markov models as well as (Markov modulated) Weibull renewal processes is presented. Simple estimators for future link error probabilities are derived from these models.

Based on measurements, the dynamic resource allocation of the UMTS is analysed and models for the dynamic bearer type switching are proposed which together with the link error models can provide a proper representation of the dynamic radio link behaviour.

In this thesis, novel cross-layer algorithms are presented which make use of the discovered predictability of the link errors and consider the service characteristics. It is shown that high gain in the quality of H.264/AVC video streaming is reached by introducing such network aware cross-layer algorithms in the UTRAN (UMTS Terrestrial Radio Access Network).

Kurzfassung

Aufgrund der relativ hohen Fehlerwahrscheinlichkeit und der spezifischen Eigenschaften der Übertragungsfehler (z.B. Fehlerhäufung) wird die Qualität der übertragenen Dienste speziell in der mobilen Funkkommunikation stark beeinträchtigt.

In dieser Arbeit wird aufgezeigt, dass eine genaue Analyse und ein entsprechendes Modellieren des Übertragungsfehlerverhaltens essentiell wichtig ist, nicht nur um die Protokolle der höheren Schichten zu evaluieren und zu optimieren. Die Kenntnis der Fehlereigenschaften ist auch eine notwendige Grundlage für die Entwicklung von neuen schichtenübergreifenden Methoden, die auf die momentanen Zustände im Übertragungskanal Rücksicht nehmen. Dadurch sind diese neuen Prozesse in der Lage, die speziellen Eigenschaften der Fehlerstatistik (wie z.B. die Vorhersagbarkeit der Übertragungsfehler) explizit auszunützen.

Diese Doktorarbeit präsentiert die gründliche Analyse der Übertragungsfehler in der Funkschnittstelle basierend auf Messungen in operativen UMTS (Universal Mobile Telecommunication System) Netzen. Es wird gezeigt, dass aufgrund der resultierenden Fehlereigenschaften in der Funkschnittstelle grundsätzlich nur zwei verschiedene Szenarien unterschieden werden müssen: statisches Szenario und Szenario mit Bewegung (unabhängig von der Art der Bewegung). Die Analyse zeigt außerdem, dass – neben den Effekten des zeitvariablen Schwundes im Übertragungskanal – auch die qualitätsbezogene Regelung der Sendeleistung Korrelationseffekte zwischen den Übertragungsfehlern verursacht, die in Fehlerhäufung und der Vorhersagbarkeit der Übertragungsfehler resultieren.

Es wird gezeigt, dass bekannte klassische Fehlermodelle nicht in der Lage sind, die gemessenen Fehlereigenschaften des UMTS DCH (Dedicated Channel) richtig zu beschreiben. Speziell die Vorhersagbarkeit der Übertragungsfehler kann durch die erzeugten geometrischen Verteilungen der Modelle nicht reproduziert werden. Darum werden in dieser Arbeit neue Modellierungsansätze vorgestellt, die auf Semi-Markov Modellen und (Markov-modulierten) Weibull-Erneuerungsprozessen aufbauen. Von diesen Modellen können auch einfache Schätzer für die bedingte Fehlerwahrscheinlichkeit abgeleitet werden.

Anhand von Messergebnissen wird die dynamische Ressourcenzuteilung in UMTS analysiert und daraus resultierende Modelle für das dynamische Umschalten zwischen verschiedenen Träger-

typen werden vorgestellt. Diese Modelle bilden gemeinsam mit den Übertragungsfehlermodellen eine geeignete Beschreibung des dynamischen Verhaltens der UMTS Funkschnittstelle.

In dieser Arbeit werden neue schichtenübergreifende Algorithmen präsentiert, die die in den Analysen vorgefundene Vorhersagbarkeit der Übertragungsfehler ausnutzen und gleichzeitig die speziellen Eigenschaften der übertragenen Dienste berücksichtigen. Es wird gezeigt, dass durch die Anwendung dieser neuen schichtenübergreifenden Methoden im UTRAN (UMTS Terrestrial Radio Access Network) große Verbesserungen in der Qualität von übertragenen H.264/AVC Videodaten erreicht werden können.

Acknowledgements

There is a large number of persons I would like to thank for their contribution to the success of my thesis.
First of all, special thanks are due to my advisor Professor Markus Rupp for his unlimited support on my thesis and for providing a highly instructive and unconstrained ambiance for work and research.
I also wish to express my gratitude to Professor Roberto Verdone for agreeing to act as my second supervisor.
Furthermore, I am grateful to Professor Ernst Bonek, who invited me to join his research group and gave me the motivation to start working towards this thesis.
Many thanks to my colleagues at the mobile communications group for support, fruitful discussions and research cooperations.
I also would like to thank mobilkom austria AG for technical and financial support of this work[1] as well as for their valuable feedback.
Especially, I would like to thank my family for their constant help and support.
Last, but not least, I am grateful to Sibylle and all my friends, who helped me not to forget about life outside research, science and technology.

[1]The views expressed in this thesis are those of the author and do not necessarily reflect the views within mobilkom austria AG.

Contents

Chapter 1

Introduction

Contents

1.1 Motivation

Third generation mobile communication systems like the UMTS (Universal Mobile Telecommunication System) [1, 2] already offer relatively high data rates and great flexibility, making the systems very attractive for many new services. Despite the fact that these new services can be delivered to the end users, the quality of most of such services (especially multimedia services like video calls, conferencing, streaming, broadcast or multicast) is heavily suffering from the error characteristics of the underlying networks.

The overall goal of this thesis is to show that detailed knowledge about the specific link error characteristics, gained by measurement based link error analysis, forms the essential basis for developing new and optimizing existing services and protocols. Substantial gain in end-to-end quality is reached when considering error characteristics of the underlying link.

Especially in wireless communications, applications are facing high link error probabilities as the frequency spectrum is a scarce resource and there is power and interference limitation in the system. Furthermore, fading effects due to multipath propagation and mobility, together with certain properties of the air interface, lead to correlation between single error events. The resulting error bursts and the high error probability are causing significant degradations of the perceived QoS (Quality of Service) [3–6].

As a matter of fact, new demanding services (e.g. in terms of real-time requirements or error-proneness due to the high compression ratio in case of mobile multimedia) are the most challenging, but also services which have originally been designed for being used at very low link error probability in wired communication networks are considerably affected.

Thus, in order to be able to optimize existing services or to develop new services in an optimum way for the transmission over wireless mobile communication networks, a thorough understanding of the specific error characteristics of the particular link is necessary. In this work the analysis of the link error characteristics is presented for the UMTS DCH (Dedicated Channel). Since the main error source is the radio link, the error analysis primarily focuses on the wireless access part of the UMTS network.

The knowledge about the link error characteristics can not only be used for the performance evaluation or optimization of services or protocols, but the very specific error characteristics (e.g. the error predictability) can explicitly be exploited in cross-layer methods. In this thesis new cross-layer algorithms are proposed and it is shown that high gain in the quality of H.264/AVC (Advanced Video Coding) [7] video streaming is reached by introducing network aware cross-layer algorithms in the UTRAN (UMTS Terrestrial Radio Access Network).

Despite the fact that the analysis in this document is performed for the UMTS DCH, similar error behaviour (predictability) may also be expected for novel and future mobile communication systems if they comprise a quality-based control mechanism.

For the evaluation of new services and protocols, as well as for the investigation of the performance of novel (cross-layer) methods, often the measured traces do not provide sufficient statistics and thus stochastic models of the observed error characteristics of the link are required. Furthermore, for efficient operation of the proposed cross-layer algorithms themselves, the link error characteristics have to be available in the form of stochastic models (providing an analytic expression). In this work a new modeling approach based on the specific properties of the UMTS DCH is presented, which in contrast to well-known existing models is capable of describing the required link error characteristics (especially the error predictability) properly.

1.2 Outline and Contributions

In the following the organization of this thesis is presented and the author's contributions to this work are explained. Relevant publications (co)authored by the author of this thesis are highlighted in this section by additionally citing the author's name. A comprehensive view on the complete problem and a detailed representation of the topic of this thesis is given in [8, W.Karner et al.].

Chapter 1: Introduction

The remaining sections of Chapter 1 give an overview of the UMTS network architecture with a special focus on the UTRAN protocol architecture and on the physical layer processing in the WCDMA (Wideband Code Division Multiple Access) radio interface.

Chapter 2: Measurement Based Analysis of Link Layer Error Characteristics

After presenting the measurement setup, the specific error characteristics of the UTRAN are analysed based on measurements in live UMTS networks of three different operators in Vienna, Austria [9, 10, W.Karner et al.]. The investigation of the link errors in various measurement scenarios with several different mobility levels is presented. The results of this investigation show that basically only the static case and the scenario with mobility (regardless of which kind of mobility) have to be distinguished for further work [10, 11, W.Karner et al.].

In order to use radio resources in cellular wireless communication systems efficiently, the minimum transmit power for each link — considering the quality target of the link — has to be found by a quality-based power control algorithm. In the UMTS DCH this task is performed by the OLPC (Outer Loop Power Control) mechanism. Whereas in scenarios with mobility the link error characteristics are dominated by fading effects, in the static case the influence of the power control becomes significant and results in a cyclic error behaviour [12, W.Karner et al.] as shown in this chapter of the thesis.

Chapter 3: Modeling of Link Layer Error Characteristics

There exists already a lot of literature addressing various approaches for modeling erroneous links in telecommunications [13]–[41]. Most of these models are kept as general as possible and were designed to minimize the computational effort (calculation time for generating an error trace). Just in contrast to generality (entailing a high number of parameters and complex estimation methods), new models are required to offer a good usability, by only a small number of parameters which can easily be determined from measured data. Furthermore, the goal of the models is not minimum computational effort like e.g., offered by models based on Markov chains, but to represent specific details of the measured error characteristics (e.g. the error burst- and gaplengths) at a high accuracy. However, in this work it is shown that despite their generality, the well-known models are not capable of describing the observed specific error characteristics of the UMTS wireless link properly [12, W.Karner et al.]. Thus, a novel modeling approach is presented [9, 10, W.Karner et al.] which keeps the number of required parameters small and offers low estimation complexity by utilizing special properties of the UMTS DCH. As the basic idea behind this modeling approach is to utilize the very specific characteristics of the underlying channels, similar modeling strategies can be followed for finding models of other new or future wireless mobile communication systems.

Naturally, the error characteristics of the channel and thus also the model properties do have a great impact on the performance of methods and algorithms (e.g. for cross-layer optimization of wireless systems) which are considering or even exploiting the special error characteristics (e.g. error predictability) of the underlying channels. In [5] it was shown that there is also a strong impact of the second-order error statistics of the channel onto the performance of the higher layer protocols. In this work the impact of link error modeling directly on the quality of streamed H.264/AVC (Advanced Video Coding) [7] video over a plain UMTS protocol stack is investigated [4, W.Karner et al.] by comparing the link- and network-layer characteristics as well as the resulting video quality simulated with various link error models and measured traces. The results lead to the conclusion that the higher order statistics of the link errors do have a significant impact on the quality of the streamed video due to the difference in resulting network layer error probability. In contradiction to that there is no impact of the error correlation properties in the network layer on the average video quality as long as the resulting network-layer error probability remains the same. This is due to the error propagation when using predictive coding in the streamed video sequence.

Chapter 4: Dynamic Bearer Type Switching

Its high flexibility is one of the main advantages offered by UMTS. 3GPP (3rd Generation Partnership Project) [1] specifications allow to change the physical resources allocated to the

users dynamically to optimize the radio resource utilization. This means the data rate in UMTS DCH can be changed every 10 ms, either by changing the transport channel type or by varying its dynamic or semi-static parameters. Such switching of the bearer properties optimizes the use of the radio resources and at the same time facilitates providing the required QoS (Quality of Service) for the user. It can be triggered by admission control, congestion control, soft handover, required throughput or the radio channel quality (power threshold). The analysis in this work [42, W.Karner et al.] focuses on the dynamic bearer type switching due to fading and coverage reasons, where the decision of up or down switching is based on the link quality only.

Based on measurements in live UMTS networks within typical mobility scenarios for using UMTS services, a modeling approach for dynamic bearer type switching is presented [42, W.Karner et al.]. This model can then be combined with the error model of Chapter 3 to be used as a complete representation of the lower UTRAN layers in a system level simulation tool.

Chapter 5: Analysis of Link Error Predictability in the UTRAN

The detailed analysis of the measured link error traces reveals that the power control mechanism of the UMTS DCH (Dedicated Channel) induces correlation properties and thus memory in the link error characteristics, which in turn offers the prediction of the link errors. Especially, systems with a quality-based power control mechanism (e.g. the OLPC in UMTS) provide the possibility to predict future link errors due to a cyclic error behaviour [12, W.Karner et al.]. In order to utilize radio resources efficiently and at the same time keep the required QoS (Quality of Service) for the end user, a quality-based (power) control mechanism is required also in new or future mobile communication systems. Thus, despite the fact that the analysis in this document is performed for the UMTS DCH, similar error behaviour (predictability) may also be expected for novel and future mobile communication systems if they comprise a quality-based control mechanism.

In order to predict the link behaviour, the transmitter has to be informed of past link errors in the forward link via a feedback of the error status of the received data. In case of the UMTS this is accomplished by the RLC (Radio Link Control) AM (Acknowledged Mode). In this chapter of the thesis it is shown that in spite of the feedback delay in the RLC AM, intervals of low transmission error probability can be predicted [43, 44, W.Karner et al.] and furthermore, the time variable instantaneous failure rate [12, W.Karner et al.] can be estimated, both at a high accuracy. It is presented, that from the models of the UMTS DCH link error characteristics, low complex estimators for future error probability can be derived.

Chapter 6: Cross-Layer Optimization for Video Streaming over UMTS

An improvement in the quality of services like video streaming can be achieved either by optimiz-

ing the service to cope with the special link error characteristics, by adapting the characteristics of the wireless link to meet the requirements of the service or by installing a cross-layer design to connect the characteristics of the link with the service properties in an optimum way. All the approaches highly depend on consolidated knowledge about the specific error characteristics of the considered link. In this chapter it is presented how the detailed link error analysis and modeling (Chapter 2-5) can be utilized in cross-layer methods for improving H.264/AVC video quality over UMTS DCH.

In this work the term cross-layer design refers to protocol design by actively exploiting the dependency between protocol layers to obtain performance gains, according to the definition in [87]. This is unlike layering, where the protocols at the different layers are designed independently. As defined in [87], cross-layer design is protocol design by the violation of a reference layered communication architecture, with the seven-layer OSI (Open Systems Interconnect) model as reference architecture in this thesis; the OSI reference model [88] was jointly defined by ISO (International Organization for Standardization), IEC (International Electrotechnical Commission) and ITU-T (International Telecommunication Union - Telecommunication Standardization Sector).

In [45, W.Karner et al.] a cross-layer error detection method for H.264/AVC video streaming is presented. The performance evaluation of the proposed design showed that the method is capable of reaching great improvements in video quality in spite of having the special burst error characteristics of the UMTS DCH.

Of course, the knowledge of the link error characteristics as well as the special link error models are not only required for performance evaluation of new (cross-layer) methods but the very specific error properties of the link can also explicitly be exploited for improving the end-to-end service quality. In this chapter it is shown that the predictability of the link errors can be used for network-aware cross-layer processing [46, W.Karner et al.] in order to provide substantial gain in video quality with respect to the required data rate. The proposed method transmits additional redundancy information for decoding of the streamed video only in case of expected high link error probability.

Cross-layer processing algorithms are achieving higher improvements in QoS when considering service characteristics in addition to the awareness of network characteristics. In that sense a network and data-priority aware cross-layer scheduling algorithm [43, 47, W.Karner et al.] is presented which provides significant quality improvements in the decoded video by rescheduling the transmission of IP packets containing streamed coded video data according to their priority level. High priority packets (containing parts of I video-frames) are transmitted in time intervals with lower predicted error probability.

Considering more priority levels for smaller data packets (RLC PDUs — Radio Link Control

Protocol Data Units) and setting the priority according to the distortion the loss of the packet would cause, results in even higher video quality. The highest quality in the streamed video is reached with a network-aware, distortion-based scheduling algorithm [44, W.Karner et al.] capable of exploiting the predictability of the instantaneous link failure rate and being video-content sensitive.

As presented in Chapter 2 of this thesis, the main error source of the network is the radio link. Therefore, especially when considering the quality of the services during service usage (without e.g., considering errors during call setup and call blocking or dropped calls), the radio link is the only 'network' related error source and thus the naming "network-aware approach" as a generalisation of being aware of errors in the upper physical layer or equivalently in the data link layer. Meaning, by considering the error characteristics of the radio link, the system is completely aware about the erroneous network behaviour as far as it is relevant for the service quality during service usage (e.g., during a video streaming sequence).

For the network-aware cross-layer processing methods presented in this chapter of the thesis, information flow (link error information) from OSI layer 1 (CRC is performed in the upper physical layer) to layer 2 and to layers 6/7 (video decoder) is required in order to forward the link error location information within one video packet. Furthermore, information flow from the layers 6/7 (video encoder) to the cross-layer scheduling mechanism in layer 2 is required in order to know about data priority for the video decoding process at the network and data-priority aware cross-layer scheduling mechanism.

In this sense the definition of the used cross-layer approach includes the requirement of a cross-layer information flow as well as the cross-layer protocol design coupling (modification) and cross-layer calibration according to [87], as for example, the upper layers have to be modified in order to make use of the correct parts of the video slices.

Chapter 7: Conclusions

This chapter summarises the thesis, provides some general conclusions and final remarks.

1.3 UMTS Overview

As a part of the IMT-2000 (International Mobile Telecommunications at 2000 MHz) standards of the ITU (International Telecommunications Union), the UMTS (Universal Mobile Telecommunication System) is specified within 3GPP (3rd Generation Partnership Project) where its main radio access technologies based on WCDMA (Wideband Code Division Multiple Access) are called UTRA (Universal Terrestrial Radio Access) FDD (Frequency Division Duplex) and TDD (Time Division Duplex). The UTRA TDD mode up to now only possesses little practical relevance, thus in this work the focus is on the FDD mode only.

3GPP is specifying the UMTS system in several steps, from Release '99/4 offering theoretical bit rates of up to 2 Mbit/s, followed by Release 5 and 6 reaching higher bit rates beyond 10 Mbit/s with the introduction of HSDPA (High Speed Downlink Packet Access) and HSUPA (High Speed Uplink Packet Access).

Whereas 2G (2nd Generation) systems like GSM (Global System for Mobile communications) were designed for voice communications, UMTS as a 3G (3rd Generation) communication system with its high data rates, low delay and high flexibility is designed for the delivery of multimedia services.

1.3.1 UMTS Network Architecture

Basically, the UMTS system is built according to the same well-known architecture that has been used by all main second generation systems. At a high-level, the UMTS network consists of three parts [48]. These are the mobile station, also called UE (User Equipment) as the interface between the user and the radio part, the UTRAN (UMTS Terrestrial Radio Access Network) containing radio-related functionality and the CN (Core Network) responsible for the connection to external networks. This high-level system architecture, as well as the most important nodes and interfaces are presented in Fig. 1.1.

The UE consists of the physical equipment used by a PLMN (Public Land Mobile Network) subscriber; it comprises the Mobile Equipment (ME) and the Subscriber Identity Module (SIM), called UMTS Subscriber Identity Module (USIM) for Release '99 and following. The ME comprises the Mobile Termination (MT) which, depending on the application and services, may support various combinations of Terminal Adapter (TA) and Terminal Equipment (TE) functional groups to provide end-user applications and to terminate the upper layers.

Within the UTRAN, several Node Bs (Base Stations) — each of them controlling several cells — are connected to one RNC (Radio Network Controller). The main task of the Node B is the performance of physical layer processing including channel coding, interleaving, rate adaptation, spreading, etc. Furthermore, some RRM (Radio Resource Management) operations like the inner

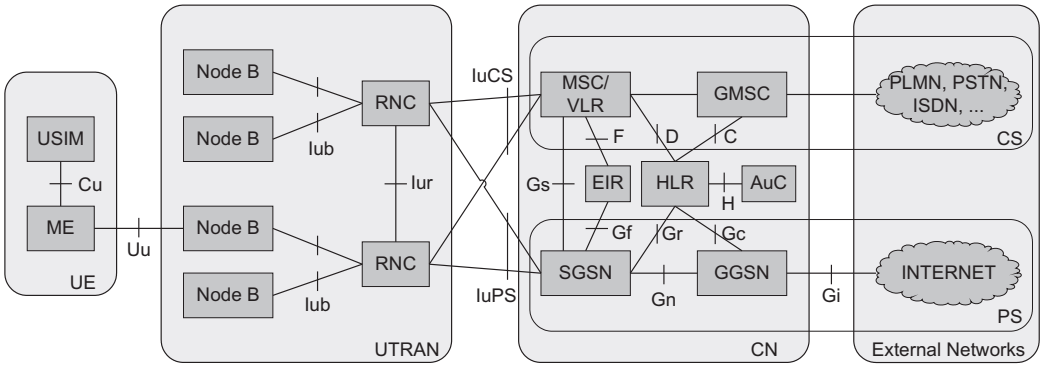

Figure 1.1: UMTS network architecture.

loop power control as well as fast HARQ (Hybrid Automatic Repeat Request), scheduling and priority handling for HSDPA have to be performed in the Node B. The RRM tasks performed in the RNC are the load and congestion control of its own cells, admission control and code allocation for new radio links to be established in those cells as well as handover decisions and the outer loop power control. The RNC performs the layer two processing of the data to/from the radio interface and macrodiversity combining in case of soft handover.

While the UE and the UTRAN contain new specific protocols as well as a new radio interface (WCDMA), the Release '99 UMTS CN was inherited from the GSM system and both UTRAN and GERAN (GSM Edge Radio Access Network) connect to the same core network. As presented in Fig. 1.1, the core network consists of the CS (Circuit Switched) domain for the real time data and the PS (Packet Switched) domain for non-real time packet data. In the CS domain the MSC (Mobile Switching Center) including the VLR (Visitor Location Register) connects to the RNCs. It switches the CS data transactions and stores the visiting user's profiles and location. The GMSC (Gateway MSC) connects UMTS to external networks like e.g. the PSTN (Public Switched Telephone Network). In the HLR (Home Location Register) the user's service profiles and the current UE locations are stored and the EIR (Equipment Identity Register) is a database for identification of UEs via their IMEI (International Mobile Equipment Identity) numbers. The SGSN (Serving GPRS Support Node) is the equivalent to the MSC but for the PS domain. It is responsible for the user mobility and for security (authentication). With the GGSN (Gateway GPRS Support Node) the connection to external networks (internet) is realized.

According to the presented network architecture, Fig. 1.2 shows the layered UMTS bearer architecture, where each bearer on a specific layer offers its individual services using those provided by the layers below. To realize a certain network QoS (Quality of Service), a bearer service with clearly defined characteristics and functionality is to be set up from the source (TE) to the destination (TE) of a service, passing MT, RAN (Radio Access Network), CN edge

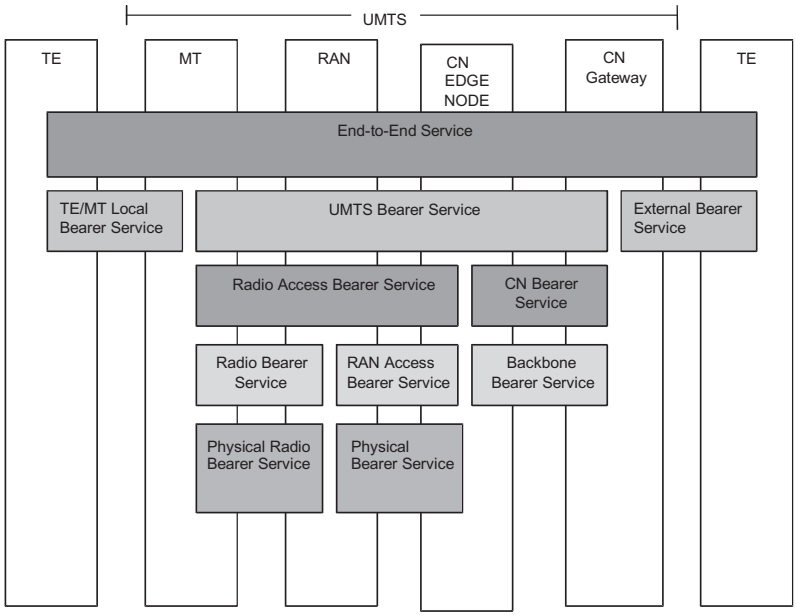

Figure 1.2: Architecture of UMTS bearer services.

node (SGSN) and CN gateway (GGSN). Details of the UMTS QoS concept and architecture can be found in [49] and the interaction and QoS negotiation with other neighbouring networks is specified within [50].

Note, the presented CN architecture is considering Release '99/4. Further details of the UMTS CN architecture and its evolution within Release 5, 6 and 7 can be found in [2] as well as in the corresponding versions of [48].

1.3.2 UTRAN Protocol Architecture

A general overview over the UMTS radio interface protocol architecture [51] is presented in Fig. 1.3. The radio interface is layered into the three protocol layers: L1 (physical layer), L2 (data link layer) and L3 (network layer), and L2 is further split into the following sublayers: MAC (Medium Access Control), RLC (Radio Link Control), PDCP (Packet Data Convergence Protocol) and BMC (Broadcast/Multicast Control). Vertically, L3 and RLC are divided into control and user planes, where the control plane is used for all UMTS-specific control signalling including the RRC (Radio Resource Control) as the lowest L3 sublayer.

In Fig. 1.3 the SAPs (Service Access Points) for peer-to-peer communication are marked with ellipses at the interface between sublayers. The service provided by Layer 2 is referred to as the RB (Radio Bearer). The control plane RBs, which are provided by RLC to RRC, are denoted as signalling RBs.

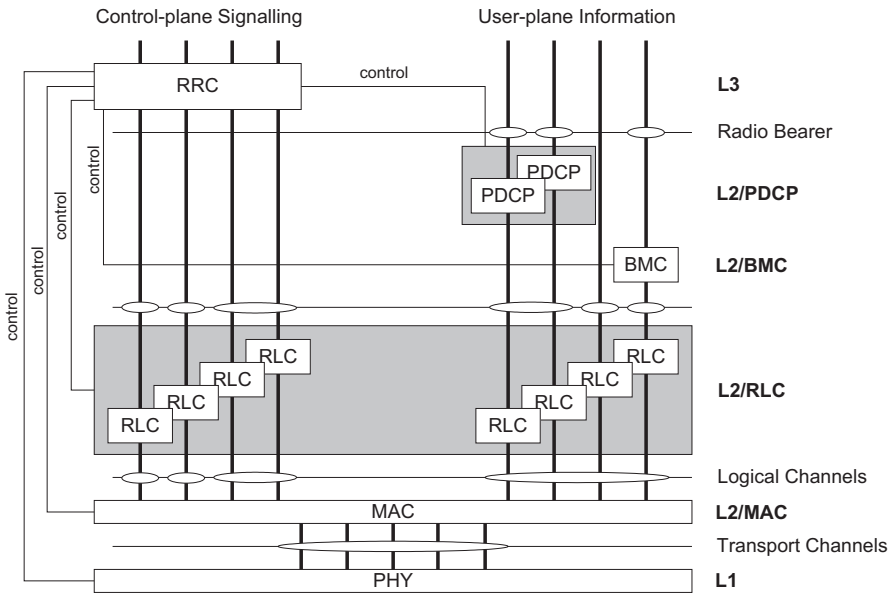

Figure 1.3: UMTS radio interface protocol architecture.

A fundamental part in the UTRAN architecture is the channel concept — the different functions of the channels and the channel mapping. The logical channels are providing an interface for the data information exchange between the MAC protocol and the RLC protocol. There are two types of logical channels: control channels for the transfer of control plane information and traffic channels for the transfer of user plane information. Table 1.1 gives an overview of available logical channels in UTRAN.

Control Channels (CCHs):	Broadcast Control Channel (BCCH)
	Paging Control Channel (PCCH)
	Dedicated Control Channel (DCCH)
	Common Control Channel (CCCH)
	Shared Channel Control Channel (SHCCH)
	MBMS point-to-multipoint Control Channel (MCCH)
	MBMS point-to-multipoint Scheduling Channel (MSCH)
Traffic Channels (TCHs):	Dedicated Traffic Channel (DTCH)
	Common Traffic Channel (CTCH)
	MBMS point-to-multipoint Traffic Channel (MTCH)

Table 1.1: Logical channels.

Whereas the logical channels are separated by the information they are transporting, the transport channels are separated by how the information is transmitted over the air interface (in a shared connection or via a dedicated link). The transport channels provide the bearers for the information exchange between the MAC protocol and the physical layer. In contrast to the logical channels which can be bidirectional, all transport channels are unidirectional. A list of transport channels as well as the possible mapping to the logical channels is presented in Fig. 1.4. The arrows show whether the mapping is for DL (Down Link) and UL (Up Link) or unidirectional only.

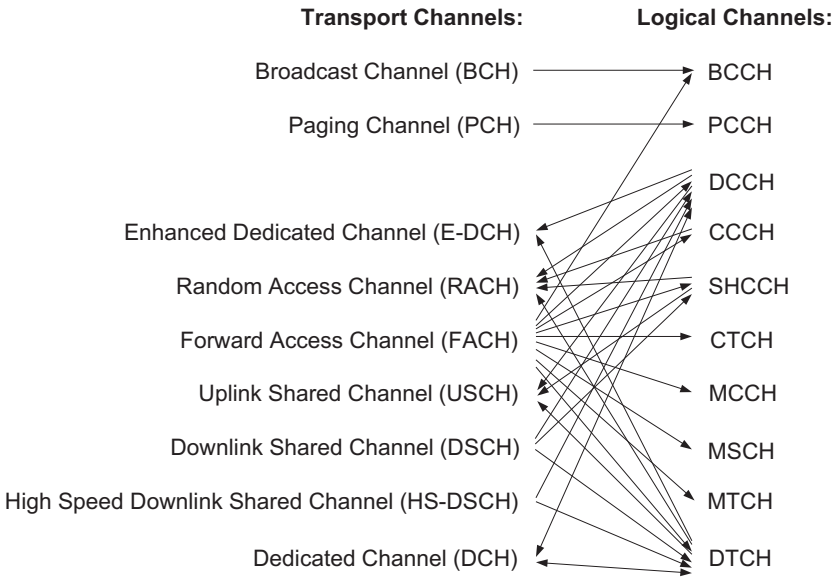

Figure 1.4: Mapping between transport channels and logical channels (seen from UE side).

The RRC [52] is the central and most important protocol within the UTRAN protocol stack as it controls most of the UE, Node B and RNC protocols and configures the physical layer through the transfer of peer-to-peer RRC-signaling messages. The main tasks of the RRC are the establishment, maintenance and release of an RRC connection between the UE and UTRAN as well as paging, QoS control, UE measurement reporting and the OLPC (Outer Loop Power Control).

The PDCP [53] performs header compression and decompression of IP data streams (e.g., TCP/IP and RTP/UDP/IP headers) at the transmitting and receiving entity, respectively.

Several RLC instances are placed in the control and user planes without differences. As a classical data link layer (L2) application the main function of the RLC [54] is the exchange of higher layer PDUs between RNC and UE. Further tasks of the RLC are the segmentation and de-segmentation of higher layer PDUs, overflow protection via discard of SDUs (e.g. after a

maximum number of retransmissions or timeout), error correction by retransmissions and in-sequence delivery in case of retransmissions. The RLC can work in three different modes. In RLC TM (Transparent Mode) the RLC-protocol simply conveys higher layer SDUs (Service Data Units) to the peer RLC-entity without error detection and correction mechanisms, ciphering or in-sequence delivery. The RLC UM (Unacknowledged Mode) enables in-sequence delivery and error detection but no retransmissions and thus error correction mechanisms need to be taken care of by higher layers. The RLC AM (Acknowledged Mode) guarantees the error-free transmission (by means of retransmissions) and in-sequence delivery of upper layer PDUs to the peer entity.

In the MAC-layer [55] the Logical Channels are mapped to the Transport Channels and an appropriate TF (Transport Format) is selected from the TFCS (Transport Format Combination Set) for each Transport Channel, depending on the instantaneous source rate. Further functions of the MAC protocol are priority handling between data flows of one UE and also between UEs by means of dynamic scheduling (for FACH and DSCH), service multiplexing for RACH/FACH/CPCH and DCH, ciphering in case of RLC TM and dynamic Transport Channel type switching.

The processing of the layer three packets within the UTRAN protocol stack can also be seen in Fig. 1.5, where the data flow for non-transparent RLC and non-transparent MAC is shown.

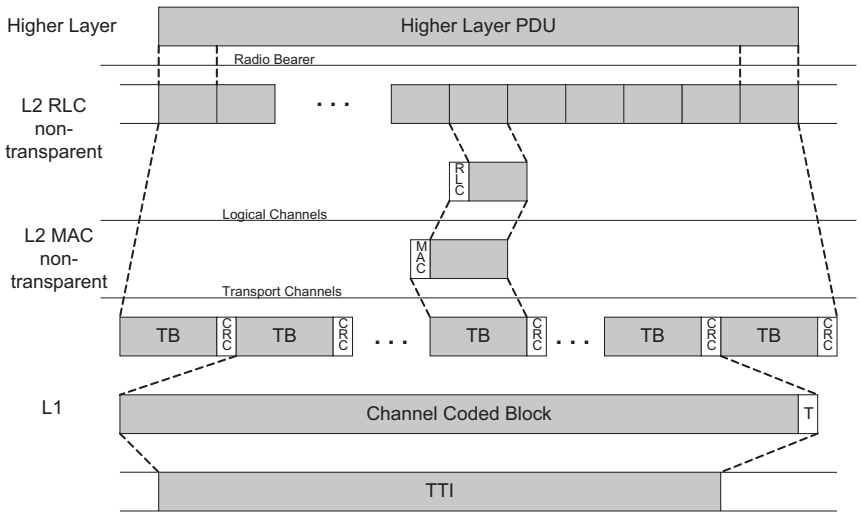

Figure 1.5: Schematic illustration of a packetization example for the transmission over UMTS.

One higher layer PDU (e.g. IP packet) coming from the user plane in layer three will be delivered via the according radio bearers to layer two where it can be processed either by PDCP to perform header compression, the packet can be handled by the BMC or it can directly be

delivered to the RLC layer. There, the higher layer PDUs may be segmented into smaller packets. In case of a bearer with data rate below 384 kbit/s, usually 320 bits (40 bytes) payload within the RLC packets is used [56]. For non-transparent RLC (RLC AM/UM) a header will be added to the packets then forming RLC PDUs. The header size for the RLC AM is 16 bits, whereas the RLC UM packet header contains 8 bits. These RLC PDUs then are transported via the logical channels to the MAC layer where a MAC header is added if transport channel multiplexing (non-transparent MAC) is used in the system. After that, the TBs (Transport Blocks = MAC PDUs) are sent via the transport channels to the physical layer.

1.3.3 Physical Layer Data Processing in the UTRAN Radio Interface

The first process in the physical layer after delivering the so called TBs (Transport Blocks = MAC PDUs) via the transport channels, is a CRC (Cyclic Redundancy Check) of the packet data [57, 58].

Then, after attaching the CRC bits to the TBs [59], these are segmented or concatenated in order to fit to the block size of the channel coding [60]. For packet oriented applications usually turbo coding is used with a coding rate of 1/3, which can further be punctured to match the rate with the physical resources.

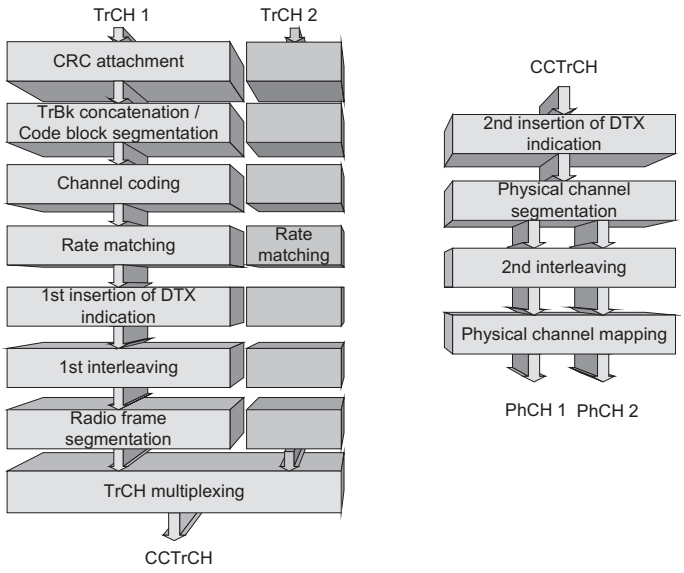

Figure 1.6: UTRAN physical layer procedures.

Fig. 1.6 shows a sequential illustration of the physical layer processes like rate matching, first DTX (Discontinuous Transmission) insertion indication, first interleaving (over one coded block), radio frame segmentation and then multiplexing of the various transport channels. The CCTrCH

(Coded Composite Transport Channel) is then, after a second insertion of DTX indication, segmented into the appropriate physical channels. After a second interleaving (over one radio frame) the data bits are mapped onto the correct physical channel.

In order to illustrate the physical layer procedures and their sequential processing, the data flow of an example for a bearer with 64 kbit/s user data throughput is presented in Fig.1.7.

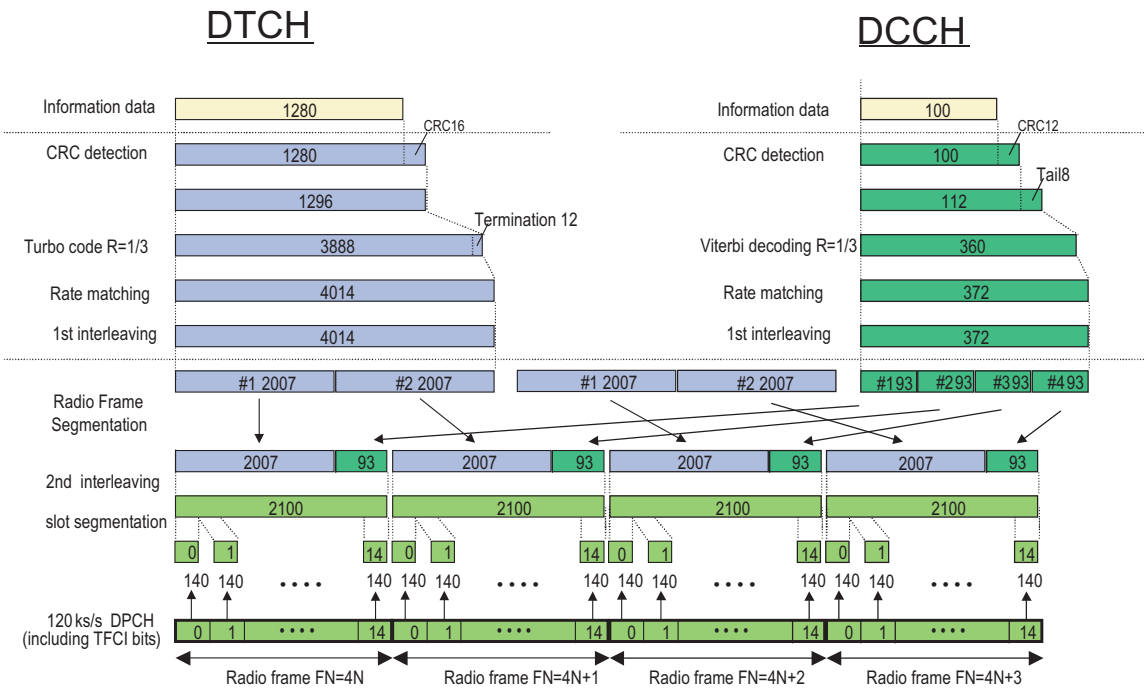

Figure 1.7: Data flow in UTRAN physical layer for 64 kbit/s reference channel [61].

In this example user data from a DTCH enters the physical layer in form of one TB with the size of 1280 bits. It is shown that this TB with 1280 bits is transmitted within two radio frames (10 ms per radio frame) which gives the required 64 kbit/s. The first processing step in the physical layer is the adding of a CRC information to the TB followed by the coding of the resulting data block via turbo code with rate 1/3. Rate matching has to be performed so that the coded data bits fit into the associated radio frames. After a first interleaving and radio frame segmentation, the bits of the DTCH get multiplexed with the information data of the DCCH which transmits TBs of 100 bits over four radio frames (40 ms) and thus reaches a throughput of 2.5 kbit/s. After the transport channel multiplexing, a second interleaving over one radio frame (10 ms) is performed. The resulting data blocks with 2100 bits are transmitted within one radio frame, reaching a bitrate of 210 kbit/s at that point. Every data block with 2100 bits then is segmented into 15 blocks with 140 bits each for fitting into the 15 slots per radio frame.

Considering QPSK[1] (Quadrature Phase Shift Keying) modulation of the DPCH (Dedicated Physical Channel) together with physical layer control data, is resulting in 120 ksymbols/s which become 3.84 Mchip/s due to the spreading operation with a SF (Spreading Factor) of 32 [63].

In Fig. 1.8 a table for the mapping of transport channels onto corresponding physical channels is shown. In case of the DCH, the data from DTCH and DCCH is mapped onto the DPDCH (Dedicated Physical Data Channel) and multiplexed with the DPCCH (Dedicated Physical Control Channel) which contains TPC (Transmit Power Control), TFCI (Transport Format Combination Indicator) and Pilot bits.

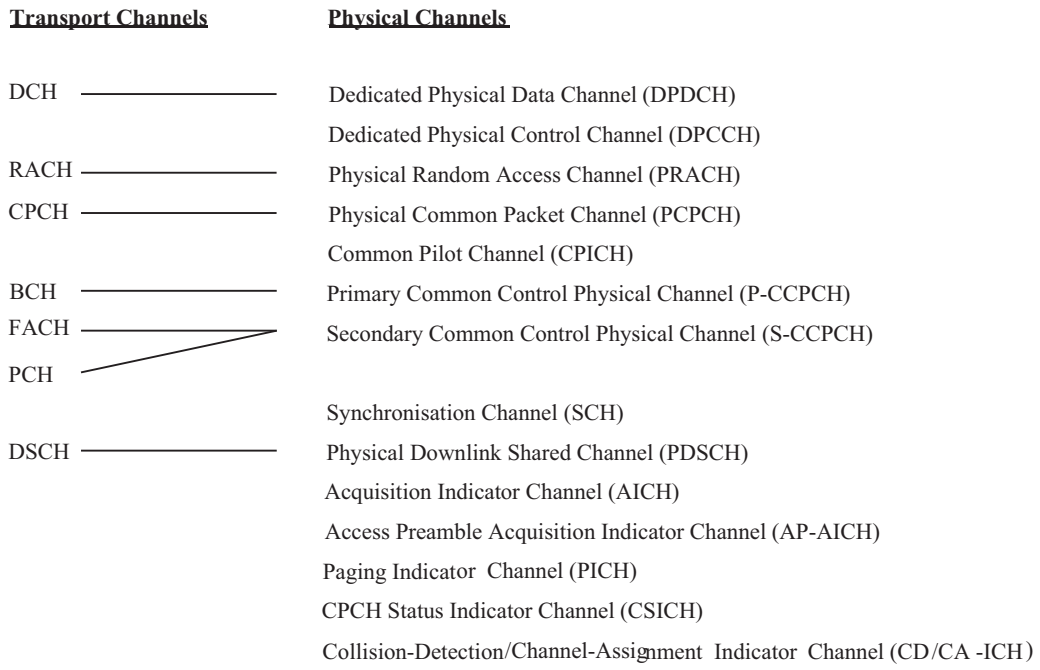

Figure 1.8: Mapping of transport channels onto physical channels [64].

In the UMTS DL the DPDCH and the DPCCH are time multiplexed and modulated via QAM (Quadrature Amplitude Modulation) whereas in the UL the DPCCH and the DPDCH are modulated according to two orthogonal PAM (Pulse Amplitude Modulation) schemes separately in order to prevent an interference of the transmitting UL signal with audio equipment like in GSM. An exemplary illustration of the slot structures in UL and DL is presented in Fig. 1.9.

[1]Despite the fact that QPSK (Quadrature Phase Shift Keying) and BPSK (Binary Phase Shift Keying) is mentioned throughout the documents in 3GPP, as modulation scheme AM (Amplitude Modulation) in form of QAM (Quadrature Amplitude Modulation) and PAM (Pulse Amplitude Modulation) with RRC (Root Raised Cosine) transmit pulse-shaping filter (roll-off factor $\alpha = 0.22$ [62]) is used while the terms QPSK or BPSK just indicate the symbol constellation [63].

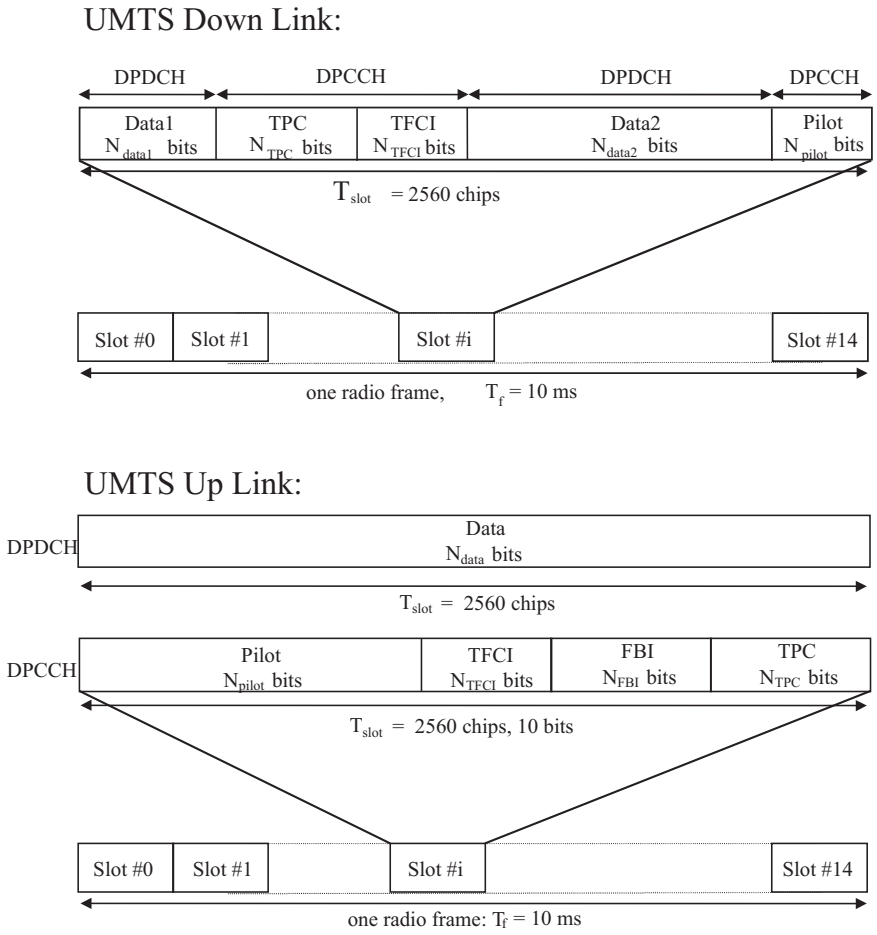

Figure 1.9: DL and UL slot structure example [64].

Chapter 2

Measurement Based Analysis of Link Layer Error Characteristics

Contents

The following link error analysis and the resulting modeling was performed based on measured link layer error traces. The measurements have been realized in live UMTS networks of three different operators in the city center of Vienna, Austria.

Due to similar adjustments of the relevant radio parameters in the different networks the measured results out of the three networks [10] lead to the same conclusions. Thus, we are focussing on the measured error traces out of only one operator's live network (with the UMTS radio network elements from Ericsson [65]) in the following in this document.

Moreover, the measurements for this work have been performed with several different mobile stations and radio bearers as well as in various mobility scenarios. Throughout this work the following structure for presenting the results from these measurements is maintained: As long as the resulting conclusions from the measurements with the different mobiles/bearers/scenarios are the same, the measured results from only one mobile/bearer/scenario are presented with the specific name of the mobile/bearer/scenario given as reference. In case more information or further conclusions become available from the comparison of the results measured with different mobiles/bearers/scenarios, these results are presented in parallel.

2.1 Measurement Setup

2.1.1 General Setup

In Fig. 2.1 a schematic illustration of the measurement setup for the measurements in the live networks is given. For the measurements a UDP (User Datagram Protocol) data stream with bit rates of 360 kbit/s, 120 kbit/s and 60 kbit/s (which are 372 kbit/s, 125,6 kbit/s and 62,8 kbit/s including UDP/IP overhead respectively) was sent from a PC over the UMTS network to a notebook using a UMTS terminal as a modem via a USB (Universal Serial Bus) connection.

Additionally, measurements have been performed in a reference network which is a separate network for acceptance testing of the same operator and with equal parameter settings than in the corresponding live UMTS network. As illustrated in Fig. 2.2, in the reference network the radio link (Uu interface) is replaced by a cable connection (attenuation ≈ 60 dB) between the Node-B antenna connector and a RF (Radio Frequency) shielding box (Willtek 4920 [66]) where the mobile is enclosed to avoid interference, multipath propagation and fading effects.

In order to be capable of tracing the internal measurements of the mobiles, WCDMA 'TEMS' mobiles[1] were used as terminals in connection with 'TEMS Investigation' software as offered by Ericsson [65]. In this document we refer to the used mobiles as 'mobile 1' to 'mobile 4' as listed in Table 2.1.

[1]Modified UMTS mobiles for TEMS data logging are offered by Ericsson [65]. The mobiles are modified in a way that they provide the internal measurements via the interface to the notebook.

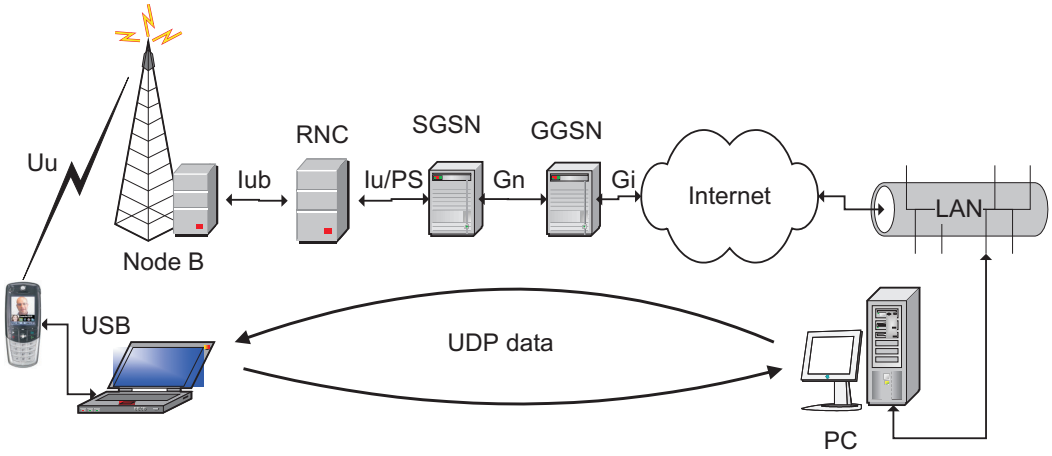

Figure 2.1: Scheme of the measurement setup (live network).

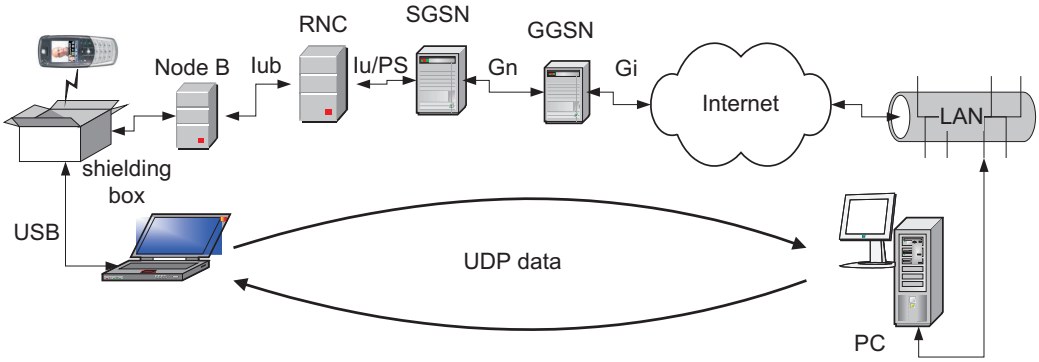

Figure 2.2: Scheme of the measurement setup (reference network).

'mobile 1'	Motorola A835
'mobile 2'	Motorola E1000
'mobile 3'	Sony Ericsson Z1010
'mobile 4'	Sony Ericsson V800

Table 2.1: Mobile equipment used for the measurements.

Throughout this work the following structure for presenting the results from different mobile terminals is maintained: As long as the resulting conclusions from the measurements with the different mobile terminals are the same, the measured results from only one are presented with

the specific mobile name given as reference. In case more information or further conclusions can be given from the comparison of the results measured with different mobiles, the results are presented in parallel.

After parsing the export files of the 'TEMS Investigation' software tool, various parameters of the UMTS system can be analysed. One of these parameters is the CRC information of the received TBs which is used for the analysis of the UMTS link error characteristics as presented further on in this document.

In the considered UMTS networks turbo coding with a coding rate of 1/3 was used for the DTCH with a TB size of 336 bits, consisting of 320 bits for the RLC payload and 16 bits for the RLC AM header. Although using the RLC AM with its error detection and feedback mechanism, the link error analysis was performed without consideration of retransmissions, thus offering independence of selecting RLC AM/UM and the specific RLC AM parameter adjustments like discard timer or the maximum number of retransmissions [54]. The settings of SF, TTI and the number of TBs which are jointly coded and transmitted per TTI is shown in a schematic illustration in Fig. 2.3 for the three UMTS DL radio bearers which have been used during the measurement campaign.

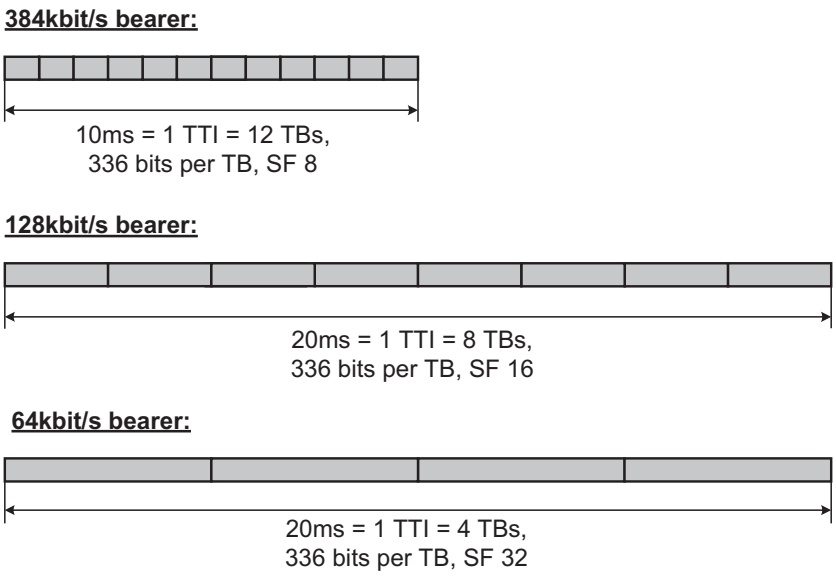

Figure 2.3: Illustration of bearer parameters.

The focus throughout this work mainly is on the 384 kbit/s bearer as it represents the most demanding of the available bearers. Like already mentioned for the measurements from different networks and different mobile stations, also the measurement results for the other bearers are presented just in case there arise new conclusions out of it. Otherwise, only the results for the

384 kbit/s bearer are shown in this work and the conclusions for the other bearers can be derived similarly.

Another very important parameter of the UTRAN for evaluating the error characteristics of the DCH is the BLER (BLock Error Ratio) quality target value for the OLPC (Outer Loop Power Control) mechanism. This target value was set to 1 % in the networks used for the measurements. As a consequence, the OLPC tries to adjust the SIR (Signal to Interference Ratio) target for the ILPC (Inner Loop Power Control) mechanism in a way that the required link quality (1 % TB error probability in this case) is satisfied. It will be shown further on in this chapter that this quality target is missed significantly in all the scenarios.

2.1.2 Mobility Scenarios

For the analysis of the UMTS DCH link error characteristics, we have considered several scenarios with different mobility characteristics to which in this work we refer as **'static'**, **'small-scale movements'**, **'walking indoor'**, **'tramway'**, **'car-city'**, **'car-highway'** and **'reference'** (see also Table 2.2).

reference network (within shielding box) - without movement:	'reference'
live network - without movement:	'static'
live network - with movement ('dynamic'):	'small-scale movements' 'walking indoor' 'tramway' 'car-city' 'car-highway'

Table 2.2: List of measurement scenarios with different mobility characteristics.

The measurements for the **'static'** case were performed in an office room in the city center of Vienna/Austria, with the UMTS terminal lying on the table in a typical Viennese office environment. Due to few movements of persons or other objects around the mobile station, there were only little variations in the channel. The **'small-scale movements'** measurements were performed by a person sitting at the table and randomly tilting and moving the UMTS mobile with his hands. In the **'walking indoor'** scenario, as the label indicates, the measurement results were obtained while walking around inside the building.

The rest are outdoor scenarios with the measurements performed in a tramway going round the city center of Vienna (**'tramway'**) and going by car either on a street in Vienna (**'car-city'**) with moderate speed up to 50 km/h or on a highway with higher speeds up to 100 km/h (**'car-**

highway'). The speed distribution for the measurements in the last two scenarios is presented in Fig. 2.4.

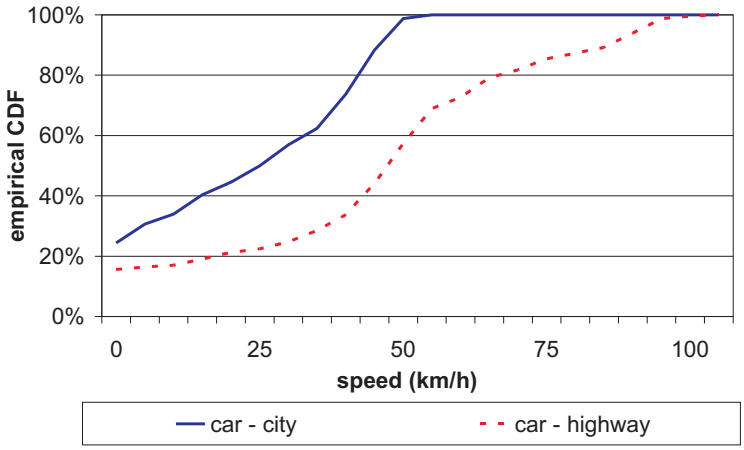

Figure 2.4: Comparison of speed in the scenarios 'car-city' and 'car-highway'.

As already mentioned, for the **'reference'** scenario the mobile was enclosed within a shielding box to avoid interference, multipath propagation and fading effects, with the direct signal from the Node-B antenna connector fed into the box via a planar antenna coupler at the bottom inside the box.

In order to get sufficient statistics of the measured link error characteristics, several data traces have been recorded in each of the mentioned scenarios, all with a length of about one hour.

Of course, the measured link error statistics within all the scenarios heavily depend on the actual system and service coverage in the network — if the mobile station moves towards or even over the cell edge and thus the link transmit power at the base station reaches its limit, the target SIR value at the receiver cannot be met anymore and the link error probability increases. Therefore, measured network performance statistics in general are to be seen as a snapshot at the actual state of the UMTS radio network deployment at the time when the measurements have been performed. This especially is the case for the scenarios with movement. It can be observed in Figs. 2.6 and 2.8 that the CPICH (Common Pilot CHannel) E_c/I_0 (chip energy to noise and interference ratio) and CPICH RSCP (Received Signal Code Power) are reaching very low values during the measurements in the 'tramway' and 'car-highway' scenarios, measured with 'mobile 1'. Usually, a CPICH E_c/I_0 level of about -12 dBm is defined as the cell edge [2].

On the other hand, it is shown in Figs. 2.5 and 2.7 that in the 'static' and the 'reference' scenario, the CPICH E_c/I_0 levels indicate to be well inside network coverage. Note, 'static 1' to

'static 3' are measurements in 'static' scenarios at different locations, on different days and at different times of day, thus representing different propagation and network load conditions. It will be presented in the next section of this document that all these measurements in the 'static' scenarios are resulting in the same link error characteristics despite having different absolute values of CPICH E_c/I_0 and CPICH RSCP. Furthermore, measurements in heavy loaded cells have shown that the cell load has no impact on the DCH error characteristics — meaning there is a well working admission control mechanism in the system. Due to the mentioned arguments we can conclude that the measured results from the 'static' scenarios are widely independent of the current state of network deployment and network load.

However, Figs. 2.7 and 2.8 show great differences between the variance of the measured CPICH RSCP values of the 'static' and 'reference' scenarios and the movement case which is due to the differences in the fading properties. It will be shown in the next section of this document, that these differences in the CPICH RSCP variance caused by different fading effects are leading to different link error characteristics due to the non optimality of the power control algorithm.

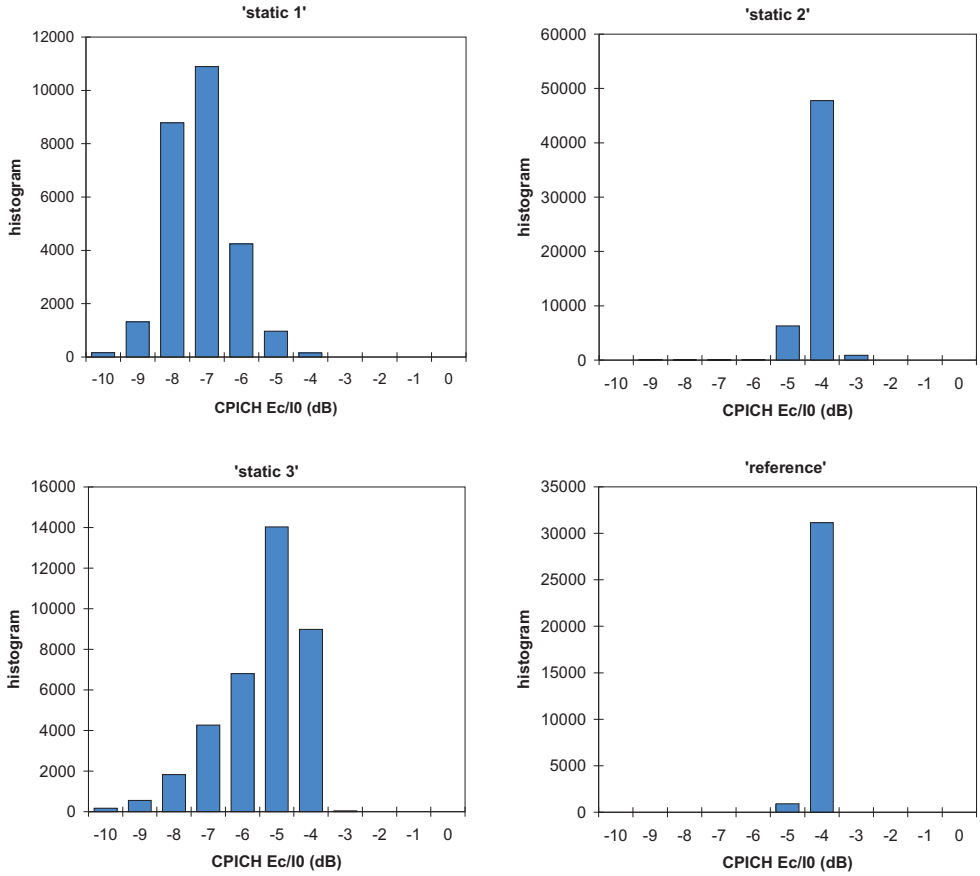

Figure 2.5: CPICH E_c/I_0 with different propagation and network load conditions in the live network ('static') and in the reference network ('reference'), 'mobile 1'.

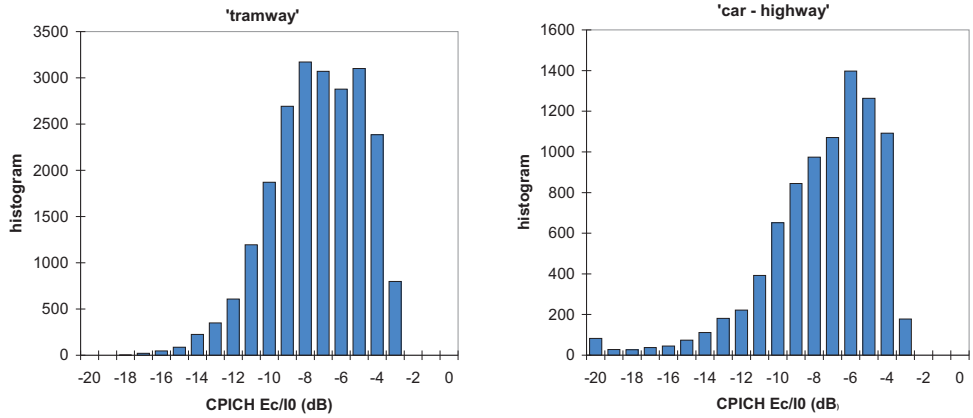

Figure 2.6: CPICH E_c/I_0 for 'tramway' and 'car-highway' scenarios, 'mobile 1'.

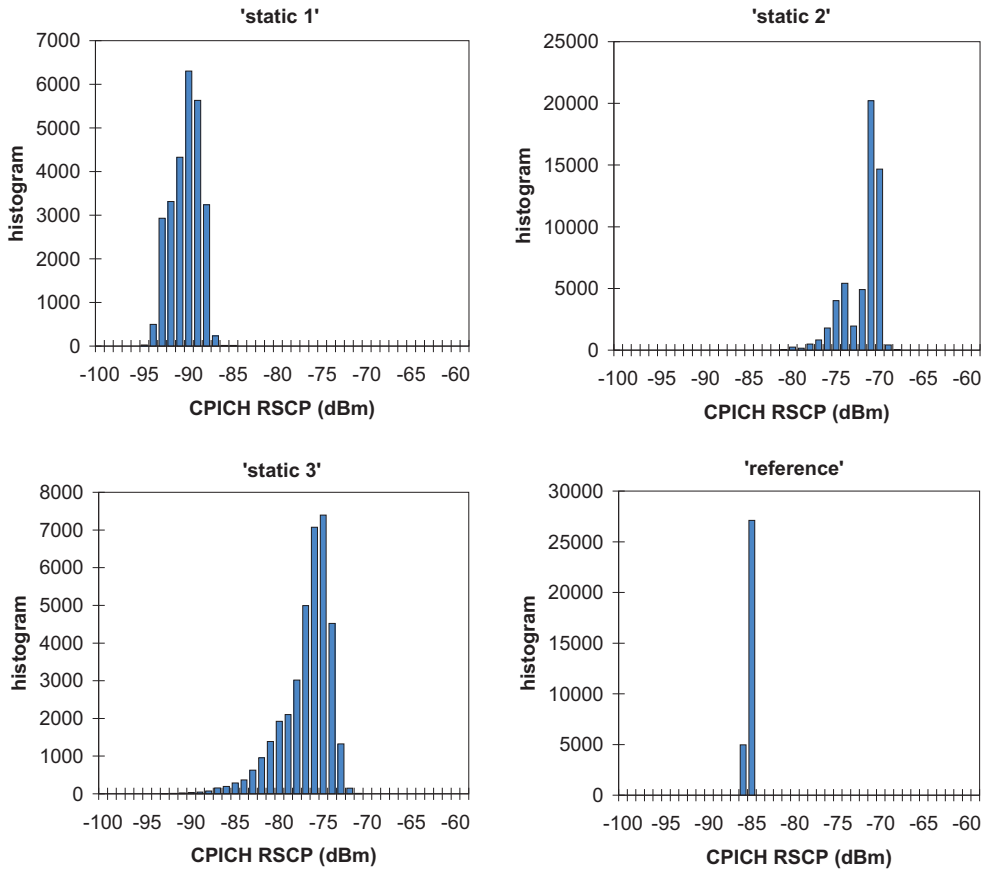

Figure 2.7: CPICH RSCP with different propagation and network load conditions in the live network ('static') and in the reference network ('reference'), 'mobile 1' .

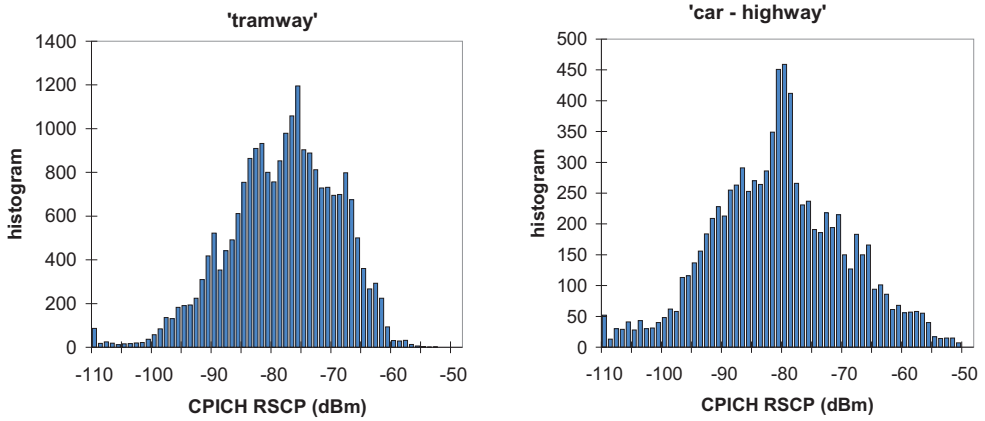

Figure 2.8: CPICH RSCP for 'tramway' and 'car-highway' scenarios, 'mobile 1'.

2.2 Link Error Analysis

When transmitting data over wireless mobile communication systems, most of the errors are originating from the radio link, which will also become the main or even only error source in the transmission chain as long as the frequency spectrum is a scarce resource and there is interference limitation in the system. Since the highest error probability is found in the radio link, we are focussing the error analysis primarily on the wireless access part of the UMTS network.

As presented in Section 1.3, the UMTS standard specifies an error check in the radio link in form of a CRC per TB in the physical layer [57]. This information can perfectly be used in our link error analysis, as all the important error sources within the radio link (fading, interference, multipath propagation, power control, as well as channel coding and interleaving effects) are included in the resulting link error statistics.

2.2.1 Link Error Probability

A first parameter to be considered for the analysis of the link error characteristics is the estimated total TB error probability[2] $\overline{P}_e$(TB) (over the complete measured trace) presented in Table 2.3 for the different scenarios, measured with 'mobile 1' and a 384 kbit/s PS (Packet Switched) bearer. Similar results were obtained with the other mobiles and bearers.

Table 2.3: Estimated probabilities of TB and TTI errors (384 kbit/s bearer).

Scenario	$\overline{\mathbf{P}}_\mathbf{e}$**(TB)**	$\overline{\mathbf{P}}_\mathbf{e}$**(TTI)**
static	$2.66 \cdot 10^{-3}$	$4.72 \cdot 10^{-3}$
small-scale movements	$2.22 \cdot 10^{-2}$	$2.34 \cdot 10^{-2}$
walking indoor	$1.98 \cdot 10^{-2}$	$2.44 \cdot 10^{-2}$
tramway	$1.44 \cdot 10^{-2}$	$1.70 \cdot 10^{-2}$
car-city	$2.06 \cdot 10^{-2}$	$2.48 \cdot 10^{-2}$
car-highway	$2.34 \cdot 10^{-2}$	$2.63 \cdot 10^{-2}$

In all the scenarios with mobility, the estimated TB error probability is around $2 \cdot 10^{-2}$ whereas in the static case the estimated error probability is smaller by one order of magnitude. Due to these results, not only we can conclude that the TPC (Transmit Power Control) algorithm

[2] As there is no proof of ergodicity, error ratios cannot be equivalently seen as error probabilities and thus in this work generally the terms 'estimated error probability' or 'error ratio' are used for measured values. Despite these arguments, sometimes also the term 'error probability' is used for measured results in order to prevent confusion with expected future probabilities, which then of course also refers to an 'estimated error probability' or 'error ratio'.

of the UMTS is not capable of adjusting the required BLER target value (1 %), but we can also see that there are coarsely only two different link error characteristics with respect to estimated total link error probability: static and dynamic (regardless of which kind of movement). An interesting and new conclusion, contradicting common researcher opinion is that e.g., the small-scale movements performed by a person just sitting and moving the mobile with his hands result in the same error probability as when going by car.

The work in this thesis is based on measurements in live UMTS networks. Unfortunately, the analysis of the received signal within one TTI was not possible with the used measurement equipment and therefore, the different underlying fading effects could not be analysed. Furthermore, it was outside the scope of this thesis to perform an analysis with a system level simulator, which would have to include propagation effects, channel coding, interleaving as well as a feedback channel and the inner closed loop power control and quality-based outer closed loop power control mechanisms. However, the very important conclusion can be drawn that the resulting channel (fading and interference effects in connection with power control, channel coding, interleaving,...) behaves the same for all the considered scenarios with movement regardless of which kind of movement.

It may be assumed that the similarity in error characteristics e.g., of the 'small-scale movements' scenario and the scenario 'car-highway' is due to the spatial small scale fading in the indoor scenario caused by multipath propagation. Together with the small scale movements (changing the position of the mobile station by hands) the mobile is affected by fading in time which seems to be equivalent to the large-scale fading which occurs when going by car with up to 100 km/h. Another reason can be the antenna characteristics of the mobile station in connection with the small scale movements (tilting and turning the mobile station) again resulting in a time variable channel. The resulting fading of received code power (or equally the changing of SIR) in time, together with the non optimality of the power control mechanism is leading to the high error probability in the small-scale movement scenario.

Moreover, the differences between the 'static' case, where there is only little movement of objects or persons around the mobile station, and the 'dynamic' case with the time variant channel, may be assumed to be due to the fact that in the 'static' case the influence of the power control algorithm dominates in the error characteristics while in the 'dynamic' scenarios the fading effects dominate.

When comparing $\overline{P}_e(\mathrm{TB})$ in Table 2.3 with the estimated total TTI error probability $\overline{P}_e(\mathrm{TTI})$ (from the same measurements) we note that the results are almost equal especially in the scenarios with mobility. With 12 TBs per TTI in case of the 384 kbit/s bearer and the assumption of a memoryless BSC (Binary Symmetric Channel) one would expect a much higher estimated total TTI error probability (e.g. $\overline{P}_e(\mathrm{TTI}) = 1 - (1 - 1.44 \cdot 10^{-2})^{12} = 0.159$ in the 'tramway'

scenario) and thus we can conclude that there is a high correlation between the error states of the received TBs within one TTI leading to bursty TB error behaviour of the UMTS DCH which is further analysed in the following sections of this document.

2.2.2 Number of erroneous TBs in TTIs

The analysis of the number of erroneous TBs within erroneously received TTIs presented in Fig. 2.9 for the 384 kbit/s bearer and measured with 'mobile 1' leads to the same conclusions — especially for the scenarios with mobility, where the ratio of having all (twelve) TBs erroneously received within one erroneous TTI is between 0.8 and 0.9. On the other hand, the ratio of having only one out of twelve TBs received erroneously within one TTI is increasing, the less movement is disturbing the transmission. We observe ratios of up to 0.7 with the mobile in the shielding box of the reference network where there is no interference, no multipath propagation and no fading in the propagation environment. This brings up the conclusion that the small variations of transmit power caused by the UMTS DCH TPC only, lead to a very small TB error probability within one TTI (meaning within one jointly turbo-coded and interleaved data block), whereas the deep fades of the scenarios with movement are heavily influencing the channel and thus leading to a high ratio of TB errors within an erroneous TTI. This conclusion is approved by using a low (0.05) and a high (0.98) error probability for Bernoulli experiments, resulting in comparable distributions (binomial) of the number of erroneous TBs per TTI (Fig. 2.10) for the reference network and the case with movement, respectively. The static case can be regarded as a mixture of both cases. These results can be used for refinement of the link error models especially in 'dynamic' scenarios, as presented in Section 3.2.1.

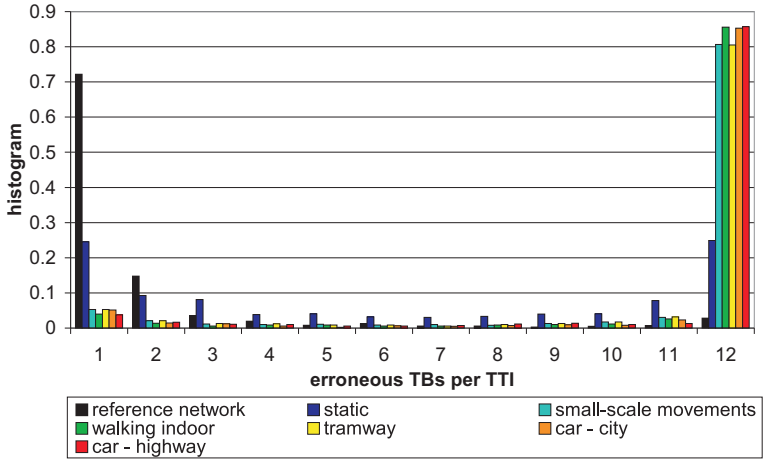

Figure 2.9: Number of erroneous TBs within erroneously received TTIs (384 kbit/s bearer, 'mobile 1').

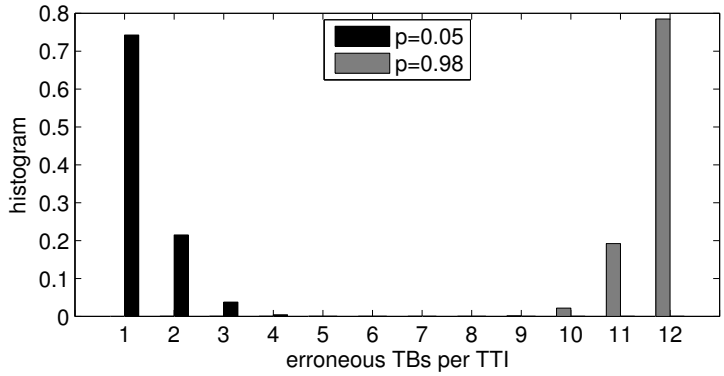

Figure 2.10: Theoretic analysis of the number of erroneous TBs within one TTI: binomial distributions with 0.05 and 0.98 TB error probability.

2.2.3 TTI-Burstlength, TTI-Gaplength

For further analysis of the correlation properties of the link errors outside one TTI (one jointly turbo coded and interleaved data block) we are building the statistics of the TTI-gaplengths and the TTI-burstlengths (see Def. 2.2.3.1 and Def. 2.2.3.2).

Definition 2.2.3.1 *The **TTI-gaplength** is the number of subsequently received error-free TTIs, while the number of subsequent erroneous TTIs is called **TTI-burstlength**.*

Definition 2.2.3.2 *According to [67], an **error burst** is defined as a group of successive units (bits, packets or TBs and TTIs in our case) in which two successive erroneous units are always separated by less than a given number L_c of correct units. For our work L_c is equal to zero for bursts.*

From the presentation of the TTI-gap- and TTI-burstlengths in Fig. 2.11 we observe that the sequences of error-free TTIs are up to 700 and there are up to ten TTIs subsequently erroneously received. Again we recognize that the statistics for the scenarios with movement are similar while the 'static' scenario shows a considerably different distribution of the TTI-gaplengths and much shorter TTI-burstlengths of up to two only. The reason for this difference between 'static' and 'dynamic' scenarios is that in the scenarios with movement the influence of fading-effects is dominating in the link error characteristics while in the 'static' scenario the impact of the OLPC algorithm becomes significant.

The fact, that the TTI-gap- and TTI-burstlengths of the scenarios with only slow movement are similar to those measured in the scenarios with fast movement of up to 100 km/h also

becomes interesting when analysing the expected coherence time of the channel. With a speed of 100 km/h and the assumption of a coherence length of half a wavelength, the coherence time of the channel due to small-scale fading is shorter than one TTI (10 ms), whereas with walking speed the coherence time would be of the length of 10 to 20 TTIs. Therefore, it may be concluded that in the case of high speeds it is the large-scale fading with an assumed coherence length of about 20 m which results in the same error characteristics than the small-scale fading in case of slow movement.

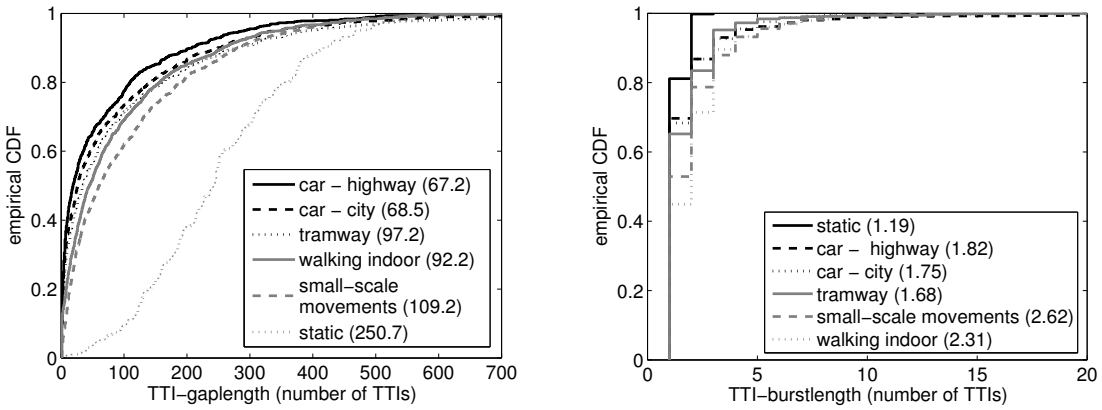

Figure 2.11: Statistics of TTI-gaplengths and TTI-burstlengths (mean values in the legends), 384 kbit/s bearer, 'mobile 1'.

2.2.4 TB Error Bursts, TB Error Clusters

Due to the high ratio of having all TBs within one TTI erroneously received for the scenarios with movement (as shown in Fig. 2.9) the gap and burst analysis in terms of TTIs would be sufficient, whereas in the 'static' case the granularity of the analysis has to be refined to TB level in order to catch correlation properties of TB errors within one TTI.

Thus, we define TB error gaps and TB error bursts with the according gaplength and burstlength in Def. 2.2.4.1, again with the definition of a TB error burst from Def. 2.2.3.2. Additionally, we define groups of erroneously received TBs as error clusters in Def. 2.2.4.2.

Definition 2.2.4.1 *The **gaplength** is the number of subsequently received error-free TBs, while the number of subsequent erroneous TBs is called **burstlength**.*

Definition 2.2.4.2 *An **error cluster** is a group of erroneously received TBs, if, for L_c equal to the number of TBs per TTI, the TB error bursts are separated by $\leq L_c$ error-free TBs.*

In Fig. 2.12 we present a schematic illustration of the TB error bursts, gaps and the error clusters.

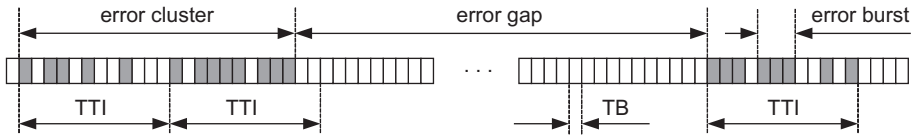

Figure 2.12: Schematic illustration of TB error bursts, gaps and clusters.

Fig. 2.13 shows the results of the analysis of the gaplengths and burstlengths (in number of TBs) for the 'static' scenario measured with a 384 kbit/s bearer and 'mobile 1' at three different locations as well as at different dates and times of day ('static 1' to 'static 3') in order to have different propagation, interference and network load situations [9]. We observe that there is only little difference between the statistics.

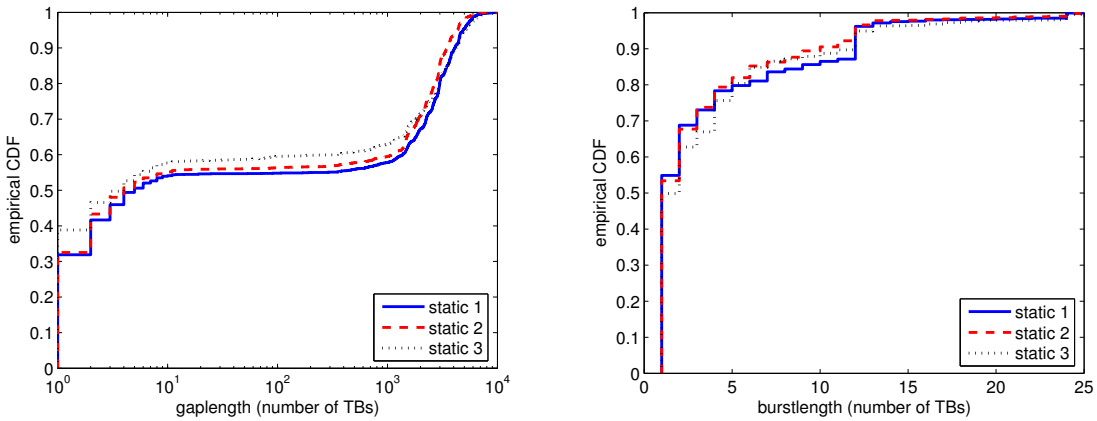

Figure 2.13: Statistics of TB gaplengths and burstlengths, 384 kbit/s bearer, 'static' scenario, 'mobile 1'.

Furthermore, in measurements with additional DCH and also HSDPA (High Speed Downlink Packet Access) users for creation of intra- and inter-cell interference we found out that additional interference in the DL does not change the measured statistics as long as the UMTS call admission control is working properly to keep the required link power off the power limit.

We can see in Fig. 2.13 that there are short gaps of $\leq$ 12 TBs which are the gaps within one or two successive TTIs. The statistics also include long gaps from 200 to 7000 TBs. The burstlengths are between one and 25 with two higher steps at 12 and 24 TBs caused by the ratio of having all TBs erroneously received within one TTI (there are 12 TBs within one TTI in case of the 384 kbit/s bearer) which is still not negligible even in the 'static' case.

Of course, the turbo coding as well as the interleaver algorithms of the UMTS radio access are affecting the link error characteristics, especially within one radio frame, TTI or coding block. However, the measurement-based analysis of the positions of the erroneous TBs within one TTI has shown that all possible TB locations exhibit almost the same estimated error probability (presented in Fig. 2.14). Similar results are observed from the analysis of the output of turbo code simulations in [68].

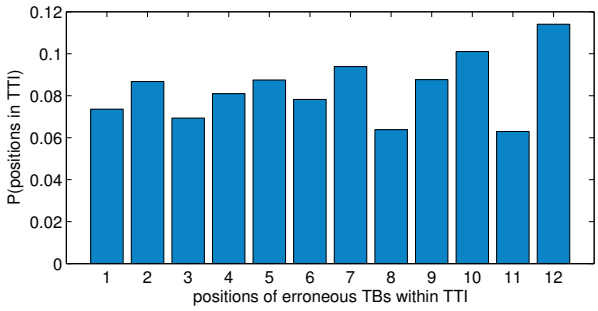

Figure 2.14: Estimated probabilities of TB error positions, 'static' scenario, 'mobile 1', 384 kbit/s bearer.

The uniform distribution of the positions of the erroneous TBs within one TTI, together with the high ratio of having not all the TBs within one TTI received with error, results in short error bursts (separated by short gaps) within one and even two successive TTIs.

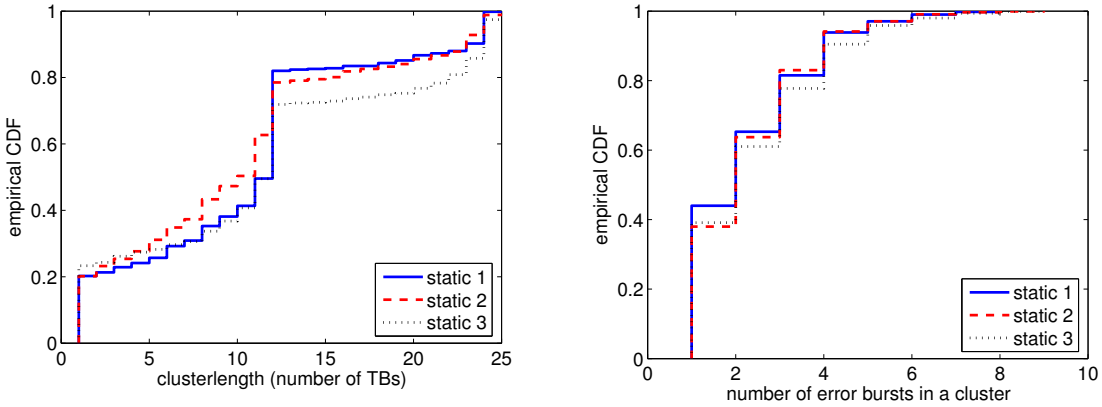

Figure 2.15: Statistics of TB clusterlengths and number of error bursts per cluster, 384 kbit/s bearer, 'static' scenario, 'mobile 1'.

Thus, in order to analyse the error properties within one or more successive TTIs — meaning the characteristics of and within an error cluster — it is also interesting to present the statistics of

the clusterlength and the number of error bursts in a cluster as shown in Fig. 2.15 for $L_c = 12$. There we show three 'static' measurements which have a mean clusterlength of 10.84, 11.19, 14.98, a mean number of erroneous TBs per error cluster of 7.97, 7.74, 9.81, and a mean number of burst per cluster of 2.19, 2.23, 2.38, respectively. We observe that the clusterlength is ≤ 25 in most cases and we do have one to eight error bursts within one such error cluster. Since the influence of the quality based power control algorithm on the link error characteristics becomes significant in the 'static' scenarios, the error clusters are separated by long error gaps as shown in the following.

2.2.5 The Influence of TPC on Link Error Characteristics

In order to use radio resources in cellular wireless communication systems efficiently, the minimum transmit power for each link has to be found while meeting the quality target of the link. In the UMTS this task is performed by the OLPC algorithm [69] — a slow (100 Hz) closed loop control mechanism which sets the target SIR for the fast ILPC in order to converge to a required link quality given by the network (1 % in our work). The ILPC then regulates the link power level with a rate of 1500 Hz to meet the target SIR.

Within 3GPP (3rd Generation Partnership Project) the OLPC is not entirely specified. A common way of implementing the control procedure is using CRC information feedback of the transmitted TBs together with a sawtooth algorithm as described in [2] and [70]. If a TB was received in error, the target SIR value is increased by $K \cdot \Delta$, whereas in case of error-free transmission it is reduced by Δ, where Δ is the step size in dB. If K is chosen to be $1/\mathrm{BLER}_{\mathrm{target}} - 1$, the algorithm tries to keep the BLER always less than or equal to $\mathrm{BLER}_{\mathrm{target}} = 1/(K+1)$. Thus, by selecting a proper value K a quality level for the wireless link can be set.

For this work there was no information available about the used TPC algorithms of the three different manufacturers of the used equipment in the three considered different UMTS live networks. Despite this fact, the analysis of the measured SIR signal and SIR target values lead to the conclusion that a TPC algorithm, like the one described, is used in the systems. This TPC algorithm is also the only one which is very commonly used in literature. Of course, this is an assumption and minor differences in the actually used TPC algorithms to the assumed one are possible.

Fig. 2.16 shows a schematic illustration of this OLPC algorithm for the case when an error occurs in the link with a probability of one if the $\mathrm{SIR}_{\mathrm{target}}$ and thus also the SIR at the receiver falls below a certain threshold. Obviously, in this special case there is an erroneous TB after K error-free TBs.

As we have seen, in the UMTS DCH, error clusters are generated due to joint interleaving and coding over one TTI and also due to the stochastic nature of the channel. The power control

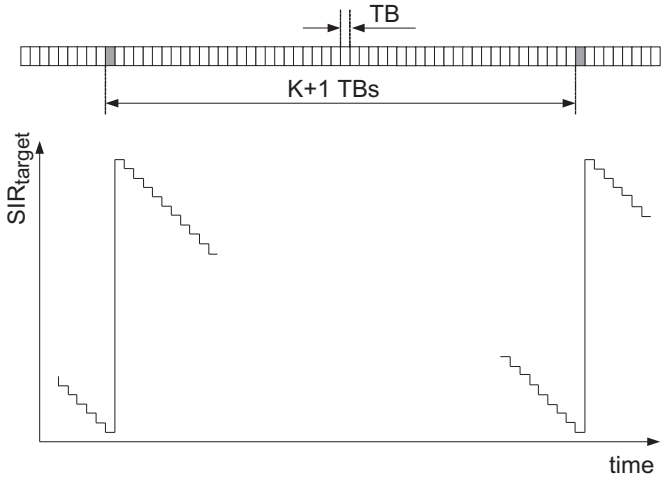

Figure 2.16: Quality based power control (ideal).

algorithm, which is still trying to meet the quality target for the link, increases the target SIR by $K \cdot \Delta$ times the number of errors in the cluster. This leads to a multiple of K error-free TBs (long gap) between two error clusters.

From these arguments we can conclude that especially in the static case, the link error characteristics are a result of the quality-based closed outer loop power control mechanism of the UMTS DCH. Compared to the static case, in the scenarios with movement, fading effects are dominating the reasons for radio link errors but nevertheless, the TPC still affects the link error characteristics in a way that the error process becomes recurrent.

Furthermore, we can conclude that the predictability of the link errors is due to the fact that the quality-based control mechanism introduces recurrence to the error behaviour. Because of having recurrence, there is e.g., a predictive end of an error free run and therefore, there is memory in the error process and the error gaps are non-geometrically distributed. In a mobile communication system without quality-based power control mechanism, there would for example be a high probability of having a very long error free run with non-predictive end in case of a very good channel.

In the static case, stationarity of the UMTS DCH error process can be assumed, as the statistics (e.g., of the gap and burstlengths) are constant over time due to call admission control and also due to the closed loop power control mechanism which adjusts the same SIR and also the same link error characteristics for different received power levels and therefore for different cells and propagation scenarios. On the other hand, in scenarios with movement, as we are moving from cell to cell, the results depend on the current status of network deployment. But, it is presented in this thesis that there are no significant differences even between different scenarios with movement and it can be shown that the statistics are also constant over one for this thesis

relevant time period. Therefore, stationarity can be assumed also for scenarios with movement.

Generally, the stochastic process of received power levels in a mobile communication network is a non ergodic process as means are not fixed when performing measurements in live networks under mobility conditions. On the other hand, the error process of the UMTS DCH may be seen as an ergodic process due to the fact that the closed loop power control mechanism (consisting of the inner loop power control and the quality-based outer loop power control mechanism) introduces recursiveness to the error process. Together with noise and fading the process is also aperiodic which would support the assumption of having ergodicity.

2.2.6 Statistical Dependency between successive Gaps/Bursts

Having several error bursts separated by short gaps forming an error cluster and the occurrence of long gaps following error clusters means to have correlation not only between successive TBs but to have statistical dependencies also between neighbouring gaps and bursts as indicated in the schematic illustration in Fig. 2.17.

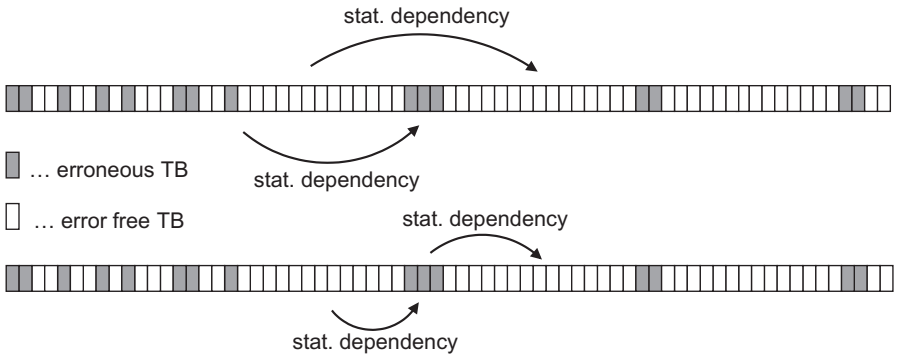

Figure 2.17: Schematic illustration of correlation between error gaps and bursts.

The validation of this conclusion is given in Fig. 2.18 where we see that the estimated probability of having short (≤ 12 TBs) and long (> 12 TBs) gaplengths does depend on previous gaplengths. Furthermore, we see from the autocorrelation function of the sequence of gaplengths (Fig. 2.18 right) that the memory in the channel outreaches two successive gaps. Similar results can be shown for bursts and also for the dependencies between gaps and bursts.

2.2.7 Block Error Ratio (BLER)

For many applications as well as for the analysis of codes another important measure of the link error characteristics is the statistics of the estimated error weight probability $P(m, n)$, the estimated probability of having exactly m errors occurring in a block of n (bits, packets or TBs

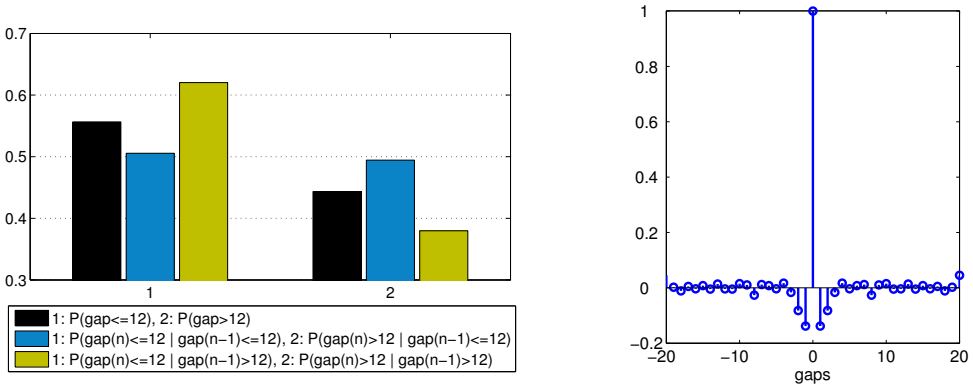

Figure 2.18: Statistical dependency between successive gaps (left) and autocorrelation of gaplengths (right), 384 kbit/s bearer, 'static 2' scenario, 'mobile 1'.

in our case). This error weight probability is also well-known as a performance indicator of the wireless link in mobile communication systems like the UMTS. There it is called BLER, which is the number of m erroneously received TBs out of n TBs. In our work we have selected $n = 2400$ TBs to meet the tradeoff between providing accurate results also for BLER $< 1\,\%$ and offering BLER values in fine time granularity. Moreover, $n = 2400$ TBs $= 200$ TTIs is adjusted for the performance measures in the considered UMTS networks for the 384 kbit/s bearer.

The measurements have been performed in all the scenarios with a duration of ≈ 1 h each. This leads to the fact that $\geq 3 \cdot 10^6$ TBs have been available in each of the measured traces. Thus, there have been considered ≥ 1000 blocks of $n = 2400$ TBs for the analysis of the BLER.

Fig. 2.19 shows measurement results for the 'static' scenario and also for scenarios with movement ('tramway' and 'small-scale movements'), measured with 'mobile 2' and 'mobile 3' and with all the available bearers for the DCH in DL. We can see that in the 'static' scenario the TPC of the UMTS is capable of keeping the BLER around the target value of 1 %, while in the scenarios with movement the radio link with a very high probability is experiencing high BLER due to fading effects, insufficient precision and lack of speed of the TPC algorithm. Again we notice the similarity of the results for the scenarios with movement regardless of which kind of movement. E.g. the small-scale movements performed by a person sitting at a table and tilting and turning the mobile station with his hands are producing the same link error characteristics as when going by tramway. The empirical CDFs of the BLER in Fig. 2.19 furthermore support the conclusion for a high ratio of having all TBs within one TTI erroneously received in case of an error (error bursts), visible through the steps in the CDFs.

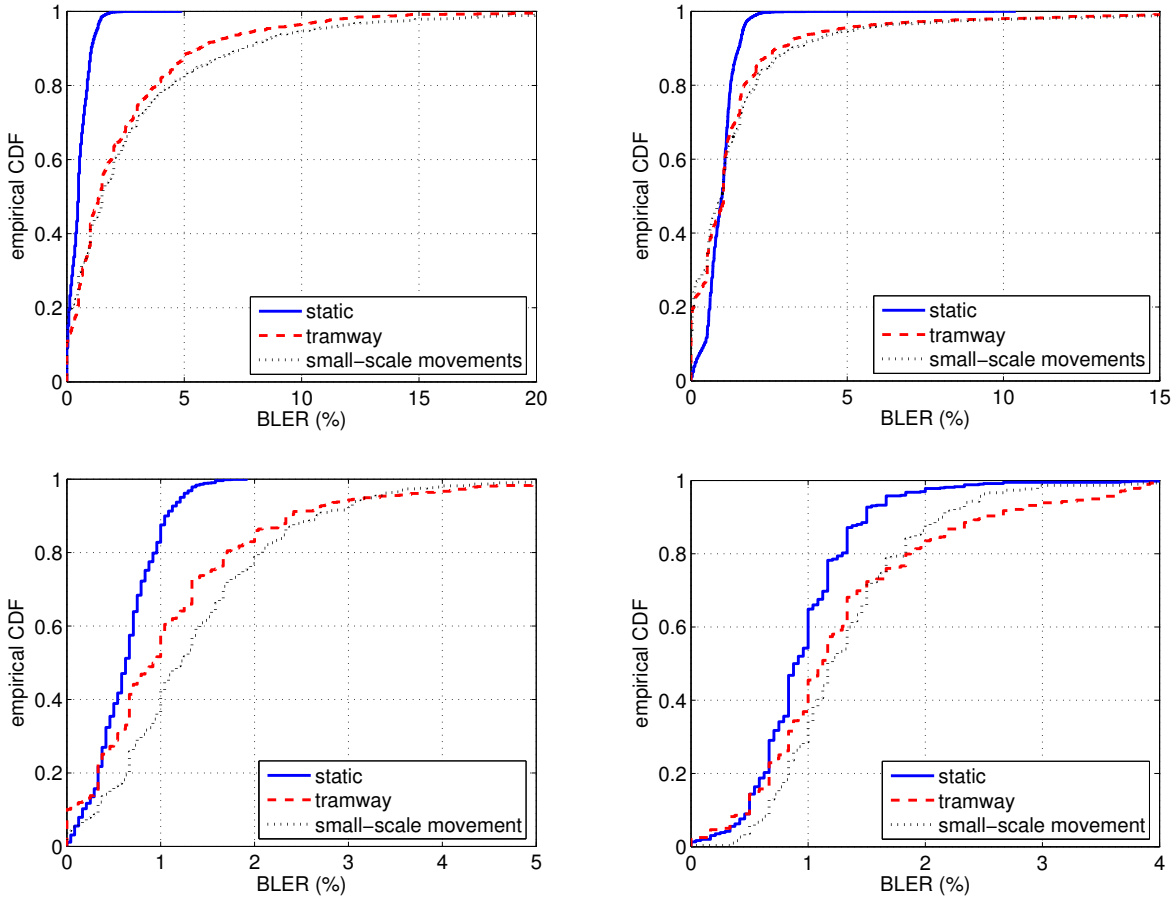

Figure 2.19: Comparison of DL BLER (%) for the 384 kbit/s bearer measured with 'mobile 2' (upper left) and 'mobile 3' (upper right) and for the 128 kbit/s (lower left) and 64 kbit/s (lower right) bearers measured with 'mobile 2'.

Chapter 3

Modeling of Link Layer Error Characteristics

Contents

Nowadays, simulations have become an important tool for the analysis of communication systems. In order to keep the computing effort within a manageable dimension, different abstraction levels of the system are introduced in the simulator, where models are representing lower layer functionalities and characteristics. With increasing complexity in coding, protocols and applications, increasingly large amounts of processed data are required for reaching sufficient statistics in the analysis. Thus, simulations with stochastic link models are becoming necessary instead of e.g. using simulations with measured traces.

This work focuses on link error models for the radio access part of the UMTS network as it represents the most relevant error source in the communication chain. The presented models are situated in the link layer of UTRAN and thus incorporate complete physical layer functionality and characteristics like propagation effects, interference, power control, transmitter and receiver properties, channel coding and interleaving. On the other hand this modeling approach offers the possibility of performance evaluation of link layer algorithms such as retransmission mechanisms and scheduling as well as the analysis of all higher layer protocols and applications.

3.1 Modeling Erroneous Channels — A Literature Survey

The simplest way of modeling an erroneous channel is a communication link where errors occur with probability p. Considering the transmission of bits, this channel is called BSC (Binary Symmetric Channel) [71, 72] as in case of an error the '0' bit is changed to '1' and vice versa (see Fig. 3.1). This channel model is frequently used in coding and information theory as it is one of the simplest channels to analyze.

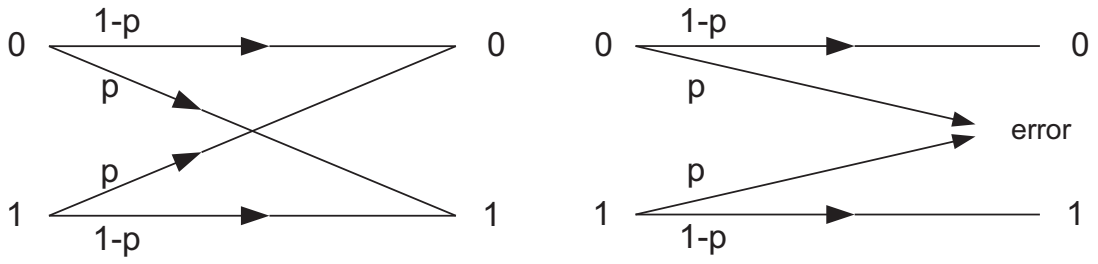

Figure 3.1: BSC (Binary Symmetric Channel) and BEC (Binary Erasure Channel).

Similarly to the BSC, also the BEC (Binary Erasure Channel) [73] is often used in information theory analyses. In the BEC, instead of 'flipping' the bits like in the BSC, the transmitted bit either is received correctly or an 'erasure' message is sent to the receiver in case of an error.

The corresponding memoryless error source model can be illustrated as shown in Fig. 3.2 comprising one state only, in which either an erroneous unit (1) with probability $P(1) = p$ or an error free unit (0) with probability $P(0) = 1 - p$ is produced at each discrete time instant. Of course, the probability to return to the state is one.

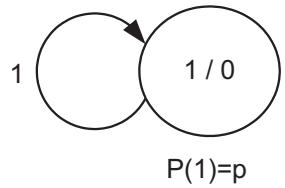

Figure 3.2: Schematic illustration of the memoryless error source model.

Like it is the case when comparing the BSC with the BEC, the handling of the error information from the error source model is different and depends on the application e.g. whether bit 'flipping' is applied in the channel or packets are marked as erroneous or discarded. Thus, in the following in this work, different error source models are presented without discussing the error handling in detail.

Beginning in the 1960s with the work of Gilbert [13], researchers understood that a memoryless BSC or BEC is not capable of describing the bursty nature of real communication channels. Gilbert introduced an error source model where the error probability of the current unit depends on the error state of the last unit — meaning there is correlation between two successive error events and thus there is memory and burstiness in the channel. Gilbert's model is a first-order two-state Markov model with one 'good' and one 'bad' state, where no errors occur in the 'good' state but in the 'bad' state the error probability takes some value > 0. In Fig. 3.3 Gilbert's model is presented.

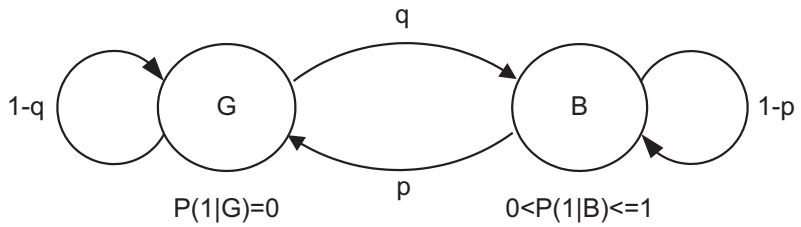

Figure 3.3: Schematic illustration of Gilbert's model.

The probability to be in the bad state is $P(B) = \frac{q}{p+q}$ and thus the resulting total error probability is $\bar{P}_e = P(1|B) \cdot \frac{q}{p+q}$. Gilbert's model is a renewal model, meaning the current

burstlength or gaplength is independent of the previous burstlength or gaplength, and it is a HMM (Hidden Markov Model) as the current state of the model (G or B) cannot be determined from the model output. Furthermore, the model parameters (p, q, $P(1|B)$) are not directly observable from training data and thus have to be estimated via trigram statistics or curve-fitting as proposed in [13].

Due to the straightforwardness of the parameter estimation, a simplified version of Gilbert's model has often been used in subsequent studies, where the error probability in the 'bad' state is one. Then the model changes from an HMM (Hidden Markov Model) to a first-order two-state Markov chain. Thus, the two parameters (p,q) for this 'simplified Gilbert' model can be calculated directly from the measured error trace by using the mean error burstlength ($p = 1/\bar{L}_{burst}$) and the mean gaplength ($q = 1/\bar{L}_{gap}$) or the total error probability ($q = \frac{\bar{P}_e}{\bar{L}_{burst} \cdot (1-\bar{P}_e)}$).

Enhancements to Gilbert's model can be found in the work of Elliott [14] where errors can occur also in the 'good' state as shown in Fig. 3.4.

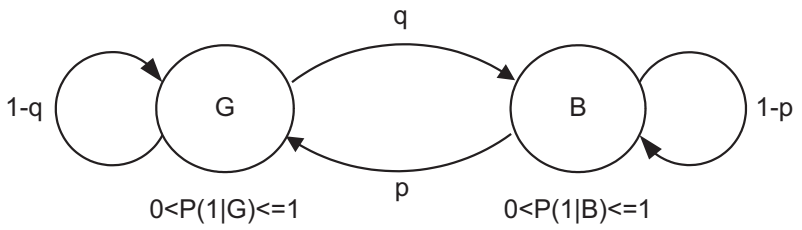

Figure 3.4: Schematic illustration of the Gilbert-Elliott model.

This model, also known as GEC (Gilbert-Elliott Channel), overcomes the limitation of Gilbert's model in having geometric distributions of burstlengths only and beside being an HMM it is also non-renewal, that is for example, the current burstlength is not statistically independent of the previous burstlength. This of course brings opportunities for channel modeling but also complicates the parameter estimation. The parameters for non-renewal HMMs like the GEC have to be estimated as for example using the Baum-Welsh algorithm [15].

Later on, also in the 1960s, J.M. Berger [16], S.M. Sussman [17], E.O. Elliott [18] and B. Mandelbrot [19] proposed to use renewal processes to model the error characteristics of communication links with the suggestion in [16] to use independent Pareto ($f(t|a) = a \cdot t^{-a}$, $t \geq 1$, $0 < a < 1$) distributions for the intervals between successive errors.

Further enhancements to Gilbert's model were published by Fritchman [20], proposing partitioned Markov chains with several error-free and error states as shown in Fig. 3.5. With the restriction of forbidden transitions between error states and also between error-free states, the model parameters can be estimated via separate fitting of polygeometric distributions to the

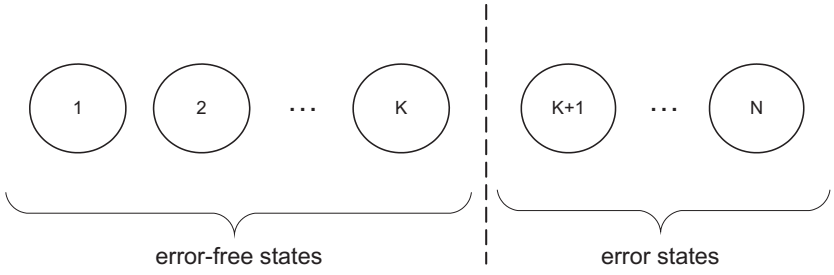

Figure 3.5: Partitioned Markov chain.

CCDF of gaplengths and to the CCDF of burstlengths. The polygeometric CDFs are given by

$$F(x|K)_{\text{polygeometric,gaplength}} = 1 - \sum_{i=1}^{K} \mu_i \lambda_i^x, \tag{3.1}$$

$$F(x|K,N)_{\text{polygeometric,burstlength}} = 1 - \sum_{i=K+1}^{N} \mu_i \lambda_i^x, \tag{3.2}$$

with the constraints of $0 < \mu_i < 1$ and $0 < \lambda_i < 1$ (μ_i and λ_i correspond to the probabilities of transition to the state and within the state, respectively). K is the number of error-free states and N is the total number of error states.

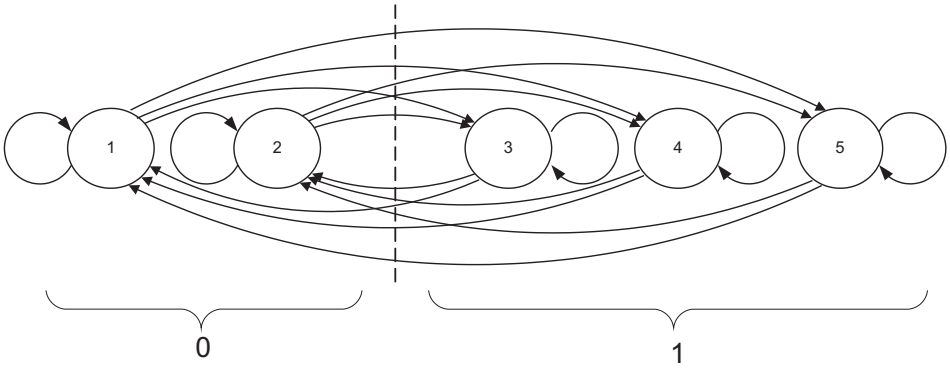

Figure 3.6: Partitioned Markov chain with two error-free states and three error states.

Using this model configuration, as presented in Fig. 3.6 for two error-free states and three error states, there is still statistical dependency between the current gaplength and the previous burstlength and vice-versa, and also between the current gaplength (burstlength) and the previous gaplength (burstlength). Therefore, for a full definition of the model also these dependencies have to be considered. However, with the additional restriction of having constant

proportions of allowed transition probabilities from one state to the successive states the model becomes renewal. For example in case of the 2/3 configuration of Fig. 3.6 this would mean $p_{13} : p_{14} : p_{15} = p_{23} : p_{24} : p_{25}$ and $p_{31} : p_{32} = p_{41} : p_{42} = p_{51} : p_{52}$.

The well-known 'Fritchman' model is a special case of the partitioned Markov chains and has one error state only as presented in Fig. 3.7. In this configuration the error-free run (gaplength) distribution uniquely specifies the model and the model parameters can be found via curve-fitting [20].

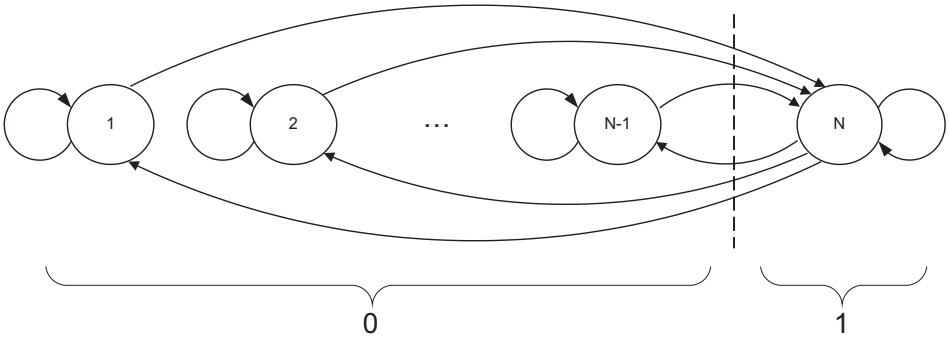

Figure 3.7: The well-known 'Fritchman' model configuration.

Note, each state of the 'Fritchman' model represents a memoryless error model and therefore, Fritchman's model is limited to polygeometric distributions of the gaplengths as well as to burstlengths according to a single geometric distribution.

Table 3.1 presents a comparison of the main properties of the above mentioned 'well-known' and frequently used classical models: the two-state Markov chain ('simplified Gilbert' model), the 'Gilbert' model, the GEC and the 'Fritchman' model.

Model	**Properties**
Two-state Markov chain ('simplified Gilbert' model)	renewal, geometric distributed burst- and gaplengths
'Gilbert' model	renewal, HMM, geom. burstlengths, approx. polygeom. gaplengths
GEC (Gilbert-Elliott Channel)	non-renewal, HMM, approx. polygeom. burst- and gaplengths
'Fritchman' model	renewal, HMM, geom. burstlengths, polygeom. gaplengths

Table 3.1: Comparison of frequently used classical models.

These previous models (see [21] for a survey of proposed models until 1978) have attracted a great deal of attention from researchers, who have used the models or tried to improve the models' characteristics and usability. In [22] and [23] methods for the parameter estimation of Gilbert's and Fritchman's models are presented.

Further models: The limitation of the 'Fritchman' model to polygeometric distributions led J.A. Adoul, B.D. Fritchman and L.N. Kanal to the proposal of slowly spreading Markov chains for error modeling in [24]. In [25], [26], [27] and [28] HMMs in general are addressed. But also new models to describe special error characteristics of (wireless) communication links have been developed up to the present, including a three-state model for Poisson distributed burstlengths [29], a two-state model with segmented exponential distributed burst- and gaplengths [30], two- and three-state Markov models [31] and an FSMC (Finite State Markov Channel) [32]. Other recently published bipartite models [33] are similar to Fritchman's model but they partition the burst- and gaplength statistics in subintervals. A bit error model for IEEE 802.11b using chaotic maps is presented in [34]. In [35] a two-state process is used in which the gaps and bursts are generated from measured distributions. An RLM (Run Length Model) is presented in [36], generating the error trace with a two-state renewal process with a mixture of geometric distributions which then is shown to be equal to a four-state Markov model. A three-state RLM is presented in [37]. Another RLM which uses half-normal distributions for burst- and gaplengths is shown in [38] and aggregated Markov processes are presented in [39], both for modeling the packet errors in DVB-H (Digital Video Broadcasting - Handheld).

3.2 Link Error Models for the UMTS DCH

Recent developments show that in most of the cases new models have been proposed to describe special properties of the links instead of taking general HMMs. The reason is that just in contrast to general models which need a high number of parameters and complex parameter estimation methods, the new models are required to offer a good usability by a small number of parameters which can easily be determined and adapted to changing system configurations. Furthermore, the goal in designing a new model is to represent the specific details of the measured error characteristics (e.g. the error predictability) with high accuracy, whereas a small computational effort to generate an error trace has become less significant. Often, (as it is the case in this work) the specific link error properties observed in the measurement-based analysis already lead to a particular approach for designing a link error model. The well known modeling approaches are considered in this work (like the famous 'Gilbert' model, the GEC or Fritchman's partitioned Markov chains), but it will be demonstrated in the following that these models are not appropriate for describing the error statistics of the UMTS DCH.

As the link error characteristics of the UMTS DCH may change when adjusting different parameters in the UTRAN (like different TF (Transport Format), TTI-length, SF, and so on) the goal is to present a universal UMTS DCH modeling approach resulting from the specific link error characteristics of the UMTS DCH seen in the analysis in Chapter 2. In this document the focus is particularly on the characteristics of the 384 kbit/s UMTS DCH PS bearer in DL which currently represents the DCH PS bearer with the highest throughput available in the considered networks.

Despite the fact that the analysis in this document is performed for the UMTS DCH, the basic idea behind the modeling approach, meaning to consider the specific properties of the underlying network, may as well be applied for other new or future mobile communication systems.

When modeling the UMTS DCH link error characteristics, only the two cases 'static' and 'dynamic' have to be considered, as already observed. Due to the fact that the ratio of having all TBs erroneously received within one TTI (given this TTI is erroneous) in the scenarios with movement is very high (≈ 0.9), we propose to model the link error characteristics based on TTI granularity in the 'dynamic' case. On the other hand, in the 'static' scenario, the measured ratio of having all TBs erroneously received within one erroneous TTI is just around 25 %. Furthermore, the ratio of having only one erroneous TB within an erroneous TTI is also 25 %. Therefore, in the 'static' case, modeling in TB granularity is necessary.

3.2.1 Link Error Modeling — 'Dynamic' Case

We have shown the existence of error bursts (up to ten TTIs long) and therefore the simplest way of modeling the link errors via a memoryless BSC is not possible. One slightly more sophisticated modeling approach is to extend the single-state error model of the memoryless BSC (Bernoulli trials) to a two-state Markov chain, where in one state error-free TTIs are generated and in the other state erroneous TTIs are generated, as in Gilbert's model [13] with an error probability of one in the bad state ('simplified Gilbert' model). This two-state Markov chain is uniquely specified by two parameters, namely the packet loss rate and the average TTI-burstlength (or the average TTI-gaplength) which can be directly determined from the measured error trace as the model is not an HMM.

As shown in Fig. 3.8(right), the geometric distribution of TTI-burstlengths, as generated by this simplified Gilbert model (two-state Markov chain), fits quite well to the measured empirical CDF of the lengths of TTI bursts (here shown for the 'car-city' scenario). On the other hand, when comparing the simulated statistics (geometric distribution) of TTI-gaplengths to the measured statistics (Fig. 3.8 left), it can be observed that the two-state Markov chain is not capable of representing the correct TTI-gaplength distribution.

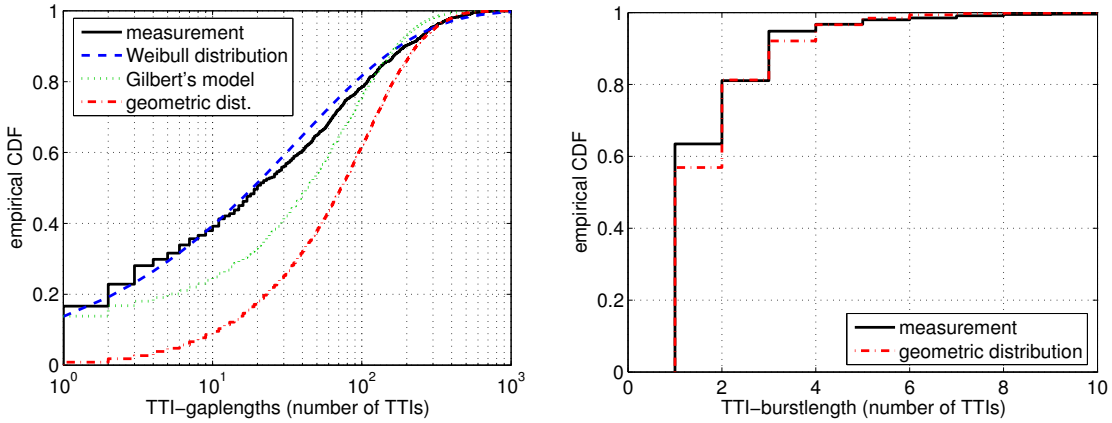

Figure 3.8: Comparison of empirical CDFs of TTI-gaplengths and TTI-burstlengths; measurements vs. two-state Markov chain, Gilbert's model and Weibull distribution; 384 kbit/s bearer, 'car-city' scenario, 'mobile 1'.

Enhancing the model from the simplified version of Gilbert's model (two-state Markov chain) to Gilbert's hidden Markov model with three parameters ($p = 0.384$, $q = 0.00139$, and $P(1|B) = 0.756$ for the 'car-city' scenario, estimated by using trigram-statistics as proposed in [13]) shows improvements but it does also not bring sufficient accuracy. Further improvements within the same modeling strategy could be reached via Fritchman's partitioned Markov chains with more than three states ($\gg 4$ parameters). As we want to keep the model parameters as small as possible and because a Weibull distribution perfectly fits the measured TTI-gaplength statistics (see Fig. 3.8), a change in the strategy towards other renewal processes like semi-Markov models or RLMs seems appropriate. The two-parameter Weibull [74] PDF and CDF are given by

$$f(x|a,b) = ba^{-b}x^{b-1}e^{-(\frac{x}{a})^b}, \tag{3.3}$$

$$F(x|a,b) = 1 - e^{-(\frac{x}{a})^b}, \tag{3.4}$$

where a and b are scale and shape parameters, respectively.

The Weibull distribution was selected due to the following reasons. Waloddi Weibull proposed this distribution in 1939 for modeling the breaking strengths of materials. Current usage also includes reliability and lifetime modeling. When modeling the breaking strengths of materials, the model has in a way take into account that always the weakest link in the chain is breaking first. This idea can analogically be used when modeling the link error characteristics. Also in the wireless channels the error occurs at the time instant where the smallest SIR value is seen at the receiver. Especially, when having a quality-based closed loop power control algo-

rithm like the one for the UMTS DCH. The TPC algorithm continuously reduces the transmit power level until the weakest point breaks meaning there is an error in the transmission.

Moreover, the (two parameter) Weibull distribution is very flexible in a way that extremely different distributions can be modeled when varying the parameters. For example, shape parameter equal to one results in an exponential distribution, shape parameter equal to two gives a Rayleigh distribution and shape parameter equal to 3 generates a distribution similar to a Normal distribution (Gauss). This flexibility is also very welcome when modeling the link error characteristics.

For the estimation of the parameters for the Weibull distributions in this work, the following methods have been considered: Weibull plots, least squares parameter estimation, maximum likelihood parameter estimation and manual-visual curve-fitting.

To generate the TTI-gaplengths via a Weibull distribution and the TTI-burstlengths in a discrete sense with a Markov chain (geometric distribution), a mixed semi-Markov/Markov model (M-SMMM) with two states is proposed (see Fig. 3.9).

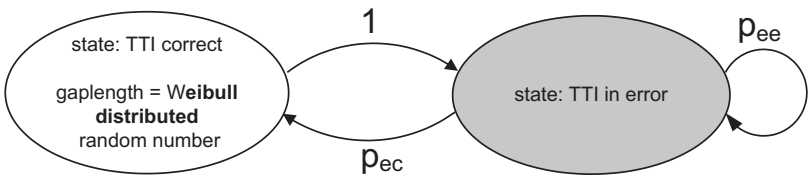

Figure 3.9: Mixed semi-Markov/Markov model (M-SMMM) for the UMTS DCH 'dynamic' case.

The state in which the error-free TTIs are generated is formed according to a renewal process with Weibull-distributed gaplengths, whereas erroneous TTIs are produced one by one in the discrete state. After generating an error gap, a single erroneous TTI is generated which may again be followed by a state change according to the transition probability $p_{ec} = 1 - p_{ee}$. Therefore, three parameters are required for defining the model: two parameters for the Weibull distribution (scale parameter $a = 43.4$, shape parameter $b = 0.593$) and $p_{ee} = 0.43$ (values for the 'car-city' scenario, 'mobile 1'), which are determined by curvefitting to the TTI-gaplength statistics and direct calculation from the mean TTI-burstlength, respectively.

The ratio of receiving all TBs in error within one erroneous TTI in the 'dynamic' scenario is around 85 % (see Fig. 2.9. Therefore, in the remaining 15 %, refinements of the model are possible by considering the number of TB errors within the erroneous TTIs via the measured ratio of having x TBs erroneously received within one TTI [10]. In Fig. 2.9 the additionally required eleven parameters are presented. Instead we could make use of the fact that these values follow a binomial distribution (only one additional parameter required) as shown in Fig. 2.10.

3.2.2 Link Error Modeling — 'Static' Case

For modeling the link error characteristics of the UMTS DCH in the 'static' scenario, a model based on TTI granularity is not sufficient due to the high probability of receiving less than all TBs in error within one erroneous TTI. Thus, the goal of a first modeling approach is to generate a sequence of TBs with the correct statistics of gaplengths and burstlengths.

In cases of the memoryless channel and the 'simplified Gilbert' model (classical two-state Markov chain), the produced gaplengths follow a geometric distribution [71]. As it was shown in Section 2.2 (see Fig. 2.13), there are two main areas with occurrence of gaplengths. Obviously, these statistics of gaplengths cannot be met via a single geometric distribution; therefore, the memoryless channel and the 'simplified Gilbert' model are not adequate tools in this case.

When using Fritchman's partitioned Markov chains with two error-free states (four parameters) and following the proposal of [22] (estimating the model parameters by separate curve fitting for different parts of the distribution function), the measured distribution of gaplengths is represented as demonstrated in Fig. 3.10.

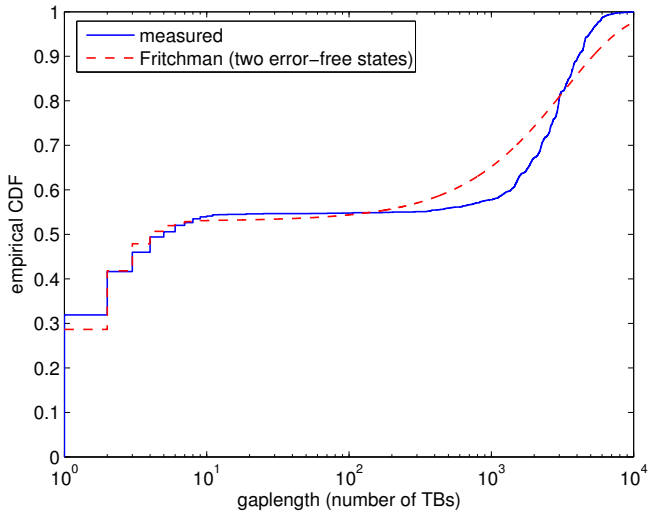

Figure 3.10: Gaplength distribution, measured vs. Fritchman's model (two error-free states), 384 kbit/s bearer, 'static' scenario, 'mobile 1'.

One of the error-free states is responsible for the distribution of the short gaps whereas the second state builds the form of the distribution for the long gaps. It can be shown that the fitting for the short gaps can be improved by adding additional error-free states to the model, but further states do not add accuracy to the fitting of the distribution for the long gaps. The reason can be seen in Fig. 3.11, where the empirical CDFs of the measured data and a CDF of

a geometric distribution are shown in linear scale. Note that in Fig. 3.11 only the CDFs of the long gaplengths (> 12 TBs) are plotted.

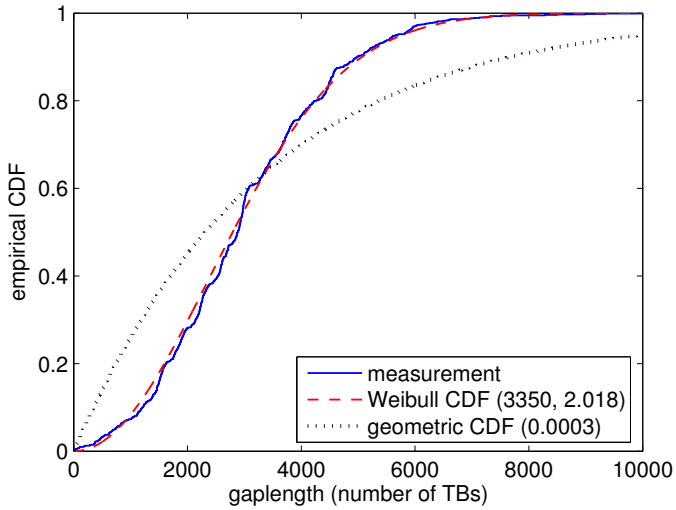

Figure 3.11: Distribution of long gaps, measured data vs. Weibull and geometric distribution, 384 kbit/s bearer, 'static' scenario, 'mobile 1'.

The curve of the measured data is convex between zero and the inflection point (mode). The geometric distribution on the other hand is concave all over its support. Therefore, fitting a geometric distribution (as formed by one state of Fritchman's partitioned Markov chains) to the measured distribution is not possible. Furthermore, it becomes clear that the best fit is already reached with only one geometric distribution (meaning only one error-free state in Fritchman's model for the long gaps). Fritchman's model offers the possibility of having more than one error-free state and thus to have polygeometric distributions for the gaplengths, but every additional geometric distribution in a linear combination with only positive weighting factors adds concavity to the curve, and thus, increases the fitting error. Therefore, all models producing polygeometric distributions of gaplengths like Gilbert's model, Fritchman's model (with any number of states N) or the GEC (Gilbert-Elliott-Channel) are not useful for modeling UMTS DCH link error characteristics in 'static' scenarios.

Moreover, as the best fit of Fritchman's model to the measured distribution of the long gaplengths is reached with one error state and due to the memoryless property (constant failure rate) of the geometric distribution, Fritchman's model is also incapable of describing the predictive nature of the measured link error characteristics (see Chapter 5 for details). The geometric distributions of gaplengths provided by Gilbert's model and by the GEC lead to the same conclusions.

As we can see in Fig. 3.11, a Weibull distribution with scale parameter $a = 3350$ and shape parameter $b = 2.018$ perfectly fits the measured distribution of the long gaps. Furthermore, a Weibull distribution ($a = 1.2$, $b = 0.7$) can be used to describe the distribution of the short gaps as shown in Fig. 3.12 (left) and another Weibull distribution ($a = 1.1$, $b = 0.55$) with two additional steps at 12 and 24 TBs represents the statistics of burstlengths properly for the 384 kbit/s bearer (see Fig. 3.12 (right)).

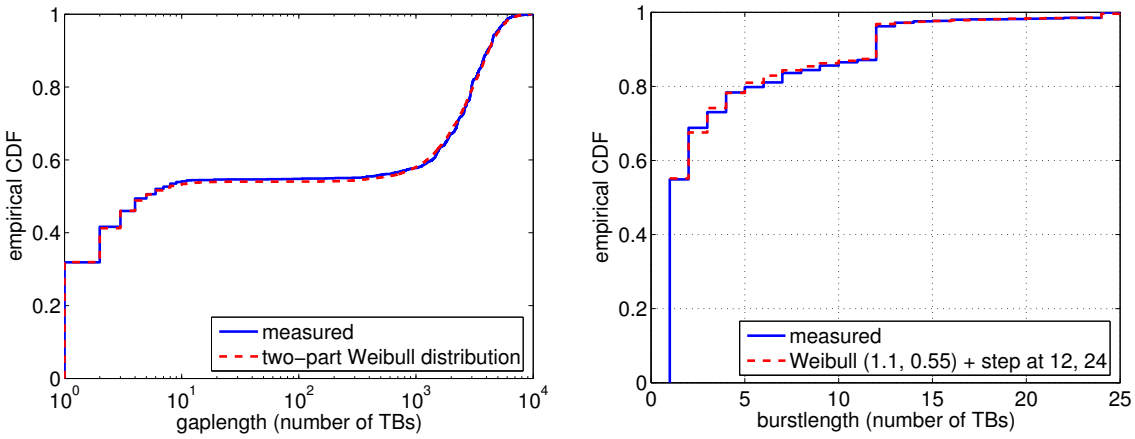

Figure 3.12: Comparison of empirical CDFs of gap- and burstlengths; measurements vs. two-state Weibull renewal process; 384 kbit/s bearer, 'static' scenario, 'mobile 1'.

Following the explained method of describing adequate statistics for gap- and burstlengths we arrive at a two-state model (two-state alternating Weibull renewal process) as shown in Fig. 3.13, where in one state correct TBs and in the other state erroneous TBs are generated. After each calculation of either burst- or gaplength, the state is changed.

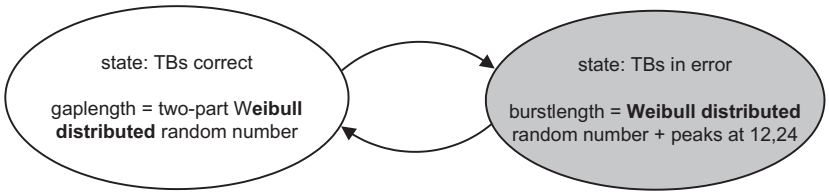

Figure 3.13: Two-state model for the UMTS DCH 'static' scenario — two-state alternating Weibull renewal process ('static-model 1').

In the correct state, the number of subsequent error-free TBs (gaplength) is calculated via a two-part Weibull distributed random number with a probability of $p_{sg} = 0.55$ to produce short gaps. The number of subsequent erroneous TBs (burstlength) in the error state is also

calculated via a Weibull distributed random number (with a probability p_{Wb}) but with additional burstlengths of 12 and 24 with estimated probabilities of $p_{12} = 0.09$ and $p_{24} = 0.01$, respectively. For bearers other than the 384 kbit/s bearer or for different TFs (Transport Formats) these peaks are found at other burstlengths (e.g., at 8 and 16 TBs for the 128 kbit/s bearer used in this work).

This two-state model can be equivalently presented with five states as shown in Fig. 3.14, which clearly shows that for modeling the error characteristics of the UMTS DCH for the 'static' scenario in this way, a total number of nine parameters is required. These include: four parameters for the two Weibull distributions of the correct states (error gaps) and one parameter for their separation. Additionally, two parameters are necessary to select burstlengths of either 12 or 24 or according to a Weibull distribution which also adds two parameters.

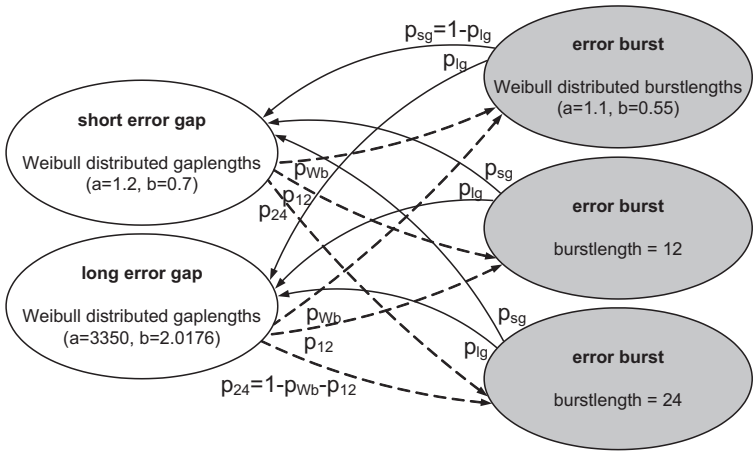

Figure 3.14: Equivalent five-state model illustration ('static-model 1').

From Fig. 3.14, it becomes clear that this model represents a renewal process, meaning that the probability e.g. of producing a short (long) gap does not depend on the length of the last error burst and is also independent of previous gaplengths. On the other hand — as shown in Chapter 2 — in the measured link error characteristics there is statistical dependency between subsequent gaps, subsequent bursts and also between a gap and the following burst. Therefore, in order to capture these correlation properties, a non-renewal model like an HMM is required.

Instead of taking a general HMM, again special error characteristics of the UMTS DCH are considered to keep the model as simple as possible (as few parameters as possible and simple parameter estimation). Nonetheless, the model captures a major part of the dependencies between gaps and bursts by utilizing conditional probabilities of having long and short gaps and the conditional probabilities for burstlengths of 12 and 24, as shown further on in this chapter.

Fig. 3.15 presents a non-renewal two-layer model, where dependency between gaps and bursts is introduced via the upper layer two-state Markov chain which modulates the parame-

ters of the lower layer process. This model represents an HMM, since from the model output (erroneous/error-free TB) the current state of the model cannot be determined.

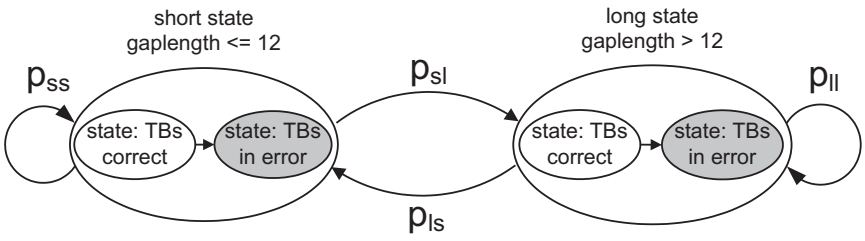

Figure 3.15: Non-renewal two-layer Markov model — Markov modulated Weibull renewal process ('static-model 2').

The model has one 'short state' where gaps with $\leq$ 12 TBs are generated and one 'long state' for gaps with $>$ 12 TBs. The corresponding transition probabilities are the estimated conditional probabilities out of Chapter 2 (according to Fig. 2.18). After entering a certain state, first the gaplength is calculated via a Weibull distributed random number for either the small or the large gaplengths according to the separation in Chapter 2 and depending on the current upper layer state. Then, while staying in the same state of the upper-layer Markov chain, a burst is calculated with different probabilities of either creating a Weibull-distributed burstlength or a burst with a length of 12 or 24, again, depending on the actual upper layer state. After that, the state of the upper layer Markov chain possibly changes and the procedure begins again with calculation of the gaplength. Therefore, via the upper layer Markov chain, the model introduces dependency between subsequent gaps and also between a gap and the following burst. In Fig. 3.16, a detailed illustration of this two-layer model is presented whereof it becomes clear that the following 11 parameters are required for a complete definition of the model.

Two parameters are necessary for the determination of the transition probabilities of the upper layer Markov chain ($p_{ss} = 1 - p_{sl} = 0.53$, $p_{ls} = 1 - p_{ll} = 0.56$). Two additional parameters are used for the Weibull distributions of each of the correct states and only two parameters more for the Weibull distribution of both erroneous states as the Weibull distributions are the same for the erroneous states of both upper layer states. The probability of generating Weibull distributed burstlengths depends on the current state of the upper layer Markov chain in case of the erroneous states. For these probabilities ($p_{Wb|s} = 0.99$, $p_{Wb|l} = 0.78$) and the separation between a burstlength of 12 and 24 ($p_{12|x}/p_{24|x} = 9$) three additional parameters are required ($p_{Wb|x} + p_{12|x} + p_{24|x} = 1$).

Despite the fact that this model is of type HMM, all the parameters can directly be determined from a measured link error trace without the necessity of difficult parameter estimation

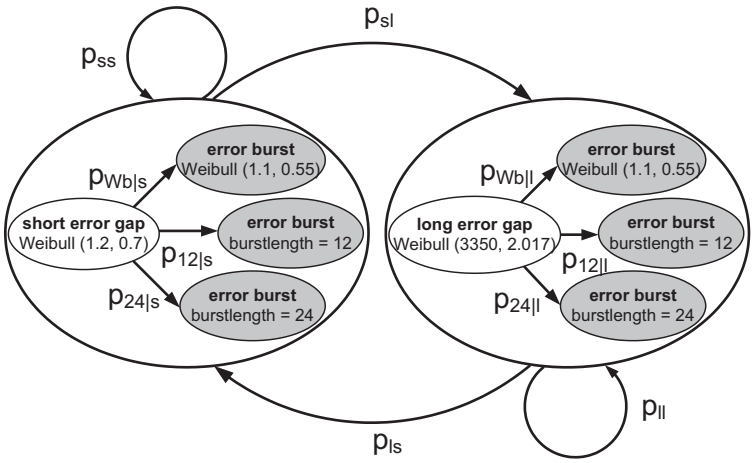

Figure 3.16: Detailed illustration of the non-renewal two-layer Markov model ('static-model 2').

algorithms.

The two-state Weibull renewal process ('static model 1') and the non-renewal two-layer Markov model ('static model 2') produce correct estimated total TB error probabilities of 0.0027 and 0.0026, respectively. Also the mean clusterlengths (10.50, 10.35), the mean numbers of erroneous TBs per error cluster (7.90, 7.83) and the mean numbers of error bursts per cluster (2.17, 2.18) are very close to the measured results for the scenario 'static 1'. Furthermore, as presented in Fig. 3.17, both models are capable of generating an error trace with an appropriate BLER. However, there is just little improvement from the non-renewal two-layer Markov model over the two-state Weibull renewal process ('static model 1') for the case of the 'static 1' scenario.

The potential of introducing correlation between subsequent gaps and bursts of the non-renewal two-layer Markov model becomes evident when analysing the autocorrelation function for example of the sequence of gaps (see Fig. 3.18) for the scenario 'static 2' which includes a higher grade of dependency between the gaps. While the two-state Weibull renewal process does not produce any dependency between successive gaps we observe in Fig. 3.18 that the non-renewal two-layer Markov model captures the correlation properties between neighbouring gaps.

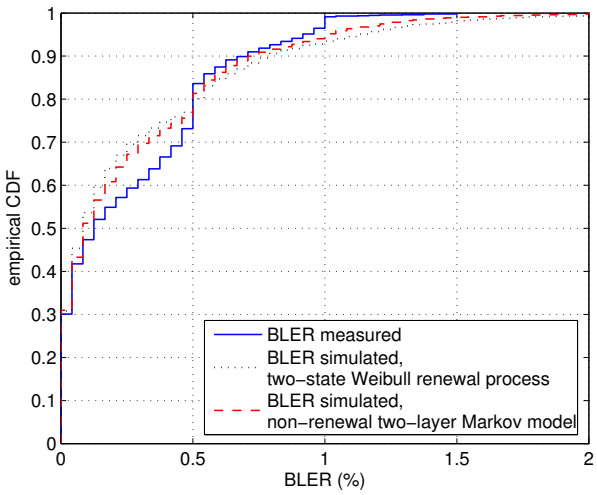

Figure 3.17: Comparison of BLERs — measurement vs. models, 'static 1', 'mobile 1', 384 kbit/s bearer.

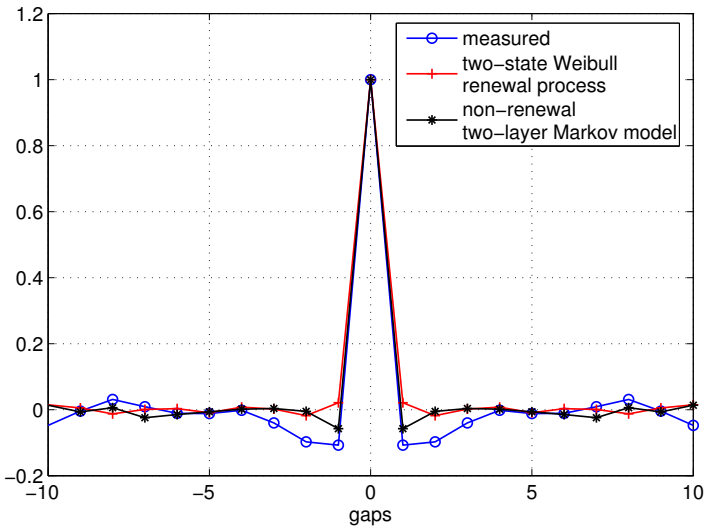

Figure 3.18: Comparison of autocorrelation functions — measurement vs. models, 'static 2', 'mobile 1', 384 kbit/s bearer.

3.3 Impact of Channel Modeling on the Quality of Services

QoS, as well as the performance of methods and algorithms within wireless communication systems are highly affected by the error characteristics of the underlying channels. Especially, systems including cross-layer algorithms which are introducing network awareness and therefore are considering or even exploiting [12, 43, 44, 46] specific properties of the link errors, are concerned. But also when utilizing a plain protocol stack there is a strong impact of the second-order error statistics of the channel onto the performance of the higher layer protocols as shown in [5].

As the error characteristics of the channel do have a great impact on the higher layer protocols, thus also different link error model properties show consequences. In this section the impact of UMTS DCH link error modeling directly on the quality of H.264/AVC streamed video [7] is investigated [4] by comparing the link- and network-layer characteristics as well as the resulting video quality simulated with link error models and with measured link layer error traces (384 kbit/s bearer, 'static' scenario, 'mobile 1').

3.3.1 Compared Models

In literature, various link error models are used for simulating video streaming over wireless networks. For example in [75], a memoryless error model was used for evaluating video streaming over UMTS DCH, motivated by simulation results in [76]. There, uncorrelated block errors are assumed for UMTS networks. However, this contradicts measured results presented in [9]. Others (e.g. [77]) are using the simplified Gilbert model [13] representing a two-state Markov chain. In many publications, the authors quote GEC, which after some assumptions for simplification also ends in the simplified Gilbert model (two-state Markov chain). Referring to these commonly used models, for the investigation a memoryless channel, the simplified Gilbert's model (two-state Markov chain) and a Markov modulated Weibull renewal process ('static model 2') have been used.

Although all of the models are producing the same estimated link error probability (0.266%) as measured in the live UMTS network, they are showing significant differences in their error characteristics. That can be observed in Figs. 3.19, where the CDFs of the gaplengths (number of error free link layer packets between two errors) and the CDFs of the burstlengths (number of subsequently received errors) are presented.

3.3.2 Experimental Setup

For transmission of a video stream over the UMTS network the following procedure of packetization has to be performed. Each frame of the video is first subdivided into smaller parts

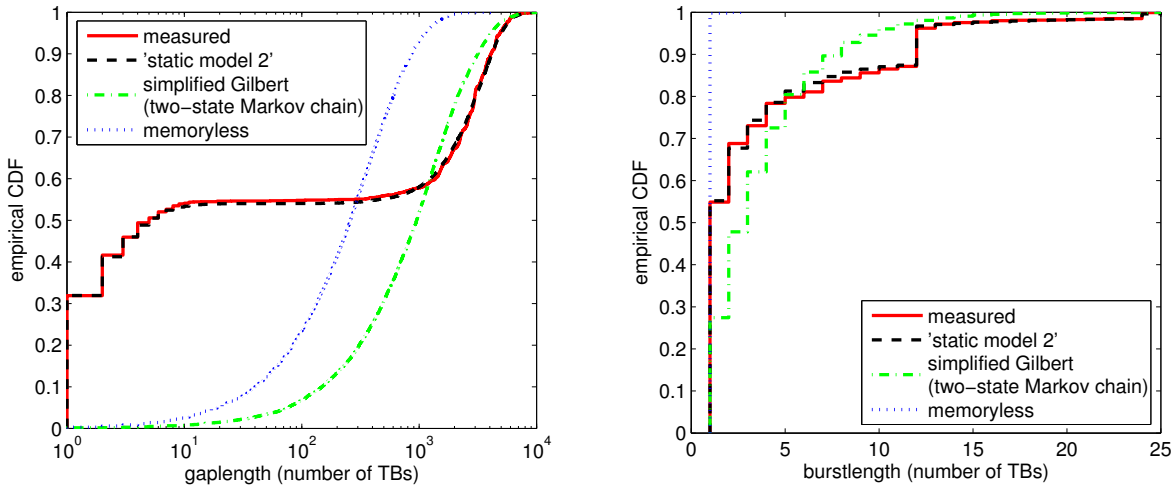

Figure 3.19: Comparison of empirical CDFs of link layer gap- and burstlengths; 'memoryless' model, 'simplified Gilbert' model, 'static model 2' vs. measured trace, 384 kbit/s bearer, 'static' scenario, 'mobile 1'.

(slices) which then are encoded. As shown in the schematic illustration in Fig. 3.20, the encoded video slices or equivalently the NALUs (Network Abstraction Layer Units) containing the slice data are encapsulated into RTP (Real Time Protocol) packets with a header of 12 bytes. Each RTP packet then is transmitted within the UDP protocol which adds a header of 8 bytes. If no segmentation is needed, the UDP packets are further packed within IP packets with a header size of 20 bytes for IPv4 and 40 bytes for IPv6. These IP packets are then transmitted over the UTRAN, where the layer three data packets are further processed as described in Section 1.3.2.

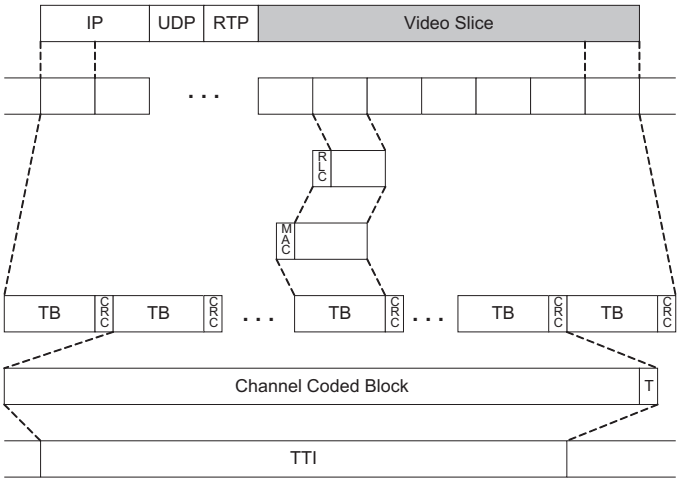

Figure 3.20: Schematic illustration of a packetization example for the transmission of a video slice over UMTS.

In Fig. 3.21 the experimental setup can be seen. The Joint Model (JM) H.264/AVC [78] was adapted by adding the interface for the IP error traces and by implementing a simple error concealment scheme at the decoder (for I frames weighted averaging [79] was used and for P frames the corresponding location was copied from the previous frame). The encoder is modified to deliver the IP packet lengths for mapping the IP packets onto the link layer packet trace.

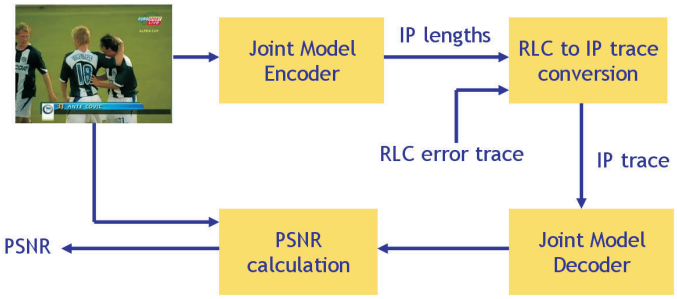

Figure 3.21: Scheme of the experimental setup for simulating H.264 encoded video over error prone links.

For the experiments a 'soccermatch' [80] video sequence was used with SIF (QVGA) picture resolution (320 × 240 pixel) and a frame rate of 10, containing a soccer match with different scenes. The sequence was encoded using I and P frames only (every 40th frame is an I frame) according to the baseline profile with slicing mode two and ≤750 bytes per slice. The quantization parameter was set to 26 and the rate-distortion optimization was disabled. To obtain reliable results, the video was decoded several times resulting in ≈10 hours of video stream.

The resulting packet sizes (sizes of the NALUs including NALU header of 1 byte) in the data stream of the video sequence, encoded with the mentioned parameters are presented in Fig. 3.22. For the resulting IP packet sizes, 40 bytes of headers (RTP 12 bytes, UDP 8 bytes, IPv4 20 bytes) have to be added.

To evaluate the improvements in end-to-end video quality, the peak signal-to-noise ratio of the luminance component (Y-PSNR) was used, given for the nth luminance frame $\mathbf{Y}_n$ by

$$\text{Y-PSNR}(n) = 10 \cdot \log_{10} \frac{255^2}{\text{MSE}(n)}, \tag{3.5}$$

$$\text{MSE}(n) = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} [\mathbf{Y}_n(i,j) - \mathbf{F}_n(i,j)]^2, \tag{3.6}$$

where $\text{MSE}(n)$ denotes the mean square error of the nth luminance frame $\mathbf{Y}_n$ compared to the luminance frame $\mathbf{F}_n$ of the reference sequence. The resolution of the frame is $N \times M$, indexes

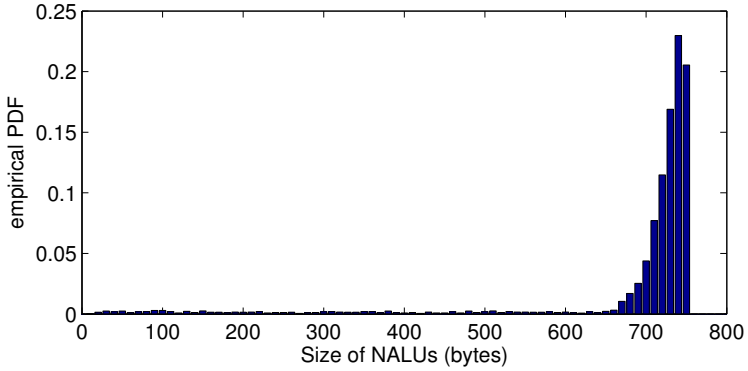

Figure 3.22: Empirical PDF of NALU sizes (size of encoded video packets).

i and j address particular luminance values within the frame. As a reference sequence we used the *non-compressed* original (non-degraded) sequence.

3.3.3 Simulation Results for H.264 Encoded Video over Error Prone Links

In Fig. 3.23 histograms of the Y-PSNR per video frame for the considered models are presented. It can be observed that despite having the same estimated link layer packet error probability[1] (see Table 3.2) only the 'static model 2' (non-renewal two-layer Markov model) meets the measured statistics.

Table 3.2: Comparison of estimated link-layer and IP packet error probability.

Model	$\overline{\mathbf{P}}_\mathbf{e}$**(TB) (%)**	$\overline{\mathbf{P}}_\mathbf{e}$**(IP) (%)**
measured	0.266	0.888
'static model 2'	0.262	0.892
simplified Gilbert	0.266	1.009
memoryless	0.266	4.501

This is due to the fact that only the 'static model 2' shows almost the same link error characteristics as the measured traces. All the other models and especially the memoryless

[1] As the memoryless channel and the simplified Gilbert model are defined by the mean error probability (and the mean error burstlength) they show exactly the same estimated mean error probability as the measured traces. The parameters of 'static model 2' are defined by curvefitting of the gap- and burstlengths and via the correlation between gaps and bursts. Thus, the resulting estimated link error probability differs slightly from the measured as a consequence of the modeling.

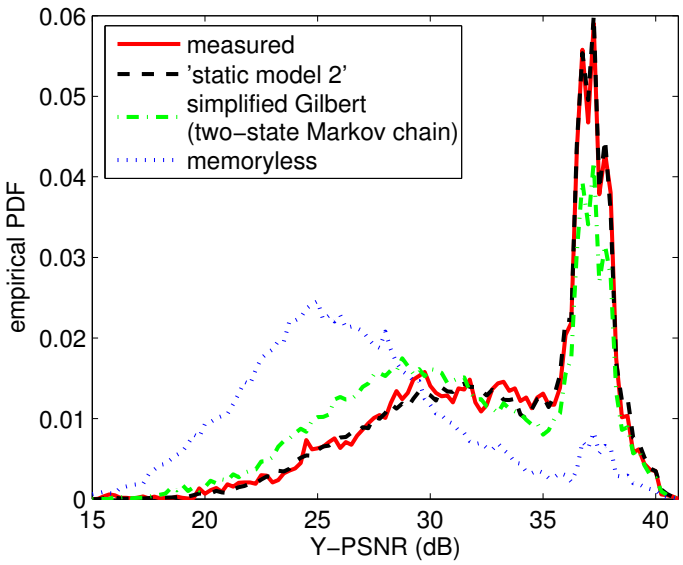

Figure 3.23: Y-PSNR per frame for the various link layer error models.

channel have completely different statistics of the burst- and gaplengths and thus result in a different number of erroneously received link layer packets (TBs — Transport Blocks) within one IP packet (see Fig. 3.24). This in turn results in a much higher IP error probability e.g. in case of the memoryless link layer characteristics as shown in Table 3.2. From these results we conclude that it is important to have the correct higher order statistics when modeling the link layer not only for video streaming simulations.

Contrary to the conclusions for the link layer modeling it will be shown in the following that for modeling of the errors in the network layer it is not important to meet the correct higher order statistics for evaluating the quality of streamed video with common parameter settings. In Fig. 3.25 we can see that by applying a memoryless channel model in the network layer with appropriate IP BLER (0.888 % instead of 4.501 % – see Table 3.2), the measured statistics of the streamed video quality (Y-PSNR per frame) is met with a memoryless model in the network layer, despite having completely different network layer error characteristics.

These differences in the network layer error characteristics can be seen in Fig. 3.26, where the PDFs of the IP-gaplengths (number of error-free IP packets between two errors) and the PDFs of the IP-burstlengths (number of subsequently erroneous IP packets) are compared. It can be observed in Fig. 3.26 (left) that the PDFs of the IP-gaplengths possess a maximum at around 150 in case of the measurements and the 'static model 2', whereas the memoryless network layer model (IP memoryless) shows the maximum at short IP-gaplengths. Furthermore, the memoryless network layer model only generates IP-burstlengths of length one, as can be seen

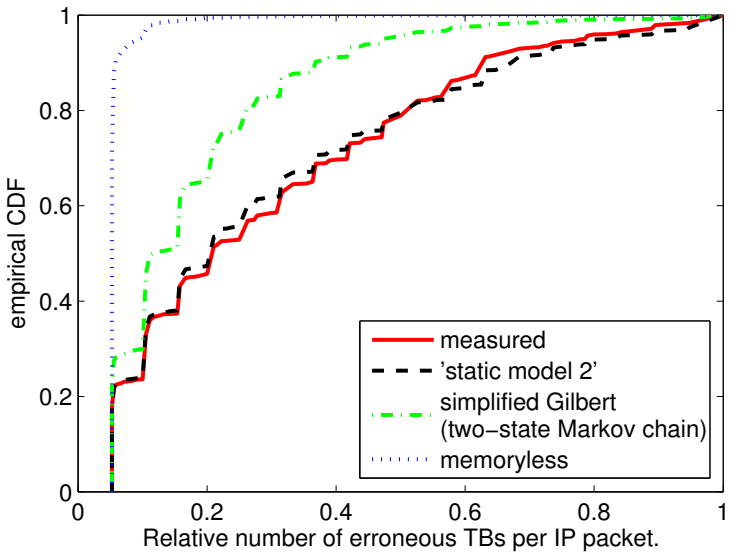

Figure 3.24: Relative number of erroneous TBs per (erroneous) IP packet.

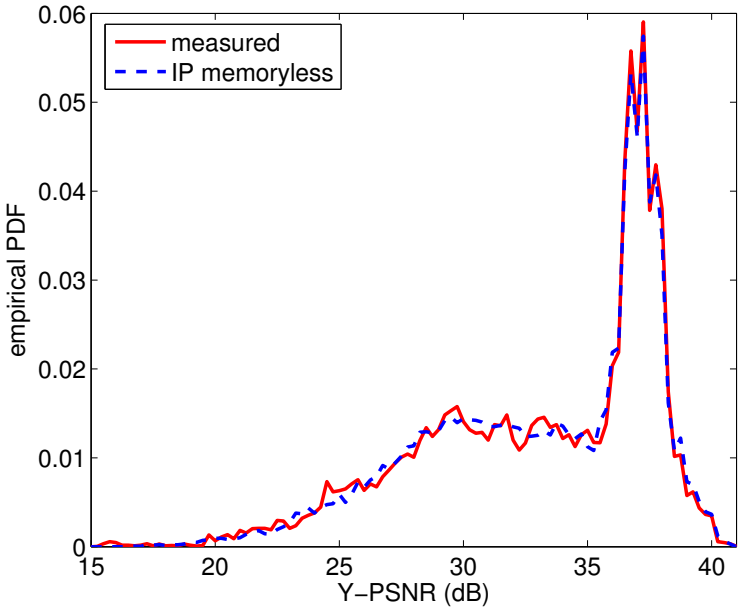

Figure 3.25: Y-PSNR per frame for IP memoryless model vs. measurements.

in Fig. 3.26 (right). On the other hand the measured statistics also show IP-burstlengths up to three subsequently erroneous IP packets.

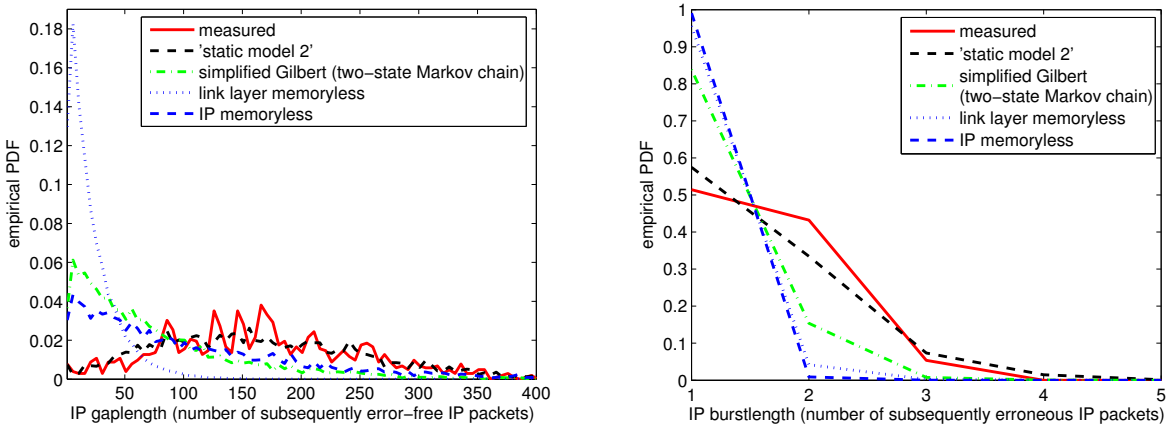

Figure 3.26: Comparison of IP gap- and burstlengths for the various error models.; 'memoryless' model, 'simplified Gilbert' model, 'static model 2' vs. measured trace, 384 kbit/s bearer, 'static' scenario, 'mobile 1'.

The fact that the different statistics of the IP-burstlengths and IP-gaplengths of the measured traces and the memoryless network layer model are resulting in the same video streaming quality is becoming even more interesting when observing the difference in the relative number of erroneous IP packets per (erroneous) video frame in Fig. 3.27. There it can be observed that the longer IP-burstlengths are indeed resulting in more erroneous IP packets per (erroneous) video frame in the measured case and the 'static model 2' compared to the memoryless network layer model (IP memoryless). However, these results are in contradiction to [3] where lower video quality for longer IP error bursts is expected.

The reason for having the same video quality regardless of having different network layer error characteristics is the error propagation within one GOP (Group of Pictures) as shown in the following.

The evaluation of the video quality was performed using H.264/AVC with a common parameter setting for video streaming over wireless networks, meaning the video stream consists of one I and 39 following P frames where the latter ones are predictively encoded. This predictive coding of the P frames leads to error propagation within one GOP which in turn is the reason for having the same video quality with different IP burstlengths and also with a different number of erroneous IP packets per video frame.

When encoding the video with I frames only, there is no error propagation in the decoded video sequence. As a consequence, a difference in the Y-PSNR values per frame between the memoryless network layer and the measured characteristics can be observed as shown in Fig. 3.28. There, PDFs of the Y-PSNR values per frame - only the part of the erroneous frames - are

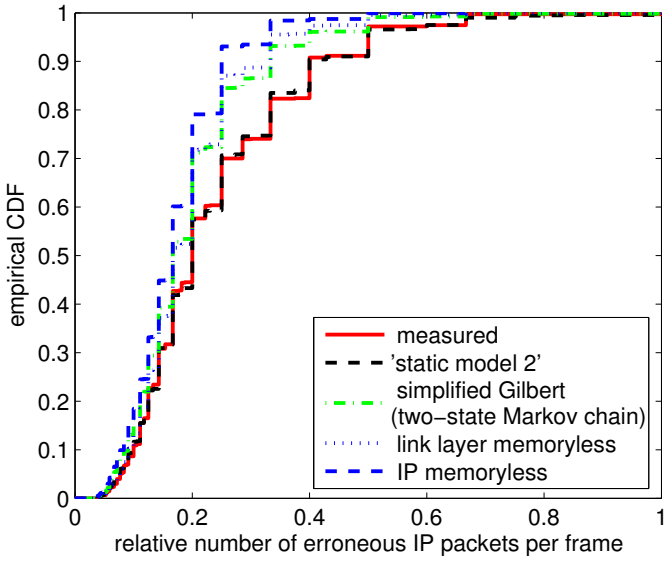

Figure 3.27: Relative number of erroneous IP packets per video frame.

presented. Note that there is a higher peak in the PDF for the error-free frames at higher Y-PSNR values.

Thus, the conclusion is that with a memoryless channel model in the IP layer (equivalently a memoryless channel model in the link layer with appropriate BLER), a plain IP-UDP-RTP protocol stack above RLC layer and common parameter settings for video streaming over UMTS DCH, despite having different higher order statistics in the network layer, the statistical end-to-end quality of streamed video data will be the same as for channels with highly correlated errors.

Therefore, when modeling the link layer, special attention has to be paid to the error correlation properties even with plain IP-UDP-RTP protocol stack, while a model in the network layer just has to meet the correct packet error probability.

Of course, it is important to mention that this analysis is only based on time independent Y-PSNR statistics and thus different results for the perceived end user quality may be expected.

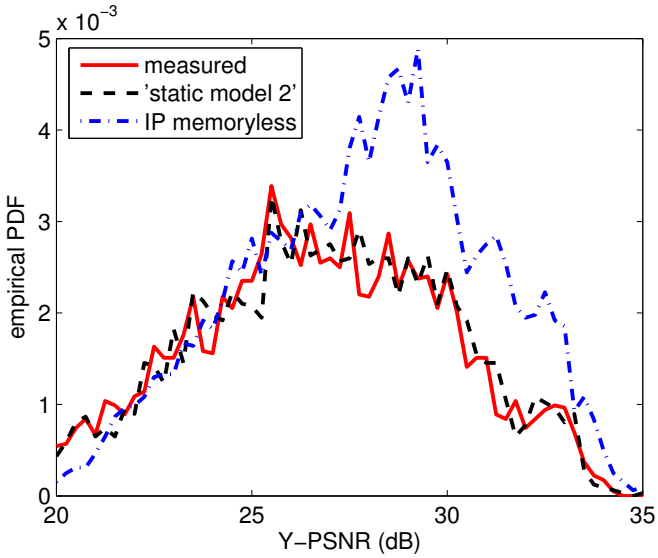

Figure 3.28: Y-PSNR per frame for only I frames (no error propagation).

Chapter 4

Dynamic Bearer Type Switching

Contents

High flexibility is one of the main advantages offered by UMTS. 3GPP specifications allow to change the physical resources allocated to the users dynamically to optimize the radio resource utilization in UMTS. This means the data rate in UMTS can be changed every 10 ms, either by changing the transport channel type or by varying the dynamic or semi-static parameters of the TF (Tansport Format) [57], [2]. Network operators can make use of this optimization feature by properly adjusting dynamic bearer type switching in their networks. Such switching of the channel characteristics optimizes the use of the radio resources and facilitates providing the required quality of service for the user. It can be triggered by admission control, congestion control, soft handover, required throughput or the radio channel quality (power threshold).

The focus of this thesis is on the dynamic switching between the three different DCH DL radio access bearers for the PS domain with 64 kbit/s, 128 kbit/s and 384 kbit/s which are characterized by different values of TTI, TBSS (Transport Block Set Size) and SF, as presented in Chapter 2 (see Fig. 2.3). The analysis focuses on the dynamic bearer type switching due to fading and coverage reasons, where the decision of up or down switching is based on the link quality only. Therefore, it is assumed having the admission for using all the bearers during the measurements and thus having a congestion-free network. This means that there were enough resources available in all the cells, so that the favourite bearer (384 kbit/s) could be assigned. Furthermore, a UDP data traffic with a constant bit rate of 372 kbit/s is enough throughput to be assigned a 384 kbit/s bearer due to throughput constraints.

For the performance evaluation of applications in system level simulators, models are needed which are capable of representing the error characteristics of the underlying channel properly. In Chapter 3 of this thesis such models are developed [10], assuming absence of dynamic bearer switching. Due to the fact that each bearer shows different characteristics, a better representation of the link can be reached by including bearer switching in the error model which is done e.g. in [81] based on simulations only. The approach in this work is to develop a model for dynamic bearer type switching based on measurements in live UMTS networks within typical mobility scenarios for using UMTS services. This model for dynamic bearer type switching can then be combined with the error model of Chapter 3 [10] and the combination can be used as a representation of the lower layers in a system level simulation tool.

4.1 Measurement Based Analysis of Dynamic Bearer Type Switching

In Fig. 4.1 the measured usage probabilities of the various bearers for 'mobile 1' are shown (based on the number of TTIs in each bearer — expressed in times the values are different). We observe that in all the scenarios with mobility, the estimated probability of being in the

384 kbit/s bearer is around 75 %, whereas in the static case it is almost 100 %. Of course, the usage probabilities of the various bearers depend on the network coverage situation and also on the quality of the receiver in the mobile.

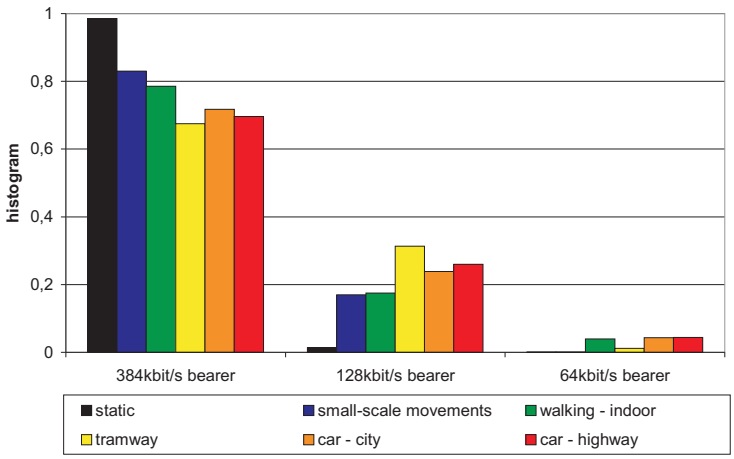

Figure 4.1: Bearer usage in TTIs with dynamic bearer type switching, 'mobile 1'.

The analysis of dynamic bearer type switching is based on the runlength distribution of each of the bearers, measured in the number of subsequent TTIs in case of 'mobile 1' and in seconds between the 'Transport Channel Reconfiguration' messages for 'mobile 3'. For 'mobile 3' the runlengths are provided in seconds instead of TTIs, due to the fact that with 'mobile 3' it is not possible to analyze the runlengths in the accuracy of TTIs but only by the 'Transport Channel Reconfiguration' messages.

In Fig. 4.2 (left) the CDFs of the runlengths are shown for the 384 kbit/s bearer and for 'mobile 1' in various scenarios. We can observe that the minimum runlength is 200 TTIs (= 2 s in case of 384 kbit/s bearer) and that 80 % - 90 % of the runlengths are shorter than 2000 TTIs (= 20 s). Furthermore, it can be seen that less than 5 % of the runlengths are > 6000 TTIs (60 s) and the longest period to stay in the 384 kbit/s bearer is about 80 s.

The statistics of the runlengths of the 128 kbit/s bearer with subsequent upswitching to the 384 kbit/s bearer is shown in Fig 4.2 (right). Again, these results are for different mobility scenarios and measured with 'mobile 1'. The runlengths in the 128 kbit/s bearer are all > 200 TTIs which is equal to four seconds in that case. We can observe a jump in the empirical CDFs up to 50 % or even 80 % in the 'walking indoor' scenario. The reason for having such a high step lies in the hysteresis time of the switching algorithm (4 s) and the decorrelation time (coherence time) of the channel which is obviously shorter than four seconds. This means, the channel quality would allow an upswitching to the 384 kbit/s bearer earlier than 4 s after a previous

downswitching, but this is prevented by the switching algorithm (hysteresis). Therefore, an optimized hysteresis parameter can accelerate the switching to the 384 kbit/s bearer in up to 80 % of the observed cases and thus is capable of improving the throughput over the wireless link.

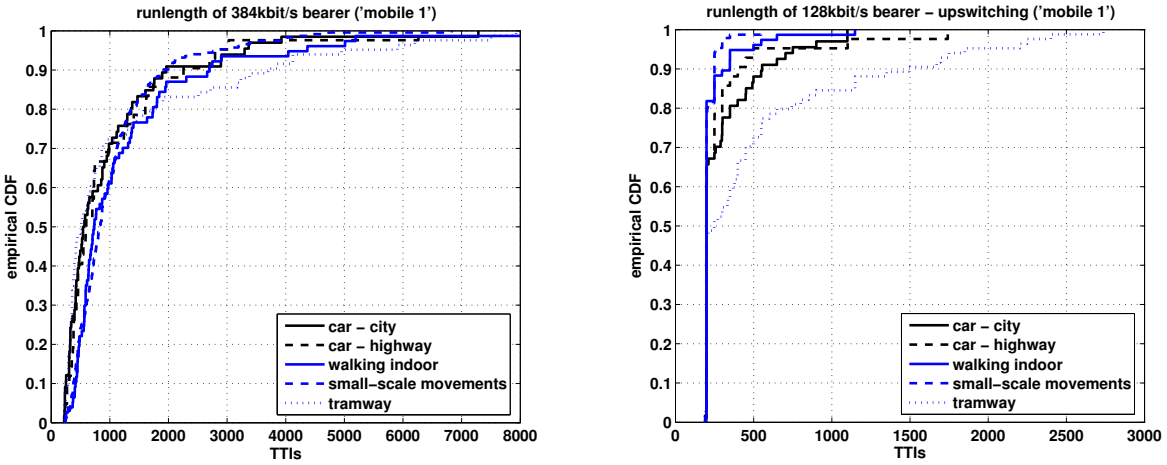

Figure 4.2: Measured runlengths of 384 kbit/s bearer (left) and 128 kbit/s bearer (upswitching) (right) in different scenarios, 'mobile 1'.

For the 128 kbit/s bearer with subsequent downswitching and in case of the 64 kbit/s bearer, the sum[1] of the statistics of the different mobility scenarios is provided in Fig. 4.3, measured with 'mobile 1'.

Figure. 4.3 (left) shows that in contrast to the runlength statistics of the 128 kbit/s bearer with subsequent upswitching, there is no high step at 200 TTIs in the runlength statistics of the 128 kbit/s bearer when being followed by a downswitching to the 64 kbit/s bearer. That difference in the runlength distributions is the reason for splitting the 128 kbit/s bearer in two parts for analysis and modeling.

The measured runlength distribution for 'mobile 3' with the 384 kbit/s and the 128 kbit/s bearer is shown in Fig. 4.4, again as the sum of the measurements of all different mobility scenarios. From Fig. 4.4 (left) it can be seen that with 'mobile 3' the runlengths of the 384 kbit/s bearer are > 2 s, and there are runlengths with a duration of up to 600 s which is much longer than the measured runlengths of 'mobile 1' in the same bearer. The reason for the difference in the runlength distribution between 'mobile 1' and 'mobile 3' is that 'mobile 3' is less sensitive to fading effects and achieves a better service coverage in the network due to a better receiver.

[1] As there is less usage of the 128 kbit/s and the 64 kbit/s bearer compared to the 384 kbit/s bearer, the sum of all mobility measurements are used for the smaller bearers to reach sufficient statistics.

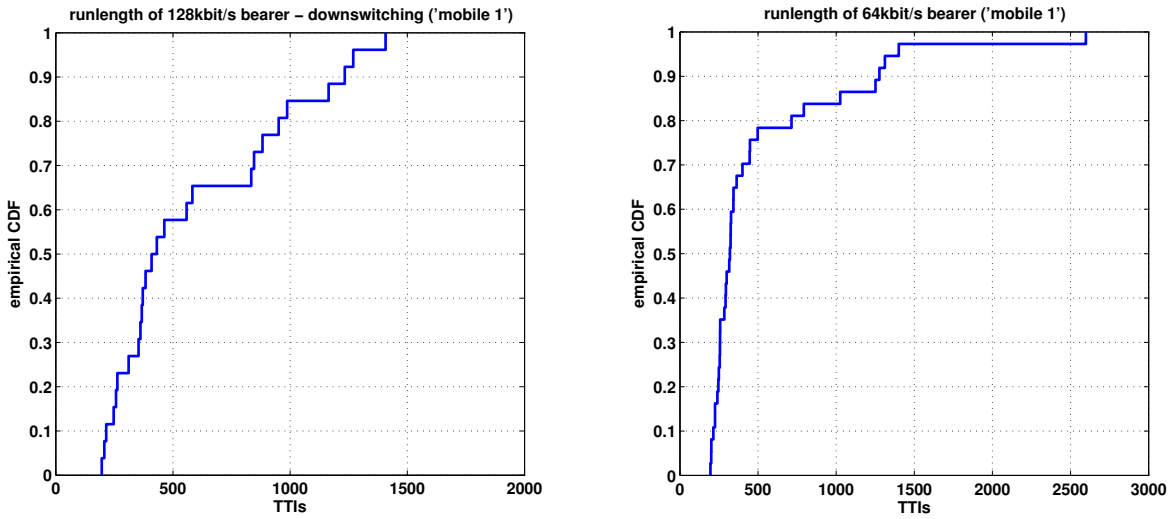

Figure 4.3: Measured runlengths of 128 kbit/s bearer (downswitching) (left) and 64 kbit/s bearer (right), sum of different scenarios, 'mobile 1'.

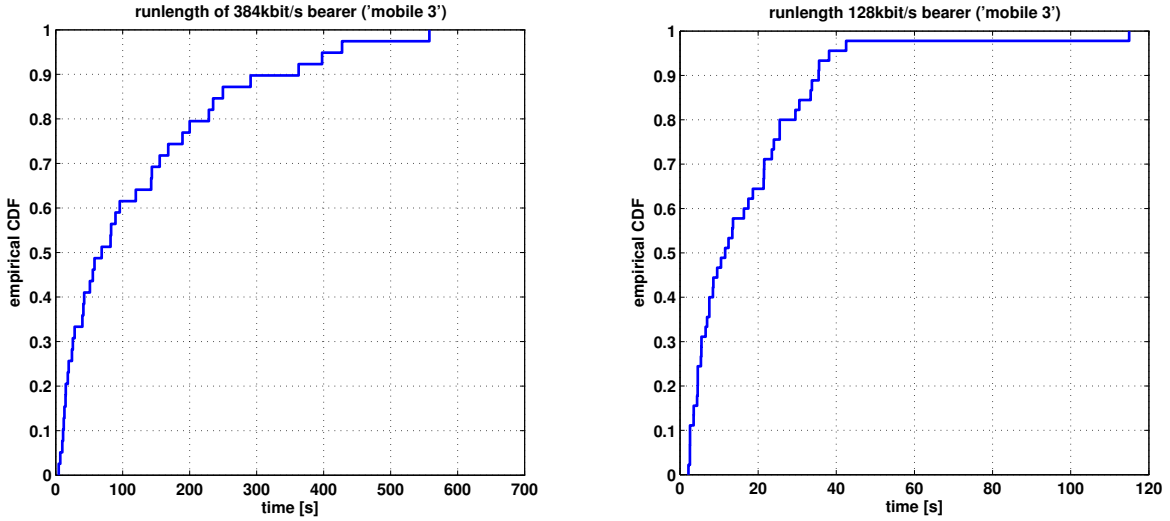

Figure 4.4: Measured runlengths of 384 kbit/s bearer (left) and 128 kbit/s bearer (right), sum of different scenarios, 'mobile 3'.

4.2 A Dynamic Bearer Type Switching Model

4.2.1 Four-state Markov Model

Based on the measurement results, the dynamic bearer switching is modeled by a four-state homogeneous Markov chain [71] [28] as presented in Fig. 4.5, where the valid state transitions are shown by arrows. The corresponding probability transition matrix $\mathbf{\Pi}$ has the following form:

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{64,64} & \pi_{64,128u} & \pi_{64,128d} & 0 \\ 0 & \pi_{128u,128u} & 0 & \pi_{128u,384} \\ \pi_{128d,64} & 0 & \pi_{128d,128d} & 0 \\ \pi_{384,64} & \pi_{384,128u} & \pi_{384,128d} & \pi_{384,384} \end{bmatrix}.$$

Its elements denote the stationary probabilities $\pi_{i,j}$ of one-step transitions from state i to state j. The probability vector $\mathbf{p}$ having the state probabilities p_k as elements, can be obtained by solving the system of linear equations $\mathbf{\Pi^T p} = \mathbf{p}$, with constraint $\sum_k p_k = 1$. These probabilities (in our case $\mathbf{p} = [p_{384}, p_{128u}, p_{128d}, p_{64}]$) correspond to the measured values in Fig. 4.1 (taking p_{128u} and p_{128d} as one state).

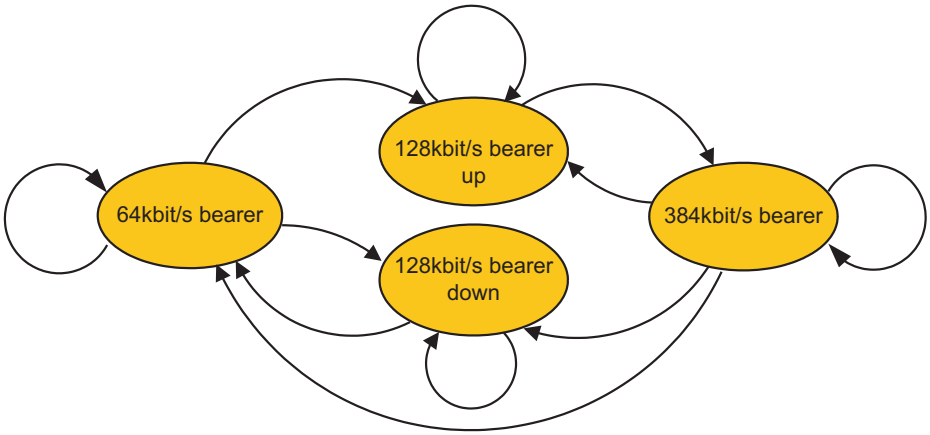

Figure 4.5: Four-state Markov model for dynamic bearer type switching.

As the discretization of time in our discrete-time Markov chain is in TTIs, the jumps to the next states have to be determined every TTI via the corresponding transition probabilities. For the complete description of the model we need eight parameters. Seven parameters are for defining the transition probabilities and one parameter is required to specify the hysteresis time of the switching algorithm (200 TTIs).

In Figs. 4.6 to 4.8 the simulated runlengths (dotted black line) together with the measured distributions (solid blue line) are shown for the different bearers and the two mobiles. In case of Fig. 4.6 the measured statistics of the 'tramway' scenario are presented as example. It can be observed that the geometric distributions, produced by the Markov chain, are not capable of meeting the measured statistics with sufficient accuracy for all cases - for example the runlength distribution of the 384 kbit/s bearer of 'mobile 1' (Fig. 4.6 (left)). Due to that fact we propose to model the dynamic bearer switching via a Weibull renewal process [74] as shown in the following.

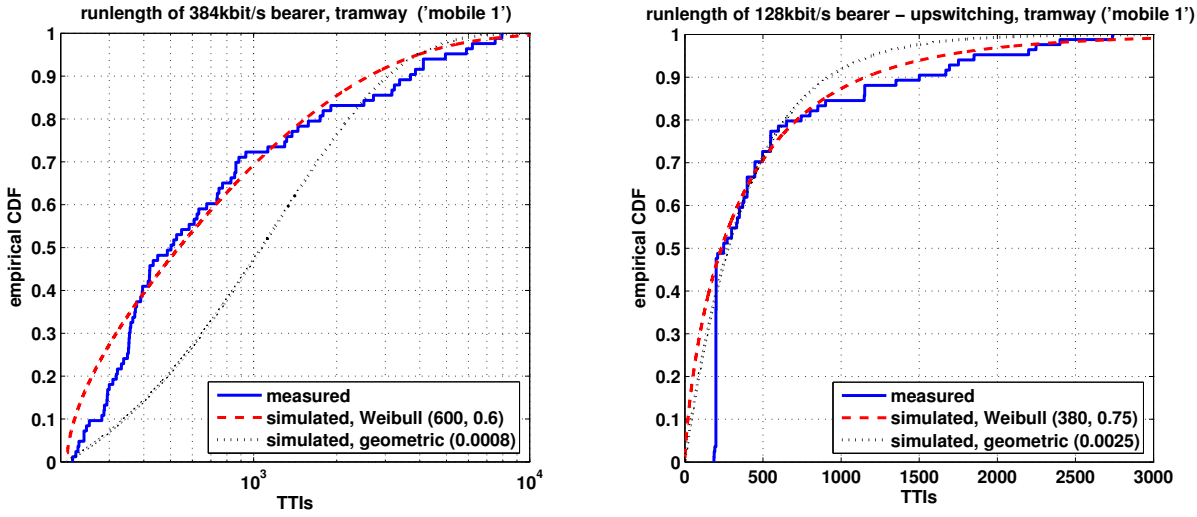

Figure 4.6: Runlength distribution of 384 kbit/s bearer (left), 128 kbit/s bearer - upswitching (right), measured vs. simulated, 'tramway' scenario, 'mobile 1'.

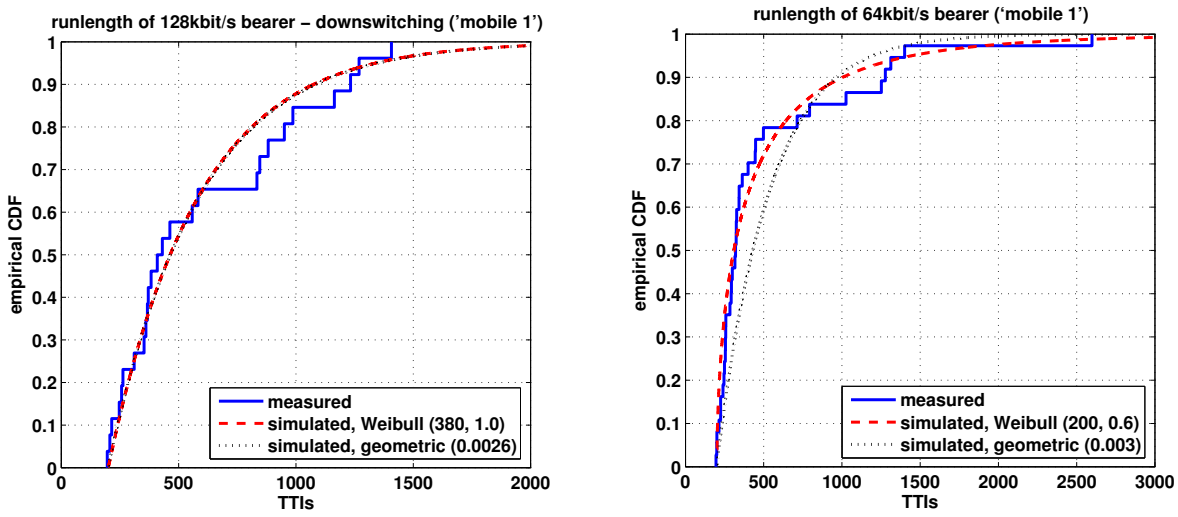

Figure 4.7: Runlength distribution of 128 kbit/s bearer - downswitching (left), 64 kbit/s bearer (right), sum of different scenarios, measured vs. simulated, 'mobile 1'.

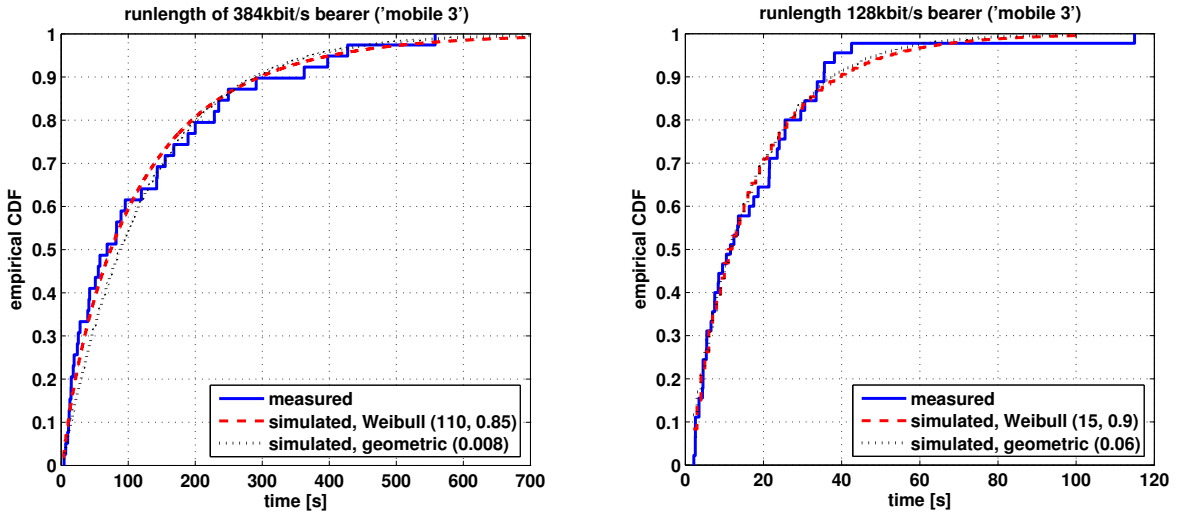

Figure 4.8: Runlength distribution of 384 kbit/s bearer (left), 128 kbit/s bearer (right), sum of different scenarios, measured vs. simulated, 'mobile 3'.

4.2.2 Enhanced four-state Model

In Fig. 4.6 (left) the empirical CDF of the runlengths of the 384 kbit/s bearer, 'tramway' scenario, is shown. The blue solid line shows the measurement result, whereas the dotted black line shows the simulated result of the Markov model representing a geometric distribution, given by:

$$f(n|p) = (1-p)^n p, \quad n = 1, 2, ... \tag{4.1}$$

with parameter p as extracted out of the measured trace ($p = 1 - \pi_{384,384}$). We can see that the geometric probability distribution is not capable of describing the measured distribution with sufficient accuracy. However, Weibull distributed runlengths with the parameters as denoted in the figure meets the measured statistics properly (dashed red line in Fig. 4.6 (left)).

Due to that fact a renewal process with four states and Weibull distributed runlengths is proposed as an enhanced model for the dynamic bearer type switching for UMTS DL DCH in the PS domain. With the high flexibility of the Weibull distribution we are capable of describing all the observed runlength distributions with adequate precision. The two-parameter Weibull CDF is given by

$$F(x|a,b) = 1 - \exp\left[-(x/a)^b\right], \tag{4.2}$$

where a and b are scale and shape parameters, respectively.

In Fig. 4.9, a schematic illustration of the model with the allowed transitions between the different states is shown. Such a Weibull renewal process [74] calculates the runlengths in each state via a two-parameter Weibull distributed random number.

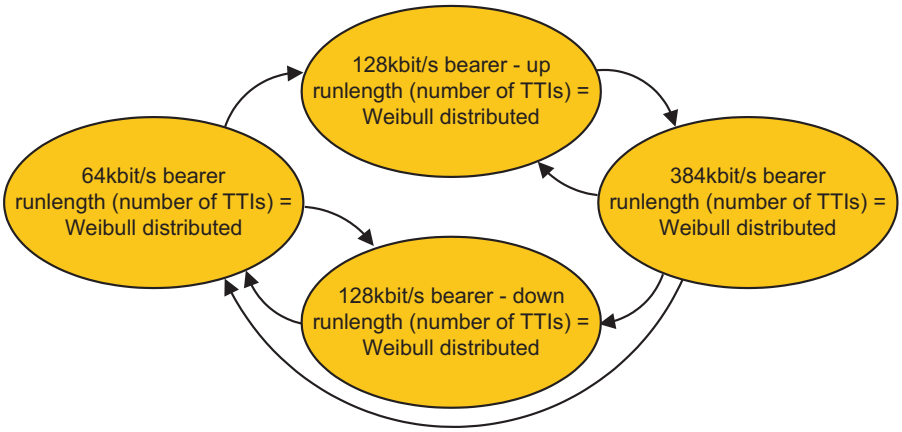

Figure 4.9: Schematic illustration of the enhanced four-state Model (Weibull renewal process).

The eight parameters for the four Weibull distributions (one in each state) together with the three parameters which determine the jumping between the states and one additional pa-

rameter for specification of the processing time of the switching algorithm result in a total of 12 parameters for a complete specification of the model.

Note, the measured distributions of the runlengths and therefore also the model parameters depend on the network load, the service coverage and the quality of the receiver in the mobile station. E.g., if there are some congested cells within the measurement route, there is less switching to the 384 kbit/s or to the 128 kbit/s bearer. In case of the varying quality of the receiver (varying service coverage) it was shown that the proposed model is flexible enough to describe the different runlength distributions.

Chapter 5

Analysis of Link Error Predictability in the UTRAN

Contents

The detailed knowledge about the link error characteristics as presented up to now in this work is necessary for the optimization of services (for example their codecs or protocols) for the usage over wireless mobile communication links [3–5]. Link error statistics can also be used for performance evaluation of new (cross-layer) mechanisms [45] as for example, for streaming video data transmission in the UMTS network.

Statistics of link error characteristics can be used to optimize the end-to-end transmission chain in a static sense and even dynamically and adaptively to predict future link errors based on past error occurrence and to use this predictability together with cross-layer optimization procedures (as shown in Chapter 6).

The analysis of the link error characteristics of the UMTS DCH in DL presented in Chapter 2 shows that there is correlation between the error states of subsequently transmitted TBs, causing error bursts. There is also correlation between successive bursts and gaps resulting in error clusters. Generally, these error correlation properties of the channel can be used to predict future link errors.

To be able to predict link errors based on past error events, the transmitter has to be aware of past link errors in the forward link, meaning there has to be a feedback link for the error status of the received data from the receiver back to the transmitter. For the UMTS DCH, this is accomplished by utilizing the error feedback of the RLC AM. For delay sensitive services we can adjust the maximum number of allowed retransmissions for the RLC AM to zero or equivalently adjust the discard timer properly. The RLC AM in the considered live UMTS networks uses 16 bits of CRC information for each TB with a size of 320 bits to detect transmission errors.

In case there was no or only very small (< 1 ms) feedback delay, the single TB errors within an error cluster could be predicted at the transmitter. Unfortunately, the feedback delay d_{FB} in the UMTS RLC AM ($d_{\mathrm{FB}} \geq 30$ ms also in case of 10 ms TTIs) is in the order of the cluster size; therefore, no error prediction within an error cluster is possible.

For this reason, the focus in this work is on the prediction of error clusters rather than single error events. Moreover, a first approach detects points in time with the least probability for the occurrence of an error cluster.

5.1 Prediction of Low Error Probability Intervals

From the statistics of the gaplengths in the 'static' scenario (see Fig. 2.13) it became clear that there are short gaps between the error bursts within an error cluster and long gaps between successive error clusters. Furthermore, there is a region in the CDF between the short and the long gaps with a very low probability for the occurence of a gap. This is because the CDF of the long gaps is convex between zero and the inflection point (mode) (see Fig. 3.11). Moreover, long

error gaps in case of the 'static' scenario are not geometrically distributed (Weibull distribution with shape parameter $a \neq 1$) resulting in memory in the link error characteristics. These properties can be used to detect intervals with low link error probability as shown in Fig. 5.1.

5.1.1 Detection of Start of Intervals

In case there is no error report via the feedback information within a certain time interval $d_{\min}$, which has to be larger than the longest of the short gaps, we can conclude that the following TBs are outside of an error cluster and the current transmission takes place within one of the long error gaps. Due to the convex distribution of the long gaps there is an interval with very low transmission error probability and we can detect its start at the minimum delay $d_{\min}$ after the last error. Of course, the total detection delay also includes the feedback delay d_{FB}. Figure. 5.1 illustrates the time series of the transmitted TBs (Tx) and of the error feedback information (FB) available at the transmitter after d_{FB}. Clusters of TB errors are marked and the start of the interval with low error probability with length L_{i} is shown at $d_{\min}$ after an error cluster.

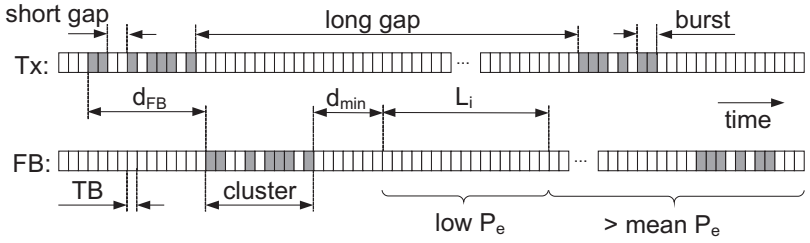

Figure 5.1: Schematic illustration of the interval with low link error probability.

5.1.2 Interval Length L_{i}

To determine the length L_{i} of the intervals with lower error probability, we make use of a theoretic analysis with long gaps separating single error events only. In Fig. 5.2 the local mean of the simulated estimated error probability $P_{\mathrm{e}}(x)$ at a point x TBs after an error event (in practice corresponding to the occurrence of an error cluster) is presented for Weibull-distributed gaplengths as fitted to the measured distribution in Chapter 3. We conclude from this curve that the error probability is very small right after an error event (practically after an error cluster), staying below the total mean error probability for approximately 2500 TBs. In practice, we may estimate L_{i} via the intersection of this conditional mean curve with the unconditional total mean $\overline{P}_{\mathrm{e}}$.

To obtain an analytic expression for L_{i} we first calculate the steady-state error probability $P_{\mathrm{e,ss}}$ which is equal to the total mean $\overline{P}_{\mathrm{e}}$. Let L be the random variable measuring the time

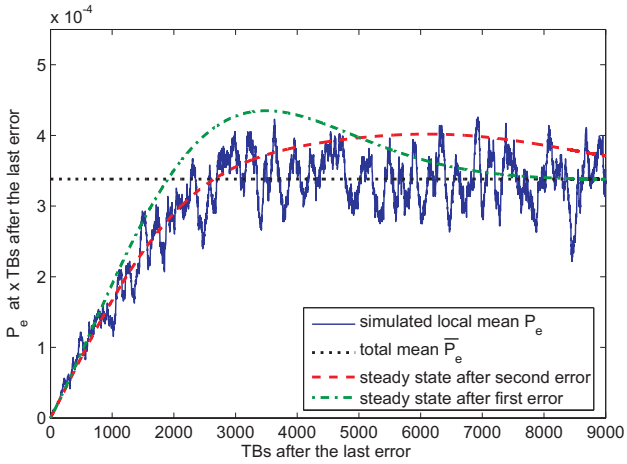

Figure 5.2: P_e at x TBs after the last error event for Weibull distributed long gaps (Weibull scale parameter a=3350, shape parameter b=2.176).

(distance in number of packets = gaplength) between two error events. Assuming that the error process is stationary, Kac's lemma [82] implies that

$$P_{\mathrm{e,ss}} = \frac{1}{\mathrm{E}\{L\}}, \tag{5.1}$$

where $\mathrm{E}\{L\} = \sum_{i=1}^{\infty} i \cdot P(L = i)$ is the average recurrence time corresponding to the average gaplength with $P(L = i)$ being the probability of having a gaplength of length i.

It turns out that the stationarity assumption holds over longer ranges, but does not lead to correct results in immediate succession to a known error event, where the (measured) lokal estimated error probability is actually lower than the total mean estimated error probability $\overline{P}_\mathrm{e}$. Therefore, one way to approximate the probability $P_\mathrm{e}(l)$ of having an error at the lth TB after an error event is to assume that the steady-state is reached before the third error after the initial error event:

$$\begin{aligned} P_\mathrm{e}(l) \quad &\approx P(L_1 = l) + P(L_1 + L_2 = l) \\ &\quad + P_{\mathrm{e,ss}} \cdot P(L_1 + L_2 < l). \end{aligned} \tag{5.2}$$

Here, L_1 and L_2 denote the gaplengths before the first and second errors, respectively. The result (red dashed curve in Fig. 5.2) shows that taking the intersection between this curve and the total mean as the end of L_i gives a good approximation, whereas the assumption of reaching the steady-state after the first error, calculated by

$$P_\mathrm{e}(l) \approx P(L_1 = l) + P_{\mathrm{e,ss}} \cdot P(L_1 < l), \tag{5.3}$$

leads to an underestimation of L_i (green dash-dotted curve).

5.2 Estimation of Expected Failure Rate

For some cross-layer methods (see Chapter 6.1), it is not sufficient to detect intervals with lower error probability; rather, an estimation of the instantaneous probability of having transmission errors is required. As previously mentioned, an exact error prediction within the error clusters is not possible due to the feedback delay but the location of the error clusters can be predicted. Therefore, the instantaneous link error probability is estimated by using the expected failure rate resulting from the distribution of the long gaps between successive error clusters as explained in the following.

Throughout this document the following notation is used to describe the error process. An erroneously received TB is indicated by '1' while '0' means error-free transmission. A positive integer in the exponent determines the number of consecutive erroneous or error-free TBs (e.g. the sequence '000001' can be written as '$0^5$1'). A gap with length m is defined as the number of 0's between two 1's and

$$p_M(m) = P(0^m 1|1) \tag{5.4}$$

for all positive integers m is the PMF (Probability Mass Function) of the gaplengths. The conditional probability $P(B|A)$ means the probability of sequence B, following sequence A. By definition,

$$\sum_{m=0}^{\infty} P(0^m 1|1) = 1 \tag{5.5}$$

and $P(0^0 1|1) = 0$, as gaps with length zero are not considered as gaplengths. The CDF of the gaplengths is then defined as

$$F_M(m) = \mathrm{P}\{M \leq m\} = \sum_{k=0}^{m} P(0^k 1|1). \tag{5.6}$$

The conditional link error probability $P(1|10^m)$ (the error probability conditioned to the number of error-free TBs since the last error) can be expressed by

$$P(1|10^m) = \frac{P(10^m 1)}{P(10^m)} = \frac{P(0^m 1|1) \cdot P(1)}{P(0^m|1) \cdot P(1)} = \frac{P(0^m 1|1)}{P(0^m|1)}, \tag{5.7}$$

where $P(0^m|1)$ denotes the probability of having a gaplength of at least length m and can be written in terms of CCDF (Complementary Cumulative Distribution Function) and PMF of the gaplengths

$$P(0^m|1) = \sum_{k=m}^{\infty} P(0^k 1|1) = 1 - F_M(m) + p_M(m). \tag{5.8}$$

Thus, the conditional link error probability $P(1|10^m)$ can be expressed as

$$P(1|10^m) = \frac{p_M(m)}{1 - F_M(m) + p_M(m)} \approx \frac{p_M(m)}{1 - F_M(m)}, \tag{5.9}$$

also presenting the approximation via the expected failure rate [71].

In Chapter 3 of this thesis it is shown that the long link error gaps can perfectly be fitted via a Weibull distribution [9,12]. Thus, after inserting the Weibull PMF and the Weibull CCDF in (5.9), we obtain the estimated conditional error probability

$$\widehat{P}(1|10^m)_{\text{Weibull}} = \frac{ba^{-b}m^{b-1}e^{-(\frac{m}{a})^b}}{e^{-(\frac{m}{a})^b}} = ba^{-b}m^{b-1}. \tag{5.10}$$

In this work we estimate $\widehat{P}(1|10^m)$ by using $b = 2.018$, corresponding to the statistics of the long gaps measured in [9]. With $b \approx 2$, Weibull distribution becomes Rayleigh and the estimated expected failure rate can be expressed by the much simpler term

$$\widehat{P}(1|10^m) = \frac{2m}{a^2}, \tag{5.11}$$

with $a = 3350$. Thus, $\widehat{P}(1|10^m)$ increases linearly with the error-free run length m (the number of error-free TBs received so far after the last error). It can be shown that (5.9) leads to an approximately linear increase in the considered region of $\leq$ 10000 TBs even without the mentioned approximation.

The measured conditional link error probability $P(1|10^m)$ from the 'static' scenario can be seen in Fig. 5.3 on the left. The linear estimator for the failure rate perfectly predicts the transmission error probability based on the number of error-free TBs since the last error (error-free runlength). In [12] further analysis of the influence of the feedback delay d_{FB} and the quality of the error prediction due to the modeling error of the gaplength statistics is presented.

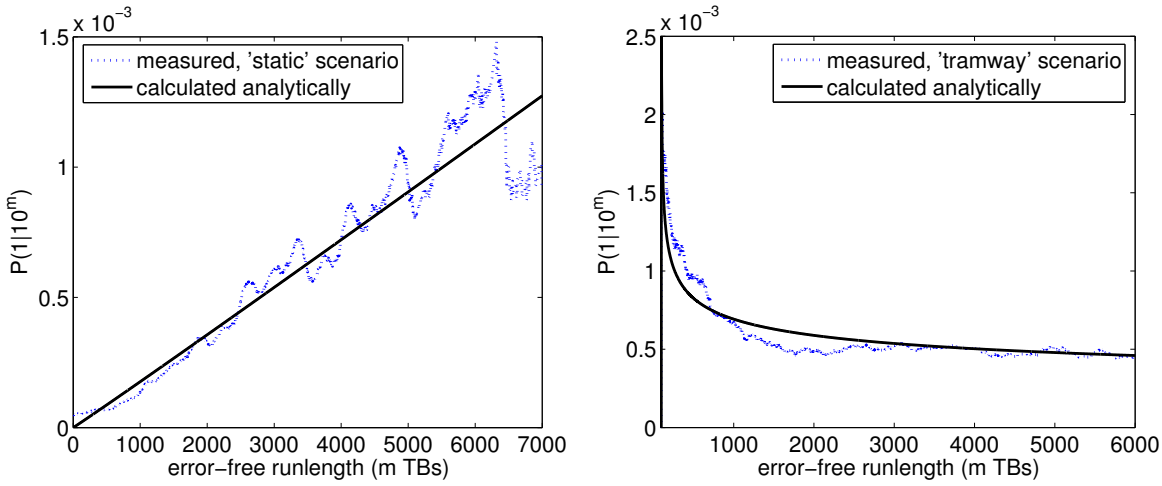

Figure 5.3: Conditional link error probability, 384 kbit/s bearer, 'static' scenario (left), 'tramway' scenario (right), 'mobile 1'.

While in the 'static' scenario the conditional link error probability linearly increases with the error-free runlength, in the 'dynamic' case the conditional link error probability has its maximum just after an error burst and cannot be estimated in a linear way. This can be seen in the right graph of Fig. 5.3, representing the measurement results of the 'tramway' scenario, the estimator according to (5.10) and with Weibull scale parameter $a = 1202.9$ and shape parameter $b = 0.782$. However, it can be observed that the estimator meets the measured conditional link error probability properly.

Fig. 5.4 shows the performance of the error estimators for different decision thresholds γ, for the 'static' scenario according to (5.11) on the left and for the 'dynamic' scenario according to (5.10) (right).

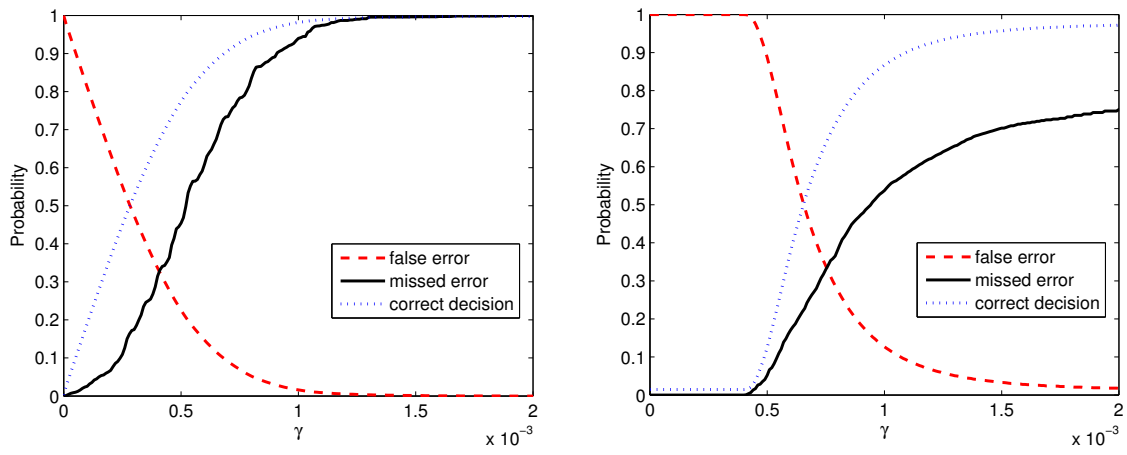

Figure 5.4: Prediction performance, 'static' scenario (left), 'tramway' scenario (right).

These figure were obtained by applying the following 'hard' decision: if $\widehat{P}(1|10^m) \geq \gamma$ an error is predicted. The probability of false error P_f is the conditional probability that $\widehat{P}(1|10^m) \geq \gamma$ if no error occurred. The probability of missed error P_m is the conditional probability that $\widehat{P}(1|10^m) < \gamma$ if an error occurred. The probability of a correct decision $P_\mathrm{c} = 1 - [(1 - P_\mathrm{err}) \cdot P_\mathrm{f} + P_\mathrm{err} \cdot P_\mathrm{m}]$ is the probability that an error was predicted and occurred plus the probability that an error was not predicted and did not occur; P_err denotes the probability of a link error.

In most applications (cross-layer processing algorithms) both, low P_m and low P_f is desired. Therefore, thresholds around the crossing of P_m and P_f are relevant.

There are different optimal thresholds for the 'static' and the 'dynamic' scenario. When using this prediction for both cases, either a suboptimal threshold can be selected to fit for the two scenarios or an additional detection for the discrimination of 'static' and 'dynamic' situations can be included in the Node B. This can be realized via a monitoring of the signal fading effects.

Predictability and general validity:

Note, the predictability of the conditional link error probability, as well as the predictability of the intervals with low error probability is existing due to the fact that there is a quality-based closed-loop power control algorithm (the OLPC) in the UMTS DCH which adjusts the SIR target value for the ILPC, as described in Chapter 2.

The quality-based power control mechanism introduces recurrence to the error process in order to meet the required quality target. For example, ideally there is a repeating sequence of 99 error free TBs after one erroneous TB in order to reach the target BLER of 1%. Of course, the real system does not behave like this, as there is fading and noise as well as error bursts and error clusters, but the idea of error predictability can be seen analogously.

Therefore, due to the quality-based control mechanism, there is memory in the error process in a sense that the occurrence of the next error event depends on the number of already subsequently received error-free TBs since the last error event. This means that the error gaplengths are not geometrically distributed.

In contrast to that, in a mobile communication system without quality-based control algorithm, there is not necessarily the same short-term recurrence in the error process as in a system with quality-based control algorithm. Therefore, for example in case of a very good channel, no link errors might be expected for a very long time interval in such systems and there is no predictive end of the error-free runlength as the end is caused by uncorrelated noise.

The mentioned arguments lead to the assumption that the predictability of link errors in the manner as described in this chapter of the thesis may also be expected for currently new or future mobile communication systems, if they comprise any quality-based control mechanism.

Chapter 6

Cross-Layer Optimization for Video Streaming over UMTS

Contents

The 3rd Generation Partnership Project (3GPP) standardizing UMTS has approved the inclusion of H.264/AVC (Advanced Video Coding) as an optional feature in Release 6 of its packet oriented mobile multimedia telephony [83] and streaming services [84] specifications. Although H.264/AVC is designed for robust delivery of streamed video over error prone mobile communication channels, the link errors severely affect the video quality. In the following it is shown how the specific link error characteristics, especially the predictability of the link errors, can be exploited in cross-layer processing methods to optimize the transmission of H.264/AVC video streams.

Whereas "traditional" (opportunistic) scheduling algorithms, such as those presented in [85, 86] concentrate on the multiuser diversity gain from different user behaviour and from the channel variations, on the multiservice gain, or on the gain from multiple video streams, in this work the video quality of a single video stream for one user is optimized. Instead of having a shared wireless channel, the proposed cross-layer processing algorithms focus on the UMTS DCH (dedicated to a single stream of one user only) and the exploitation of the error characteristics of this channel. Moreover, the presented cross-layer scheduling algorithms consider the priority of streamed video data packets and furthermore, the finally presented network and video-content aware cross-layer scheduling method explicitly defines the data priority according to the video content.

Note, the proposed 'in-stream' scheduling algorithms can also be included in multiuser, multiservice, multistream and multichannel (opportunistic) schedulers of new or future mobile communication systems with shared channels in order to further increase the QoS or to save radio resources.

In this work the term cross-layer design refers to protocol design by actively exploiting the dependency between protocol layers to obtain performance gains, according to the definition in [87]. This is unlike layering, where the protocols at the different layers are designed independently. As defined in [87], cross-layer design is protocol design by the violation of a reference layered communication architecture, with the seven-layer OSI (Open Systems Interconnect) model as reference architecture in this thesis; the OSI reference model [88] was jointly defined by ISO (International Organization for Standardization), IEC (International Electrotechnical Commission) and ITU-T (International Telecommunication Union - Telecommunication Standardization Sector).

It is presented in the following, that for the network-aware cross-layer processing methods, information flow (link error information) from OSI layer 1 (CRC is performed in the upper physical layer) to layer 2 and to layers 6/7 (video decoder) is required in order to forward the link error location information within one video packet. Furthermore, information flow from the layers 6/7 (video encoder) to the cross-layer scheduling mechanism in layer 2 is required in order

to know about data priority for the video decoding process in the network and data-priority aware cross-layer scheduling mechanism.

In this sense the definition of the used cross-layer approach includes the requirement of a cross-layer information flow as well as the cross-layer protocol design coupling (modification) and calibration according to [87], as for example, the upper layers have to be modified in order to make use of the correct parts of video slices.

As presented in Chapter 2, the main error source of the network is the radio link. Therefore, especially when considering the quality of the services during service usage (without e.g., considering errors during call setup or call blocking and dropped calls), the radio link is the only 'network' related error source and thus the naming "network-aware approach" as a generalisation of being aware of errors in the upper physical layer or equivalently in the data link layer (errors in the link between Node B and mobile station). Meaning, by considering the error characteristics of the radio link, the system is completely aware about the erroneous network behaviour as far as it is relevant for the service quality during service usage (e.g., during a video streaming sequence).

6.1 Network Aware Cross-Layer Processing

As already shown in Chapter 5, the coded video data is transmitted by using the UDP protocol, providing a 16 bit checksum for error control mechanism. The UDP packets containing parts of video slices are further segmented into smaller RLC PDUs (320 bits) in the UTRAN. These RLC PDUs together with the 16 bit CRC are forming TBs, again providing a possibility for error detection.

In case of a link error (leading to one erroneous TB), the whole UDP packet is usually discarded due to a wrong 16 bit UDP checksum, leading to significant quality degradation in the streamed video. In [45], we proposed and evaluated a cross-layer error detection mechanism which allows for a utilization of the correct parts of the video slice within the erroneous UDP packet.

Without any change of standards and without adding data overhead, the correct parts of the UDP packet until the first erroneous TB can be utilized for decoding the streamed video data by simply forwarding the error information from layer two to the higher layers (video decoder) — e.g. including UDP Lite [89], providing the possibility of forwarding the erroneous UDP data packets, and CUDP (Complete UDP) [90], where also error location information is forwarded between the protocol layers.

From the results of the analysis with the link error characteristics of the UMTS DCH shown in Fig. 6.1, it can be concluded that in more than 73 % or 55 % of cases, this method provides

gain in video quality in the 'static' scenario and 'dynamic' scenario, respectively. Obviously, the shorter error bursts in the 'static' scenario lead to a lower probability of receiving the first RLC PDU within an erroneous UDP packet in error, which increases the efficiency of the proposed method.

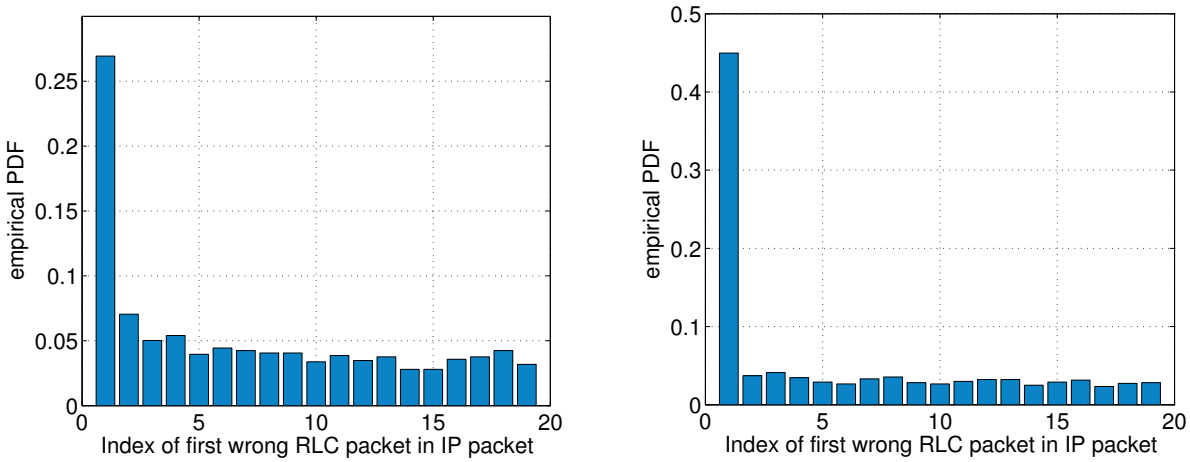

Figure 6.1: Distribution of the index of the first erroneous RLC packet (TB) within IP packet, left: 'static' scenario, right: 'dynamic' scenario, 'mobile 1'.

To evaluate the improvement of the end-to-end video quality, again the peak signal-to-noise ratio of the luminance component (Y-PSNR) is used according to (3.5) and (3.6).

In Fig. 6.2, it is shown that with the proposed method, up to 15 dB gain in video quality (Y-PSNR averaged over erroneous frames) can be reached when performing spatial error concealment for the missing parts of the erroneous slices (in case of an I frame). For this analysis the 'foreman' sequence with a 144×176 pixel resolution (QCIF) was encoded with a QP (Quantization Parameter) of 28, and the I frame frequency was set to 20. Video slices of a maximum of 700 bytes have been used because this value provides a good trade-off between resulting slice error probability and data overhead [91]. Figure 6.2 starts at position three as the first two link layer packets (TBs) contain RTP/UDP/IP header and the slice header. Without them, the slice cannot be detected correctly.

The proposed method still gains up to 6 dB when using temporal error concealment methods like simple copying from the same position in the last frame (in case of a P frame). The Y-PSNR for Fig. 6.2 was calculated comparing the degraded/concealed sequence to the original non-compressed one.

The gain of the proposed method (decoding the video up to the first erroneous link layer packet within the UDP packet) for a 'soccer match' [80] video sequence with 10 frames/s sim-

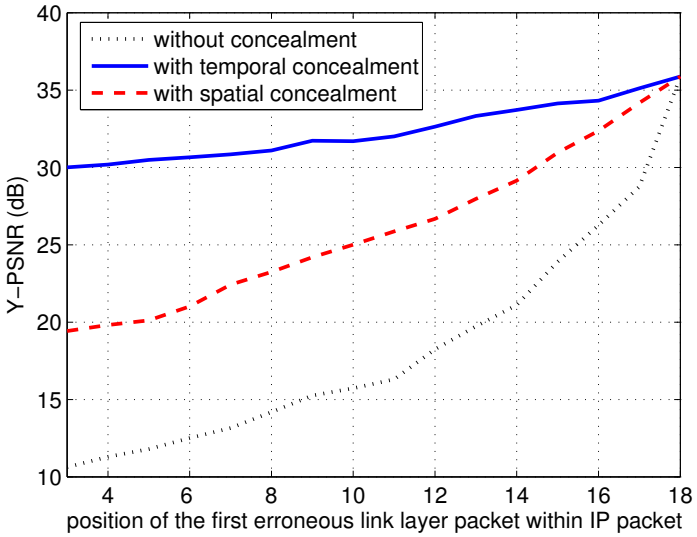

Figure 6.2: Y-PSNR vs. index of first erroneous link layer packet within the UDP packet.

ulated over a measured link error trace ('static' scenario) is presented in Fig. 6.3. For the video simulations H.264/AVC reference software Joint Model (JM) [78] was used for encoder and decoder in the baseline profile (targeting the video conferencing, streaming and especially mobile applications). Context Adaptive VLC (CAVLC) was used, with slicing mode two and a limited maximum number of bytes per slice of 800. The I frame frequency was 50, QP=26 and no B frames were used. Missing parts of the video were concealed by copying the spatially corresponding areas from the previous frames.

Although there is a low overall probability of a TB error of 0.266 % (resulting in a UDP/IP packet error rate of 0.888 %) there is considerable improvement in comparison to the usual practice of discarding the whole erroneous UDP packet (video slice).

The utilization of the correct parts within one erroneous UDP packet after an erroneous RLC PDU of course would provide further increase in the video quality. Due to the VLC (Variable Length Coding) in H.264/AVC this can only be accomplished either by changing the encoding process or by transmitting additional side information. The reason is that after a bit error, CAVLC may easily 'desynchronise', making the correct distinction between neighboring codewords impossible. The decoding of such a desynchronized stream may result in considerable visual impairments, or may become even impossible due to non-existing code-words or too many/few bits left for decoding. To overcome this problem a reversible VLC is proposed in [92] resulting in a higher data rate due to lower compression gain. Another possibility is to transmit additional synchronization information either within the VLC stream [93] or outside

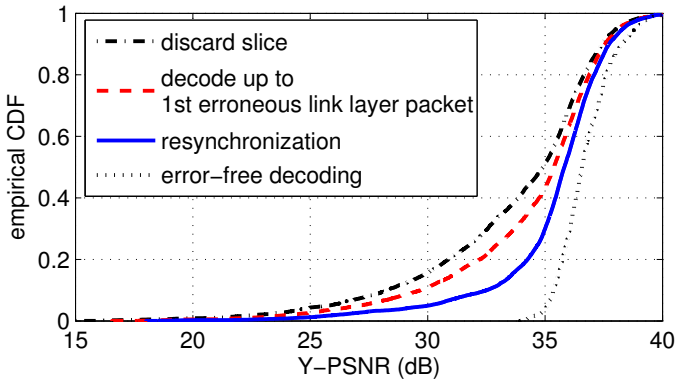

Figure 6.3: Quality of video reconstruction for different error handling mechanisms, 'soccer match' video sequence, link error trace from 'static' scenario, 'mobile 1'.

of the stream as we proposed in [45]. The resynchronisation information can be advantageously represented as the position of the start of the first MB (Makro Block) contained in the RLC PDU. For RLC PDUs with 320 bits and fixed length encoding, 9 bits per MBPI (MB Position Indicator) are required [46].

Figure. 6.3 shows that the gain from decoding all error free parts of the UDP packet (achievable with the additional resynchronisation information) compared to the decoding until the first error occurrence is still more than 5 dB in Y-PSNR for some erroneous frames, which is significant from the user experienced quality point of view.

Of course, sending the side information necessary for the VLC resynchronization requires a higher rate.

In order to decrease the required rate for sending the resynchronization information we can make use of the link error predictability from Chapter 5 in a proposed redundancy control mechanism [46] as presented in the following.

Since the MBPI information is used by the decoder only if an error in the radio interface occurs, we can reduce the required rate by sending the side information packets only if a high error probability is predicted. The decision about low/high error probability is based on a single threshold γ in the estimated expected failure ratio $\widehat{P}(1|10^m)$ according to (5.11) where the error probability is estimated via the instantaneous error-free run distribution. It turns out that $\gamma = 2 \cdot 10^{-4}$ leads to good results.

Figure. 6.4 illustrates the average quality at the decoder normalized by the required rate. There is an improvement of approximately 0.25 dB by using the error prediction as compared to the full resynchronization (sending all MBPIs). This is a considerable improvement, since the averaging was performed over the entire video sequence, containing error-free packets and

having a small TB error probability (0.266 %).

Equivalently, it can be concluded from Fig. 6.4 that the proposed methods (full resynchronisation and resynchronisation based on link error prediction) provide the possibility of significantly reducing the rate at the radio interface while keeping the same quality at the decoder.

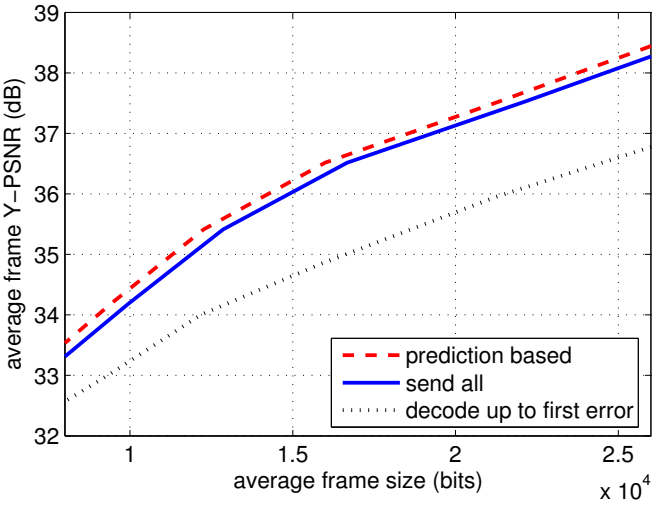

Figure 6.4: Average Y-PSNR over the rate. The rate is expressed as average size of a frame including side information in bits.

6.2 Network and Data-Priority Aware Cross-Layer Scheduling

Currently there is great effort in research to optimize services like video streaming for transmission over wireless communication networks with their higher error probability and bursty link error characteristics. The introduction and continuous improvement of H.264/AVC is the best example. Mobile communication links are also being optimized, as their properties are adapted to meet the requirements of the services for providing better QoS. For example, in [94], a truncated power control is introduced to improve the video quality and ensure efficent transmission of the video stream.

Another approach is to apply a cross-layer algorithm which connects the service properties with the network properties in an optimum way by making use of the continuous information flow from higher layers (data priority information from application layer) and from the channel or the network (channel quality status information from link layer or physical layer). In [95] and [96], (opportunistic) scheduling algorithms are presented which make use of the characteristics of the streamed video data. In [95], the more important parts of the video stream are transmitted

prior to the less important ones to ensure more opportunities for retransmissions in case of an error. On the other hand, the priority-based scheduling presented in [96] exploits the diversity gains embedded in the channel variations when there is more than one stream. The gain from having a multi-user and multi-service scenario is also exploited in another approach in [97], where prediction of the link errors is used in connection with call admission control and scheduling algorithms to avoid the system of being overloaded and thus improving the quality of the services with higher priority.

The focus in this work is on the transmission of a single video stream within one channel which is dedicated to this particular stream. The proposed mechanism makes use of the ability to predict the link errors of the UMTS DCH to improve scheduling of the individual layer-three packets of the video stream with different priority levels. Unlike in [94], where transmission is stopped in times of bad channel quality, the proposed scheduling algorithm makes use of all the available bandwidth, but delays the packets with higher priority to a position where least error probability is predicted [43]. This novel method allows the improvement of the quality of streamed video without reducing the quality of other services as occurs in [97].

The start of intervals with low error probability can be predicted, as presented in Chapter 5. It was shown that after the occurrence of an error cluster the transmission takes place within one of the long gaps; thus we can predict a very low error probability beginning at $d_{\min}$ after the last error of the cluster.

In Fig. 6.6, we present the estimated packet error probability of layer three packets with a size of 720 bytes (= 18 TBs), transmitted with delays d_{r} to the end of $d_{\min} = 37$ TBs as illustrated in Fig. 6.5. The packet size of 720 bytes was selected as it represents a typical size for video streaming layer three packets (including overhead) for an adjusted maximum slice size of ≈ 700 bytes in the video codec setting, which in turn provides a good trade-off between slice error probability and data overhead. Furthermore, 720 bytes can be divided by 320 bits (RLC PDU size). The simulations for Fig. 6.6 were performed using the novel UMTS DCH link error model ('static-model 2') presented in Chapter 3.

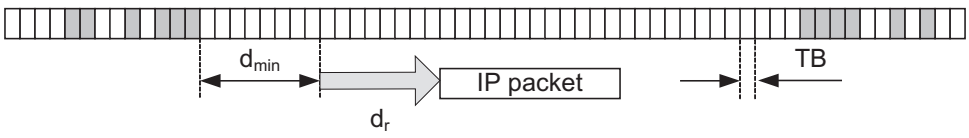

Figure 6.5: Illustration of the simulation procedure for deriving Fig. 6.6.

Figure. 6.6 shows that right at $d_{\min}$, the packet error probability reaches its minimum ($< 0.1\,\%$), exceeding the total mean (over the whole trace) of $0.925\,\%$ at $d_{\mathrm{r}} \approx 3000$ TBs.

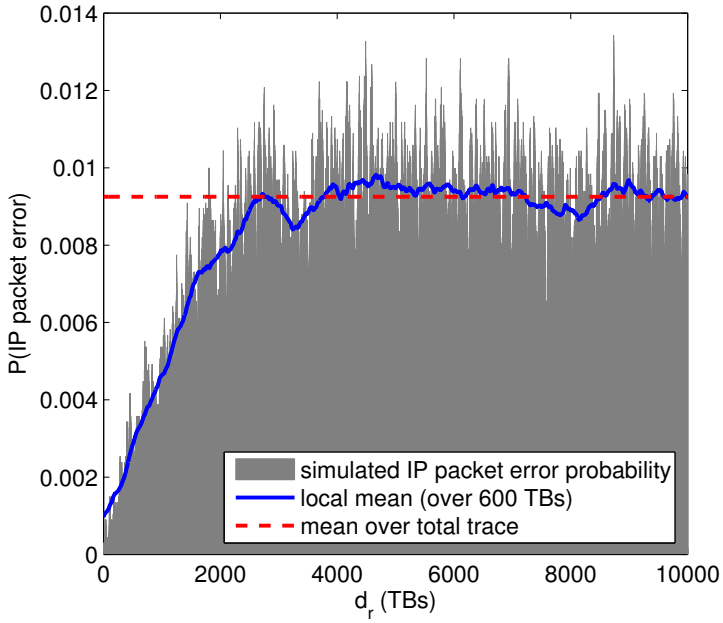

Figure 6.6: Packet error probability vs. relative position d_{r} to the end of $d_{\min}$.

The goal of the proposed scheduling algorithm is to transmit the packets with high priorities at the time instants for which low error probability is predicted.

In a video stream, the I frames are more important than the P and B frames, since they refresh the stream. If an error occurs in an I frame, it propagates over the whole group of pictures (GOP) up to the next I frame in the worst case. To prevent possible error propagation beyond an I frame, spatial error concealment is usually used for I frames, having lower performance than the temporal error concealment which is used for P frames.

In this work, layer three packets containing I frames are given higher priority than packets containing P frames (no B frames are used).

The proposed scheduling algorithm is illustrated schematically in Fig. 6.7, where in the upper part, the time series of the transmitted TBs is shown with the erroneous TBs marked. The scheduling algorithm tries to map the I packets (layer three packets containing parts of I frames) onto TBs which are to be transmitted within the time intervals in which low error probability is predicted. This is accomplished by buffering the incoming I packets and delaying their transmission to a point after the next occurrence of an error cluster. After detecting $d_{\min}$ consecutive error free TBs following an error report (performed by resetting the counter if there is an erroneous TB within $d_{\min}$), that is after the expected end of the error cluster, all the I packets within the buffer are transmitted. The rest of the TBs (RLC PDUs) are filled with the

incoming P packets (layer three packets containing parts of P frames), which may be transmitted in times of high expected link error probability as they are less important for decoding the video stream.

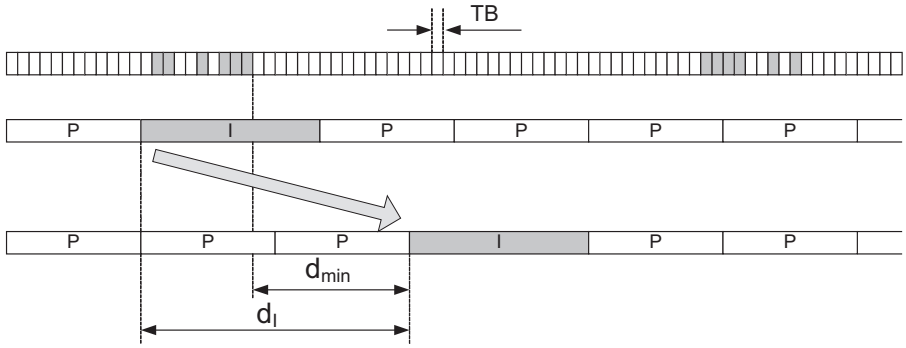

Figure 6.7: Schematic illustration of the proposed network and data-priority aware cross-layer scheduling algorithm.

This method causes additional transmission delay (d_{I}) for the I packets and thus also for the I frames. However, such delay does not cause any deterioration to the quality as long as it remains within the storage capacity of the playout buffer at the receiving terminal. In Fig. 6.8 the resulting I-frame transmission delay in number of TBs (12 TBs = 10 ms) is shown for a real video streaming sequence ('soccer match' [80]), simulated for the measured link error trace ('static' scenario).

Figure. 6.8 shows that without a limitation, the maximum resulting transmission delay for the I frames is about 7000 TBs (≈6 seconds). Thus a playout buffer of 6 seconds - which is a common value for video streaming services - would be sufficient for a full utilization of the proposed method. Current video streaming applications usually use 5 to 20 seconds playout buffers (pre-roll buffers) to cope with the channel fluctuations and the inherent variable-bit-rate nature of coded video sequences [95].

However, if the maximum allowed I frame transmission delay ($\leq$ playout buffer size) is exceeded, the concerned I frames are transmitted immediately. It can be seen in Fig. 6.8 that with a limitation of the playout buffer to 3 seconds, the proposed method still can be used successfully in 85 % of cases.

The resulting gain in video quality can be seen in Fig. 6.9 where the empirical PDFs of Y-PSNR values per frame are presented. The results have been generated in simulations with the well-known 'foreman' video sequence in QCIF resolution (144 × 176 pixel) encoded with I and P frames only, in slicing mode two with 700 bytes per slice, 15 frames per second, and an I frame frequency of 20. With a quantization parameter of 25, a video stream with an average

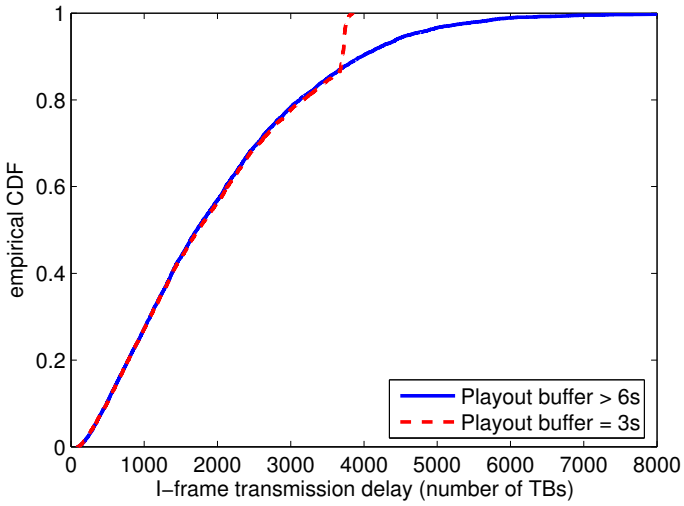

Figure 6.8: Resulting I frame transmission delay in number of TBs, 'soccer match', 384 kbit/s bearer, 'mobile 1', 'static' scenario.

bit rate of 300 kbit/s was obtained.

The empirical PDFs in Fig. 6.9 show that the presented cross-layer scheduling method reduces the number of frames with lower Y-PSNR values, whereas the number of frames with higher Y-PSNR values increases, thus, resulting in a shift of approximately 5 dB to better video quality. Note, Fig. 6.9 only contains the lower Y-PSNR range. In the higher range, there is a peak for the Y-PSNR values corresponding to the error-free frames.

For the results of Fig. 6.9, the decoder used temporal error concealment for the inter-predicted frames (P frames) and spatial error concealment for the intra-coded frames (I frames). To see the benefit of the proposed scheduling algorithm without assuming a particular error concealment method, the reduction in the number of erroneous I and P frames was evaluated, which was 83.4 % and 4.2 %, respectively.

The proposed method does not require changes in the standard method of encoding the video stream and can be applied in UMTS networks without changes to 3GPP specifications. The only required modification is to forward both the content of the layer three packets (I or P) and the error status of the received TBs in the protocol stack. At the scheduler (transmitting side) the information about the error status of the TBs at the receiving terminal is needed. This feedback information can be acquired by using the RLC AM mode, for example with the maximum number of retransmissions set to zero. In UMTS UL, all the necessary information for error prediction and scheduling is already available at the mobile terminal. In the UMTS DL, the scheduler would be situated at the RNC. As there is no IP layer available in the protocol

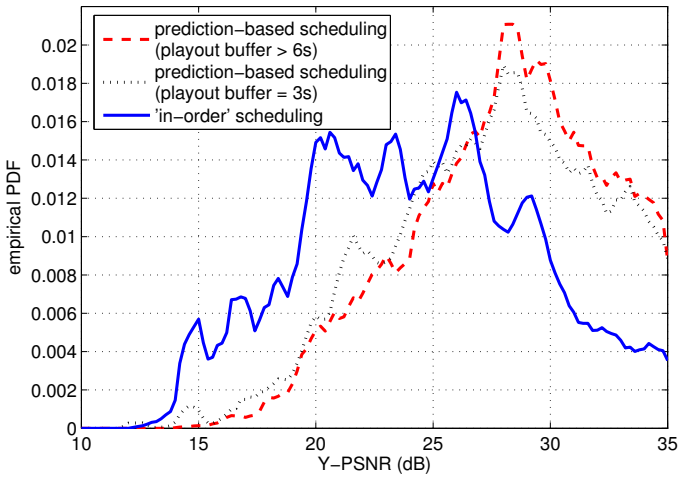

Figure 6.9: Gain of network and data-priority aware cross-layer scheduling, empirical PDFs of lower Y-PSNRs per frame, 'foreman' sequence, 384 kbit/s bearer, 'static' scenario, 'mobile 1'.

stack of the RNC, the content information for the layer three packets would have to be extracted out of the payload, included within the header (e.g., in the RTP protocol header a marker bit is available) or sent on an extra link in the UMTS core network to the RNC.

The proposed method can also be applied when allowing a higher number of retransmissions in RLC AM. Of course, with more retransmissions in the RLC layer the gain of the scheduling algorithm becomes smaller as more retransmissions would result in a lower error probability in the link.

Of course, the feedback delay of RLC AM (≈ 30 ms if one TTI = 10 ms) imposes delayed detection and less efficient utilization of the intervals with low error probability. However, a delay of ≈ 36 TBs only causes negligible loss in the efficiency of the method.

6.3 Network and Video-Content Aware Cross-Layer Scheduling

The proposed cross-layer processing method presented in the previous section exploits the specific link error characteristics and the different priority levels of streamed data packets, where the priority of the data packets is set according to their importance for the video decoding process but regardless of the particular video content.

In addition to the unequal importance of I and P frames for the decoding process, the efficiency of the coding and the performance of the error concealment methods heavily depend

on the video content. For example, in scenes with movement, packet loss could result in severe degradation in video quality due to spreading of the error over the picture, and temporal error concealment after the loss of a packet at a scene cut is not capable of reproducing the subsequent video frames properly.

Therefore, in order to minimize the distortion in the decoded video stream caused by corrupted data packets, in this work a network-aware cross-layer scheduling method [44] is proposed, which treats the individual data packets according to their priority, based on the distortion their loss would cause. In this way, the scheduler considers coding characteristics as well as video content properties.

Rate-Distortion Optimized (RDO) video streaming was proposed in [98], where a model for delay and loss in the Internet is used to perform scheduling at the IP layer. In this work, the distortion model required to define the priority levels of transmitted data packets considers losses at the RLC PDU level as presented in [44]. The impairments caused by the loss of a PDU are modeled as *primary distortion* (which is the distortion in the frame of the lost PDU) and *distortion propagation* (the propagated distortion in the subsequent frames until the end of the GOP), including error concealment. The model assumes that the video decoder can use the information in a NALU (Network Abstraction Layer Unit, typically encapsulated further in an RTP/UDP/IP packet) up to the first missing RLC PDU as proposed in Section 6.1. The distortion estimation in the model considers the position and amount of lost data, as well as the size of the corresponding picture area (affected by the loss of the single PDU), thus introducing awareness of video content like movement, scene cuts, and so on. The primary distortion $\widehat{\varepsilon}_0$ is represented as a function of the number of affected RLC packets (PDUs) N and the corresponding number of affected macroblocks M:

$$\widehat{\varepsilon}_0 = e^{a \cdot M + \frac{b}{N} + c} \tag{6.1}$$

with parameters a, b, and c determined by least squares fitting. In the model data set, the following values were obtained: $a = -0.041$, $b = -6.371$, and $c = 4.482$. The number of lost/damaged MBs required by the model can be obtained without fully decoding the video as a proportion of the MBs per NALU packet (if it is fixed) corresponding to the proportion of lost RLC PDUs, by reading of the entropy encoded stream (without decoding the video) or by extra signaling.

The error propagation is modeled as an exponential decrease of the primary distortion

$$\widehat{\varepsilon}_k = \widehat{\varepsilon}_0 \cdot e^{-s \cdot k}, \tag{6.2}$$

where k is the frame index and s determines the speed of decay, which depends strongly on the nature of the motion in the sequence. Since we assume that only M and N are available for

the scheduler, the error propagation has to be estimated by an average over all sequences in the model data set. The decay thus obtained is $s = 0.08$.

The proposed network and video-content aware cross-layer scheduling mechanism then makes use of the estimation of the *cumulative distortion* $\widehat{\varepsilon}_\Sigma$ that would be caused by the loss of a particular RLC PDU. This is calculated for every RLC PDU in the scheduling buffer as

$$\widehat{\varepsilon}_\Sigma = \sum_{k=n}^{m} \widehat{\varepsilon}_{k-n}, \tag{6.3}$$

where n denotes the number of the frame where the RLC PDU under consideration is located, and m is the number of frames in the GOP.

Figure 6.10 gives a schematic illustration and shows how the distortion model and the estimated distortion is included into the scheduler to introduce content awareness for the streamed video data transmitted over the UMTS DCH.

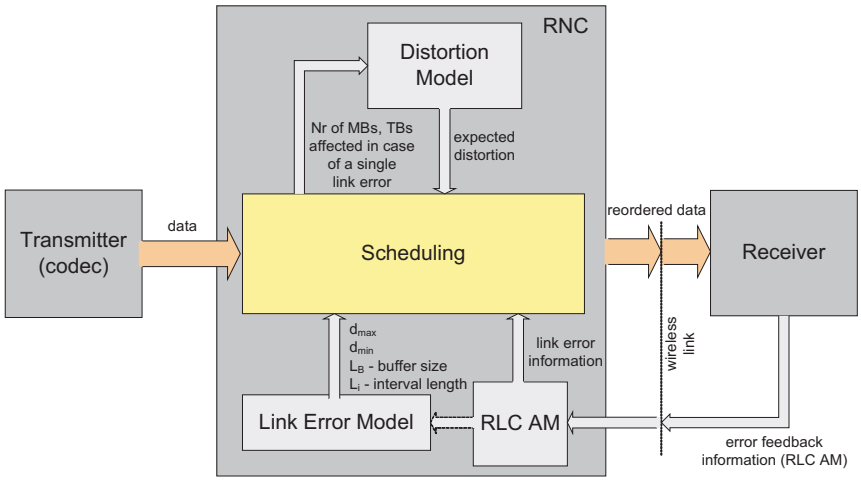

Figure 6.10: Illustration of the network and video content aware cross-layer scheduling mechanism applied to the UTRAN.

In Fig. 6.10 it is also presented that the network awareness is provided to the cross-layer scheduling algorithm by a model of the link error characteristics and the error feedback information from the UMTS RLC AM.

This scheduling algorithm again makes use of the predictability of intervals with low error probability as presented in Chapter 5. The estimated length L_i of these intervals (calculated according to (5.2)) is also required as will be shown in the following.

Figure 6.11 shows the proposed cross-layer method in detail from the segmentation of the video slices to the buffering of RLC PDUs, the selection of the RLC PDU to transmit and the transmission of the TBs (Tx) as well as the error feedback information (FB) as time series.

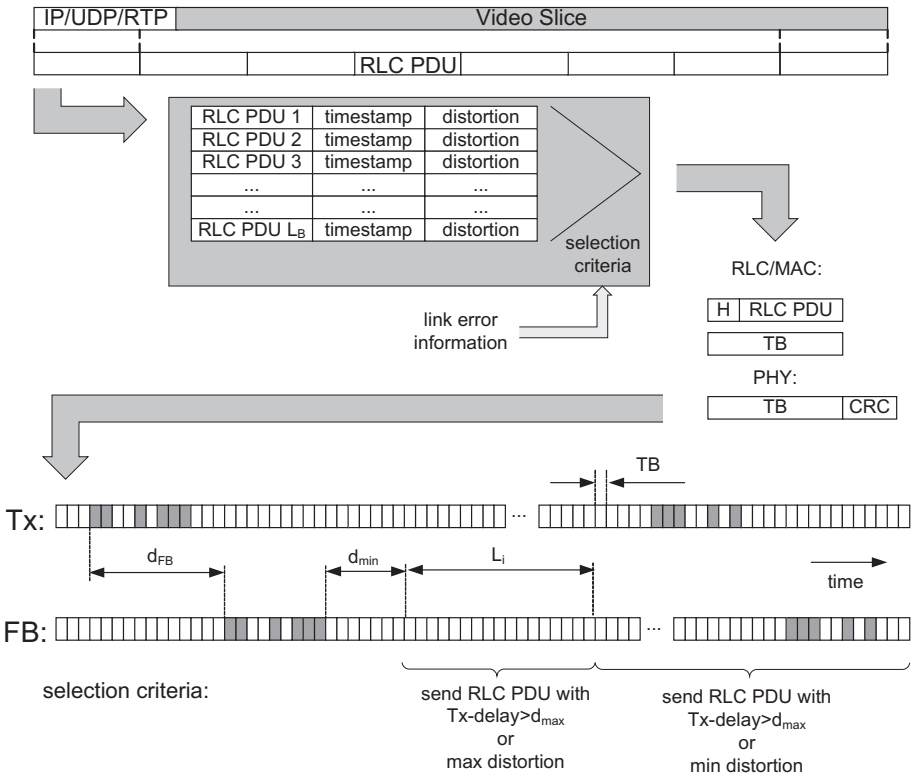

Figure 6.11: Details for the functioning of the network and video-content aware cross-layer scheduling mechanism.

At each step, if an RLC PDU remains in the scheduling buffer longer than the maximum acceptable delay $d_{\max}$, then it is scheduled immediately. Otherwise, at each step within the time interval L_{i}, the RLC PDU associated with the highest estimated cumulative distortion $\widehat{\varepsilon}_{\Sigma}$ is chosen from the scheduling buffer. Outside the interval L_{i}, the RLC PDU with the lowest $\widehat{\varepsilon}_{\Sigma}$ is scheduled. If several RLC PDUs have the same (lowest or highest) value of $\widehat{\varepsilon}_{\Sigma}$, the oldest among them is scheduled.

The performance of the scheduler can be controlled by the scheduling buffer length $L_{\mathrm{b}} \geq L_{\mathrm{i}}$ and $d_{\max}$. The delay $d_{\max}$ depends on the requirements of the service. It should be chosen smaller than or equal to the application's play-out buffer length. Moreover, the maximum gap length (see Fig. 2.13) provides a guiding value for determining the size of $d_{\max}$.

For evaluation of the gain of the proposed scheduler, the music video clip sequence 'videoclip' [80] was chosen as a test sequence, since it contains a variety of different scenes separated by cuts and gradual transitions. The sequence was encoded with 7.5 frames per second, a GOP size of 40 frames, and a QP of 26. The size of the slices was limited to 650 bytes, and it had only one reference frame. CAVLC was used with the remaining settings corresponding to the baseline profile. The transmission of the sequence over a measured link error trace of the 'static' scenario was simulated 1500 times (representing ≈ 3 h of video).

Figure. 6.12 shows the empirical CDFs of frame PSNR values of the content-aware scheduling method compared to the data-priority aware algorithm as proposed in the previous section and to the transmission without a special scheduling method (with in-order scheduling). Note, all CDFs approach one at PSNR $= \infty$ (error-free frames), and Fig. 6.12 only represents the relevant part of the CDF for this analysis.

Figure. 6.12 shows that the data-priority aware scheduling method globally reduces the number of erroneous frames since it moves the errors from the first frame (I frame) in the GOP (Group of Pictures) to the later frames (P/B frames), making the error propagation chain shorter. The proposed content aware scheduling on the other hand reduces the number of frames with higher distortion and consequently increases the number of frames with lower distortion. The simulations have been performed with $d_{\max}$ of 8400 TBs which is equal to 7 seconds and the RLC PDU buffer size was set to $L_{\mathrm{b}} = 3600$ TBs $= 3$ seconds (12 TBs per 10 ms TTI).

As shown in Fig. 6.12, it can even happen that the total number of frames containing even a small error can increase after applying the content aware scheduling algorithm. The reason is a possible spreading of (TB) burst errors onto multiple frames in order to reach minimum distortion (caused by a single TB error) in the video.

Note, for finding the minimum distortion in the video, the combination of errors in time has to be considered; for example, if an error occurs followed by another error in the next frame at the same position in the picture, then the cumulative distortion is smaller or equal to the sum

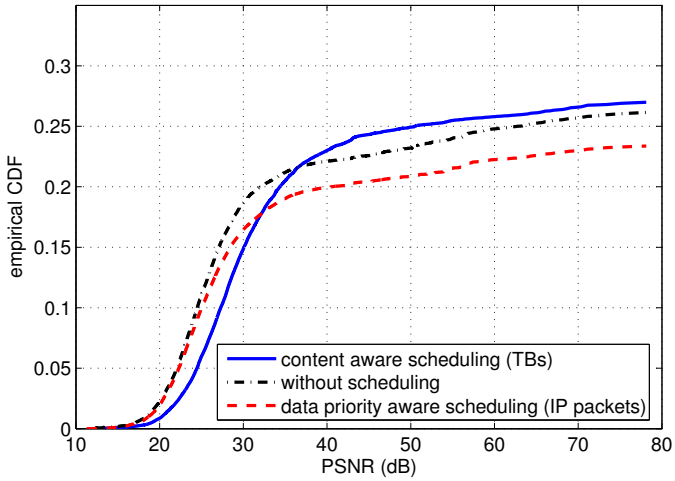

Figure 6.12: Empirical CDFs of frame PSNR, video-content aware scheduling vs. data-priority aware scheduling, 'videoclip' sequence, 'static' scenario, 'mobile 1'.

of the two distortions. However, considering all possible combinations of two and more errors would increase the complexity of the scheduler considerably. Thus, the distortion caused by single error events has been considered independently of others in this work.

Due to the different mechanisms of the two scheduling methods, we cannot compare their performance by Fig. 6.12; therefore, the mean PSNR values have to be used instead.

The resulting mean PSNR for transmission without scheduling was 29.29 dB. With data-priority aware scheduling (I frame scheduling), 30.32 dB was reached; with video-content aware scheduling, 32.03 dB was reached.

From these results, it can be concluded that considering the specific link error characteristics (predictability of intervals with lower error probability) within cross-layer algorithms provides up to 2.7 dB in quality improvement, which is significant from the users' perceived quality point of view. These improvements are reached with simultaneous awareness of the importance of the streamed data packets (UDP/IP) for the decoding process (offering 1 dB) and with consideration of the priority of the transmitted data packets (RLC PDUs) with respect to the video content, resulting in 1.7 dB more. The improvements are considerable as the simulated link error ratio was very small (0.26 % based on the measurements in the live UMTS networks), which, therefore, offers small improvement margins only. Note, the gain compared to common processing of streamed video data over wireless networks is even higher, as in the last analysis the usage of the cross-layer processing method of Chapter 6.1 (decoding of the NALUs up to the first erroneous link layer packet) is assumed.

Chapter 7

Conclusions

Nowadays, cellular wireless mobile communication systems are providing great potential for the transmission of a wide range of services. Unfortunately, the quality of demanding (quasi) real-time services like video telephony or video streaming, but also the quality of services which have originally been designed for wired networks, are heavily suffering from the error characteristics of the wireless links.

Relatively high error probability and bursty link error behaviour comes with the transmission over state of the art cellular wireless systems, and may also be expected for future mobile communication networks as the frequency spectrum is a scarce resource and interference is the major limitation.

The overall goal of this thesis is to show that detailed knowledge about the specific link error characteristics, gained by measurement based link error analysis, forms the essential basis for developing new and optimizing existing services and protocols. Substantial gain in end-to-end quality is reached when considering error characteristics of the underlying link.

Due to the absence of relevant limitations in wired communication links, the radio link will stay the main or even become the only error source in the communication chain. Thus, in order to optimize the end-to-end quality of services the awareness of the error characteristics of the radio access part of the networks becomes significant.

This thesis presents the link error analysis of the UMTS DCH based on measurements in live networks with different levels of mobility. It is shown that due to the link error characteristics basically two scenarios have to be distinguished: static and dynamic (regardless of which kind of mobility). Furthermore, the analysis shows that the quality-based power control mechanism induces correlation to the link error characteristics, resulting in error bursts, error clusters and the predictability of the link errors.

Due to the feedback delay of the error information, the errors within the bursts or clusters cannot be predicted. However, this thesis presents methods to predict intervals with low error

probability and the instantaneous failure rate, which rely on the fact that the lengths of the error-free intervals between the error clusters are not geometrically distributed. Especially in static scenarios, the link error statistics follow a Weibull distribution due to the properties of the quality-based power control mechanism and thus, error prediction is possible.

Generally, it can be concluded that also in novel or future mobile communication systems the possibility of link error prediction can be expected if the systems comprise any quality-based control mechanism.

Often measured traces do not provide sufficient statistics for system level simulations and thus stochastic models of the observed error characteristics of the link are required.

In this document it is shown that 'well-known' classical error models are not capable of describing the measured specific error characteristics of the UMTS DCH properly. Especially, the error predictability cannot be described at all due to the geometric distributions in the output of the models. Thus, a new modeling approach based on semi-Markov models as well as (Markov modulated) Weibull renewal processes is presented, providing good usability and high accuracy with low complexity. As the basic idea behind this modeling approach is to consider the specific properties of the underlying network, the approach may as well be applied for other new or future mobile communication systems.

The direct impact of modeling the link error characteristics on the quality of streamed video over a plain protocol stack is evaluated in this thesis. The results show the great importance of correct modeling of first and second order moments of the link error characteristics while only the mean packet error probability is sufficient for a network layer model in this special case.

For a complete representation of the link in a system level simulator, beside modeling of the link error characteristics of the available bearers, also the dynamic switching between these bearers has to be considered. The UMTS system offers such a changing of the physical resources allocated to the users dynamically to optimize the radio resource utilization. In this thesis the dynamic bearer switching is analysed based on measurements in live UMTS networks and within various scenarios with different types of mobility. Based on the measured statistics a four-state Markov model and a four-state Weibull renewal process is presented which can be combined with the proposed link error models for a proper description of the dynamic system behaviour.

It is important to consider the specific link error characteristics for the performance evaluation of new services, (cross-layer) designs or higher layer protocols. However, even further this work shows that a thorough understanding of the specific link error characteristics becomes significant for the development of cross-layer algorithms which explicitly exploit the particular properties of the underlying link.

A novel cross-layer processing design is presented for H.264/AVC video streaming over UMTS where a considerable gain in video quality is reached introducing network awareness by making

use of the capability of predicting instantaneous failure rates. Redundancy information for decoding the streamed video data is sent only when high error probability is predicted.

Further improvements in video quality are achieved by adding awareness of the priority of the encoded video data packets. The presented network and data-priority aware scheduler transmits packets with high importance for the decoding process in time intervals where low error probability is expected.

Finally, a novel network-aware scheduler is presented which considers the priority of the layer two packets of the encoded video stream according to the distortion they would cause in case of their loss. Thus, via a distortion model the scheduler additionally is aware of the particular video content like scene cuts or movement.

It is shown in this thesis that significant gain in video quality is achieved when combining network awareness with the awareness of service characteristics in a cross-layer approach.

One of the crucial points of this work which makes this thesis unique but at the same time is one of the main limiting factors is that the described work, especially in Chapters 2 to 5, is based on measurements in live UMTS networks. The analysis was performed on the UMTS DCH with the parameter settings as they are adjusted by the operators or manufacturers. Therefore, many of the results and conclusions are directly valid just for the UMTS DCH with the selected parameter settings and in most of these cases the only possible conclusion is that similar behavior of other systems or with other parameter settings may be expected due to the presented arguments. Of course, further simulation-based analysis could strengthen those arguments and verify the conclusions, but building a system level simulator, which has to consider e.g., propagation effects, channel coding, interleaving and not to forget the complete quality-based power control mechanism with the feedback link, would be the work for another thesis. The arguments for general validity have been presented in this work, like for example the predictability of the link errors which can also be expected in other systems with quality-based control mechanisms, the reasons for having error bursts or error clusters and that the main idea behind the presented modeling approach does have general applicability. Furthermore, the ideas for the cross-layer processing in Chapter 6 generally can also be applied in connection with other systems.

However, the main result of this thesis is valid in a general sense, which is the conclusion that detailed measurement-based knowledge about the specific link error characteristics forms the essential basis for the optimization of services and protocols and that substantial gain in end-to-end service quality can be reached when considering or even exploiting the error characteristics of the underlying link.

Appendix A

Introduction to H.264/AVC Video Streaming

H.264/AVC (Advanced Video Coding) is the newest and currently the most efficient video compression standard for low data rate applications and error prone environments. Generally, video compression is the process of converting video into a format that requires less capacity when it is stored or transmitted. Therefore, video compression (video coding) is a fundamental technology for applications such as digital television, DVD-Video (Digital Video Disc), mobile TV, video conferencing and internet and mobile video streaming. Standardising the video compression algorithms guarantees inter-operability for products from different manufacturers (e.g. encoders, decoders and storage media).

H.264/AVC was jointly defined by the ITU-T (International Telecommunication Union - Telecommunication Standardization Sector) VCEG (Video Coding Experts Group) together with the ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) MPEG (Moving Picture Experts Group). In December 2001, VCEG and MPEG formed the JVT (Joint Video Team) to define the standard now known as ITU-T Recommendation H.264 and ISO/IEC International Standard 14496-10 (MPEG-4 part 10) Advanced Video Coding [7, 99, 100].

The H.264/AVC standard was first published in 2003. H.264/AVC provides greater flexibility in compressing, transmitting and storing video — it offers the potential for having better video quality at the same data rates compared to previous standards like H.261, H.262 (MPEG-2), H.263 and MPEG-4 part 2, with only a moderate increase of complexity.

As has been the case for all ITU-T and ISO/IEC video coding standards, the scope of the standard is the central decoder as shown in Fig. A.1. The standard imposes restrictions on the bitstream and syntax, and defines the decoding process of the syntax elements in a way that all

decoders which are standard conform produce similar video output from an encoded bitstream that meets the constraints of the standard.

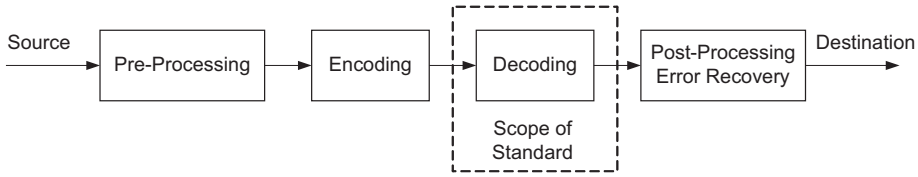

Figure A.1: Scope of (H.264) video coding standardization.

In other words, the H.264/AVC standard defines a format (syntax) for compressed video data and a method for decoding this format to reproduce a displayable video sequence. The standard document does not actually specify how to preprocess and encode (compress) digital video, neither it specifies the post processing methods. This is left to the manufacturers. However, in practice, the encoder is likely to mirror the functions of the decoding process.

As shown in Fig. A.2, the H.264/AVC design is conceptually separated into the VCL (Video Coding Layer), designed to efficiently encode the video content, and the NAL (Network Abstraction Layer), which formats the VCL representation of the video for a variety of transport layers or storage media [101]. The encoded video data produced by the VCL is segmented by the NAL in a stream of information units called NALUs (Network Abstraction Layer Units).

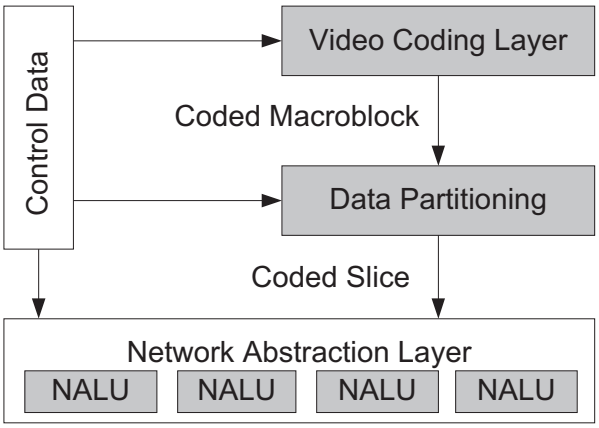

Figure A.2: Structure of H.264/AVC video encoder.

A.1 Video Coding

Generally, video signals are represented via sampling in several pictures (frames) per second, defined by the frame rate, and in pixels per picture, defined by the picture resolution in number of rows (N) times number of lines (M) of pixels.

When using the interlaced scan method, even and odd lines are sampled alternately in consecutive pictures (each image now called a field), whereas the progressive scan method samples even and odd lines in all pictures. In contrast to television systems, computer monitors generally use progressive scan which therefore is also used for internet and mobile video formats.

A.1.1 Picture Resolution and Color Space

Color pictures are sampled by three $N \times M$ color component matrices with the color values quantized by q-bits (usually $q = 8$) for the digital RGB (Red, Green, Blue) representation. In Table A.1, common picture resolutions for mobile video applications are presented which are derived either from computer monitor resolutions (VGA - Video Graphics Array) or from television resolutions (PAL - Phase Alternating Line). However, all resolutions used for digital video coding are a multiple of 16×16 pixels which is the resolution of a MB (Macroblock), the basic unit in video coding as shown in the following.

Resolution	**Synonym**
640×480	VGA (Video Graphics Array)
480×320	HVGA (Half-size Video Graphics Array)
320×240	QVGA (Quarter Video Graphics Array) SIF (Standard Interchange Format)
160×120	QQVGA (Quarter QVGA)
352×288	CIF (Common Intermediate Format) 1/4 of resolution used in PAL television standard
176×144	QCIF (Quarter Common Intermediate Format)
128×96	SQCIF (Sub Quarter CIF)

Table A.1: Common picture resolution for mobile video applications.

Since the human visual system is less sensitive to color than to luminance (brightness), the bandwidth can be optimized by storing more luminance details than color details. In H.264/AVC as in prior standards, this is achieved by using a YCbCr color space together with reducing the sampling resolution of the Cb and Cr chroma information. Component Y (*luma*) represents brightness, whereas the two *chroma* components Cb and Cr represent the color difference of Y to blue and red, respectively.

The Y signal is calculated from the RGB signal by

$$Y = k_r \cdot R + k_g \cdot G + k_b \cdot B, \tag{A.1}$$

where k_r, k_g and k_b are weighting factors and $k_r + k_g + k_b = 1$. ITU-R recommendation BT.601 [102] defines $k_b = 0.114$ and $k_r = 0.299$.

The chroma components Cb and Cr are created by calculating the scaled difference between the Y signal and the blue and red components by

$$C_b = \frac{0.5}{1 - k_b} \cdot (B - Y), \tag{A.2}$$

$$C_r = \frac{0.5}{1 - k_r} \cdot (R - Y). \tag{A.3}$$

Because of the fact that the human visual system is more sensitive to luma than to chroma, H.264/AVC uses a sampling structure in which the chroma components have one fourth of the number of samples than the luma component (half the number of samples in both the horizontal and vertical dimensions). This is called 4:2:0 sampling with 8 bits of precision per sample and reduces the required data rate by 50%.

A.1.2 Macroblocks and Slices

H.264/AVC follows the so-called block-based hybrid video coding approach, in which each coded picture is represented in block-shaped units of associated luma and chroma samples called MBs (Macro Blocks). A picture is partitioned into fixed-size MBs each covering a rectangular picture area of 16×16 samples of the luma component and 8×8 samples of each of the two chroma components.

Furthermore, a picture may be split into one or several slices, a sequence of MBs which are processed in the order of a raster scan when not using FMO (Flexible Macroblock Ordering) as shown in Fig. A.3.

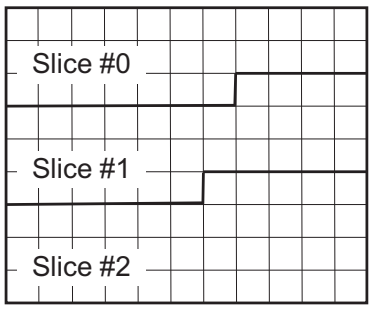

Figure A.3: Subdivision of a picture into MBs and slices (without FMO).

By introducing the concept of *slice groups* and variable MB to slice group mapping, FMO modifies the way how pictures are partitioned into slices. Each slice group can be partitioned into one or more slices. Note, the case when FMO is not in use can be viewed as the simple special case of FMO in which the whole picture consists of a single slice group. Figure A.4 shows two ways of subdividing a frame into slices with FMO, where on the left there are two 'foreground' slice groups and a 'leftover' slice group and on the right a checker-board type of mapping is presented.

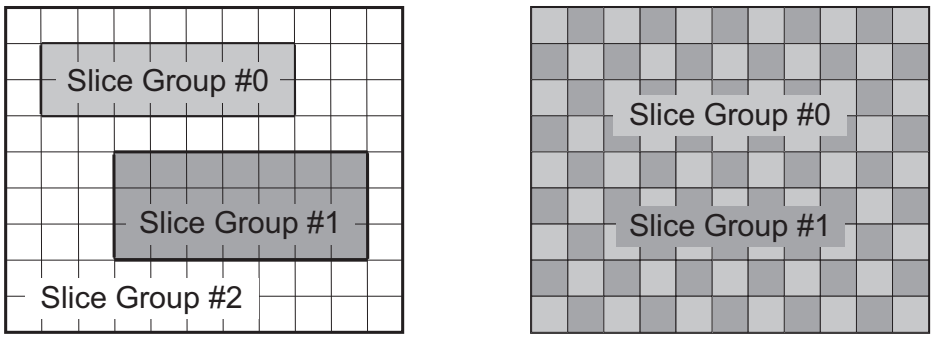

Figure A.4: Subdivision of a QCIF frame into MBs and slices with FMO.

Slices are self-contained. Therefore, the syntax elements and the sample values contained in a slice can be parsed from the bitstream and correctly be decoded without data from other slices of the picture, given the general sequence and picture parameters are known for the stream.

Each slice can be coded using different coding types:

- **I slice:** A slice in which all MBs are coded using intra prediction (spatial prediction, see Fig. A.5 (left)) without using other information than the one contained in the picture itself. Each intra coded MB is predicted using spatially neighboring samples of previously coded MBs. Which and how neighboring samples are used for intra prediction is decided by the encoder (prediction modes) and signalized within the bitstream.

- **P slice:** Some (or all) MBs of the P slice can be coded using inter prediction. These exploit the temporal correlation by referencing the MBs from previous frames and by applying motion compensation (see Fig. A.5 (right)). First, the motion estimation of each block is performed by searching the best matching region from the previous frames. The best match is taken as a prediction of the encoded block. The MV (Motion Vector) linking the position of the block within the frame and the position of its best match in the previously encoded frame has to be signalized together with the corresponding reference frame.

- **B slice:** In addition to the coding types available in a P slice, some MBs of the B slice can also be coded using inter prediction with two motion-compensated prediction signals per block. Temporally preceding as well as successive but previously coded frames can be used as a reference for the *bi-directionally predicted* B frames.

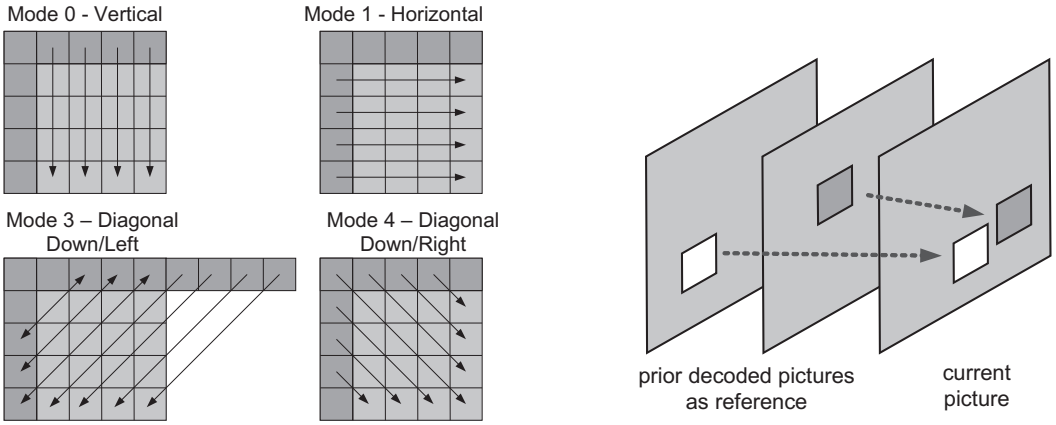

Figure A.5: Examples of 4×4 pixel intra-frame prediction modes (left) vs. inter-frame prediction by using motion compensation (right).

Frames containing I slices (I frames) are used to refresh the sequence. They enable random access and, in case of error prone transmission channels, limit the propagation of errors in time. The set of frames from an I frame up to the P frame preceding the next I frame is called GOP (Group Of Pictures). Inter-coding requires much less associated information elements than intra-coding to encode a frame. Therefore, for a given sequence, the resulting data rate depends strongly on the GOP size (I frame frequency).

Additionally to the above mentioned coding types, H.264/AVC defines two new coding types for slices which allow the exact reconstruction by using different sets of predictors or no predictor at all without the loss of efficiency resulting from sending an I picture. This can enable the switching of a decoder between different representations of the video content (e.g. with different data rates), recovery from data losses or errors, as well as enabling trick modes such as fast-forward, fast-reverse, etc. These new coding types are:

- **SP slice:** A switching P slice is coded such that it can be used by differently pre-coded P frames as a reference so that efficient switching between differently encoded streams becomes possible.

- **SI slice:** Switching I slices allow an exact match of MBs in an SP slice and thus can be used as a reference for successive P pictures but are independent of previously coded pictures. SI slices serve for random access and error recovery purposes.

There are several profiles defined in H.264/AVC as will be presented later on in detail. Note, the *Baseline profile* (recommended by 3GPP [84]) includes I and P slices, but no B, SI and SP slices, which are supported in the *Extended profile*. The *Main profile* includes I, P and B slices but not the SI and SP coding types.

A.1.3 H.264/AVC CODEC

As already mentioned, the H.264/AVC standard does not explicitly define a CODEC (enCOder / DECoder pair). However, in practice, a compliant encoder and decoder are likely to include the functional elements as shown in Fig. A.6 and Fig. A.7. While the basic functional elements (prediction, transform, quantization, entropy encoding) are the same like in previous standards (MPEG1, MPEG2, MPEG4, H.261, H.263), the important novelties in H.264/AVC are in the details of each functional element.

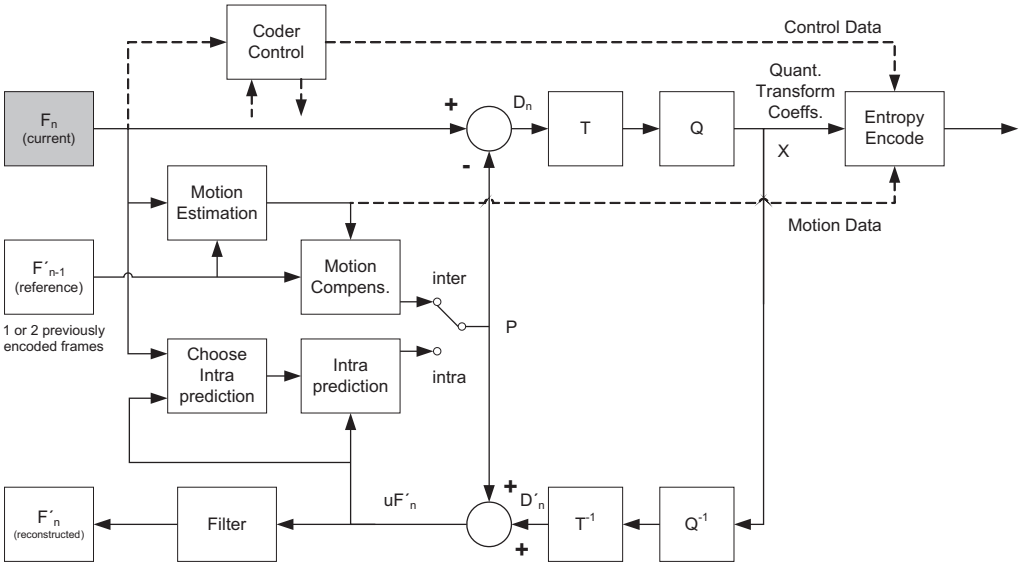

Figure A.6: H.264/AVC Encoder.

In Fig. A.6 it is shown that the encoder basically includes two dataflow paths, a forward path (left to right) and a reconstruction path (right to left).

The encoder processes incoming frames in units of MBs (F_n) where each MB can be encoded with intra prediction or using the inter prediction mode. In both cases, a prediction MB is formed (P). In intra prediction mode, P is created from samples in the current frame n which have previously been encoded and decoded as well as reconstructed (uF'$_n$). In inter prediction mode, P is created by using motion-compensated prediction from one or more previously encoded, decoded and reconstructed reference frames (F'$_{n-1}$). Whereas in the figures, the reference

frame is shown as the previously encoded frame (F'$_{n-1}$), the prediction for each MB can be performed from one or two past or future frames (in time order) that have already been encoded and reconstructed.

The prediction P is subtracted from the current MB to create a residual or difference MB (D$_n$) which is transformed and quantized. The resulting sets of quantized transform coefficients then are entropy coded together with coder control information required to decode the MB. Such side information for example includes prediction mode, QP and motion vector information describing how the MB was motion-compensated. The generated compressed bitstream is passed to the NAL for transmission or storage.

In the reconstruction path of the encoder, the quantized coefficients X are re-scaled (Q^{-1}) and inverse transformed (T^{-1}) to produce a difference MB (D$'_n$) which represents a distorted version of D$_n$ as the quantization process introduces losses.

The prediction MB P is added to D$'_n$ to create a reconstructed MB (uF'$_n$) which again due to the quantization process represents a distorted version of the original MB. After applying a filter for reducing the effects of blocking distortion, a reconstructed reference frame is created from a series of MBs (F'$_n$).

Figure. A.7 shows that the decoder basically consists of the functional elements as included in the reconstruction path of the encoder.

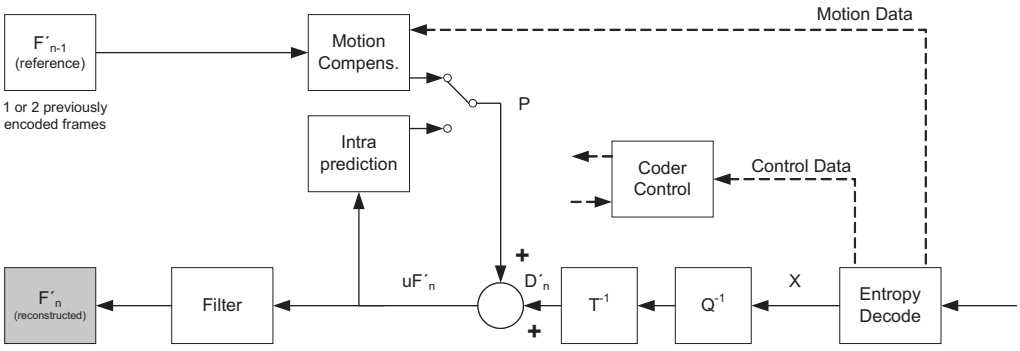

Figure A.7: H.264/AVC Decoder.

In the decoder, first the compressed bitstream received from the NAL is entropy decoded to produce a set of quantized coefficients X and to regain the coder control side information which is necessary for the reconstruction process. The rescaled and inverse transformed quantized coefficients X are then used to produce difference MBs D$'_n$ which are identical to the D$'_n$ shown in the encoder. Using the coder control information and the motion data (motion vectors) for the motion compensation the decoder creates a prediction MB (P) identical to the original prediction P formed in the encoder. P is added to D$'_n$ to produce uF'$_n$ which is filtered to create

the decoded MB (F'$_n$).

Similar to previous video coding standards, H.264/AVC utilizes transform coding of the prediction residual. However, in H.264/AVC, the transformation with similar properties as a DCT (Discrete Cosine Transform) is applied to e.g. 4×4 pixel blocks. The result of the transformation is a matrix of coefficients corresponding to different spatial frequencies. The coefficient corresponding to the lowest frequency is called DC (Direct Component), the others are denoted AC (Alternating Component) coefficients. All coefficients are further quantized. For each MB the quantization is controlled by the QP (Quantization Parameter) ranging from zero to 52 which defines the quantization step size via a mapping to the parameter Q_{step}. The quantized transformation coefficients of such a 4×4 sub MB then are scanned in zig-zag order from the DC component to the AC component representing the highest frequencies as illustrated in Fig. A.8.

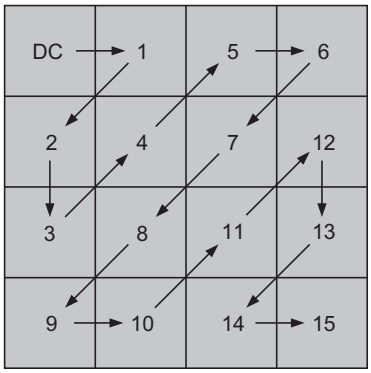

Figure A.8: Zig-zag scan used to scan samples in a submacroblock.

Table A.2 presents some examples of parameters which have to be encoded and transmitted or stored in H.264/AVC.

Parameter	**Description**
Sequence-, picture- and slice-layer syntax elements	Contained in SPS (Sequence Parameter Set) and PPS (Picture Parameter Set)
MB type	Prediction method for each coded MB
QP	Transmitted as a delta value from the previous value of QP
Reference frame index	Identifies reference frame(s) for inter prediction
MV (Motion Vector)	Transmitted as a difference from predicted MV
Residual data	Coefficient data e.g. for each 4×4 sub MB

Table A.2: Examples of parameters to be encoded and transmitted (stored) in H.264/AVC.

H.264/AVC (baseline profile) basically makes use of three different types of entropy coding.

Exponential Golomb code: Syntax elements and other coder control information is coded using an exponential Golomb code [103]. The exponential Golomb code (or simply exp-Golomb code) is a parametric (k) universal code. H.264/AVC uses a special type of the exp-Golomb codes with the parameter k set to zero, also known as Elias-γ encoding [104]. Each exp-Golomb codeword embodies the following elements:

$$\underbrace{0_1 \dots 0_M}_{M} 1 \underbrace{b_1 \dots b_M}_{M}.$$

The first M zeros and the middle one are regarded as *prefix* while the following M bits represent the *info* field. The value of the length M is unary encoded in the prefix. The exp-Golomb coded number is obtained as

$$codeNum = 2^M + \text{info} - 1.$$

Note, an error affecting the leading zeros or the middle '1' affects the decoding by modifying the value of M, therefore causing the misinterpretation of the codeword boundaries and the desynchronization of the decoding process. An error in the *info* field causes deviation of the decoded parameter value, but does not cause desynchronization directly [105].

CAVLC: For compressing quantized transform coefficients, an efficient method called the Content Adaptive Variable Length Coding (CAVLC) is used in baseline profile. The levels of the transform coefficients are encoded using a VLC-N procedure, where N is a parameter from zero to six depending on the value of the previously encoded levels. The first level of each MB is encoded using the VLC-0, the resulting codeword has the following structure:

$$\underbrace{0_1 \dots 0_M}_{M} 1.$$

A codeword encoded with a VLC-N ($N > 0$) procedure has the form:

$$\underbrace{0_1 \dots 0_M}_{M} 1 \underbrace{i_1 \dots i_{N-1}}_{N-1} s.$$

Similarly to exp-Golomb codes, the codeword starts with a sequence of M leading zeros followed by a one, an *info* field consisting of $N-1$ bits and one explicit sign bit s. The encoded value is then obtained as

$$(-1)^s \cdot ((M) << (N-1) + 1 + \text{info}),$$

where << represents the left bit-wise shift operation.

The VLC-0 codewords are highly susceptible to errors since the whole information is contained in the leading zeros. For VLC-N the first $M+1$ bits are critical, errors lying in the *info* field or sign do not cause desynchronization but affect only the decoded level. Errors in the *info* field may also cause the use of a false VLC procedure for the next decoded items [105].

In profiles of H.264/AVC other than the baseline profile, the efficiency for coding the transform coefficients can be improved when using **CABAC (Content-Adaptive Binary Arithmetic Coding)**.

Tabled codewords: H.264/AVC defines several VLC (Variable Length Coding) tables for different syntax elements and contexts. Errors may result in both deviation of the decoded value and decoding desynchronization [105].

A.2 Structure of Video Streams

The NAL of H.264/AVC is specified in a way to enable simple and effective use of the VCL for a broad variety of transport systems or storage media [106]. With the "network friendliness" of the H.264/AVC NAL the VCL data can be mapped onto transport layers such as:

- RTP/IP for real-time conversational and streaming wireline and wireless IP services,
- different file formats, e.g., ISO MP4 for MMS (Multimedia Messaging Service),
- H.32X for wireline and wireless conversational services,
- MPEG-2 systems for broadcasting services, etc.

In this work the focus is on RTP/IP streaming services and the according data structure.

After entropy coding of the video data, the encoded data is partitioned into NAL units. The first byte of such an NALU is a header byte indicating the type of data in the NALU and the payload data is interleaved with *emulation prevention bytes* to avoid the occurrence of a particular data pattern called *start code prefix*. Generally, NALUs can be formed in a byte stream format or for the usage in packet transport systems. The byte stream format is required by some systems (e.g., H.320, MPEG-2/H.222.0) which take the entire NALU stream as a stream of bits where the start of NALUs must be identified by a unique start code prefix. In packet transport systems (e.g., IP/RTP systems) the inclusion of start code prefixes would unnecessarily increase the data overhead as the NALUs are carried in separate packets.

According to the type of data, H.264/AVC builds VCL and non-VCL NALUs, where the VCL NALUs carry the coded video picture data and the non-VCL NALUs contain information associated to the sequence characteristics. Figure A.9 presents the structure of a VCL NALU which contains the encoded MBs belonging to one slice after the slice header. Note, the VLC decoding is restarted at the beginning of each NALU since there is no resynchronization point within an NALU. This also means, an error in the NALU can cause desynchronization of the entropy code in the NALU after the error position up to the end of the slice.

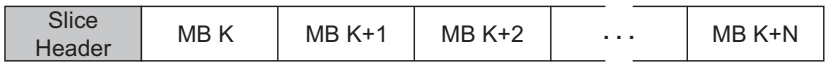

Figure A.9: Structure of a VCL NAL unit.

In order to be capable of decoding the data in VCL NALUs, the information contained in non-VCL NALUs is required as it defines the profile, resolution and information associated to the whole sequence (Sequence Parameter Set - SPS) and the type of entropy coding as well as slice group mapping and quantization properties (Picture Parameter Set - PPS). Each SPS applies to a series of consecutive coded video pictures called a video sequence while the PPS applies to the decoding of one or more individual pictures within a coded video sequence as presented in Fig. A.10. For this purpose, each VCL NALU contains an identifier that refers to the content of the according PPS and each PPS contains an identifier that refers to the relevant SPS content.

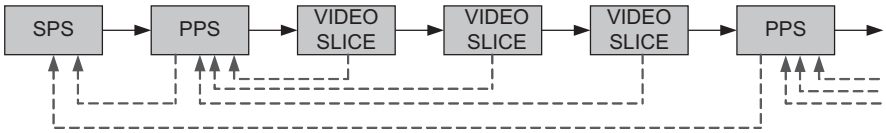

Figure A.10: NALU sequence with PPS (Picture Parameter Set) and SPS (Sequence Parameter Set).

In the mentioned IP/RTP packet oriented environment, each NALU is handled by RTP, UDP and IP protocols and the respective headers are attached to the payload. Figure A.11 shows the encapsulation of a NALU for transmission over IP.

Figure A.11: Encapsulation of a NAL unit for transmission over IP/UDP/RTP systems.

A.3 Profiles and Applications

An H.264/AVC profile includes a set of coding tools and algorithms which are used to compress the video and to generate the standard conform encoded bitstream. Profiles are designed in order to facilitate interoperability between various applications that have similar requirements. Decoders which are conform to a specific profile must support all features of that profile while encoders are not required to make use of all the provided features of a profile but have to create a bitstream conform to the profile.

In H.264/AVC several profiles are defined. The most important ones are the *Baseline Profile*, the *Main Profile* and the *Extended Profile*. Figure A.12 presents the three profiles with the provided features and the overlapping between the different profiles.

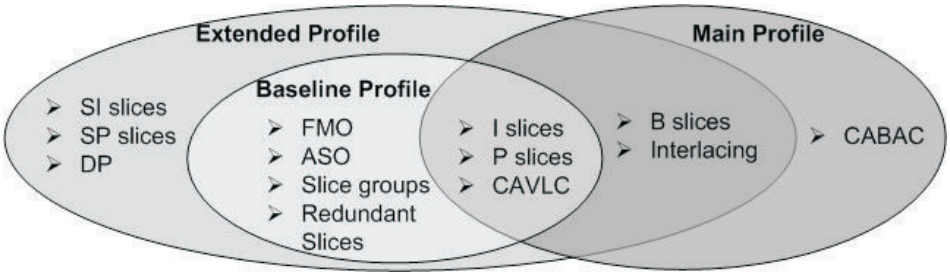

Figure A.12: H.264/AVC *Baseline Profile*, *Main Profile* and *Extended Profile* with the provided features.

The ***Baseline Profile*** provides encoding in I and P slices, enhanced error resilience tools like FMO (Flexible Macro Block Ordering), ASO (Arbitrary Slice Ordering) and redundant slices as well as CAVLC for entropy coding. This profile was designed for applications demanding less computing resources e.g., in video conferencing and mobile services.

The ***Extended Profile*** additionally to the tools offered by the *Baseline Profile* provides B, SP and SI slices as well as DP (Data Partitioning) and offers the possibility of interlaced frames. It was originally intended as the mainstream consumer profile for broadcast and storage applications but now this purpose was taken by the new *High Profiles*.

CABAC is only included in the ***Main Profile*** which was designed to offer the highest possible coding efficiency. It provides features for stream switching and to increase the robustness against data loss.

Additionally to the three mentioned profiles, there are four new profiles for high definition pro-

fessional applications:

The ***High Profile*** was particularly defined for broadcast and disc storage applications with the particular focus on high-definition television. This profile for example was included in HD (High Definition) DVD and Blue-ray Disc media.

The ***High 10 Profile*** features go beyond today's mainstream consumer product capabilities. The profile is based on the *High Profile* with the addition of supporting 10 bits per picture sample.

The ***High 4:2:2 Profile*** and the ***High 4:4:4 Profile*** are focussing on professional applications. Supporting 4:2:2 and 4:4:4 chroma sampling with up to 10 bits and 12 bits precision, respectively, they are built on the *High 10 Profile*.

As H.264/AVC in this thesis is used in UMTS applications, the focus in this work is on the *Baseline Profile* as this is recommended by the 3GPP [84].

Compared to former video coding standards, H.264/AVC offers increased compression efficiency and therefore provides potential for designing new or improving existing applications. A list of applications is presented in the following.

- Conversational services with less than 1 Mbit/s and low latency requirements like H.320 conversational video for ISDN (Integrated Services Digital Network) video conferencing, 3GPP conversational H.324/M services, H.323 conversational services for IP/RTP protocols with best effort strategy and 3GPP conversational IP/RTP services using SIP (Session Initiation Protocol) for session setup.

- Entertainment video applications with 1 Mbit/s to 8 Mbit/s and a latency of 0.5 s to 2 s, utilizing the *Main Profile* for broadcast applications over satellite, cable, terrestrial or DSL (Digital Subscriber Line) as well as standard and high-definition video on DVD and video on demand. The H.222.0/MPEG-2 systems are supporting these applications.

- Streaming services with 50 kbit/s to 1500 kbit/s and about 2 s of latency, making use of the *Baseline Profile* or *Extended Profile*. Applications are for example 3GPP streaming with IP/RTP and RTSP (Real Time Streaming Protocol) for session setup as well as streaming over wired Internet with IP/RTP/RTSP.

- Services at lower bitrates and without delay constraints (as they are distributed via file transfer) like 3GPP MMS (Multimedia Messaging Service) or video mail.

A.4 Video Performance Indicator

The real quality of a video naturally can only be determined by persons watching the video as it is a result of subjective impressions. However, a common objective metric for estimating the user perceptual subjective quality makes use of the MSE (Mean Squared Error) evaluating the distortion within the nth video frame $\underline{\underline{F}}_n$ with respect to the reference (distortion-free) frame $\underline{\underline{R}}_n$ [80]. The MSE is calculated by

$$\mathrm{MSE}\,[n] = \frac{1}{M \cdot N \cdot |\mathcal{C}|} \sum_{c \in \mathcal{C}} \sum_{i=1}^{N} \sum_{j=1}^{M} \left[\underline{\underline{F}}_n^{(c)}(i,j) - \underline{\underline{R}}_n^{(c)}(i,j) \right]^2, \tag{A.4}$$

where $N \times M$ is the size of the frame and $\mathcal{C}$ is the set of components, e.g., for RGB color space $\mathcal{C} = \{R, G, B\}$. The number of color components corresponds to the cardinality $|\mathcal{C}|$ of the set. The individual elements of the color component matrix are addressed by the row index i and the column index j.

For image distortion, the MSE is most commonly quoted in terms of the PSNR (Peak Signal to Noise Ratio) defined for one picture or frame as:

$$\mathrm{PSNR}\,[n] = 10 \cdot log_{10} \frac{(2^q - 1)^2}{\mathrm{MSE}\,[n]}\,[\mathrm{dB}], \tag{A.5}$$

where q represents the number of bits used for sampling the color component values. In YCbCr color space it may often be convenient to compare the distortion of the luminance component only ($\mathcal{C} = \{Y\}$) by means of Y-MSE $[n]$ or Y-PSNR $[n]$, especially for applications that handle all components in the same way.

The PSNR averaged over the whole video sequence can be defined in two ways — as an average $\overline{\mathrm{PSNR}}$ over the PSNR $[n]$ of all its frames, or as

$$\mathrm{PSNR} = 10 \cdot log_{10} \frac{(2^q - 1)^2}{\overline{\mathrm{MSE}}}\,[\mathrm{dB}], \tag{A.6}$$

where $\overline{\mathrm{MSE}}$ is the average MSE defined by

$$\overline{\mathrm{MSE}} = \frac{1}{F} \sum_{n=1}^{F} \mathrm{MSE}\,[n] \tag{A.7}$$

and F is the number of frames in the video sequence.

The formally correct definition is according to (A.6), as it averages over the linear values. On the other hand averaging over logarithmic values results in a systematic error. Jensens's inequality [107] states that if $f(\cdot)$ is a convex function and X is a random variable, then

$$\mathrm{E}\,\{f(X)\} \geq f(\mathrm{E}\,\{X\}), \tag{A.8}$$

where $\mathrm{E}\{X\}$ denotes the expectation of X. A function $f(\cdot)$ is convex over an interval (a, b) if for all $x_1, x_2 \in (a, b)$ and $0 \leq \lambda \leq 1$ applies

$$f(\lambda \cdot x_1 + (1 - \lambda) \cdot x_2) \leq \lambda \cdot f(x_1) + (1 - \lambda) \cdot f(x_2), \tag{A.9}$$

i.e., convexity means that all points of $f(x)$ between the arbitrary chosen $x_1, x_2 \in (a, b)$ lay below the line connecting them.

Due to the fact that $\log_{10}(1/x) = -\log_{10}(x)$ is a convex function, $\overline{\mathrm{PSNR}}$ in general provides results referring to higher quality than PSNR. Both ways of calculating PSNR can be found in literature. However, JM (Joint Model) software outputs $\overline{\mathrm{PSNR}}$ values which is therefore also used in this work for the sake of comparability to JM conform experiments in literature.

Appendix B

List of Abbreviations

3G	3rd Generation Mobile Communication
3GPP	3rd Generation Partnership Project
AC	Alternating Component
AICH	Acquisition Indicator Channel
AM	Acknowledged Mode
AP-AICH	Access Preamble Acquisition Indicator Channel
ASO	Arbitrary Slice Ordering
AVC	Advanced Video Coding
BCCH	Broadcast Control Channel
BCH	Broadcast Channel
BEC	Binary Erasure Channel
BER	Bit Error Ratio
BLER	Block Error Ratio
BMC	Broadcast/Multicast Control
BPSK	Binary Phase Shift Keying
BSC	Binary Symmetric Channel
CABAC	Context Adaptive Binary Arithmetic Coding
CAVLC	Context Adaptive Variable Length Coding
CBR	Constant Bit Rate
CCDF	Complementary Cumulative Distribution Function
CCH	Control Channel
CCCH	Common Control Channel
CCTrCH	Coded Composite Transport Channel
CDF	Cumulative Distribution Function
CD/CA-ICH	Collision-Detection/Channel-Assignment Indicator Channel
CIF	Common Intermediate Format (352×288 pixels)
CN	Core Network
CODEC	enCOder / DECoder
CPCH	Common Packet Channel

CPICH	Common Pilot Channel
CRC	Cyclic Redundancy Check
CS	Circuit Switched
CSICH	CPCH Status Indicator Channel
CTCH	Common Traffic Channel
CUDP	Complete User Datagram Protocol
DC	Direct Component
DCCH	Dedicated Control Channel
DCH	Dedicated Channel
DL	Downlink
DP	Data Partitioning
DPCH	Dedicated Physical Channel
DPCCH	Dedicated Physical Control Channel
DPDCH	Dedicated Physical Data Channel
DSCH	Downlink Shared Channel
DSL	Digital Subscriber Line
DTCH	Dedicated Traffic Channel
DTX	Discontinuous Transmission
DVB-H	Digital Video Broadcasting - Handheld
DVD	Digital Video Disc
E-DCH	Enhanced Dedicated Channel
ECDF	Empirical Cumulative Distribution Function
EIR	Equipment Identity Register
ETSI	European Telecommunications Standards Institute
FACH	Forward Access Channel
FDD	Frequency Division Duplex
FLC	Fixed Length Code
FMO	Flexible Macroblock Ordering
FSMC	Finite State Markov Channel
GEC	Gilbert-Elliott Channel
GERAN	GSM Edge Radio Access Network
GGSN	Gateway GPRS Support Node
Gi	UMTS Core Networ Interface
GMSC	Gateway MSC
Gn	UMTS Core Network Interface
GOP	Group of Pictures
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Repeat Request
HD	High Definition
HLR	Home Location Register
HMM	Hidden Markov Model
HSDPA	High Speed Downlink Packet Access

HS-DSCH	High Speed Downlink Shared Channel
HSPA	High Speed Packet Access
HSUPA	High Speed Uplink Packet Access
HVGA	Half-size Video Graphics Array
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ILPC	Inner Loop Power Control
IMEI	International Mobile Equipment Identity
IMT-2000	International Mobile Telecommunications at 2000MHz
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ISO	International Organization for Standardization
ITU	International Telecommunication Union
ITU-T	ITU - Telecommunication Standardization Sector
ITU-R	ITU - Radiocommunication Standardization Sector
Iub	UMTS Radio Network Interface
IuPS	UMTS Interface RAN-CN
JM	Joint Model
JVT	Joint Video Team
LAN	Local Area Network
MAC	Medium Access Control
MB	Macroblock
MBPI	Macroblock Position Indicator
MCCH	MBMS point-to-multipoint Control Channel
ME	Mobile Equipment
MMS	Multimedia Messaging Service
MPEG	Motion Picture Expert Group
MSC	Mobile Switching Center
MSCH	MBMS point-to-multipoint Scheduling Channel
MSE	Mean Square Error
M-SMMM	Mixed Semi-Markov/Markov Model
MT	Mobile Termination
MTCH	MBMS point-to-multipoint Traffic Channel
MV	Motion Vector
NAL	Network Abstraction Layer
NALU	Network Abstraction Layer Unit
Node B	UTRAN Base Station
OLPC	Outer Loop Power Control
OSI	Open System Interconnection
PB	Playout Buffer
PAL	Phase Alternating Line
PAM	Pulse Amplitude Modulation
PC	Personal Computer

PCCH	Paging Control Channel
P-CCPCH	Primary Common Control Physical Channel
PCH	Paging Channel
PCPCH	Physical Common Packet Channel
PDCP	Packet Data Control Protocol
PDF	Probability Density Function
PDU	Protocol Data Unit
PDSCH	Physical Downlink Shared Channel
PICH	Paging Indicator Channel
PLMN	Public Land Mobile Network
PMF	Probability Mass Function
PPS	Picture Parameter Set
PRACH	Physical Random Access Channel
PS	Packet Switched
PSNR	Peak Signal-to-Noise Ratio
PSTN	Public Switched Telephone Network
QAM	Quadrature Amplitude Modulation
QCIF	Quarter Common Intermediate Format (176×144 pixels)
QoS	Quality of Service
QP	Quantization Parameter
QPSK	Quadrature Phase Shift Keying
QQVGA	Quarter QVGA
QVGA	Quarter Video Graphics Array
RAB	Radio Access Bearer
RACH	Random Access Channel
RAN	Radio Access Network
RB	Radio Bearer
RD	Rate Distortion
RDO	Rate Distortion Optimization
RF	Radio Frequency
RLC	Radio Link Control
RLM	Run Length Model
RNC	Radio Network Controller
RNS	Radio Network Subsystem
RRC	Radio Resource Control
RRC	Root Raised Cosine
RRM	Radio Resource Management
RSCP	Received Signal Code Power
RSSI	Received Signal Strength Indicator
RTCP	Real-Time Control Protocol
RTP	Real-Time Protocol
RTSP	Real-Time Streaming Protocol
SAP	Service Access Point

S-CCPCH	Secondary Common Control Physical Channel
SCH	Synchronisation Channel
SDU	Service Data Unit
SF	Spreading Factor
SGSN	Serving GPRS Support Node
SHCCH	Shared Channel Control Channel
SIF	Standard Interchange Format
SIM	Subscriber Identity Module
SIP	Session Initiation Protocol
SIR	Signal to Interference Ratio
SPS	Sequence Parameter Set
SQCIF	Sub Quarter CIF
TA	Terminal Adapter
TB	Transport Block
TBSS	Transport Block Set Size
TE	Terminal Equipment
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TF	Transport Format
TFCI	Transport Format Combination Indicator
TFCS	Transport Format Combination Set
TM	Transparent Mode
TPC	Transmission Power Control
TrCH	Transport Channel
TTI	Transmission Timing Interval
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
UM	Unacknowledged Mode
UMTS	Universal Mobile Telecommunication System
USB	Universal Serial Bus
USCH	Uplink Shared Channel
USIM	UMTS Subscriber Identity Module
UTRA(N)	Universal Terrestrial Radio Access (Network)
Uu	UMTS Radio Interface
VBR	Variable Bit Rate
VCL	Video Coding Layer
VGA	Video Graphics Array
VLC	Variable Length Coding
VLR	Visitor Location Register
WCDMA	Wideband Code Division Multiple Access
Y-PSNR	Luminance Peak Signal-to-Noise Ratio

Bibliography

[1] `http://www.3gpp.org`

[2] H. Holma and A. Toskala, *WCDMA for UMTS, Radio Access For Third Generation Mobile Communications,* John Wiley & Sons, Ltd., 2004.

[3] Y.J. Liang, J.G. Apostolopoulos, B. Girod, "Analysis of Packet Loss for Compressed Video: Does Burst-Length Matter?" in Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'03), vol. 5, pp. 684–687, 2003.

[4] W. Karner, O. Nemethova, M. Rupp, "The Impact of Link Error Modeling on the Quality of Streamed Video in Wireless Networks," 3rd IEEE Int. Symposium on Wireless Comm. Systems 2006 (ISWCS 2006), Valencia, Spain, Sep. 2006.

[5] M. Zorzi, R.R. Rao, "Perspectives on the Impact of Error Statistics on Protocols for Wireless Networks," IEEE Personal Communications, vol. 6, pp. 32–40, Oct. 1999.

[6] S.A. Khayam, H. Radha, "On the Impact of Ignoring Markovian Channel Memory on the Analysis of Wireless Systems," in Proc. of IEEE Int. Conf. on Communications (ICC 2007), Glasgow, UK, Jun. 2007.

[7] T. Wiegand, G.J. Sullivan, G. Bjontegaard,A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[8] W. Karner, O. Nemethova, P. Svoboda, M. Rupp, "Link Error Analysis and Modeling for Video Streaming Cross-Layer Design in Mobile Communication Networks," ETRI Journal, vol. 29, no. 5, pp. 569–595, Oct. 2007.

[9] W. Karner, M. Rupp, "Measurement based Analysis and Modelling of UMTS DCH Error Characteristics for Static Scenarios," in Proc. of 8th Int. Symposium on DSP and Comm. Systems 2005 (DSPCS'2005), Sunshine Coast, Australia, Dec. 2005.

[10] W. Karner, P. Svoboda, M. Rupp, "A UMTS DL DCH Error Model Based on Measurements in Live Networks," in Proc. of 12th Int. Conf. on Telecomm. 2005 (ICT 2005), Capetown, South Africa, May 2005.

[11] L.M. Correia, *Mobile Broadband Multimedia Networks; Techniques, Models and Tools for 4G*, Elsevier, May 2006.

[12] W. Karner, O. Nemethova, M. Rupp, "Link Error Prediction in Wireless Communication Systems with Quality Based Power Control," in Proc. of IEEE Int. Conf. on Comm. (ICC 2007), Glasgow, Scotland, Jun. 2007.

[13] E.N. Gilbert, "Capacity of a burst-noise channel," Bell Systems Technical Journal, vol. 39, pp. 1253–1265, Sep. 1960.

[14] E.O. Elliott, "Estimates of error rates for codes on burst-noise channels," Bell Systems Technical Journal, vol. 42, pp. 1977–1997, Sep. 1963.

[15] L.R. Rabiner, B.H. Juang, "An Introduction to Hidden Markov Models," IEEE ASSP Magazine, vol. 3, pp. 4–16, Jan. 1986.

[16] J.M. Berger, B. Mandelbrot, "A New model for Error Clustering in Telephone Circuits," IBM Journal, Jul. 1963.

[17] S.M. Sussman, "Analysis of the Pareto Model for Error Statistics on Telephone Circuits," IEEE Trans. on Comm. Systems, vol. 11, pp. 213-221, Jun. 1963.

[18] E.O. Elliott, "A model for the switched telephone network fo data communications," Bell Systems Technical Journal, vol. 44, pp. 89-119, Jan. 1965.

[19] B. Mandelbrot, " Self-Similar Error Clusters in Communication Systems and the Concept of Conditional Stationarity," IEEE Trans. on Comm., vol. 13, no. 1, pp. 71–90, Mar. 1965.

[20] B.D. Fritchman, "A Binary Channel Characterization Using Partitioned Markov Chains," IEEE Trans. on Information Theory, vol. 13, no. 2, pp. 221–227, Apr. 1967.

[21] L.N. Kanal, A.R.K. Sastry, "Models for Channels with Memory and Their Applications to Error Control," Proceedings of the IEEE, vol. 66, no. 7, Jul. 1978.

[22] J-Y. Chouinard, M. Lecours, G.Y. Delisle, "Estimation of Gilbert's and Fritchman's Models Parameters Using the Gradient Method for Digital Mobile Radio Channels," IEEE Trans. on Vehicular Technology, vol. 37, no. 3, Aug. 1988.

[23] C. Pimentel, F. Blake, "Modeling Burst Channels Using Partitioned Fritchman's Markov Models," IEEE Trans. on Vehicular Technology, vol. 47, no. 3, Aug. 1998.

[24] J-P.A. Adoul, B.D. Fritchman, L.N. Kanal, "A critical Statistic for Channels With Memory," IEEE Trans. on Information Theory, vol. 18, no. 1, Jan. 1972.

[25] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, vol. 77, no. 2, Feb. 1989.

[26] W. Turin, M.M. Sondhi, "Modeling Error Sources in Digital Channels," IEEE Journal on Sel. Areas in Comm., vol. 11, no. 3, Apr. 1993.

[27] S. Sivaprakasam, K.S. Shanmugan, "An Equivalent Markov Model for Burst Errors in Digital Channels," IEEE Trans. on Comm., vol. 43, no. 2/3/4, 1995.

[28] W. Turin, *Digital Transmission Systems: Performance Analysis and Modelling*. New York: McGraw-Hill, 1999.

[29] R.P. Aldridge, M. Ghanbari, "Bursty error model for digital transmission channels," IEE Electronic Letters, vol. 31, no. 25, Dec. 1995.

[30] G.T. Nguyen, B. Noble, "A Trace-Based Approach for Modeling Wireless Channel Behavior," in Proc. of the 1996 Winter Simulation Conference, 1996.

[31] V. Tralli, M. Zorzi, "Markov Models for the Physical Layer Block Error Process in a WCDMA Cellular System," IEEE Trans. on Vehicular Technology, vol. 54, no. 6, pp. 2102-2113, Nov. 2005.

[32] H.S. Wang, N. Moayeri, "Finite-State Markov Channel — A Useful Model for Radio Communication Channels," IEEE Trans. on Vehicular Technology, vol. 44, no. 1, Feb. 1995.

[33] A. Willig, "A New Class Of Packet- And Bit-Level Models For Wireless Channels," in Proc. of 13th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Comm. 2002, vol. 5, pp. 2434–2440, Sep. 2002.

[34] A. Köpke, A. Willig, H. Karl, "Chaotic Maps as Parsimonious Bit Error Models of Wireless Channels," in Proc. of IEEE INFOCOM 2003, vol. 1, pp. 513–523, Apr. 2003.

[35] A. Konrad, B.Y. Zhao, A.D. Joseph, R. Ludwig, "A Markov-Based Channel Model Algorithm for Wireless Networks," Wireless Networks, vol. 9, pp. 189-199, Kluwer Academic Publishers, 2003.

[36] J. McDougall, J. Joseph, Yu Yi, S. Miller, " An Improved Channel Model for Mobile and Ad Hoc Network Simulations," in Proc. of Int. Conf. on Comm., Internet and Information Technology (CIIT 2004), St. Thomas, Virgin Islands, USA, Nov. 2004.

[37] J. McDougall, Yu Yi, S. Miller, "A Statistical Approach to Developing Channel Models for Network Simulations," in Proc. of IEEE Wireless Comm. and Networking Conf. (WCNC 2004), vol. 3, pp. 1660–1665, Mar. 2004.

[38] J. Poikonen, "Half-normal Run Length Packet Channel Models Applied in DVB-H Simulations," in Proc. of 3rd IEEE Int. Symposium on Wireless Comm. Systems (ISWCS 2006), Valencia, Spain, Sep. 2006.

[39] J. Poikonen, "Parameterization of Aggregated Renewal Markov Processes for DVB-H Simulations," in Proc. of 18th IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007), Athens, Greece, Sep. 2007.

[40] M. Azimi, P. Nasiopoulos, R.K. Ward, "Offline and Online Identification of Hidden Semi-Markov Models," IEEE Trans. on Signal Processing, vol. 53, no. 8, Aug. 2005.

[41] C.X. Wang, W. Xu, "Packet-Level Error Models for Digital Wireless Channels," in Proc. of IEEE Int. Conf. on Comm. (ICC 2005), vol. 4, pp. 2184–2189, May 2005.

[42] W. Karner, O. Nemethova, M. Rupp, "A Measurement Based Model for UMTS DL DCH Dynamic Bearer Type Switching," in Proc. of 1st IEEE Int. Symposium on Wireless Pervasive Computing (ISWPC 2006), Phuket, Thailand, Jan. 2006.

[43] W. Karner, O. Nemethova, P. Svoboda, M. Rupp, "Link Error Prediction Based Cross-Layer Scheduling for Video Streaming over UMTS," in Proc. of the 15th IST Mobile & Wireless Comm. Summit 2006, Myconos, Greece, Jun. 2006.

[44] O. Nemethova, W. Karner, C. Weidmann, M. Rupp, "Distortion-Minimizing Network-Aware Scheduling for UMTS Video Streaming," invited paper at EUSIPCO 2007, Poznan, Poland, Sep. 2007.

[45] O. Nemethova, W. Karner, A. Al-Moghrabi, M. Rupp, "Cross-Layer Error Detection for H.264 Video over UMTS," in Proc. of Wireless Personal Multimedia Comm. 2005 (WPMC 2005), Aalborg, Denmark, Sep. 2005.

[46] O. Nemethova, W. Karner, M. Rupp, "Error Prediction Based Redundancy Control for Robust Transmission of Video over Wireless Links," in Proc. of IEEE Int. Conf. on Comm. (ICC 2007), Glasgow, UK, Jun. 2007.

[47] W. Karner, O. Nemethova, M. Rupp, "Verfahren und Vorrichtung zum Steuern der paketweisen Übertragung von Daten," Patent: Austria, A 905/2006, submitted Mai 2006.

[48] 3GPP TS 23.002, "Network architecture," v.3.6.0, Sep. 2002.

[49] 3GPP TS 23.107, "Quality of Service (QoS) concept and architecture," v.3.9.0, Sep. 2002.

[50] 3GPP TS 23.207, "End-to-end Quality of Service (QoS) concept and architecture," v.6.6.0, Sep. 2005.

[51] 3GPP TS 25.301, "Radio Interface Protocol Architecture," v.6.4.0, Sep. 2005.

[52] 3GPP TS 25.331, "Radio Resource Control (RRC) protocol specification," v.3.21.0, Dec. 2004.

[53] 3GPP TS 25.323, "Packet Data Convergence Protocol (PDCP) specification," v.6.7.0, Sep. 2006.

[54] 3GPP TS 25.322, "Radio Link Control (RLC) protocol specification," v.6.9.0, Oct. 2006.

[55] 3GPP TS 25.321, "Medium Access Control (MAC) protocol specification," v.6.10.0, Sep. 2006.

[56] 3GPP TR 25.993, "Typical examples of Radio Access Bearers (RABs) and Radio Bearers (RBs) supported by Universal Terrestrial Radio Access (UTRA)," v.4.2.0, Sep. 2006.

[57] 3GPP TS 25.302, "Services provided by the physical layer," v.4.8.0, Sep. 2003.

[58] 3GPP TS 25.201, "Physical layer — General description," v.6.2.0, Jun. 2005.

[59] 3GPP TS 25.212, "Multiplexing and channel coding (FDD)," v.6.9.0, Oct. 2006.

[60] 3GPP TR 25.944, "Channel coding and multiplexing examples," v.4.1.0, Jun. 2001.

[61] 3GPP TS 34.121, "Terminal conformance specification; Radio transmission and reception (FDD)," v.5.4.0, Jun. 2004.

[62] 3GPP TS 25.104, "BS Radio Transmission and Reception (FDD)," v.4.9.0, Mar. 2007.

[63] 3GPP TS 25.213, "Spreading and modulation (FDD)," v.4.4.0, Dec. 2003.

[64] 3GPP TS 25.211, "Physical channels and mapping of transport channels onto physical channels (FDD)," v.3.12.0, Sep. 2002.

[65] http://www.ericsson.com/solutions/tems/

[66] http://www.willtek.com/

[67] ITU-T Rec. M.60, 3008; ITU-T Rec. Q.9, 0222.

[68] F. Navratil, "Fehlerkorrektur im physical Layer des UMTS," Master's thesis (in german language), Institute of Comm. and Radio-Frequency Engineering, Vienna University of Technology, Austria, Nov. 2001.

[69] 3GPP TS 25.101, "User Equipment (UE) radio transmission and reception (FDD)," v.6.12.0, Jun. 2006.

[70] A. Sampath, P.S. Kumar, J.M. Holtzman, "On Setting Reverse Link Target SIR in a CDMA System," in Proc. of 47th IEEE Vehicular Technology Conf., vol. 2, pp. 929–933, 1997.

[71] A. Papoulis, P.S. Unnikrishna, *Probability, random variables, and stochastic processes,* McGraw-Hill, 2002.

[72] C.E. Shannon, "A Mathematical Theory of Communication," Bell Syst. Tech. Journal, vol. 27, pp. 379–423, 623–656, 1948.

[73] P. Elias, "Error-free Coding," IEEE Trans. on Information Theory, vol. 4, no. 4, pp. 29–37, Sep. 1954.

[74] D.N.P. Murthy, M. Xie, R. Jiang, *Weibull Models,* John Wiley & Sons, Ltd., 2004.

[75] A. Lo, G. Heijenk, I. Niemegeers, "Evaluation of MPEG-4 Video Streaming over UMTS/WCDMA Dedicated Channels," in Proc. of the 1st IEEE Int. Conf. on Wireless Internet, 2005.

[76] M.C. Necker, S. Saur, "Statistical Properties of Fading Processes in WCDMA Systems," in Proc. of the 2nd International Symposium on Wireless Communication Systems (ISWCS 2005), pp. 54–58, Sep. 2005.

[77] Qi Qu, Y. Pei, J.W. Modestino, X. Tian, "Source-Adaptive FEC/UEP Coding For Video Transport Over Bursty Packet Loss 3G UMTS Networks: A Cross-Layer Approach," in Proc. of the 60th IEEE Vehicular Technology Conference (VTC2004-Fall), vol. 5, pp. 3150–3154, Sep. 2004.

[78] H.264/AVC Software Coordination, "Joint Model Software," v.10.1, available in http://iphome.hhi.de/suehring/tml/.

[79] M.T. Sun, A.R. Reibman, "Compressed Video over Networks," Signal Processing and Communications Series, Marcel Dekker Inc., New York, 2001.

[80] O. Nemethova, "Error Resilient Transmission of Video Streaming over Wireless Mobile Networks," PhD Thesis, Vienna University of Technology, Austria, 2007.

[81] A. Umbert, P. Diaz, "On the Importance of Error Memory in UMTS Radio Channel Emulation using Hidden Markov Models (HMM)," in Proc. of 15th IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2004), vol. 4, pp. 2998–3002, Sep. 2004.

[82] M. Kac, "On the Notion of Recurrence in Discrete Stochastic Processes," Bulletin of the American Mathematical Society 53, pp. 1002–1010, 1947.

[83] 3GPP TS 26.235, "Packet switched conversational multimedia applications; Default codecs (Rel. 6)," v. 6.4.0, Mar. 2005.

[84] 3GPP TS 26.234, "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs (Rel. 6)," v. 6.10.0, Dec. 2006.

[85] J. Tang, L. Zhang, C.-K. Siew, "An Opportunistic Scheduling Algorithm for MPEG Video Over Shared Wireless Downlink," in Proc. of IEEE Int. Conf. on Comm. (ICC 2006), 2006.

[86] A. Farrokh, V. Krishnamurthy, "Opportunistic Scheduling for Streaming Multimedia Users in High-Speed Downlink Packet Access (HSDPA)," IEEE Trans. on Multimedia, vol. 8, no. 4, Aug. 2006.

[87] V. Srivastava, M. Motani, "Cross-Layer Design: A Survey and the Road Ahead," IEEE Communications Magazine, vol. 43, no. 12, pp. 112–119, Dec. 2005.

[88] ISO/IEC 7498-1, ITU-T Rec. X.200, "Information technology – Open Systems Interconnection – Basic Reference Model: The Basic Model," ICS: 35.100.01, 1994.

[89] M. Welzl, "Passing Corrupt Data Across Network Layers: An Overview of Recent Developments and Issues," Eurasip Journal on Applied Signal Processing, pp. 242–247, 2005.

[90] H. Zheng, J. Boyce, "An Improved UDP Protocol for Video Transmission Over Internet-to-Wireless Networks," IEEE Transactions on Multimedia, vol. 3, no. 3, Sep. 2001.

[91] 3GPP TR 26.937, "Transparent end-to-end packet switched streaming service (PSS); RTP usage model," v.6.0.0, Mar. 2004.

[92] Y. Takishima, M. Wada, H. Murakami, "Reversible Variable Length Codes," IEEE Trans. on Comm., vol. 42, no. 2/3/4, 1994.

[93] A. Kiely, S. Dolinar, M. Klimesh, A. Matache, "Error Containment in Compressed Data Using Sync Markers," in Proc. of Int. Conf. on Information Theory, June 2000.

[94] E. Cianca, F.H.P. Fitzek, M. DeSanctis, M. Bonanno, R. Prasad, M. Ruggieri, "Improving Performance for Streaming Video Services over CDMA-Based Wireless Networks," in Proc. of Int. Symp. on Wireless Personal Multimedia Comm. (WPMC 2004), Padova, Italy, Sep. 2004.

[95] S.H. Kang, A. Zakhor, "Packet Scheduling Algorithm for Wireless Video Streaming," in Proc. of 12th Int. Packetvideo Workshop 2002 (PV 2002), Pittsburgh PA, USA, 2002.

[96] R.S. Tupelly, J. Zhang, E.K.P. Chong, "Opportunistic Scheduling for Streaming Video in Wireless Networks," in Proc. of the Conference on Information Sciences and Systems 2003, Johns Hopkins University, Baltimore, MD, Mar. 2003.

[97] P. Koutsakis, "Scheduling and Call Admission Control for Burst-Error Wireless Channels," in Proc. of 10th IEEE Symp. on Computers and Comm. 2005 (ISCC 2005), 2005.

[98] M. Kalman, B. Girod, " Rate-Distortion Optimized Video Streaming Using Conditional Packet Delay Distributions," in Proc. IEEE MMSP, Siena, Italy, Sep. 2004.

[99] ITU-T Rec. H.264 and ISO/IEC 11496-10 (MPEG-4), "AVC: Advanced Video Coding for Generic Audiovisual Services," v. 3, 2005.

[100] I.E.G. Richardson, *H.264 and MPEG-4 Video Compression, Video Coding for the Next-Generation Multimedia*, John Wiley & Sons Ltd., Mar. 2005.

[101] D. Marpe, T. Wiegand, G.J. Sullivan, "The H.264/MPEG4 Advanced Video Coding Standard and its Application," IEEE Communications Magazine, vol. 44, no. 8, pp. 134–143, Aug. 2006.

[102] ITU-R BT.601-6, Series BT: Broadcast service (television), "Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios," Jan. 2007.

[103] S.W. Golomb, "Run-length encodings," IEEE Trans. on Information Theory, vol. 12, pp. 399–401, Jul. 1996.

[104] P. Elias, "Universal Codeword Sets and Representations of the Integers," IEEE Trans. on Information Theory, vol. 21, no. 2, Mar. 1975.

[105] L. Superiori, O. Nemethova, M. Rupp, "Performance of an H.264/AVC Error Detection Algorithm based on Syntax Analysis," accepted for Journal of Mobile Multimedia, Rinton Press.

[106] T. Stockhammer, M.M. Hannuksela, T. Wiegand, "H.264/AVC in wireless environments," IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, pp. 657–673, Jul. 2003.

[107] D.J.C. MacKay, *Information Theory, Inference, and Learning Algorithms,* Cambridge University Press, fifth printing, 2006.