Bayerische Julius-Maximilians-Universität Würzburg
Institut für Informatik
Lehrstuhl für Verteilte Systeme
Prof. Dr. P. Tran-Gia

Performance Models for UMTS 3.5G Mobile Wireless Systems

Andreas Mäder

Würzburger Beiträge zur
Leistungsbewertung Verteilter Systeme

Bericht 02/08

Würzburger Beiträge zur Leistungsbewertung Verteilter Systeme

Herausgeber

Prof. Dr. P. Tran-Gia
Universität Würzburg
Institut für Informatik
Lehrstuhl für Verteilte Systeme
Am Hubland
D-97074 Würzburg

Tel.: +49-931-888-6630
Fax.: +49-931-888-6632
email: trangia@informatik.uni-wuerzburg.de

Satz

Reproduktionsfähige Vorlage vom Autor.
Gesetzt in LaTeX Computer Modern 9pt.

ISSN 1432-8801

Performance Models for UMTS 3.5G Mobile Wireless Systems

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Bayerischen Julius–Maximilians–Universität Würzburg

vorgelegt von

Andreas Mäder

aus

Würzburg

Würzburg 2008

Eingereicht am: 5.6.2008
bei der Fakultät für Mathematik und Informatik
1. Gutachter: Prof. Dr.-Ing. P. Tran-Gia
2. Gutachter: Prof. Dr. U. Körner
Tag der mündlichen Prüfung: 25.7.2008

Für Luca und Mara

Danksagung

Die vorliegende Dissertation ist das Ergebnis meiner mehrjährigen Forschungstätigkeit am Lehrstuhl für Informatik III an der Universität Würzburg. Die Arbeit wäre in dieser Form nicht ohne die tatkräftige Hilfe derjenigen Menschen möglich gewesen, denen ich nun im Folgenden ausdrücklich danken will.

An prominenter Stelle danke ich meinem Doktorvater Prof. Dr. Phuoc Tran-Gia, der mir Gelegenheit und Freiraum gab, mich in ein interessantes Thema einzuarbeiten. Seine vielfältigen Anregungen und Hinweise waren mir Inspiration und Stütze. Weiterhin danke ich Prof. Dr. Ulf Körner für die ermutigenden und konstruktiven Kommentare zu dieser Arbeit. Mein Dank gilt auch den Mitgliedern der Prüfungskommission, Prof. Dr. Schlling und Prof. Dr. Seipel.

Meine Zeit am Lehrstuhl für Informatik III war geprägt von herausfordernden Aufgaben, aber auch besonders von der freundschaftlichen Atmosphäre unter den Kollegen und Kolleginnen. Ich danke Dr. Michael Menth und Prof. Kurt Tutschku, die als erfahrene Kollegen immer ein offenes Ohr für meine Anliegen hatten, Dr. Rüdiger Martin und Dr. Andreas Binzenhöfer, die nahezu zeitgleich mit mir ihre Arbeit am Lehrstuhl aufgenommen haben, Robert Henjes und Tobias Hossfeld für die fachliche und moralische Unterstützung in vielfältiger Form, Rastin Pries und Simon Oechsner für die vielen interessanten, nicht ausschließlich fachlichen Gespräche, Barbara Staehle und Thomas Zinner für ihre Sicht der Dinge, sowie Michael Duelli, Matthias Hartmann und Frank Lehrieder, die kurz Fertigstellung dieser Arbeit am Lehrstuhl aufgenommen wurden. Besonders danke ich auch Dr. Dirk Staehle für die vielen Stunden der freundschaftlichen Diskussion, ohne die es diese Arbeit in der vorliegenden Form nicht geben würde. Schliesslich danke ich Gisela Förster, die in ihrer kompetenten und freundlichen

Art die Arbeit am Lehrstuhl deutlich vereinfacht hat. Weiterhin möchte ich Tuo Liu, Dr. Hans Barth und Dr. Bernd Schröder danken, die als Gastwissenschaftler und Projektpartner für viele fruchtbare Diskussionen zur Vefügung standen.

Nun möchte ich noch denjenigen danken, ohne die meine Arbeit nicht möglich gewesen wäre. Meinen Eltern, Dr. Uwe Mäder und Dr. Isolde Mäder-Kruse, die mich immer bedingungslos unterstützt haben, sowie meinen Schwestern Marisa, Elena und Miriam. Mein tiefer Dank gilt auch Andrea Meyerhöfer für die vielfältige Unterstützung. Ich widme diese Arbeit meinen wunderbaren Töchtern Luca und Mara.

Acknowledgments

This dissertation is the result of my research activities of several years at the Department of Distributed Systems at the University of Würzburg. In the following, I express my gratitude to the persons which supported me during the course of this work by providing inspiration and advice.

First, I want to thank my supervisor Prof. Dr. Phuoc Tran-Gia for his strong support as well as for the opportunity and academic freedom to choose an interesting and challenging topic. Furthermore, I thank Prof. Dr. Ulf Körner for his encouraging and constructive remarks on this work. I am also thankful to the members of the examination committee, Prof. Dr. Schilling and Prof. Dr. Seipel.

The years at the Department for Distributed Systems were full of challenging tasks. However, the friendly, cooperative and professional environment created by my colleagues faciliated the daily work strongly. My gratitude goes to Dr. Michael Menth and Prof. Dr. Kurt Tutschku. With their long experience at the department, they where always available for giving a novice advice. Special thanks to Dr. Rüdiger Martin and Dr. Andreas Binzenhöfer for sharing the long, sometimes rocky way from start. Many thanks to Robert Henjes and Tobias Hossfeld for the long discussions on many occasions and in many locations. The same applies to Rastin Pries and Simon Oechsner. I further would like to thank Barbara Staehle and Thomas Zinner for sharing their valuable views with me. I also thank Michael Duelli, Frank Lehrieder, and Matthias Hartmann. I am especially thankful to Dr. Dirk Staehle for hours of friendly and enriching discussions. Many lessons have been learned. Finally, I want to thank Gisela Förster for her kind and competent support.

Among the many project partners and guest scientist I had the privilege to meet during my time at the department, I especially thank Tuo Liu, Dr. Hans Barth and Dr. Bernd Schröder for the fruitful discussions and cooperative work.

My deepest gratitude is destined for my family for their unconditional support: Dr. Uwe Mäder and Dr. Isolde Mäder-Kruse, as well as my sisters, Marisa, Elena, and Miriam. Finally, I devote this work to my wonderful and beloved daughters Luca and Mara.

Contents

1 Introduction

The world of mobile communications is rapidly changing. New technologies emerge at the market with increasing pace, both from standardization bodies within the telecommunication community like the International Telecommunication Union (ITU) as well as on the initiative of the computer and chip industry. While the former wants to evolve existing third generation (3G) telecommunication standards like the Universal Mobile Telecommunication System (UMTS), the latter have the possibility to design completely new systems from scratch, like the IEEE 802.16 WIMAX standard.

The driving force behind these activities is the Internet. Mobile telecommunication operators must react on the great success of disruptive technologies like peer-to-peer communication and Web 2.0 applications. Users demand ubiquitous access to these new communication platforms. This requires more bandwidth and an optimized system design. Concurrently, considerable efforts are made to establish packet-switched communication with the Internet Protocol (IP) as a universal principle of end-to-end communication. IP-based vertical integration promises new opportunities to extend classical services like voice and to establish new services on the telecommunication market. A well known manifestation of these efforts is IP-television (IPTV), which is promoted by fixed-network operators for some time.

A further impulse is generated by the technological progress of the mobile device manufacturers. Up-to-date cellular phones integrate various devices in one, like a photo and video camera, a music player, or a GPS navigation system. Furthermore, smart phones blur the line between traditional cellular phones and computers. The increased capabilities of these new devices in terms of media,

memory and computing power enables new services. However, the data volume created by applications like for example video recording requires higher bandwidths for a satisfying service quality.

3G communication systems like UMTS or cdma2000 provide considerable support for IP traffic, but are designed and optimized for circuit-switched data like voice or video telephony. This prevents an efficient usage of radio resources and obstructs the development and deployment of new services.

Mobile telecommunication systems of the 3.5th generation (3.5G) constitute a first step towards the requirements of an all-IP world. As the denotation suggests, 3.5G systems are not completely new designed from scratch. Instead, they are evolved from existing 3G systems like UMTS or cdma2000. 3.5G systems are primarily designed and optimized for packet-switched best-effort traffic, but they are also intended to increase system capacity by exploiting available radio resources more efficiently. Systems based on cdma2000 are enhanced with 1xEV-DO (EV-DO: evolution, data-optimized). In the UMTS domain, the 3G partnership project (3GPP) specified the High Speed Packet Access (HSPA) family, consisting of High Speed Downlink Packet Access (HSDPA) and its counterpart High Speed Uplink Packet Access (HSUPA) or Enhanced Uplink. The focus of this monograph is on HSPA systems, although the operation principles of other 3.5G systems are similar.

1.1 Contribution

New opportunities also arise new challenges. HSPA systems will become integrated into existing UMTS networks. This increases their capabilities and capacity, but also their complexity. Although the HSPA specification is completed, the behavior, performance, and interaction with UMTS not fully understood. For operators, this is a barrier for the inclusion of HSPA in the network planning and optimization process, and ultimately also for deployment.

In this monograph, we approach to operators and equipment manufacturers

with tools to analyze integrated, 3.5G-enabled UMTS networks. One of the main contributions of our work are performance models which allow a holistic view on the system. The models consider user traffic on flow-level, such that only on significant changes of the system state a recalculation of parameters like bandwidth is necessary. The impact of lower layers is captured by stochastic models. This approach combines accurate modeling and the ability to cope with computational complexity. Adopting this approach to HSDPA, we develop a new physical layer abstraction model that takes radio resources, scheduling discipline, radio propagation and mobile device capabilities into account. Together with models for the calculation of network-wide interference and transmit powers, a discrete-event simulation and an analytical model based on a queuing-theoretical approach are proposed. For the Enhanced Uplink, we develop analytical models considering independent and correlated other-cell interference.

We further investigate the interaction between radio resource management schemes, 3.5G and 3G radio bearers, and its impact on user and system performance. We compare different resource assignment and reservation schemes for HSDPA and show their impact on the different types of radio bearers. We unveil the sensitivity of the system to reservation parameters and discuss the relationship between transmit power allocation schemes, interference and performance. For the Enhanced Uplink, we evaluate different scheduling disciplines and discuss the performance trade-off of a preemptive admission control scheme in terms of blocking and dropping probabilities.

Finally, we gained insights on the interaction between best-effort traffic and different HSDPA scheduling disciplines. We show that scheduling disciplines that maximize cell throughput in static scenarios are not necessarily optimal as soon as traffic dynamics are considered. This confirms our position that user behavior must not be neglected when evaluation system performance.

1.2 Outline

This monograph is structured into three main chapters. In Chapter 2, the fundamental principles of UMTS, HSDPA, as well as Enhanced Uplink are explained. The chapter starts with a short overview of the historical development of data communication in digital mobile networks. Then, UMTS fundamentals like Wideband CDMA, network architecture and radio resource management are introduced. Finally, different aspects of HSDPA and Enhanced Uplink are explained, providing the background for the subsequent performance evaluation chapters.

Chapter 3 addresses the performance of HSDPA. At first, we provide an overview of related work in this field of research. Then we explain different concepts of radio resource sharing between HSDPA and UMTS radio bearers and formulate a mathematical model for the calculation of NodeB transmit powers and interference. The HSDPA physical layer abstraction model uses these as input parameters for an approximation of the user throughput. All these components together are then used to investigate the flow-level performance with a discrete-event simulation and an analytical queuing model.

In Chapter 4, we propose performance models for the Enhanced Uplink. We first introduce a model for the calculation of the network-wide interference, which is subsequently used to explain the fundamental problem of system feasibility with constraints like transmit power at the mobile stations. Subsequently, we introduce capacity models considering uncorrelated other-cell interference to evaluate scheduling and admission control strategies. This model is then enhanced by taking the mutual dependency between other and own-cell interference explicitly into account.

Finally, this monograph is concluded by Chapter 5, where we summarize our work and draw conclusions from the insights we found during the course of this work.

2 Basics of 3.5G UMTS Networks

This chapter introduces the basic principles of the 3G Universal Mobile Terrestrial Telecommunication System and its 3.5G enhancements High Speed Downlink Packet Access and Enhanced Uplink. The first section explains the fundamentals of the UMTS air interface which implements Wideband Code Division Multiplex (WCDMA). In Section 2.2, we explain the shortcomings of UMTS regarding transportation of Internet traffic. Sections 2.3 and 2.4 explain the design and the operation principles of HSDPA and Enhanced Uplink, respectively. Finally, we illustrate network scenarios used for the performance evaluation in the next chapters in Section 2.5.

2.1 Introduction to UMTS

UMTS has been adopted by the International Telecommunication Union (ITU) in the course of the IMT[1]-2000 initiative. The goal of the initiative was the development of a global common standard for wireless mobile communications of the third generation as successor for GSM (Global System for Mobile Communication) and GPRS (General Packet Radio System). Several proposals which met the requirements of the initiative have been filed in to the ITU, one of them was UMTS, developed by the 3rd Generation Partnership Project (3GPP). The goal

[1] International Mobile Telecommunications

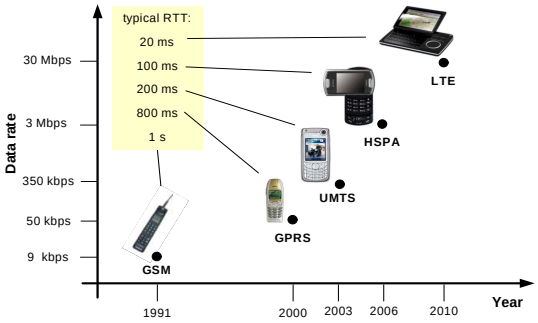


Figure 2.1: *Evolution of data services in digital mobile wireless networks.*

of the 3GPP standardization body was to evolve the existing GSM to enable services like video conferencing or streaming. Consequently, the first releases of the UMTS standard maintained a large part of the GSM architecture, but specified new air interface standards which provide higher data rates and lower latencies.

The first operational release of the UMTS standard was published in 1999 and is known as Release '99. It specifies two different modes, frequency division duplex (FDD) and time division duplex (TDD). The TDD mode is primarily used in fixed wireless deployments for Internet access as DSL substitute. Most mobile network operators deploy FDD mode equipment. This monograph refers to the FDD mode only.

Release '99 was primarily designed for real-time applications like voice and video telephony, and streaming multimedia services. Release '99 is therefore oriented along a circuit-switched paradigm, which means that radio bearers are primarily designed to provide constant a Quality of Service (QoS). Traffic is transported on Dedicated Channels (DCH) which are exclusively reserved for one connection. DCH radio bearers only have marginal capabilities to react on time-varying traffic demands which are characteristic for Internet traffic.

The increasing demand of the telecommunication market for broadband mo-

bile Internet, and the trend in the telecommunication industry towards a unification and integration of services in packet-switched transport networks, initiated a development process towards an all-IP packet-switched concept. UMTS Release 5, approved in 2002 by the 3GPP consortium, is influenced by this new paradigm. The specification enables end-to-end packet switching with the IP Multimedia Subsystem (IMS), and defines the High Speed Downlink Packet Access for broadband mobile Internet access in downlink direction. With Release 6 in 2004 the specification of the Enhanced Uplink[2] or High Speed Uplink Packet Access (HSUPA) followed. So, Release 5 and 6 completed the introduction of a set of new specifications optimized for packet-switched traffic which together form the High Speed Packet Access (HSPA) service. In 2006, UMTS Release 7 introduced refinements and additional features for HSPA like Multiple Input Multiple Output (MIMO) support. The next major step will be Release 8, which is also known as Long Term Evolution (LTE). It will introduce a completely new air interface with higher data rates (up to 100 Mbps peak rates) and lower latencies, and an evolved radio access network with a flat hierarchy.

Figure 2.1 gives a graphical impression of the evolution of data services in digital mobile communication systems. The figure shows two system properties which are most important for the performance of packet-switched applications: data rate and round trip time, the latter roughly defined as two times the packet latency to an Internet host. The values on the abscissa refer to the date of market introduction. We see that the typical data rates increase almost exponentially with the release dates.

2.1.1 UMTS Architecture

The UMTS architecture is derived from its predecessors GSM and GPRS, which allows operators to reuse and update their existing network infrastructure without the requirement to invest in a completely new network from scratch. The largest difference can naturally found in the radio access network (RAN), while the core

[2] We will use this denomination in the remainder

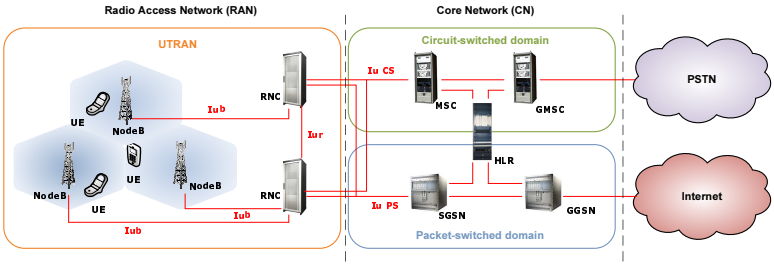


Figure 2.2: *Architecture of a UMTS-Network [30].*

network (CN) part is more or less inherited from GSM and GPRS. Figure 2.2 outlines the logical and functional elements of a UMTS network.

The core network comprises a transport network of typically considerable dimension and the functional entities which provide authentication, authorization and accounting (AAA), session management, network management, gateways to external networks, etc. It is further partitioned in a circuit-switched (CS) domain and a packet switched (PS) domain. The CS domain handles UMTS Release'99 circuit-switched connections like voice or video calls between mobile users and to users in external networks, like the public switched telecommunication network (PSTN). CS connections traverse the mobile switching center (MSC), gateway MSC (GMSC) and enter the UTRAN via the Iu CS interface. Both MSC and GMSC are connected to the home location register (HLR), which stores user and accounting data. The packet switched domain is inherited from GPRS, which is the reason why the logical entities are denoted as Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN). The former adopts the role of the MSC and the latter of the GMSC for packet switched connections. The GGSN also serves as border gateway to the Internet. In UMTS Release '99, packet-switched connections are used for connections to the Internet and are often assumed to carry best-effort traffic, regardless of the application layer protocol. This effectively prevents the mapping of any QoS requirements to

its pendants on UMTS side. In later releases, IMS is introduced to overcome this disadvantage (see e.g. [31]).

The UMTS terrestrial RAN (UTRAN) is responsible for all radio related tasks like radio resource and connection management. The radio network controller (RNC) manages the radio resources for Release '99 connections within its domain. One RNC controls several NodeBs (base stations) which are connected to the controlling RNC via the Iub interface. The RNCs are further connected to adjacent RNCs via the Iur interface for signaling of hand-over information. Finally, the NodeBs constitute the connection point for the mobile stations. In Release '99, the primary task of the NodeBs is signal processing, but with Release 5 and the introduction of HSDPA, the NodeBs are also perform radio resource management (RRM). RRM for HSDPA connections has been relocated to the NodeBs since otherwise the signaling delay would be too large to react timely on radio channel and resource fluctuations. The same argument also holds for the Enhanced Uplink, where the NodeB has a similar role as for the HSDPA.

The third functional domain, not indicated in Figure 2.2, comprises the mobile stations which are known as user equipments (UEs) in the UMTS terminology. The logical interface between UE and NodeB is the Uu interface.

The impact of the HSPA enhancements on the UMTS network architecture is with some exceptions restricted to the radio access network and the user equipments. In the RAN, the relocation of the RRM functionality for HSDPA to the NodeBs requires a flow control mechanism on the Iub interface in order to prevent buffer overflows [32]. The RNC still has influence on the radio resources for HSDPA and Enhanced Uplink connections, but more indirectly by preemption because of circuit-switched connections, or by enforcing long-term policies like resource reservation schemes. Most of these requirements can be fulfilled with software updates, although in practice signal processing capabilities (performed by channel elements, see e.g. [8]) in the NodeBs and often the Iub interface capacity may require enhancements as well. However, on UE side, new hardware is mandatory.

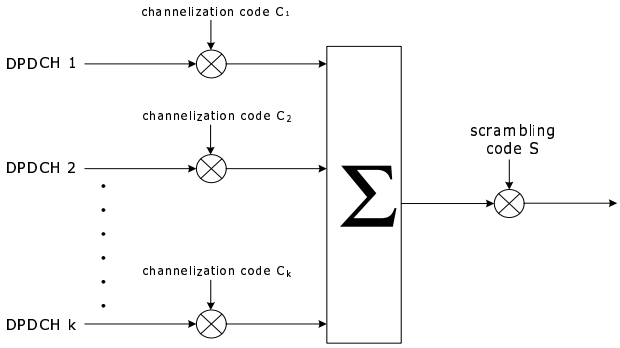


Figure 2.3: *Spreading and scrambling in WCDMA. Both operations are performed in the NodeB and in the user equipment.*

2.1.2 Wideband Code Division Multiple Access

Wideband CDMA is the air interface of UMTS. It is a direct-sequence code division multiple access scheme (DS-CDMA). Channel signals are separated by multiplication of the original symbols with "chip"-sequences generated from CDMA spreading codes. This process is called spreading, because the narrowband signal power is spread over a wider frequency band depending on the system chip rate, i.e. the number of chips per time interval. In WCDMA, the chip rate is 3.84 Mcps, which corresponds approximately to a system bandwidth of 5 Mhz.

Spreading in UMTS comprises two steps: first the multiplication of a symbol with a channelization code, and second the randomization of the resulting chip sequence with a pseudo-random scrambling code. Figure 2.3 shows a schematic view of the spreading process. Spreading and scrambling is in principle performed equally for all types of physical channels in uplink and downlink direction.

The result of the first operation is a sequence of chips, which represents one symbol. Channelization codes are orthogonal $+1/-1$-sequences which are as-

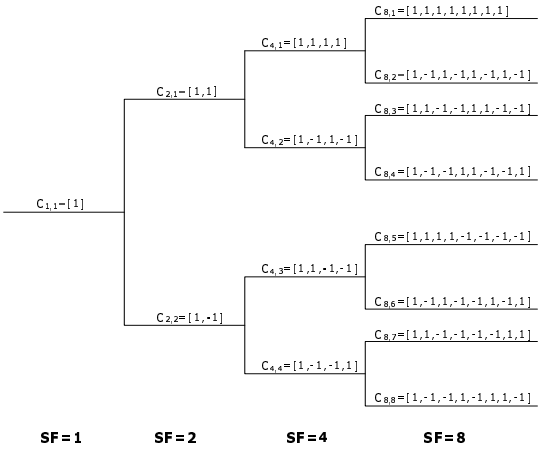


Figure 2.4: *A part of the OVSF code tree. Orthogonality does exist between disjunct sub-branches only.*

signed to a physical transport channel. Orthogonality means that a scalar multiplication of two or more code sequences C_x results in zero: $C_1 \cdot C_2 \cdot \ldots \cdot C_n = 0$. Channelization codes are orthogonal variable spreading factor (OSVF) codes which can be constructed from Walsh matrices. The length of a channelization code corresponds to its spreading factor, i.e. the number of chips which are used to represent one symbol. Different gross data rates in Release '99 UMTS are realized by using different spreading factors. For example, a spreading factor of $SF = 8$ means that for one symbol 8 chips are used. With a system chip rate of 3.84 Mcps and binary phase shift keying (BPSK) modulation, the resulting gross data rate is then $3.84\,\text{Mcps}/8\,\text{chips} = 480\,\text{kbps}$.

OSVF codes are often represented in form of a tree as shown in Figure 2.4, accurately reflecting some of their properties. First, the number of orthogonal codes corresponds to their spreading factor. This means that for example 8 codes with a spreading factor of $SF = 8$ exist. Second, orthogonality does only exist in

disjunct paths of the tree starting from the root to the leafs. That means if a code is used, all successors in root and leave direction are effectively blocked. OSVF codes are used in the UMTS downlink and uplink direction. In the downlink, they are used to separate physical channels within one cell from each other. In the uplink, OSVF codes are used to separate different channels in a *single* mobile.

Scrambling is performed to randomize the chip sequence resulting from channelization, and for signal identification. Randomization prevents the reception of correlated signals, which would lead to significant performance losses due to interference. Scrambling codes are quasi-random chip sequences which are pseudo-orthogonal to each other. Unlike channelization codes, the inner scalar product of two scrambling codes is not necessarily zero. However, for long sequences of length n it holds that on average $E[S_1 \cdot S_2] = 0$ and VAR$[S_1 \cdot S_2] = n$. Additionally, low auto-correlation of scrambled chip sequences reduce inter-symbol interference due to multipath reception, since a scrambled signal is "nearly orthogonal" to a time shifted copy of itself.

In the downlink, a unique primary scrambling code and 15 secondary scrambling codes are assigned to each cell. The scrambling codes used in the downlink have a length of 38400 chips, which are taken from a set of 8192 available codes from which 512 are primary scrambling codes, and the rest associated secondary scrambling codes [33]. The limited number of downlink scrambling codes leads to the necessity of scrambling code planning [34], although due to the relatively large number of codes this task not as complex as frequency planning for GSM networks. In the uplink, scrambling codes are used to distinguish signals from different users, which means that every UE uses a different scrambling code. There are 2^{24} long and 2^{24} short scrambling codes available, whereas the long codes have a length of 38400 chips like in the downlink, and the short codes are 512 chips long.

User signal recovery at the receiver is done by reversing the transmission chain: the incoming signal is de-scrambled and then correlated (i.e. multiplied) with the appropriate channelization code. With perfect orthogonality, the sum of the appropriate chips per symbol is either zero, if the code of the received signal

and the multiplied code does not match, or the code weight (i.e. the spreading factor), if both codes are identical. With imperfect or pseudo-orthogonality or in the presence of noise, the result is not equal to the code weight, and the receiver decides which symbol was transmitted by determining in which decision region the result falls [35].

2.1.3 Radio Resource Management for Dedicated Channel Radio Bearers

In UMTS Release '99, DCH radio bearers are the most commonly used transport channel for all types of data, including packet-switched traffic. DCH radio bearers are circuit-switched, which means that a certain amount of radio resources are exclusively reserved for a connection. Each DCH radio bearer is associated with Quality of Service (QoS) attributes which are defined in [36]. These attributes are for example frame error rate, residual bit error rate, guaranteed bit rate, and transfer delay. The attributes itself and their values depend on the traffic class the bearer belongs to. UMTS distinguishes between four service classes (conversational, streaming, interactive and background) which are intended for different types applications (for example voice for conversational).

A DCH connection is therefore defined by a set of attributes which are static or semi-static during the life time of a connection. The QoS attributes define physical attributes like channelization codes, ratio of information bits to total bits and transport time interval (TTI), i.e. the time it takes to transmit one MAC frame. The task of radio resource management for DCH radio bearers is now to provide a connection quality that complies with the negotiated QoS attributes. This is accomplished by ensuring that the received energy-per-bit-to-interference ratio, E_b/N_0[3], for each user is sufficiently high. For this reason, each DCH connection is associated with a target-E_b/N_0 value. *Power Control* ensures that received

[3]Note that for historical reasons, N_0 in this "formula" comprises multiple access interference *and* thermal noise and should not be confused with the thermal noise density introduced in Equation (2.1).

E_b/N_0 values and target-E_b/N_0 values match. The E_b/N_0 is defined as

$$\varepsilon_k = \frac{W}{R_k} \cdot \frac{T_k \cdot d_k}{W \cdot N_0 + I_0}, \tag{2.1}$$

where ε_k is the received E_b/N_0-value for a mobile k, R_k is the information data rate, T_k is the transmit power and d_k is the path gain between UE and NodeB. Note that the second fraction on the right hand side constitutes the signal-to-interference ratio (SIR), while the first fraction is the *processing gain* which includes the *spreading gain* from signal despreading. Both path gain and interference are time-varying random variables.

Power Control Transmit power control is implemented in the uplink and in the downlink. It consists of two control loops, the *inner loop power control* (or fast power control) and *outer loop power control.*

Inner loop power control compares the instantaneously received E_b/N_0 value with the target E_b/N_0 value and adjusts the transmit power accordingly. This is performed fifteen times per frame, i.e. with a rate of 1.5 kHz. Fast power control enables compensates fast fading if the mobile moves with moderate speed. Signaling consists of power control commands (UP, DOWN) which indicate whether the transmitter should increase or decrease transmit power by a discrete value, typically ± 1 dB. The granularity and the signaling rate leads to a certain power control error such that the received E_b/N_0 values fluctuate around the target E_b/N_0 value. This is called imperfect power control. *Perfect power control* describes the assumption that received and target E_b/N_0 values always match exactly. Inner loop power control keeps the received signal power at the receiver at a predefined level and thus neutralizes the "near-far-effect" in the uplink, which occurs if a mobile station close to the antenna has the same transmit power as a mobile farther away and therefore generates more interference.

Supplementary, *outer loop power control* adjusts the target E_b/N_0 to the predefined QoS requirements. Outer loop power control is performed in the RNC by comparing performance measurements like residual bit error rate to the QoS requirements. The connection quality may for example diminish if the mobile

speed, and therefore also the fast power control error, increases. Outer loop power control would then compensate by increasing the target E_b/N_0 value of the inner loop power control. Analogously, the target E_b/N_0 value is decreased if the QoS requirements are overachieved, thus preventing waste of radio resources. Specifically, outer loop power control leads to lower target E_b/N_0 values if the mobile is in soft or softer hand-over.

Admission and Load Control *Admission Control* (AC) is performed if a user wants to establish a new connection to the network. Generally, AC is responsible for keeping the consumption of radio resources within the perimeters of the system. DCH admission control must therefore consider interference, power and channelization codes at the NodeBs depending on the direction of the bearer.

Interference and path loss determines the amount of transmit power which is required to match the target-E_b/N_0 value. If the interference becomes too high, the required transmit power may exceed the maximum allowed transmit power which leads to signal impairments and outage. In the downlink, the amount of interference at a UE depends on its location in the cell, on the propagation channel and the resulting orthogonality of channelization codes, and on the transmit power of adjacent NodeBs. In the uplink, the interference at the NodeB antenna depends on the number of UEs in the own cell and on the spatial configuration of UEs in adjacent cells. Note that interference and transmit power in own and adjacent cells depend on each other both in uplink and downlink due to the utilization of the same frequency band. Interference variations over time can also be induced by user mobility in the whole network, transmission gaps and radio propagation variations. Interference is random over time, therefore the capacity of a UMTS cell cannot be calculated in a deterministic fashion. The term *soft capacity* reflects this property by stating that a certain capacity, for example in terms of Erlang per cell, is only meaningful with an associated probability that enough radio resources (transmit power or interference) are available.

In downlink direction, transmit power is under direct control of the NodeB,

which is the reason why it is often taken as admission criterion. It is commonly restricted to 10 W or 20 W depending on the type of the NodeB. However in the downlink, another capacity limiting resource are the channelization codes due to their orthogonality property as we explained in the previous section. In fact, they often are the dominant capacity limiting resource [37]. Other resources must be considered as well, like available Iub interface and hardware element capacities [8].

In the uplink, interference limits the capacity due to pseudo-orthogonality of the scrambling codes. The absolute interference value can be represented as cell load by relating the multiple access interference (MAI) generated by UMTS mobiles to the ubiquitous thermal noise:

$$\eta = \frac{I_0}{I_0 + W \cdot N_0}, \tag{2.2}$$

where I_0 is the multiple access interference (MAI) from all users in the network, W is the system bandwidth and N_0 is the uncolored ("white") thermal noise density [30]. The pole capacity of a single cell is then computed by calculating the number of sustainable radio bearers with η approaching 1. However, in practice operators define load thresholds much lower than 1 in order to absorb load overshoots and to increase the coverage area (the relation between coverage and capacity is explained e.g. in [38]).

Admission control strategies vary from simple number or throughput-based methods over measure-based and effective bandwidth approaches to priority-based and optimum schemes [39, 40]. In [25], a more general definition of the cell load is provided that includes transmit power, interference and the channelization code occupation [25].

Load control is related to admission control, but ensures that the resource requirements of *existing* connections stay within the system perimeters. This may for example happen if several users move to the cell edge, which leads to higher transmit power requirements due to increased path loss and other-cell interference. For circuit-switched connections like DCH radio bearers, load control may drop connections in order to keep the system stable. Rate controlled radio bearers

may be slowed down in order to decrease their resource consumption. However, in UMTS Release '99, rate control is performed in the RNC on a time-scale of several seconds [16].

Hand-Over Procedures Hand-overs within a UMTS network are initiated if a mobile moves from one cell to another. Three types of hand-overs are supported: soft, softer and hard hand-over. A UE in soft hand-over state establishes connections to several NodeBs at the same time. Whether a UE is in soft hand-over is determined by the *Active Set* (AS), that contains all NodeBs with a sufficiently good signal quality. NodeBs are added and removed to and from the AS with an addition and removal hysteresis relative to the strongest pilot signal in the AS, implicating that at least one NodeB is always member of the AS. In uplink direction, the received signals at all NodeBs are combined in the RNC (for example using maximum ratio combining), exploiting signal diversity [41, 42]. In downlink direction, signal combination is done in the user equipment. Related to soft hand-over is softer hand-over, which occurs if the UE is on the boundary between two sectors. Similarly to soft hand-over, the UE establishes two connections with different scrambling codes, but signal combining is performed in the RAKE receivers at the UE and NodeB, respectively. Finally, hard hand-overs are necessary if the UE switches to another frequency (inter-frequency handover) or system (like GSM) [43]. For HSDPA, only hard hand-overs are implemented due to the independent scheduling at the NodeBs.

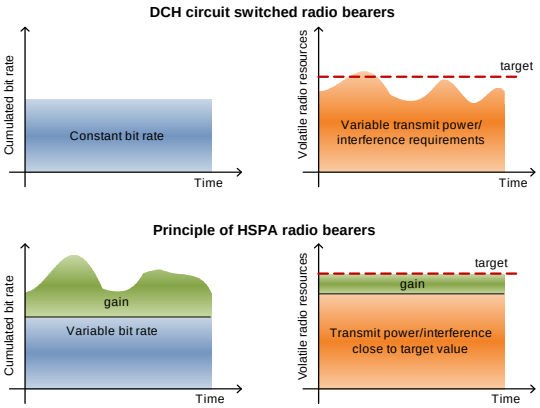


Figure 2.5: *Radio resource utilization with Release '99 and HSPA.*

2.2 Radio Resource Management for Elastic Internet Traffic

Internet traffic is data traffic which uses the Internet Protocol (IP) for transport. Traffic objects are fragmented into smaller packets which are routed independently from each other from host to host, hence it is packet-switched. Internet traffic can be roughly divided into two groups: streaming traffic and elastic traffic [44]. Streaming traffic is generated by interactive applications like voice over IP, video streaming, conferencing, or web radio. It has certain requirements on end-to-end delay, jitter and data rate which must be fulfilled for an acceptable perceived service quality. On transport layer, mostly connection-less protocols like the real-time transport protocol (RTP) are used for transportation.

In contrast, elastic traffic refers to non-interactive applications like file transfers, web browsing, or peer-to-peer file sharing that do not have stringent requirements on bandwidth and delay. These applications use in most cases the

transmission control protocol (TCP) for transport, which adapts to the current bandwidth situation due to the TCP flow control mechanism. As a consequence, also the time it takes to download a document adapts to the bandwidth, hence this type of traffic is “elastic”. Note that also certain web video players like the flash video player uses TCP and the hypertext transfer protocol (HTTP) for transport.

Especially web traffic is bursty: Relatively short periods of activity where a web page and related inline objects are transmitted alternate with long periods of inactivity where the user reads the content of the page [45, 46].

So, while Internet best-effort traffic is packet-switched, elastic and bursty, it is transported in UMTS Release ’99 with DCH radio bearers, which are circuit-switched, static and resource-persistent. This leads to resource inefficiencies in the following way:

- Admission Control for ciruit-switched radio bearers works under the premise to provide constant bit rates. This means that for volatile radio resources like interference and transmit powers, admission control has to reserve a large “safety” margin in order to prevent load overshoots. This margin decreases the system capacity.

- Channelization codes are exclusively assigned to a radio bearer during the lifetime of a DCH connection. This means that with bursty, low activity traffic like web traffic, a large fraction of time the code resources are occupied but not used, and therefore wasted. If the idle timer which controls the release DCH connections in the RNC is set to a small value, code resource are released more quickly. However, due to the long setup time for DCH radio bearers, this would also lead to significant delays and decreased QoS for the next packet burst.

- Constant bit rate DCH radio bearers cannot provide higher data rates even if transmit power and interference situation would allow for higher signal-to-interference ratios.

HSDPA and Enhanced Uplink have been introduced to overcome these afore-

mentioned shortcomings. Figure 2.5 clarifies the gain for HSPA on the example of radio resource efficiency. Circuit switched radio bearers with constant bit rates lead to fluctuating requirements on volatile radio resource like transmit power or interference, illustrated in the upper right figure. This means that an efficient utilization of the radio resources is not possible. This is visible in the difference between the target (maximum) resource utilization and the actual resource utilization. One reason for the performance gain of HSPA over Release '99 is the exploitation of these unused resources, as we can observe in the lower half of the figure. In this idealized diagram, HSPA utilizes the radio resources completely. Accordingly, the cumulated bit rate of all connections in the lower left figure fluctuates. Other factors like higher-order modulation also contribute to the performance gain of HSPA, as we will see in the next two sections where we explain the fundamentals of HSDPA and Enhanced Uplink.

2.3 High Speed Downlink Packet Access

The purpose of HSDPA is primarily to provide high data rates and to overcome the shortcomings of DCH radio bearers for packet-switched best-effort traffic. HSDPA implements a shared channel for all user per cell such that channelization codes are occupied only if traffic is sent. It further supports higher order modulation such that good radio channel conditions can be more effectively exploited. Therefore, spare resources which are not used by DCH connections can be used to increase the system capacity. Additionally, it reduces the latency induced by the air interface, such that also interactive applications can benefit from HSDPA. The impact of HSDPA on the network architecture is mainly restricted to the RAN, where hardware and software updates at the NodeBs and RNCs as well as new user equipment are necessary.

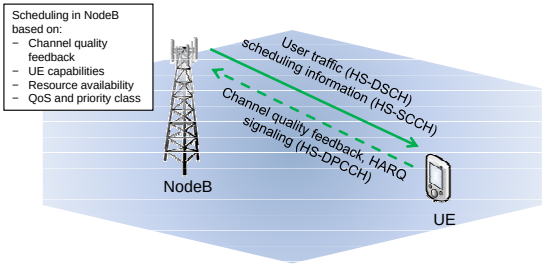


Figure 2.6: *Operation principle of the HSDPA with downlink transport channel (HS-DSCH) and signaling channels.*

2.3.1 Basic Principles

DCH connections provide nearly constant QoS in terms of data rate by means of fast power control, which adapts the transmit power in order to keep the signal-to-interference ratio at an appropriate target level. HSDPA breaks with this principle. Instead, HSDPA uses Adaptive Modulation and Coding (AMC) to adapt the instantaneous data rate to the channel quality, which allows for higher data rates in case of good radio conditions. The transportation of data is done on the High Speed Downlink Shared Channel (HS-DSCH), which implements a mixed TDMA/CDMA scheme in contrast to pure CDMA as for DCH radio bearers. Resource assignment is done in the NodeB scheduler with help of Channel Quality Feedback (CQI) values which are signaled by the UEs. In order to reduce packet latency and SIR requirements, Hybrid Automatic Repeat Request (HARQ) has been introduced which handles retransmissions on MAC-layer. In addition to the HS-DSCH, two new signaling channels have been introduced. In the downlink, the HS-SCCH (High Speed Signaling Control Channel) carries information about the UE to be scheduled in the next subframe and its code rate and modulation scheme. In the uplink, the HS-DPCCH (High Speed Dedicated Physical Control Channel) carries information about the HARQ reception status, which is ACK or NACK, and the CQI values. Figure 3 gives a graphical impression.

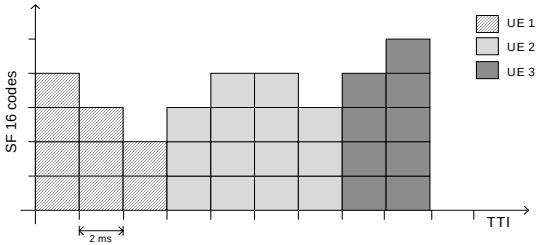


Figure 2.7: *Schematic view of the HS-DSCH. Multiplexing is supported in code and time domain. Each box constitutes an HS-PDSCH and an SF 16 channelization code.*

The HS-DSCH is used for the transport of user data. It is on the time axis subdivided into Transport Time Intervals (TTIs) of 2 ms length, and on the code axis in HS-PDSCHs (High Speed Physical Downlink Shared Channel), which correspond to a channelization code with spreading factor $SF = 16$. The amount of bits which can be transported within one TTI is defined by the Transport Block Size (TBS), which is chosen by the NodeB scheduler according to the channel quality and the HARQ retransmission number. The HS-DSCH enables a mixture between CDMA and TDMA, since the HS-PDSCHs in one TTI can be assigned to different UEs. This may be useful for very time-critical traffic where the data volume is small, but the packets arrive with small inter-arrival times. Another scenario is that a single UE cannot use the available HS-PDSCHs because of bad channel conditions, than the remaining resources could be consumed by a second UE [47]. However, multi-user scheduling in one TTI leads to multiple access interference like in conventional WCDMA, so it is not always beneficial for overall performance. Therefore, and because of the increased complexity of a scheduler with code-multiplexing, scheduling is done mostly one-by-one as shown in 2.7, where three UEs are scheduled one after another.

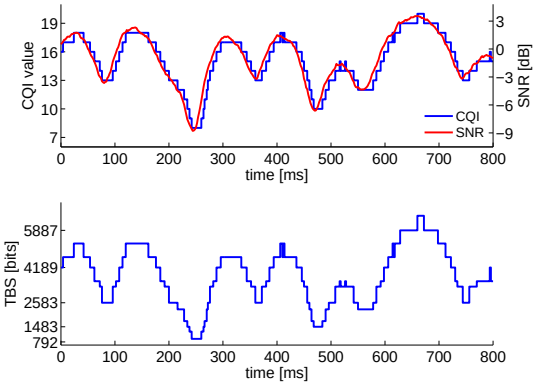


Figure 2.8: *Trace with CQI values and corresponding SNR (above), Transport Block Sizes (below).*

2.3.2 Adaptive Modulation and Coding

The perceived channel quality at each UE is permanently changing, a phenomenon which is known as fading. Fading on millisecond time scale is called fast fading and is the result of multi-path propagation. Instead of compensating fast fading with fast power control as for DCH connections, the HS-DSCH adapts the instantaneously modulation and code rate to the channel quality. This scheme, called adaptive modulation and coding, enables an effective exploitation of the channel capacity [48], since the number of information bits is adapted to the current theoretic capacity. Choosing the correct modulation and code rate requires knowledge of the throughput curves for different modulation orders and code rates. Since the curves are system specific, which means that they depend on the used coding scheme (for the HS-DSCH, turbo codes are used) and on implementation issues like the decoding algorithm, they are generated by extensive simulations [49, 50].

For the HS-DSCH this means that the number of bits which can be transported

CQI	TBS	# HS-PDSCH	Modulation
0	Out of range		
1	127	1	QPSK
2	173	1	QPSK
3	233	1	QPSK
⋮	⋮	⋮	⋮
27	21754	15	16-QAM
28	23370	15	16-QAM
29	24222	15	16-QAM
30	25558	15	16-QAM

Table 2.1: *Simplified CQI table*

within one TTI, can change every 2 ms. Figure 2.8 shows an example trace of a slowly moving UE. The signal-to-noise ratio (SNR) varies in a range from -9 dB to 3 dB. The UE measures the SNR and signals the corresponding CQI back to the NodeB. Note the small time difference between actual SNR and measured CQI due to signaling delays. The NodeB translates then the CQI value to a TBS value, as seen in the lower part of the figure. The CQI values are chosen such that on average, the frame error rate of the first HARQ transmission does not exceed 10 %.

The CQI value defines the Transport Block Size, the number of HS-PDSCHs and the modulation order. Table 2.1 shows a simplified example. The TBS implicitly defines the ratio between information bits and bits after turbo coding (code rate): Coded bits are punctured or repeated to match the number of gross channel bits which can be carried by the HS-PDSCH. This example is an excerpt of the CQI table for UE class category 10, which allows data rates up to 12.8 Mbps with 15 codes in parallel. The UE categories describe the capabilities of the UEs and differ in the maximum number of parallel codes, the modulation order (16 QAM is optional) and the minimum inter-TTI scheduling interval. The relation between

CQI and TBS is defined in the specification document [51].

The CQI tables define the maximum number of bits per TTI which can be transmitted with the current CQI, but this value does not necessarily represent the TBS which is then actually chosen. The reason is that the number of available codes may be lower than the number of required HS-PDSCHs, for example if some codes are already occupied by DCH connections. In that case the maximum TBS value for the available number of codes is selected. Another reason may be that the number of bits in the transmit buffer is lower than the TBS value. Then, the lowest TBS which can carry the buffer content is chosen in order to decrease the frame error probability.

2.3.3 Hybrid ARQ

Hybrid Automatic Repeat Request (HARQ) combines error correction by retransmissions with information coding techniques. In UMTS Release '99, retransmissions are handled by the Radio Link Control (RLC) protocol which is one layer above MAC in the UMS protocol stack, with end points in the Radio Network Controller (RNC) and the UE. That means that a retransmission travels over two hops, making it time consuming and complex to handle. RLC retransmissions are therefore expensive and should be avoided, which in turn requires higher transmit powers to achieve lower frame error rates.

Hybrid ARQ is located on MAC layer between NodeB and UE. It is implemented as an N-stop-and-wait protocol, where N is the number of HARQ processes. The HARQ processes are called in a round-robin manner, such that each process is responsible for one TTI in a cycle. The number of processes is chosen such that during the time it requires to acknowledge a MAC frame, no idle time is created. In HSDPA, $N = 4$.

Figure 2.9 illustrates the principle: If the UE receives an erroneous MAC frame (in the figure frame number 2), it will not be discarded but the UE stores a copy of the Turbo-decoder output ("soft-bits") in the HARQ soft-buffer. The second transmission is then combined with the first transmission (for example with max-

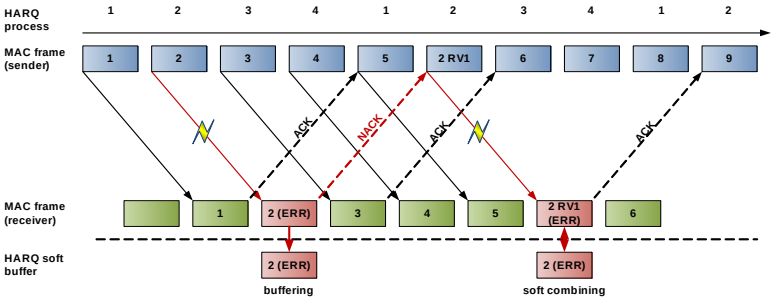
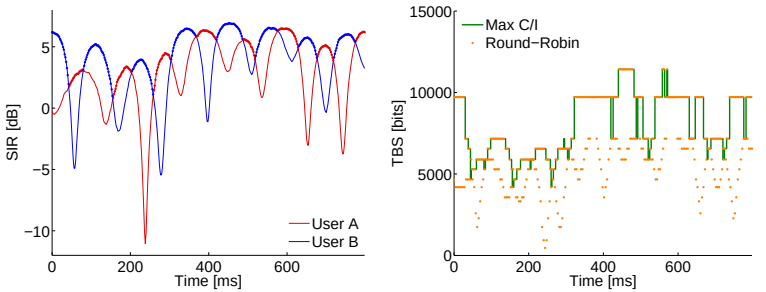


Figure 2.9: *Operation principle of Hybrid ARQ. Erroneous frames are buffered and soft-combined with a retransmitted version.*

imum ratio combining) and is again decoded. In the example figure, also the self-decodable retransmission is erroneous, but the decoding of the combined frame succeeded. Note that always the same HARQ process (here number 2) is responsible for retransmissions of the same frame. With the HS-DSCH up to three transmissions are possible. The code rate may be decreased with the retransmission number, i.e. the redundancy is increased ("incremental redundancy"), which gives a slight performance gain over a scheme with identical retransmissions ("Chase combining") [52, 53, 54]. In the figure, this is indicated by the "RV1" suffix at the retransmitted frame. However, the total number of bits transmitted must not change for retransmissions.

2.3.4 Scheduling

The task of the scheduler in the NodeB is to assign the available resources to the users such that if possible some performance goals are fulfilled. The performance metrics depends on the carried user traffic type and on the viewpoint (network or user centric). Typical performance metrics from user perspective are throughput, packet delay, and jitter. Also, the resources should be assigned "fair", which



(a) SIR-traces and scheduling decisions

(b) Instantaneous TBS for Max C/I and Round-Robin scheduling

Figure 2.10: *Channel-aware scheduling exploits multi-user diversity.*

means that the performance differences between the users should not be too large. From the view of the network, the total throughput should be optimized, which can be achieved by serving always the currently best users. The attempt to maximize one metric may lead to lower performance for other metrics, for example in case of total throughput and user fairness.

The scheduling disciplines can be classified into channel-aware schemes, which take the channel qualities at the UEs into account, and ´channel-blind disciplines like Round-Robin. Channel-aware schedulers can exploit multi-user diversity, which describes the fact that since in HSDPA the current user data rates are random variables, the probability to see a user with good channel quality increases with the number of users in the cell [55, 56, 57]. A scheduler which maximizes the instantaneous cell throughput is therefore the Max C/I scheduler (or MaxTBS-Scheduler), which always selects the user with the currently best CQI/TBS. Figure 2.10 clarifies multi-user diversity and channel-aware scheduling with an example. Figure 2.10(a) shows SIR traces of two users which are scheduled with a Max C/I scheduler: The best user is chosen for transmission, which is indicated by the thickness of the curves. Figure 2.10(b) shows the corre-

sponding instantaneous TBS for Max C/I scheduling and Round-Robin scheduling. We observe that Max C/I scheduling forms the upper envelope of the Round-Robin trace, since the latter also schedules the user with a currently low SIR. The gain of Max C/I over Round-Robin scheduling is the multi-user diversity gain which grows with an increasing number of users.

The Max C/I scheme, often described as "riding the top", has the disadvantage that it will lead to starvation of users which are on the cell edge. A scheduling scheme which combines channel-awareness with user fairness is Proportional-fair scheduling, which aims at maximizing not the cell throughput, but the ratio between current and past data rate or data volume [58, 59]. Thus, a user j is selected at TTI t if:

$$j = \arg\max_j \left\{ w_j \cdot \frac{\text{TBS}_j(t)}{\overline{\text{TBS}}_j(t)} \right\}, \tag{2.3}$$

where w_j is a scaling factor, TBS_j is the current TBS and $\overline{\text{TBS}}_j$ is the average perceived TBS, which is often calculated with a sliding window approach as:

$$\overline{\text{TBS}_j} = (1-\alpha) \cdot \delta_j \cdot \text{TBS}_j(t) + \alpha \cdot \overline{\text{TBS}}_j(t-1), \tag{2.4}$$

with α as weight factor and δ_j as indicator whether the current user has been served or not. In the literature, multitudes of versions and modifications exist of the Proportional-fair scheduler. In [60], several variants are analyzed which allow different sets of users to compete for a frame: all, only the one with non-empty queue or only the one which can use the whole TBS. Further variants differ in whether the actual TBS or the number of actually served bits is used for the average throughput update. Another question is whether the throughput average should be updated right after the scheduling decision or whether the scheduler should wait for a positive acknowledgement such that only the goodput is considered [61]. An important issue is also the interaction between TCP and scheduling: although the proportional-fair scheme is a good choice for best-effort traffic over TCP, the inter-scheduling times for the different users should be not too large in order to avoid TCP timeouts [62].

Round-Robin and proportional-fair schedulers assign radio resources on average "fair" to the user. However, they cannot guarantee QoS like a guaranteed data rate or delay. Multimedia traffic like Voice over IP or streaming video traffic that requires such guarantees therefore require special schedulers, which are also often modifications of the proportional-fair principle. Schedulers that try to guarantee data rates often use a barrier function which gives over-proportionally priority to users which fail the guaranteed data rate [63, 64, 65]. The basic idea is to adapt the scaling factor according to the difference between the guaranteed data rate and the actual received data rate. One example used in [63] is:

$$w_j = 1 + \beta \cdot \exp\left[-\gamma \cdot (\overline{\mathrm{TBS}} - \mathrm{TBS}_j^g)\right], \tag{2.5}$$

where TBS_j^g is the guaranteed data rate of user j, and β, γ are parameters which control the "aggressiveness" of the barrier function. The parameters can be used to tune the trade-off between multi-user diversity gain and effectiveness of the GBR mechanism.

Multimedia traffic like Voice over IP or streaming video also has more stringent requirements on the packet delay than best-effort traffic. For this type of traffic, delay-aware schedulers have been developed which try to use the benefits of AMC and to minimize packet delay. A well known scheduler of this type is the Modified Largest Weighted Delay First (M-LWDF) scheduler [66, 67]. This scheduler tries to ensure that the probability that the packet delay D_j exceeds a certain target value D^* does not exceed a target probability ξ, i.e. $P(D_j > D^*) < \xi$. For that reason, the scaling factor is modified such that:

$$w_j = -\log(\xi) \cdot \frac{D_j}{D^*}. \tag{2.6}$$

The last term of the scaling factor approaches one if the delay comes close to the target delay, such that users with long queues are prioritized. Other approaches try to ensure the delay constraints more aggressively, like the channel-dependent earliest deadline first (CD-EDD) scheduler [68], or the exponential rule (ER) scheduler that introduces a barrier function for the delays [69].

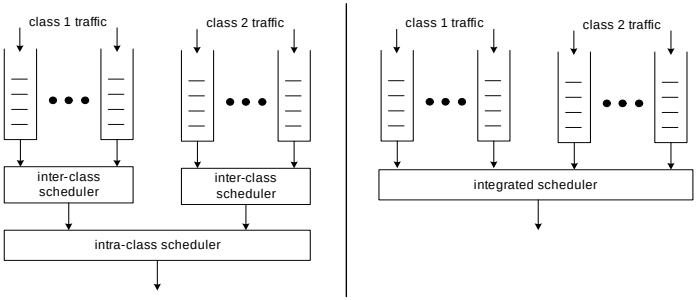


Figure 2.11: *Hierarchical (left) and flat (right) class-aware scheduling schemes.*

All these schedulers are designed for a specific class of traffic, like best-effort, streaming services or delay-sensitive traffic. However, it is also intended to transport different types of traffic and therefore also different service classes over HSDPA, which makes the scheduling problem even more complex. Besides the mapping of the QoS parameters of the UMTS service classes to HSDPA as elaborated in [70], the question is how to design a scheduler which meets all requirements of the different traffic types. Generally, two approaches can be distinguished. The first is a flat design in which one scheduler is responsible for the queues of all users, regardless of their service class. This implicates that the scheduler can be parameterized and considers all relevant QoS variables. Although many of the QoS-aware schedulers have a parameter for service class differentiation, they do mostly not consider data rate and delay at the same time. Examples of such schedulers are given in [63, 71]. Another option is to perform hierarchical scheduling with intra-class and inter-class schedulers. The intra-class schedulers select the UE for transmission within their service class, while the inter-class scheduler selects the between the classes [72].

Figure 2.11 clarifies the two concepts. The advantage of the hierarchical scheduling approach is that it allows different scheduling concepts for the different service classes, and allows easy implementation of priority scheduling (ser-

vice priority is signaled via the service priority indicator – SPI – to the NodeB) by the intra-class scheduler. It is also more convenient to parameterize due to the functional split between inter- and intra-class scheduling. However, if the inter-class scheduler should take the scheduling metrics of the intra-class schedulers into account, the metrics between the classes must be comparable.

2.4 Enhanced Uplink

The Enhanced Uplink or HSUPA is the pendant to the HSDPA in the UMTS uplink. Its main purpose is to also to provide higher data rates, lower latencies and better resource utilization. For this reason, a fast rate control mechanism is implemented which is located in the NodeB. This allows rapid reactions on interference fluctuations and traffic dynamics, such that interference overshoots can be avoided and the system capacity is increased. Additionally, Hybrid ARQ and smaller transmission time intervals are introduced.

2.4.1 MAC and Physical Layer

With the Enhanced Uplink, some new transport and signaling channel have been introduced. The most prominent is the Enhanced DCH (E-DCH) and the corresponding physical channel, the E-DPDCH, which carries user data. Additionally, a new uplink control channel, the E-DPCCH and the corresponding physical channel, E-DPDCCH, have been defined. In the downlink, the E-DCH HARQ Indicator Channel (E-HICH) is used to indicate whether the actual received HARQ transmission has been decoded successfully. Scheduling information in the form of grants are sent over the E-DCH Relative Grant Channel (E-RGCH) and the E-DCH Absolute Grant Channel (E-AGCH). While the former is a dedicated channel which is co-located to a DCH, the latter is a shared channel for all E-DCH users within a sector. Feedback information is sent to the NodeB via the E-DCCH control channel. Figure 2.12 gives an overview of the involved data and signaling channels.

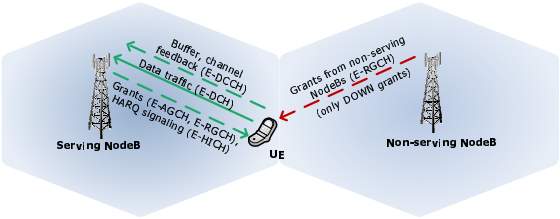


Figure 2.12: *Enhanced Uplink signaling and data channels.*

In Enhanced Uplink, high data rates are achieved on physical layer by using multi-code transmissions with low spreading factors. The highest configuration uses two SF 2 codes and two SF 4 codes in parallel, which means a gross channel bit rate of 5.76 Mbps. The number of parallel codes and the TBS, which corresponds to the number of information bits [73], define together the E-DCH Transport Format Combination (E-TFC). Enhanced Uplink supports two different transport time intervals, 2 ms and 10 ms. The highest data rates are only available with the 2 ms option.

On MAC-layer, the MAC-e and MAC-es entities have been introduced. One UTRAN-side, MAC-es only exists in the RNC where it is responsible for macro diversity selection and reordering. An MAC-es PDU (packet data unit) contains one or more MAC-d PDUs. MAC-e is on UTRAN-side is located in the NodeB, where it is responsible for scheduling and HARQ. MAC-e PDUs contain one or more MAC-es PDUs and constitute the final data frame which is then transmitted on physical layer to the UE. On UE-side, both entities exist. Besides the reciprocal tasks for the UTRAN-side, additionally E-FTC selection and restriction takes place here.

The Hybrid ARQ functionality is nearly identical to the HSDPA implementation. Differences are in the number of supported HARQ processes, which depends on the chosen transport time interval. With a TTI of 10 ms, 4 HARQ processes are supported, while with a TTI of 2 ms, the number of HARQ processes is

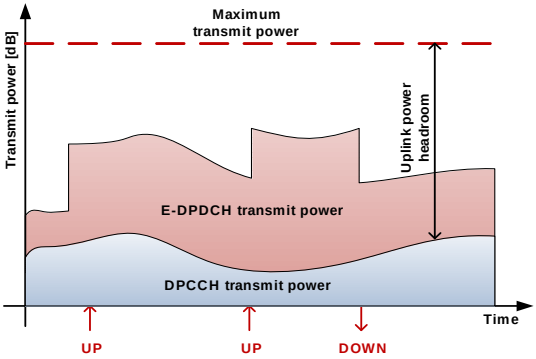


Figure 2.13: *Power assignment for the E-DPDCH. Power grants increase or decrease the E-DPDCH power relative to the DPCCH physical control channel.*

8. Similar to HSDPA, retransmissions must use the same E-TFC as for the initial transmission.

2.4.2 Rate Assignment

The most prominent feature of the Enhanced Uplink is the rate control or resource assignment mechanism. Rate control is realized indirectly by transmitting scheduling grants, which denote the maximum allowed power offset the UE may use for the transmit power of the E-DPDCH physical data channel over the DPCCH physical control channel. Thus, the received power S_k of a UE k at a NodeB can be expressed as

$$S_k = S_{c,k} + \Delta_{G,k} \cdot S_{c,k}, \tag{2.7}$$

where $S_{c,k}$ is the received power of the DPCCH and $\Delta_{G,k}$ is the scheduling grant. The possible values for $\Delta_{G,k}$ are defined in [73] in a table with 38 entries,

where the entry with index 0 is the zero grant which means that the UE pauses its transmission. Figure 2.13 shows an example scenario. The power offset of the E-DPDCH over the DPCCH is constant, until it is increased and decreased by UP and DOWN grants. Grants can be either set as absolute value via the absolute grant channel or as UP/DOWN/HOLD commands on the relative grant channel. The relative grant channel exists for all NodeBs in the active set of the UE, but only the Serving NodeB is allowed to send UP commands. Grants from non-serving NodeBs (other-cell grants) are used to avoid the flooding of an adjacent cell with interference so that here, only DOWN and HOLD commands are allowed. The condition for sending DOWN commands is defined by the total received power at the NodeB and by the ratio between other-cell E-DCH power to own-cell E-DCH power and can be chosen by the operator. Mobiles which are not in soft handover only have one associated RGCH, while the one at the cell border may additionally receive DOWN commands.

The selection of the E-TFC (and therefore of the instantaneous bit rate) according to the signaled power grant is task of the UE. This is done in two steps: First, the current power grant is compared with the uplink power headroom (UPH) which indicates the maximum *possible* power offset. The UPH is defined as $U = \frac{T_{max,k}}{T_{c,k}}$, where $T_{max,k}$ is the maximum allowed transmit power and $T_{c,k}$ is the power of the DPCCH. The result is the power offset which is chosen for transmission:

$$\Delta_{max,k} = \min\left\{\Delta_{G,k}, U\right\}. \tag{2.8}$$

In the second step, the E-TFC restriction algorithm identifies the set of E-TFCs which require a power offset lower than $\Delta_{max,k}$. This is done by estimating the power offset requirements of all E-TFCs according to a reference E-TFC. A state machine with two states decides which E-TFCs are in “restricted” and which are in “allowed” state. If the $\Delta_{max,k}$ is lower than the required power offset of an E-TFC for a certain number of TTIs, the E-TFC is marked as “restricted” and cannot be used for transmission. Otherwise the E-TFC remains “allowed”. The opposite direction, i.e. from “restricted” to “allowed”, works analogously. After the restriction process, the UE may then choose an E-TFC in “allowed” state

which is suitable for the data volume in the transmit buffer. For a more detailed discussion of the E-TFC restriction and selection process see [74].

2.4.3 Scheduling

The E-DCH scheduler decides on when, to which mobiles, and how much resources (corresponding to scheduling grants) are assigned to the mobiles. The scheduler can use the following feedback information from the UEs for its decision: Uplink power headroom, buffer status (total buffer and per priority class) and the "happy bit" which indicates whether the incoming traffic at the UE can be transmitted with the current resource assignment. The UPH is an indicator for the propagation loss and can therefore be used for channel-aware scheduling: A lower UPH means that the a larger fraction of the possible transmit power is required to meet the E_b/N_0-requirements of the power controlled signaling channel.

Theoretically, the scheduler can assign grants every 2 ms to the UEs, however this would induce a significant signaling overhead on the downlink. It is therefore expected that scheduling is performed in larger time intervals [75], which leads to a trade-off between efficiency and signaling overhead. The scheduler must operate within the resource constraints of the system. In WCDMA uplink, the resource to assign is the interference or an deviated measure like cell load or noise rise. Power grants and interference are related as following: Let $\mathcal{E}$ be the set of E-DCH users in a cell, and I_E^* the interference share that can be consumed by E-DCH radio bearers. Then, with Equation (2.7) and summing over all received powers, the cumulated power grants are:

$$\sum_{k\in\mathcal{E}} \Delta_{G,k} = \frac{I_E^*}{S_c} - |\mathcal{E}|. \tag{2.9}$$

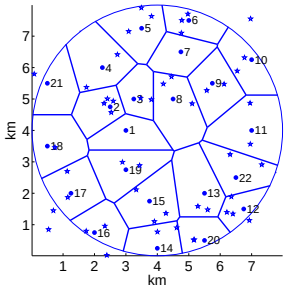
Note that in this equation it is assumed that the received powers of the control channels of all users are approximately equal due to fast power control. Grants may now be assigned individually to UEs. For example, a channel-aware scheduler that favors users which are on average closer to the serving NodeB antenna

may use the following expression [12]:

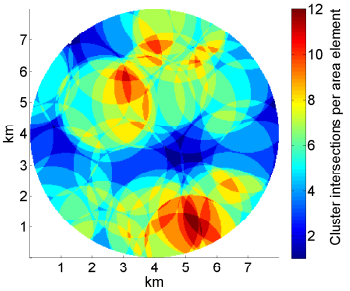
$$\Delta_{G,k} = \frac{U_k}{\sum_{j \in \mathcal{E}} U_j} \cdot \left(\frac{I_E^*}{S_c} - |\mathcal{E}| \right). \tag{2.10}$$

Uplink CDMA scheduling is in some aspects more complex than downlink scheduling. A general rule is that favoring users which are close to the cell center maximizes the instantaneous cell throughput. In [76] it is shown that with no transmit power constraints, a greedy one-by-one scheduling algorithm that always favors the user with the best channel condition is optimal regarding cell throughput. With transmit power constraints, the same holds for a water-filling approach with parallel scheduling. However, these results are only applicable in the single-cell case. In a multi-cell environment, an optimal solution has to consider all involved UE. For a further discussion, see Section 4.1.

Literature on practical E-DCH scheduling is sparse. In [77], the authors propose a fast scheduler for the WCDMA uplink based on the past selected data rates of the mobile. It is therefore a channel-blind scheduler which implicitly takes the traffic dynamics into account. Another approach is taken in [78], where a channel-aware scheduler is proposed with a similar concept as in HSDPA scheduling. The scheduler evaluates the channel quality of each mobile (denoted as uplink CQI – UCQI), and assigns scheduling grants in an one-by-one fashion. However, this approach collides with the already mentioned high signaling load and also with the fact that the UMTS uplink is unsynchronized. In [75], the authors describe the constraints for designing an E-DCH scheduler, but do not explicitly propose a scheduler. A comparison of different channel-aware and channel-blind schedulers is performed in [79]. In [12], a channel-aware scheduler and a buffer-aware scheduler are proposed and compared with the performance of an equal-rate scheduler.



(a) NodeB (circle) and cluster center (star) positions



(b) Heterogeneous arrival densities as the result of cluster intersections

2.5 Network Scenarios

Throughout this monograph we use different network scenarios. The radio access network of a UMTS network consists of a number of NodeBs and RNCs, whereas we ignore the latter and concentrate on the layout, i.e. the spatial configuration of the NodeBs. In practice, the cell layout depends on the characteristics of the terrain and radio environment, on the expected number of users, their service and movement profile, and on practical issues like community regulations regarding antenna positions. Every NodeB controls a coverage area which we will refer to as cell in the reminder. In our scenarios, the coverage area is further partitioned into quadratic *area elements* with a certain edge length, for example $25\,\mathrm{m}$. Each area element is associated with a user arrival density λ_f, where f indicates the area element. Note that the partitioning also allows to use an arbitrary propagation map which covers the properties of different terrains like streets, wood, buildings, etc.

We consider two types of networks: a “traditional” 19-cell hexagonal layout and a heterogeneous layout which is the result of a Voronoy tessellation of a circular network area. For the hexagonal layout, we assume that the user arrival

densities are the same for all area elements. The distance between the NodeB antennas is 1.2 km. For the heterogeneous layout, the network area is superimposed with a number of circular *clusters* which consist of area elements of equal user arrival densities. In the intersection areas of the clusters the arrival densities of the clusters are summed up to the total arrival densities of the specific area elements, c.f. to [40]. Figure 2.14(a) shows the Voronoi tessellation and the resulting NodeB coverage areas of the heterogeneous cell layout. The position of NodeB antennas are indicated by circles and the centers of the arrival clusters are indicated by stars. The number of cluster intersections in Figure 2.14(b) illustrate the relative arrival densities. The absolute values depend on the offered load scenario and are given in Section 3.6.

For both the HSDPA and Enhanced Uplink performance evaluation we use the modified COST 231 Hata pathloss model for a small urban environment [80]. We assumed an uplink carrier frequency of 1.95 GHz, a downlink carrier frequency of 2.14 GHz, a mobile antenna height of 1.8 m and a NodeB antenna height of 30 m. For the HSDPA evaluation, multi-path propagation profiles follow the definitions in [81].

3 Flow-Level Performance Models for HSDPA

In this chapter, we develop a framework for the performance evaluation of integrated UMTS networks with HSDPA and DCH users in coexistence. The motivation for such a framework is described in Section 3.1. The framework enables system models on flow-level, a concept which is explained in Section 3.2. A key factor which influences the system performance is radio resource management. In Section 3.3, we describe and formulate which schemes can be used to share radio resources, i.e. downlink transmit power and orthogonal channelization codes, between DCH and HSDPA connections. The calculation of transmit powers for the proposed schemes is then explained in Section 3.4. Transmit power and available channelization codes are input variables for the physical layer abstraction model for the calculation of the HSDPA user throughput, which is introduced in Section 3.5. The model provides a stochastic representation of small time-scale effects during the life-time of a data flow. Also in this section, we develop analytical expressions for different channel-aware and channel-blind scheduling disciplines.

In the second part of the chapter, a flow-level simulation is used to investigate different aspects of the HSDPA and DCH performance subject to the radio resource management schemes and scheduling disciplines introduced before. In Section 3.6, we firstly demonstrate the impact of a volume-based traffic model on the spatial user distribution within the network. Then, we look at the average user and cell throughput for different scheduling disciplines and explain how and why the results stand in contradiction to models which do not take flow-dynamics into

account. Finally, we investigate the impact of radio resource management and power allocation schemes.

In the third part (Section 3.7), we develop an analytical model based on the flow-level framework. The challenge in this case is the proper modeling of spatial user heterogeneity, which we solved with a closed-form approximation for Round-Robin scheduling. We conclude the chapter with some remarks on the results in Section 3.8.

3.1 Motivation and Related Work

Performance evaluation of an integrated, HSDPA-enabled UMTS network is a difficult task due to the complex operations and interactions on different timescales, protocol layers and involved system entities. On physical and MAC layer, the relevant time period is in a range from less than one microsecond on symbol level up to several milliseconds for scheduling and retransmission handling. Radio resource management procedures such as reservation and preemption operate on time scales where user behavior, for example arrivals, departures, and movements, takes place. These traffic dynamics occur in a scale of several seconds to minutes. Nevertheless do the processes on higher layers influence lower layer operations and vice versa. A second dimension is opened due to the dynamic structure of WCDMA networks. Since all NodeBs are transmitting in the same frequency spectrum (frequency reuse factor 1), the NodeB transmit powers are mutually depending on each other. This means that the considered network scenarios must be large enough to ensure the proper reproduction of other-cell interference effects. Also, in an integrated scenario HSDPA and DCH users influence each other, either directly by radio resource requirements or indirectly by interference. In such a scenario, steady-state statistics on performance measures like blocking probabilities require long simulation runs since relatively seldom events like long-living DCH users with high data rates at the cell edge have an disproportional influence on the network performance.

Performance evaluation requires therefore tools which on the one hand model small-time scale physical layer effects with adequate accuracy, but on the other hand allow for either long simulation runs in reasonable time, or analytic methods. In this chapter, we provide a framework for such tools on ground of a decomposition approach. Higher network layers are modeled on flow-level, such that the effects of traffic dynamics are captured. Lower layers are modeled with approximative stochastic methods.

In the literature, the performance on packet-level is subject of a vast amount of papers, which often concentrate on a certain aspect of the whole system. For example in [54, 82, 83, 84, 85], the impact and modeling of Hybrid ARQ with different redundancy schemes (incremental redundancy, Chase Combining) is investigated. Link-level performance of AMC specifically for the HSDPA is considered e.g. in [50, 86, 49]. Another aspect that has been extensively investigated is the performance of different scheduling disciplines, often also in conjunction with QoS-guarantees for streaming services or with mixed service classes [65, 67, 87]. For an overview of scheduling options for the HSDPA the reader is referred to Section 2.3.4. There are many more facets, but all these studies have in common that they use detailed packet-level simulations to concentrate on specific aspects of the system performance.

A frequently used approach to speed up simulations is to use physical layer traces of CQI or TBS values, which are generated offline by specialized link-level simulators. The traces are then used as input for higher-level network simulators. While this approach only requires memory for the traces, it is inflexible regarding channel quality variations which are not primarily due to the radio link itself but due to external factors like available code resources, transmit power and other-cell as well as own-cell interference. This is a characteristic that is common to most simulators which use physical layer traces. A well known simulation tool relying on this approach comes from the EURANE project, which resulted in the development of a packet-level HSDPA simulator based on the Network Simulator-2 (NS-2) [88]. The EURANE simulator uses pre-generated physical layer traces which provide block error rate (BLER) and current band-

width in each TTI. Further examples for trace-based simulators can be found in [72, 89, 90, 91, 92].

Other approaches [93, 94, 95] use Monte-Carlo techniques. While Monte-Carlo simulations are generally time-efficient and easy to implement, it is difficult to capture the impact of flow-dynamics. Especially when using a volume-based traffic model for the HSDPA users, the mutual dependency between sojourn time and available bandwidth can only be approximated as in [11, 93] for Round-Robin scheduling. Finally, in [96], the impact of traffic dynamics on the HSDPA performance is described for a single cell scenario. The HSDPA bandwidths are obtained via Monte-Carlo simulations which are then used as input for time-dynamic simulations and an analytical model.

3.2 The Flow-Level Concept

The term flow-level refers to the modeling of an ongoing data transfer as a continuous *flow* with a certain data volume or holding time. We have to distinguish between QoS flows which require a fixed bandwidth, as for voice calls over DCH transport channels, and between best-effort or elastic flows which adapt their bandwidth requirements to the currently available bandwidth. Such a flow may be an FTP transfer or the combined elements of a web page including inline objects such as embedded pictures and videos, so it may consist of overlapping TCP connections. A flow can be loosely defined as a coherent stream of data packets with the same destination address [97]. An important distinction between the two flow types is that QoS flows typically follow a time-based traffic model, which means that the user wants to keep the connection for a certain time span. In contrast, elastic flows are volume based, i.e. the user is satisfied as soon as a certain data volume is transmitted.

In the context of simulation and analytical models, the flow level concept is an abstraction of the packet level. A flow is thus a single traffic object which is described by statistic properties. This properties can be classified into properties

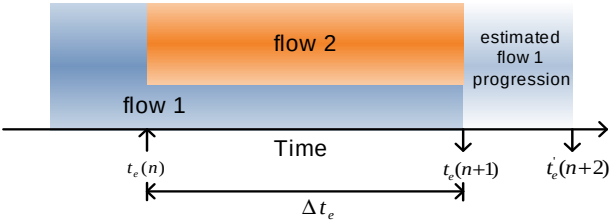


Figure 3.1: *Flow progressions.*

describing the attributes of the object itself, like flow size or life time distribution, and properties which describe the structure of the underlying packet process within the flow. Examples of the latter are packet inter-arrival times or jitter. Thus, flow-level modeling allows to find a balance between modeling accuracy and required computational effort. In this work we are only interested in statistical results on flow level, i.e. we do not consider packet-level statistics. This means in the context of HSDPA simulations or analytical approaches, that it is sufficient to consider events which change the *average* throughput of a flow. This avoids the explicit calculation of physical and MAC layer effects like fluctuation of the channel quality, and leads to a drastic reduction of the number of events.

An event is characterized by its type and its event time t_e. We denote the time between two events with Δt_e. Events are generated if flows arrive into or depart from the system, or if it is required to recalculate radio resource requirements, interference or user bandwidth. Events are therefore predominantly generated at the starting- and ending-points of a flow. However, other events are also possible e.g. if a user significantly changes his location, or if radio resource management performs operations such as power-ramping or power adaptation within ongoing flows. Depending on the type of the event, different actions are performed.

QoS and elastic flows are treated in different ways. While for the former the departure time is known at the beginning of its life time, the departure time of elastic flows depends on the data rate they can utilize. So at each event, the trans-

mitted data volume in the time span Δt_e and the remaining data volume is calculated. Then, the mean HSDPA bandwidths for all flows are recalculated and the new departure times are estimated according to the new data rates. Figure 3.1 illustrates the principle with two example flows. The second flow arrives at time $t_e(n)$. The thickness of the flows illustrates their bandwidth. If flow 2 departs from the system at $t_e(n+1)$, the new departure time for flow 1 is estimated as $t'_e(n+2)$.

3.3 Sharing Code and Power Resources between HSDPA and DCH

A key part of the Radio Resource Management in HSDPA enhanced UMTS networks is the sharing of code and power resources between DCHs, signaling channels, common channels, and finally channels required for HSDPA, namely the HS-DSCH and the HS-SCCH. The signaling channels and common channels mostly require a fixed channelization code and a fixed power as for the pilot channel (CPICH) or the forward access channel (FACH). The DCHs are subject to fast power control which means that their power consumption depends on the cell or system load that determines the interference at the UE. The general level of power consumption depends on the processing gain and the required target bit-energy-to-noise ratio (E_b/N_0) of the radio access bearer (RAB). Two types of DCHs are distinguished: real-time (RT) and non-real-time (NRT) DCHs. RT DCHs continuously require a certain service level which means that they keep their channelization code, processing gain, and target E_b/N_0 value throughout the connection. RT DCHs are mainly used to transport voice or video traffic. NRT DCHs are subject to a slow rate or load control which is located in the RNC. Depending on the current system situation the RNC may change the channelization code and thus also the power consumption of a NRT DCH. Measurements in a laboratory environment [16] have shown that the load control of NRT DCHs takes place in the magnitude of several seconds, and that large differences exist

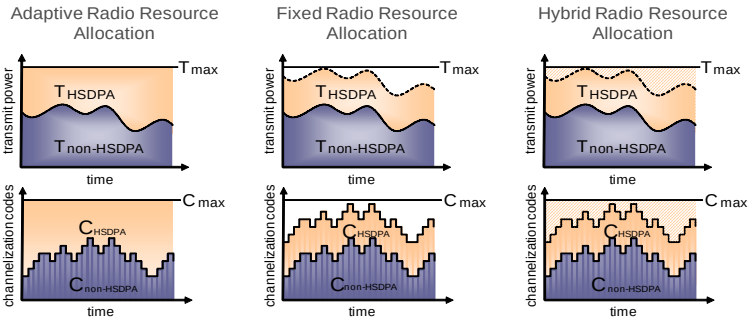


Figure 3.2: *RRM schemes for transmit power and channelization codes*

between the load control algorithms of different vendors.

Sharing resources between HSDPA and DCH may be conducted either semi-static or dynamically. The first studies on HSDPA performance [90, 91, 94] assume a fixed allocation of power and code resources to the HS-DSCH. More recent publications [98, 99] assume a dynamic allocation of code and power resources which means that the HSDPA may use all resources instantaneously not occupied by CCHs or DCHs except for a certain safety margin that has to be kept due to the fluctuations in the DCH resources and the time required for adapting to these changes.

An adaptive, dynamic allocation of code and power resources leads to an optimal utilization of the available spectrum and optimizes HSDPA performance while the DCH achieve precisely their required service quality. As a negative consequence the throughput of HSDPA connections diminishes in cells with high DCH load. There is no performance guarantee for the HSDPA. A fixed resource allocation leads to certain guaranteed HSDPA throughput but, also restricts the DCH cell capacity. Additionally, resources may remain unused if only few DCH connections are present or if the code and power resources for the HSDPA are not well balanced. The hybrid resource allocation combines the advantages of

the fixed and the dynamic allocation. Allocating fixed code and power resources to the HSDPA ensures a certain HSDPA throughput. Dynamically adding unused DCH resources to the already allocated fixed HSDPA resources avoids wasting valuable radio resources and optimizes HSDPA performance. Figure 3.2 sketches the idealized power and resource sharing for the three allocation schemes fixed (or semi-static), adaptive (or dynamic), and hybrid.

3.3.1 Code Resources

The HSDPA requires code and power resources. Codes are the channelization codes that are generated according to the Orthogonal Variable Spreading Factor (OVSF) code tree. The number of codes available for a certain spreading factor (SF) is equal to the spreading factor itself. A 384 kbps DCH occupies a SF 8 channelization code. Accordingly, the maximum number of parallel 384 kbps users per sector is theoretically 8. In practice, only 7 parallel 384 kbps users are possible since the signaling and common channels also require some code resources.

The largest spreading factor in UMTS is $SF = 512$. Let us therefore define the SF 512 code as base unit. A DCH i with SF k occupies $c_i = 512/k$ code resources. An HSDPA code with SF 16 requires $c_{HS} = 32$ code resources. Let C_{DCH} be the total code resources occupied by all DCHs, C_{CCH} be the resources occupied by signaling and common channels, and, $C_{HS} = n_{HS} \cdot c_{HS}$ be the total number of code resources used by the HSDPA where n_{HS} is the number of SF 16 codes allocated to the HS-DSCH. The total number of code resources is equal to $C_{tot} = 512$. Depending on the code allocation scheme, the number of codes available for the HS-DSCH is

$$n_{HS} = \begin{cases} n^*_{HS}, & \text{for fixed allocation,} \\ \lfloor C_{tot} - C_{CCH} - C_{DCH} \rfloor, & \text{for adaptive and hybrid allocation.} \end{cases} \tag{3.1}$$

The adaptive and hybrid allocation may be done on different time scales and additional code margins. The allocation of channelization codes to channels is

done by the RNC via the NodeB application protocol (NBAP) [100]. The HS-DSCH code allocation is signaled to the UEs via the mandatory HS-SCCH control channel, such that no additional radio resources are consumed. Changing the channelization code of a DCH, however, requires signaling from RNC to UE via the Radio Resource Control (RRC) protocol [101]. Changing a DCH code therefore consumes radio resources and introduces signaling delays. Consequently, frequent changes of DCH codes should be avoided. One possibility [98] is to allocate HS-DSCH codes at one end of the OVSF code tree and codes for signaling, common, and dedicated channels starting from the other end of the code tree. Between DCH codes and HSDPA codes a certain buffer zone is introduced in order to allow short-term allocation to DCH or signaling channels. Fragmentation of the code tree can further reduce HSDPA code resources. Fluctuation of DCH code resources results from arrival and departure of DCH users but also from varying DCH activity. The time scale on which these activity changes take place determines whether the codes are meanwhile released and available for the HSDPA or not. The delay requirement of the application determines how fast the codes have to be available again after an inactivity period. As an example, HSDPA might utilize free code resources that results from the inactivity of an Internet user on a 384 kbps DCH but might not utilizes free code resources resulting from silence periods in voice calls.

3.3.2 Power Resources

The transmit power $T_{x,tot}$ of the NodeB consists of a constant part T_{CCH} for common and signaling channels, a part T_{DCH} for DCHs, and a part T_{HS} for the HS-DSCH:

$$T_{x,tot} = T_{CCH} + T_{DCH} + T_{HS}. \tag{3.2}$$

Let T_{max} be the maximum NodeB transmit power and T^* be the target transmit power. Then, the HS-DSCH transmit power according to the different resource

allocation schemes is

$$T_{HS} = \begin{cases} T^*_{HS}, & \text{for fixed power allocation,} \\ T^* - T_{CCH} - T_{DCH}, & \text{for adaptive and hybrid power allocation,} \end{cases} \tag{3.3}$$

where T^*_{HS} is the power reserved for the HS-DSCH and T_{DCH} is the total DCH power averaged over some period of time.

The DCH admission control for fixed and hybrid power allocation has to take care that the DCH power does not exceed $T_{max} - T_{CCH} - T^*_{HS}$. This is achieved by keeping the average DCH power below $T^* - T_{CCH} - T^*_{HS}$. The margin between target and maximum power is the power control headroom reserved for fluctuations of the DCH power. In a system without HSDPA [30] the target power is by a factor K smaller than the maximum power. A typical value is $K = 0.5$. The introduction of the HSDPA allows for a better utilization of the power resource [94] even if fixed power allocation is used. As the power control headroom is chosen relative to the average DCH power ratio we obtain $T^* = K \cdot (T_{max} - T^*_{HS}) + T^*_{HS}$. Adaptive or hybrid power allocation tries to follow the fluctuations of the DCH power. The control of the HSDPA power can be located in the RNC or in the NodeB. If it is located in the RNC only a rather slow adaptation of the power is possible so it is beneficial to perform power adaptation directly in the NodeB. The faster the NodeB is able to adapt the HS-DSCH power, the smaller the power control headroom can be chosen.

3.4 Downlink Transmit Powers

We define a UMTS network as a set $\mathcal{L}$ of NodeBs with associated UEs, $\mathcal{M}_x$, with $x \in \mathcal{L}$. A DCH connection k corresponds to a radio bearer with data rate R_k and code resource requirements c_k. Since the power consumed by the DCH connection is subject to power control, the received E_b/N_0 ε_k fluctuates around a target-E_b/N_0 value ε^*_k, which is adjusted by the outer-loop power control such that the negotiated QoS-parameters like frame error rate are fulfilled. A common

approximation for the average E_b/N_0 value is

$$\varepsilon_k = \frac{W}{R_k} \cdot \frac{T_{k,x} \cdot d_{k,x}}{W \cdot N_0 + I_{k,oc} + \alpha_k \cdot T_{x,tot} \cdot d_{k,x}}, \tag{3.4}$$

where the orthogonality α_k describes the impact of the multi-path profile for DCH k. In this equation, $T_{k,x}$ is the DCH transmit power for user k, W is the system chiprate, N_0 is the thermal noise power density, $I_{k,oc}$ is the other-cell interference from adjacent NodeBs and $T_{x,tot}$ is the total transmit power of NodeB x. In this work, we assume perfect power control, i.e. the mean E_b/N_0 value meets exactly the target-E_b/N_0 such that $\varepsilon_k = \varepsilon_k^*$. The mean transmit power requirement of a DCH connection follows than as

$$T_{k,x} = \frac{\varepsilon_k^* \cdot R_k}{W} \cdot \left(\frac{W \cdot N_0 + I_{k,oc}}{d_{k,x}} + \alpha_k \cdot T_{x,tot} \right). \tag{3.5}$$

The average other-cell interference comprises the received powers of surrounding NodeBs such that

$$I_{k,oc} = \sum_{y \in \mathcal{L} \setminus x} T_{y,tot} \cdot d_{k,y}. \tag{3.6}$$

The total NodeB transmit powers can be calculated with an equation system over all NodeBs. For that reason we follow [40] and define the load of NodeB x with respect to NodeB y as

$$\begin{aligned} \eta_{x,y} &= \sum_{k \in \mathcal{M}_x} \omega_{k,y} \\ \text{with } \omega_{k,y} &= \tfrac{\varepsilon_k^* \cdot R_k}{W} \cdot \begin{cases} \alpha & \text{, if } \mathcal{L}(k) = y \\ \frac{d_{k,y}}{d_{\mathcal{L}(k),k}} & \text{, if } \mathcal{L}(k) \neq y. \end{cases} \end{aligned} \tag{3.7}$$

After some algebraic modifications, this allows us to formulate the total DCH transmit power in a compact form as

$$T_{x,DCH} = \sum_{y \in \mathcal{L}} \eta_{x,y} \cdot T_{y,tot}. \tag{3.8}$$

In this equation, we neglect the thermal noise since in a reasonably designed network it is not relevant for the transmit power requirements. Note also that

the equation includes the case $y = x$. For the total transmit power we need to introduce the boolean variable $\delta_{y,HS}$ which indicates whether at least one HSDPA flow is active in cell x. With the *adaptive* and *hybrid* RRM schemes, the total transmit power is then

$$T_{x,tot} = \delta_{x,HS} \cdot T_x^* + (1 - \delta_{x,HS}) \cdot \left(T_{x,CCH} + \sum_{y \in \mathcal{L}} \eta_{x,y} \cdot T_{y,tot} \right). \quad (3.9)$$

With the *fixed* scheme, the HS-DSCH transmit power is independent of the DCH transmit power, and Eq. (3.9) simplifies to

$$T_{x,tot} = \delta_{x,HS} \cdot T_{x,HS}^* + T_{x,CCH} + \sum_{y \in \mathcal{L}} \eta_{x,y} \cdot T_{y,tot}. \quad (3.10)$$

Introducing the vectors

$$\begin{aligned} V_{\mathcal{F}}[x] &= T_{x,c} + \delta_{x,HS} \cdot T_{x,HS}^* \\ V_{\mathcal{A}}[x] &= \delta_{x,HS} \cdot T_x^* + (1 - \delta_{x,HS}) \cdot T_{x,CCH} \end{aligned} \quad (3.11)$$

and matrices

$$\begin{aligned} M_{\mathcal{F}}[x,y] &= \eta_{x,y} \\ M_{\mathcal{A}}[x,y] &= (1 - \delta_{x,HS}) \cdot \eta_{x,y} \end{aligned} \quad (3.12)$$

for the *fixed* and *adaptive/hybrid* power allocation schemes, respectively, leads to a the matrix equation

$$T = V + M \cdot T \Leftrightarrow T = (I - M)^{-1} \cdot V, \quad (3.13)$$

that is valid for all three resource allocation schemes. The matrix I is the identity matrix, and T is the vector of NodeB transmit powers T_x. The DCH and HSDPA transmit powers are then calculated with Equation (3.8) and Equation (3.3).

In order to prevent outage, an ideal DCH admission control should accept a new user only if the total DCH transmit powers including that of the new user stay below the admission control threshold. This means that admission control for DCHs is performed under the assumption that the HSDPA power is switched on

in every cell, since otherwise the additional interference would lead to increased transmit power requirements which may exceed the maximum transmit power. Thus, even in case that the HS-DSCH power is switched off, the above equations with $\delta_{x,HS} = 1$ for all cells x are used for determining the DCH power relevant for DCH admission control, but the equations are used with the actual value of $\delta_{x,HS}$ for computing the actual cell transmit powers.

3.5 HSDPA Throughput Approximation

In the following we describe a physical layer abstraction model for the calculation of the data rate of an HSDPA user depending on the other-to-own-cell interference ratio, the HS-DSCH transmit power, the number of available codes, the number of HSDPA users, and the scheduling discipline. The model is the counterpart to the classical orthogonality factor model for CDMA downlink as it has been used in the previous section to calculate the DCH transmit powers. However, since DCH connections are power controlled, it is sufficient to calculate with a mean orthogonality factor, hence with a mean SIR. For the calculation of the HSDPA data rates, the distribution of the SIR at every location in the cell is required since the CQI values and SIR are directly related to each other. Hence, the key of our throughput approximation method is the assumption that the first moment and the standard deviation of the distribution of the CQIs in a random TTI depends mainly on the ratio between average other-cell interference to average own-cell interference (or other-to-own interference ratio in short). We further assume that within an inter-event time, the SIR evolution is a stationary process, which in essence means that the UE stays roughly within its original area element. Note however that it is also possible to model mobility by generating a new event if the UE reaches a new area element.

We assume idealized CQI reporting, i.e. perfect estimation of the channel, no feedback delay, and constant channel quality during a single TTI. These assumptions are realistic for slowly moving users with relatively slow channel variations.

In [86], the following relation between SIR and CQI has been found, which will be used later for the TBS distribution:

$$\text{CQI} = \max\left(0, \min\left(30, \left\lfloor \frac{\text{SIR[dB]}}{1.02} + 16.62 \right\rfloor\right)\right). \tag{3.14}$$

The CQI-value q defines the maximum possible transport block size (TBS) $v(q) \in \mathcal{V}_k$, that can be transmitted in one TTI. The set of transport block sizes $\mathcal{V}_k$ depends on the category of the user equipment and is defined in [73]. The CQI value and the TBS also defines the number of required parallel codes $n_{HS}(q)$. If the number of available codes n_{HS} is lower than $n_{HS}(q)$ the scheduler selects the maximum possible TBS-value according to n_{HS}. This means that an optimal usage of resources is only possible if the transmission format according to the reported CQI utilizes all available codes. If too few code resources are available, power resources are wasted, and if too few power resources are available, the CQI is too small to utilize all available codes. The reported CQI value depends essentially on the multi-path profile, the user's location, the available HS-DSCH power, and the other-cell power. The number of codes required for a certain CQI value depends on the CQI category.

In the following, we first explain our SIR model including multi-path propagation. Then we describe how to determine the resulting CQI and TBS probability distributions [28]. These are then used to model different scheduling mechanisms for the calculation of the average user data rates.

3.5.1 SIR and CQI Distribution

Consider an HS-DSCH with transmit power $T_{x,HS} = \Delta_{x,HS} \cdot T_{x,tot}$ and n_{HS} parallel codes allocated to the HS-DSCH. The variable $\Delta_{x,HS}$ describes the fraction of the HS-DSCH transmit power on the total NodeB transmit power. The propagation channel between NodeB x and mobile k is subject to multi-path propagation with a set of $\mathcal{P}_x$ independent paths. Each path experiences Rayleigh fading with an instantaneous propagation gain of

$$d_{x,k,p} = d_{x,k} \cdot \beta_p, \tag{3.15}$$

where $d_{x,k}$ is the average propagation gain and β_p is an exponentially distributed random variable with normalized means m_{β_p} such that $\sum_{p\in\mathcal{P}} m_{\beta_p} = 1$. In this work we refer to the Pedestrian A and B and Vehicular A multi-path propagation models [81]. We assume a RAKE-receiver at the UE side with a finger on each path and perfect Maximum Ratio Combining of the path signals. Accordingly, the SIR at a UE k is equal to

$$\gamma_k = \Delta_{HS} \cdot \underline{\gamma}_k, \tag{3.16}$$

where $\underline{\gamma}_k$ is the *normalized SIR*, which is defined as

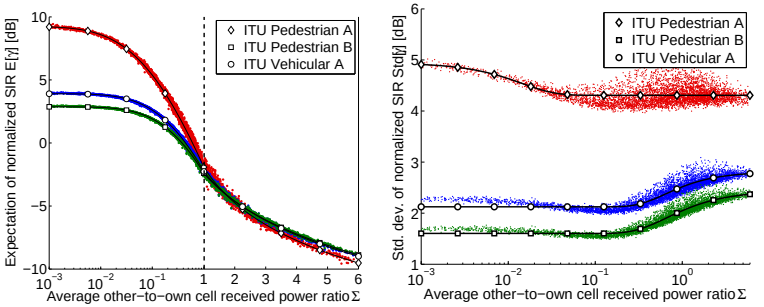
$$\underline{\gamma}_k = \sum_{p\in\mathcal{P}} \frac{\beta_p}{\sum\limits_{y\in\mathcal{L}\setminus x} \left(\frac{T_{y,tot}\cdot d_{y,k}}{T_{x,tot}\cdot d_{x,k}} \sum\limits_{q\in\mathcal{P}_y} \beta_q \right) + \sum\limits_{r\in\mathcal{P}_x\setminus p} \beta_r}. \tag{3.17}$$

The first term in the denominator of this equation describes the other-cell interference from adjacent NodeBs, which is also subject to multi-path propagation and fading. The second term is self-interference from other signal paths than the considered path p. Thermal noise is neglected since it is in well-designed networks by magnitudes smaller than the multiple access interference. All other variables are equally defined as in Section 3.4.

In Equation (3.17), the impact of the other-cell interference increases on average if the mobile is closer to the cell border. Our assumption is therefore that the mean and the variance of the SIR distribution is a function of the average other-cell received power to average own-cell received power (or "other-to-own cell power ratio" in short) Σ_k:

$$\Sigma_k = \sum_{y\in\mathcal{L}\setminus x} \frac{T_{y,tot}\cdot d_{y,k}}{T_{x,tot}\cdot d_{x,k}}. \tag{3.18}$$

Note that the other-to-own interference ratio is the reciprocal of the "G-factor" which is introduced e.g. in [102] to describe the influence of the geometry on the HSDPA performance.



(a) Mean normalized SIR

(b) Standard deviation of normalized SIR

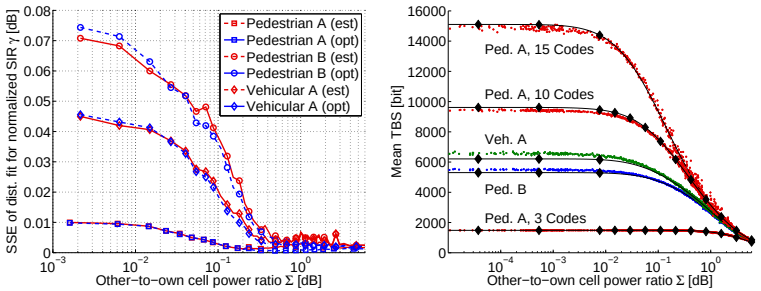
Figure 3.3: *Mean and standard deviation of the normalized SIR versus the other-to-own interference ratio.*

So, we are interested on the probability distribution of the normalized SIR and in the functions $f_E(\Sigma)$ and $f_{STD}(\Sigma)$ that map Σ_k to the mean $E[\gamma_k]$ and the standard deviation $STD[\gamma_k]$ in decibel scale. We propose to use the following four-parametric Weibull function

$$f_{a,b,c,d}(x) = a - b \cdot e^{-c \cdot x^d}, \tag{3.19}$$

both for f_E and f_{STD}. The parameters of the functions f_E and f_{STD} are found by fitting on results obtained with Monte-Carlo simulations. The simulation generates NodeBs with uniformly distributed transmit powers in a square area according to a homogeneous spatial Poisson process, and measures the received powers at the randomly located UEs.

Figures 3.3(a) and 3.3(b) show the mean and standard deviation of the normalized SIR depending on the other-to-own interference ratio for different multi-path propagation profiles defined in [81]. The dots represent simulated values, and the solid lines the fitted Weibull functions. Note that the x-axis of Figure 3.3(a) is divided in two intervals in order to improve the visibility of the results. On the left side, the average other-to-own ratio is in linear scale, while on the right side



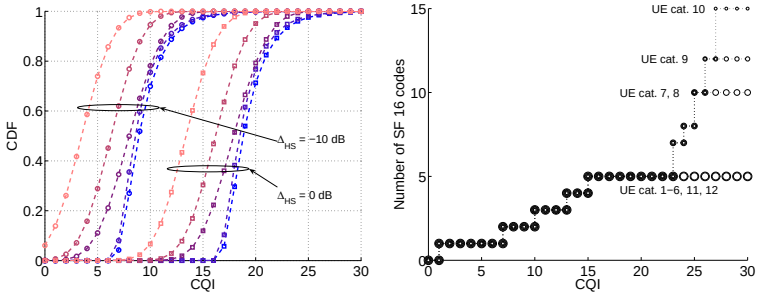
(a) SSE plot for distribution fit

(b) Estimated mean TBS and simulation results

Figure 3.4: *Comparison between estimated and simulated SIR and TBS results.*

it is in decibel scale. We see that while the functions are fairly in the center of the simulation results, the results for the mean normalized SIR are more function-like than the results for the standard deviation, especially for higher values of Σ_k. An interesting observation is that the standard deviation is nearly independent of Σ_k, while the mean varies from $+9$ dB to -9 dB.

In the next step, the SIR distribution is approximated by fitting a suitable probability distribution to the results of the simulation with help of the mean and standard deviation values. We experimented with several distributions both in linear and decibel domain. The results show that the normal distribution in decibel domain is a good candidate for all three multi-path profiles, as shown in Figure 3.4(a), although in some cases it is also beneficial to apply different distributions for Σ-ranges below and above $\Sigma = 0.1$ [28]. The figure shows the maximum sum of squared error (SSE) of the normal distribution fit in decibel scale versus the average other-to-own interference ratio. Results indicated with (est) are generated with estimated distribution parameters according to the functions $f_E(\Sigma)$ and $f_{STD}(\Sigma)$. For results indicated with (opt), the actual mean and standard deviations gained from the simulation samples have been used. We observe that the SSE becomes smaller with an increasing other-to-own received power ratio, and



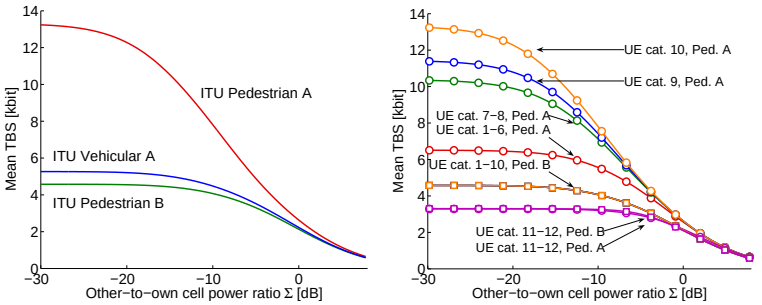
(a) CDF of the CQIs for different values of Σ and Δ_{HS}

(b) CQI vs. required codes for different UE categories

Figure 3.5: *CQI distributions (left) and code requirements (right).*

that the difference for optimal distribution parameters from simulation results and estimated distribution parameters from the fitted Weibull function is small. In Figure 3.4(b) we compare the estimated mean TBS (solid lines with diamond markers) to simulation results (shown as dots). For the Pedestrian A multi-path profile, results are shown for different maximum number of parallel codes (3, 10, and 15 codes). The figure affirms the increasing accuracy of the estimation with increasing Σ-values. For small Σ-values, i.e. very close to the antenna, the model slightly underestimates the mean TBS.

The distribution of the CQI values is then obtained as discretization of the SIR distribution by applying Equation (3.14). Figure 3.5(a) shows the cumulative distribution functions (CDF) of the CQI values for several other-to-own-interference and transmit power ratios. The Σ-values range from 0 dB to -20 dB with a step size of 5 dB. Lower Σ-values are indicated by darker lines. We observe that with a lower Σ, the shape of the CDF becomes steeper due to a lower variance and an increasing positive skew. Additionally, the transmit-power ratio is set to 0 dB and -10 dB, which causes a shift of the distributions to lower CQIs, but does not change the variance nor the shape of the distributions.



(a) Impact of multi-path profile

(b) Impact of UE category

Figure 3.6: *The multi-path propagation profile has a strong influence on the UE performance in the inner cell area (left). Accordingly, UE categories are most relevant with good channel conditions (right).*

Finally the probability distribution of the TBS values of a user k, $p_{\text{TBS,k}}(q)$, is calculated from the CQI distribution. The maximum possible TBS v^* follows from the available code resources as described in Section 3.3.1. Figure 3.5(b) clarifies the relation between CQI and required codes. Truncating the CQI distribution according to the available codes for the HS-DSCH yields the probability distribution of the TBS-values for a user k as

$$p_{\text{TBS,k}}(v) = \begin{cases} p_{\text{CQI,k}}(v(q)), & \text{if } v(q) < v^*, \\ \sum_{q=q^*}^{30} p_{\text{CQI,k}}(q), & \text{else,} \end{cases} \tag{3.20}$$

where q^* is the maximum allowed CQI corresponding to the maximum possible TBS v^*. Furthermore, we denote the CDF of the CQI and TBS values with $P_{\text{CQI,k}}(q)$ and $P_{\text{TBS,k}}(v)$.

The physical layer abstraction model gives us insights into the impact of system parameters like multi-path channel profile, number of available codes and UE category. Figure 3.6(a) shows the gross data rate, i.e. the throughput a single UE would achieve, depending on the other-to-own-interference ratio for the

ITU Vehicular A, Pedestrian A and Vehicular B multi-path propagation models. A profile with a strong dominating path, like in Pedestrian A, enables indeed very high data rates up to 12.8 Mbps. In contrast, profiles with several paths with similar mean power, like for Vehicular A and Vehicular B, lead to significantly lower data rates. In fact, with these two models, it is sufficient to provide five SF 16 codes for the HS-DSCH. Figure 3.6(b) shows the gross data rates for different UE categories, which reflect the capability for 16 QAM, number of parallel codes and inter-scheduling time. It is remarkable that UEs without 16 QAM support (categories 11 and 12) have significantly lower data rates than UEs with 16 QAM, although the transport block sizes are identical (categories 1–6).

3.5.2 Scheduling

The scheduler in the NodeB has a large influence on the user-level and system-level performance of the HSDPA. Several proposals exist for HSDPA scheduling, from which we implemented the three most common ones into the framework. The Round-Robin-scheduler, although not channel-aware, is easy to implement and time-fair, which is often sufficient to prevent starvation of users on the cell edge. The MaxTBS-scheduler chooses always the user with the currently best possible TBS, including restrictions due to code resources. This may lead to starvation of users with bad channel conditions but also optimizes the instantaneous cell throughput. Finally, the Proportional-fair scheduler selects the user which has the proportionally best TBS in relation to its past throughput.

For the calculation of the mean throughput we need the conditional probability $p_{s,k}(v)$ that a user k is scheduled with TBS v. This probability depends on the scheduling discipline. According to the theorem of total probability, the expected transmitted data volume per TTI is then

$$E[V_{k,TTI}] = \frac{1}{1+p_{err}} \cdot \sum_{v \in \mathcal{V}_k} v \cdot p_{TBS,k}(v) \cdot p_{s,k}(v), \tag{3.21}$$

where p_{err} is the probability for an erroneous first transmission within the Hybrid Automatic Repeat Request (HARQ) process. Further retransmissions have a

negligible impact on the throughput approximation.

Round-Robin Scheduling

The *Round-Robin* scheduler selects the users consecutively for transmission. The probability that a user is selected is approximately

$$p_{s,k} = \frac{1}{|\mathcal{M}|} \tag{3.22}$$

for a sufficiently long time interval. This means that the user throughput in this case depends on the number of users, but not on the location or the radio conditions of the other users. Therefore, the mean transmitted data volume in a random TTI is

$$E[V_{k,TTI}] = \frac{E[v_k]}{|\mathcal{M}| \cdot (1 + p_{err})}, \tag{3.23}$$

where $E[v_k] = \sum_{v \in \mathcal{V}_k} v \cdot p_{TBS,k}(v)$ is the mean *possible* TBS of user k. This enables to speed up the computation by storing the possible mean TBS for all values of Σ, Δ, and $C_{x,h}$ in a database.

MaxTBS Scheduling

With *MaxTBS* (or Max C/I) scheduling, the user with the currently best TBS is scheduled. If two or more users have the highest TBS, a random user out of this set is chosen. We assume that the scheduler decides on behalf of the maximum *possible* TBS, i.e. the TBS that can be assigned if the available code resources are taken into account. Users with low SNR-values have therefore a higher probability to are in tie with high-SNR users, meaning that they are more often selected for scheduling. This rule introduces therefore an additional degree of fairness into the system.

In contrast to Round-Robin scheduling, the throughput of a user depends not only on its own location, but also on the location of the other users. In [96], this scheduling discipline is modeled as a priority queue, where locations closer to

the NodeB have higher priority than locations farther away. However, it is also possible to calculate the mean throughputs directly from the TBS-distributions of the users. Given a certain possible TBS v of user k, the probability that this user is scheduled can be expressed as

$$p_{s,k}(v) = \left(\prod_{m \in \mathcal{M} \setminus k} P_{TBS,m}(v) \right) \cdot \left(\sum_{n=0}^{|\mathcal{M}|-1} \frac{1}{n+1} \cdot p_{N_{k,v}}(n) \right). \tag{3.24}$$

The term in the first two brackets corresponds to the probability that the possible TBS of all other users is not higher than the TBS of user k, which is the probability that the maximum TBS of this set of users is v. The second term is the probability that user k is selected for scheduling even if some other users have an equal TBS value. The running variable n in the sum denotes the number of users with TBS values equal to v. We assume that each user is selected with equal probability, which is expressed by the probability $\frac{1}{1+n}$.

The probability that n users have TBS k under condition that the maximum TBS is v is denoted with $p_{N_{k,v}}(n)$. For the calculation, let us first define the Bernoulli random variable $\xi_i(v)$ which denotes the case that the experienced TBS V_i of user i is v, under condition that v is the maximum allowed TBS:

$$\xi_i(v) = \begin{cases} 1 & \text{if } V_i = v \text{ under condition } V_i \leq v, \\ 0 & \text{else.} \end{cases} \tag{3.25}$$

The corresponding probability $P(\xi_i(v) = X)$ is given by

$$P(\xi_i(v) = X) = \begin{cases} 1 - \frac{p_{TBS,i}(v)}{P_{TBS,i}(v)} & \text{if } X = 0, \\ \frac{p_{TBS,i}(v)}{P_{TBS,i}(v)} & \text{else.} \end{cases} \tag{3.26}$$

Let us further define with the random variable $N_{k,v} = \sum_{m \in \mathcal{M} \setminus k} \xi_m(v)$ the number of users which have additionally to user k the TBS-value v. The probability $p_{N_{k,v}}(n) = P(N_{k,v} = n)$ follows then from the convolution of $P(\xi_i(v))$ as

$$p_{N_{k,v}}(n) = \circledast_{m \in \mathcal{M} \setminus k} P\left(\xi_m(v) = n\right), \tag{3.27}$$

where $\circledast$ denotes the discrete convolution operator.

The drawback of this scheduling discipline in terms of computation time is that in contrast to the Round-Robin scheduler it is not possible to store the mean throughput values for all situations in a large database, since the scheduling probability depends not only on the own TBS distribution, but also on the other users involved. A simple implementation requires therefore the calculation of the TBS distribution on each event, which leads to significant higher computation requirements mainly due to the convolution in Equation (3.27). A more sophisticated implementation could therefore estimate the impact of certain users on the TBS distribution and neglect users which have a very low probability to get scheduled.

Proportional-fair Scheduling

Proportional-fair (PF) scheduling is a scheduling discipline which has been developed for the 1xEv-DO-system in the downlink, [59]. Its basic principle is to give each user the bandwidth proportional to its link quality and its past throughput. This is achieved by choosing the user which has the best instantaneous relative throughput over its past throughput, which is often determined with a sliding window approach. However, different versions of PF scheduling exist. The most fundamental difference is the way how the past throughput is calculated. The first variant updates the past throughput every scheduling period regardless whether the user has been scheduled or not, the second variant updates the past throughput only if the user is indeed chosen for transmission. The difference between both versions is that in the first case the mean throughput of a user is proportional to only its channel quality, while in the second case it is also related to the generated traffic. In [103] and [104] it is argued that both variants approximately lead to the same results in case of statistically identical fading and infinite backlogs. The second assumption is reasonable during the inter-event time, while the first assumption is contradicted by the fact that the shape of the CQI distribution depends on the level of received other-cell interference. For a comparison of both variants refer to [6].

Let us introduce the random variable $\theta_k(v_k) = v_k/\bar{v}_k$ as the proportional-fair scheduling index for the k-th user. We assume an infinite history for the mean TBS $\bar{v}_k$, which means that it can be calculated directly from the TBS distribution as $\bar{v}_k = E[v_k]$. The CDF of θ_k is then

$$P_{PF,k}(\theta) = P_{TBS,k}(\theta \cdot E[v_k]). \tag{3.28}$$

Neglecting the very small probability that two users have the same PF-index, the probability to be scheduled when having TBS v is given by

$$p_{s,k}(v) = \prod_{m \in \mathcal{M} \setminus k} P_{PF,m}(\theta_m(v)). \tag{3.29}$$

In fact, the only way for two users having the same PF-index is due to discretization errors in the scheduler implementation. Thus, the proportional-fair scheduler is computationally less demanding than the *MaxTBS* scheduler, since no convolution has to be performed. However, the computation still depends on the TBS distributions of all users and is therefore more time-consuming than *Round-Robin* scheduling.

3.6 Flow-Level Performance Results

In this section, we present some numerical results to show different aspects and influencing factors on the HSDPA and DCH performance. For the results, an event-based flow-level simulation is used. Users are generated according to a Poisson process with inter-arrival time $1/\lambda$ in one of the two network scenarios we described in Chapter 2. Dedicated Channel users, i.e. QoS users, are modeled with a time-based traffic model, such that they leave the system after a certain time which is determined at arrival. This time is exponentially distributed with mean $1/\mu$. HSDPA users are volume-based, i.e. they want to transmit (or receive) a certain data volume, which is also exponentially distributed with mean $E[V_{HS}]$. Note that other distributions are also possible. The HSDPA and DCH users do not change the location during their lifetime.

On each event, NodeB transmit power and interference, code resources, and HSDPA user throughput are recalculated if necessary. On a DCH arrival or departure, HSDPA code resources in the relevant cells are decreased or increased according to the DCH code requirements. Additionally, the total transmit powers are updated for all NodeBs and the interference at the user locations are calculated. Transmit powers are also recalculated if the HS-DSCH is switched on or off because of HSDPA user arrivals or departures. In all cases, the data volume transmitted by HSDPA users within the past inter-event time is subtracted from their remaining data volumes. New HSDPA data rates are calculated, taking the new radio resource and interference situation into account. Finally, the expected departure times of the HSDPA users are updated according to the remaining data volumes and data rates.

Admission control for HSDPA users is realized with a simple count-based approach, the threshold is set to 10 concurrently active users if not indicated otherwise. Admission control for DCH connections is based on code and transmit power, while the latter is calculated under the assumption of full NodeB power, as we explained in Section 3.4. The NodeB target power is $T_{target,x} = 10\,\text{W}$, from which $T_{CCH} = 2\,\text{W}$ are reserved for common and pilot channels. The network

granularity, i.e. the length of the edge of an area element, is 25 m.

We consider two DCH classes with 128 kbps and 384 kbps and target-E_b/N_0 values of 3 dB and 4 dB. The code resource requirements of these classes are 32 and 64 SF 512 code units, which corresponds to spreading factors of SF 8 and SF 16. The offered DCH load is defined with the code load as

$$\rho_c = \sum_{s \in \mathcal{S}} \frac{\lambda_s}{\mu_s} \cdot \frac{c_s}{C_{tot}}, \tag{3.30}$$

where $\mathcal{S}$ denotes the set of service classes, and c_s the code requirements of service class s. The mean DCH call time is 120 s for both service classes, the arrival rates follow then from Equation (3.30).

3.6.1 Volume-Based Traffic Model and Spatial User Distribution

An important distinction between QoS and elastic flows is that QoS flows typically follow a time-based traffic model, which means that the user wants to keep the connection for a certain time span. In contrast, elastic flows are often modeled as volume based, i.e. the user is satisfied as soon as a certain data volume is transmitted. In reality the user behavior is a mixture between both models, depending on factors like user satisfaction, pricing models, type of content. However, the two models can be seen as extremes cases of real user behavior.

A time-based traffic model implicates that the number of currently active users is independent of the perceived data rates. Moreover, the *spatial distribution* of the number of users corresponds to the spatial *arrival process*: If users arrive according to a homogeneous Poisson process with arrival rate λ, the number of concurrently active users in steady-state follows according to Little's law [105] as λ/μ, if no blocking occurs.

A volume-based traffic model means that users stay in the system until their service demands are satisfied. Therefore, the number of active users depends on the assigned data rates. In HSDPA systems, data rate depends on the channel

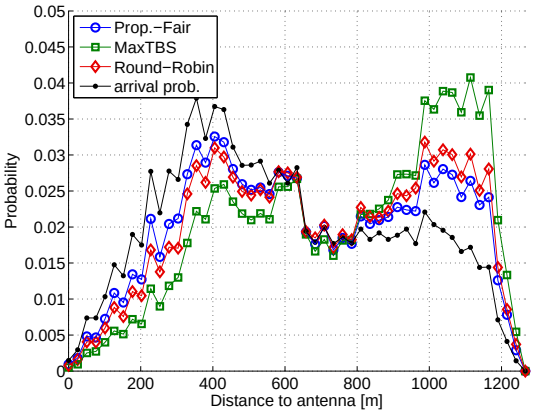


Figure 3.7: *Arrival and residence probabilities for an irregular cell layout with inhomogeneous user arrivals and DCH offered load $\rho_c = 0.4$. The line with diamond markers indicates the user arrival probability.*

quality, which means that users with low average channel qualities stay longer in the system than those with good channel qualities. Since the average channel quality is dominated by other-cell interference, users at the cell edges stay on average longer in the system than users in the center of the cell. This also implicates that the spatial arrival process and the spatial steady state distribution are not congruent anymore, a fact which makes the planning process of HSDPA networks significantly more complex. One reason is that Monte-Carlo methods [95] now have to estimate the spatial user population for every snapshot, which is difficult without knowledge of the currently ongoing flows. With Round-Robin scheduling, a direct formulation of the mean transfer time was found in [14] and [93], since in that case the data rates of the users only depend on the number of users and their positions, but are otherwise independent of each other.

We now clarify the effect of spatial heterogeneity with some example scenarios. Figure 3.7 shows the arrival probability and the residence probability vs.

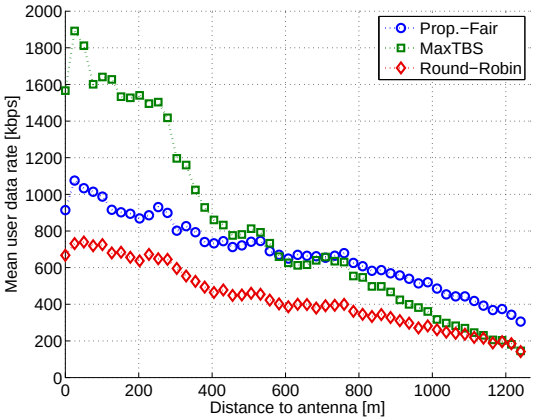


Figure 3.8: *Mean data rate vs. distance to antenna with offered DCH load* $\rho_c = 0.4$*. Proportional-fair scheduling leads to an almost constant gain over Round-Robin scheduling independent from the distance.*

the distance to the antenna in a cell which is taken from the irregular scenario. The arrival probability describes the probability that a user arrives in this cell at a certain point, while the residence probability expresses the distribution of the users in the cell how it is seen in steady state. The spiky shape of the curves stems from the discretization of the cell area into area elements. It is obvious that arrival probability and residence probability are not equal, and that the difference depends on the scheduling discipline. MaxTBS scheduling shows as expected the largest differences, since users close to the antenna leave the system much earlier than users farther away. We observe that the residence probabilities with Proportional-fair scheduling match the arrival probabilities better than with Round-Robin scheduling. The reason is that Proportional-Fair scheduling tries to assign data rates more balanced and indeed favors users at the cell edge under certain circumstances. With volume-based users and MaxTBS-scheduling,

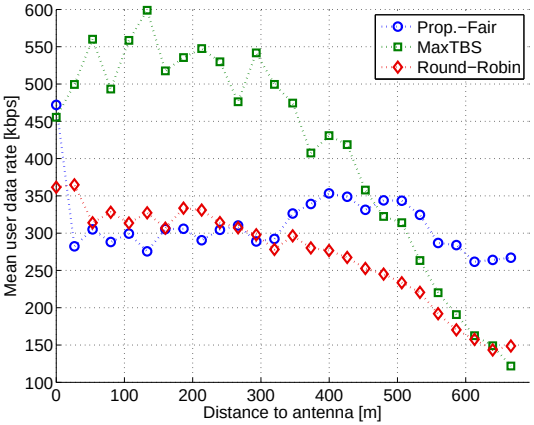


Figure 3.9: *Mean data rate vs. distance to antenna for a hexagonal cell with offered DCH load $\rho_c = 0.6$. The high DCH load leads to low HSDPA throughputs. Proportional-Fair scheduling favors user at cell edges.*

the probability to meet a user at the cell edge is four times higher than the arrival probability at the same location.

Figure 3.8 shows the average user data rate depending on the distance to the antenna. While MaxTBS-scheduling strongly favors users in the cell center, Proportional-fair and Round-Robin scheduling lead to more balanced results. The difference between Round-Robin and Proportional-Fair reflects the scheduling gain due to multi-user diversity. The gain of Proportional-fair over Round-Robin scheduling is almost constant over the whole distance range.

Finally, in Figure 3.9, the average data rate for the enter cell of the homogeneous scenario is shown, but in a scenario with a higher DCH load of $\rho_c = 0.6$. Here, the lack of resources leads to low throughputs, but more interesting is the small throughput gain of users in the outer cell area for Proportional-fair scheduling. This is caused by the higher variance of the TBS distribution of those users

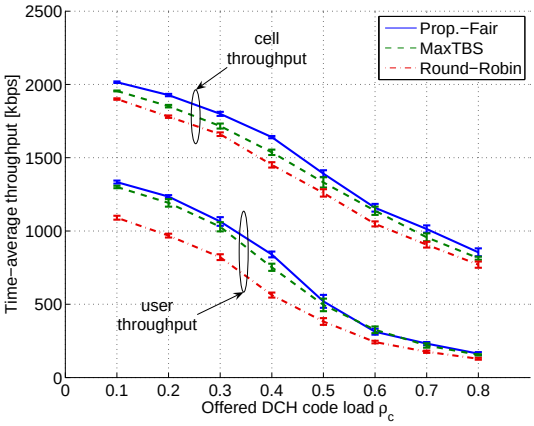


Figure 3.10: *Time-average user and cell throughput vs. offered DCH load for different scheduling disciplines.*

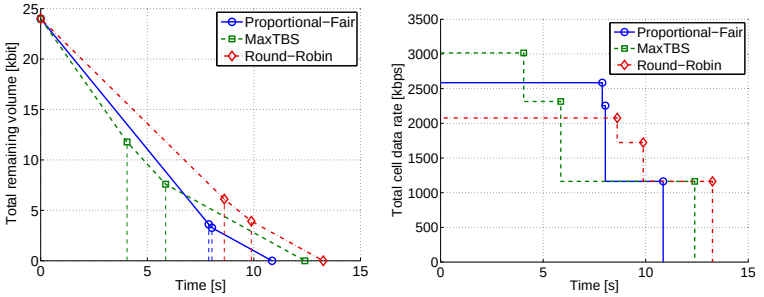
which is also reflected in the probability distribution of the proportional-fair indices defined Section 3.5.2. Due to the high DCH load, the average *possible* throughput of user close and more distant to the antenna is nearly equal. However, the higher variance of the TBS distribution of distant users lets them be more often scheduled compared to users in the inner cell area. For a further discussion of this behavior see also [106].

3.6.2 Impact of Scheduling Disciplines

Let us now investigate the impact of different scheduling disciplines on the overall performance of the network. We consider the homogeneous scenario with the adaptive RRM scheme and increase the offered DCH load from 0.1 to 0.8, which is a very high and therefore unrealistic value, but it helps to understand the system behavior better.

Figure 3.10 shows the resulting time-average cell and user throughput versus the offered DCH load. As expected, the channel-aware scheduling disciplines lead to better results than the channel-blind Round-Robin discipline, regardless of the DCH load. With higher load, the difference between the scheduling disciplines becomes smaller, since the lack of code resources prevents efficient utilization of multi-user diversity. An interesting result is that Proportional-Fair scheduling leads to higher throughput curves than MaxTBS-scheduling, which is at a first glance counter-intuitive. MaxTBS-scheduling provides sum-optimal throughput for a static scenario, i.e. for a fixed number of users and consequently also during any inter-event time. This means, if we look at an arbitrary system snapshot, MaxTBS-scheduling always leads to a higher cell throughput than would Proportional-Fair scheduling in the same snapshot. However, in cases where the differences between the mean channel conditions are large, the MaxTBS scheduler selects nearly always the best user, assigning very low data rates to the remaining UEs, such that their data volume is only marginally reduced. Therefore, users with bad channel conditions stay very long in the system which is reflected in the time-average cell and user throughput. With Proportional-Fair scheduling, "good" users stay longer in the system, but the total data volume transported within a certain time span is higher on average. Note that in principle this also holds for Round-Robin scheduling, however, channel-blindness outweighs this effect such that the average throughput is indeed lower.

In the literature, some numerical results seems to stand in contradiction to the results presented here. In [90] and [107], the system throughput for Round-Robin, Proportional-Fair and Max C/I (which is equivalent to MaxTBS) is compared, and it is concluded that MaxTBS scheduling provides the highest throughput. However, the results apply to static scenarios with persistent data flows for a fixed number of users. In such a scenario, MaxTBS scheduling is optimal, but it is not comparable with the flow-level throughput in a stable system. In [91], users arrive according to a Poisson process and request 100 kByte of data, which is the same amount of data as in our scenario. However, users are dropped from the system if they stay longer than 12.5 s in the system, so in fact the study employs



(a) Total remaining data volume vs. time

(b) Cell throughput vs. time

Figure 3.11: *Comparison between MaxTBS, Proportional-Fair and Round-Robin scheduling for a static three user scenario with fixed data volume. Vertical dashed lines indicate departures.*

a mixture between a time- and volume-based traffic model. Consequently, the results show a small performance gain for Max C/I scheduling. In [87], while it is not clear how users are generated in their simulation, they are dropped from the system if their throughput is lower than 9.6 kbps. It is not clear over which time span the throughput is measured, however, since those are exactly the users with low bandwidth which lead to the effects described above, the gain for MaxTBS scheduling in this case is reasonable.

Figure 3.11(a) demonstrates the difference between the schedulers in a example scenario for a fixed data volume and three users with Σ-values of -20 dB, -10 dB and 0 dB. It shows the remaining total data volume versus time. Figure 3.11(b) shows the corresponding data rates. With MaxTBS scheduling, the first and second users leave the system faster than with the other disciplines (indicated by the vertical dashed lines), but the remaining data volume of the "worst" user with $\Sigma = 0$ dB is so large that in total, the Proportional-Fair scheduler needs less time to transport the whole data volume. Note that it depends on channel profile and the cell layout how large the advantage of the Proportional-Fair scheduler

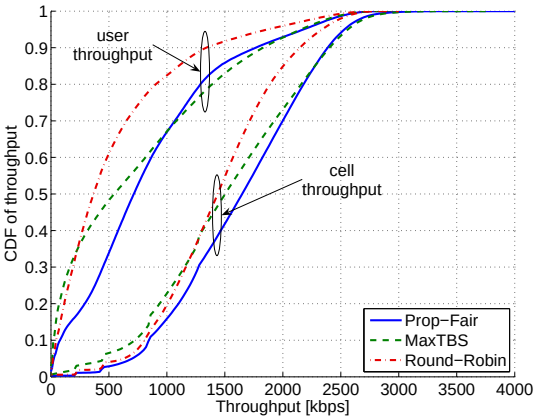


Figure 3.12: *CDF of user and cell throughput for an offered DCH load of* $\rho_c = 0.4$*. With MaxTBS scheduling, the majority of the users gets lower data rates than with Proportional-Fair scheduling*

is and whether it exists at all. With channel profiles which have a "flat" curve progression for high Σ-values (see Figure 3.6(a)), the differences in the data rate assignments of the MaxTBS scheduler are not so distinctive, such that in this case the MaxTBS scheduler is better than the Proportional-Fair scheduler.

Finally, Figure 3.12 shows the CDF of the user and cell throughputs for an offered DCH load of $\rho_c = 0.4$. The CDF of the MaxTBS scheduler confirms the time-average throughput curves: a large portion of the probability weight is on very low data rates, but in the same time the higher quantiles, e.g. for 0.8, are higher than for Proportional-Fair and Round-Robin scheduling. Also note the stair-like shape of cell-throughput CDF for low data rates, which is caused by preemption from DCH connections.

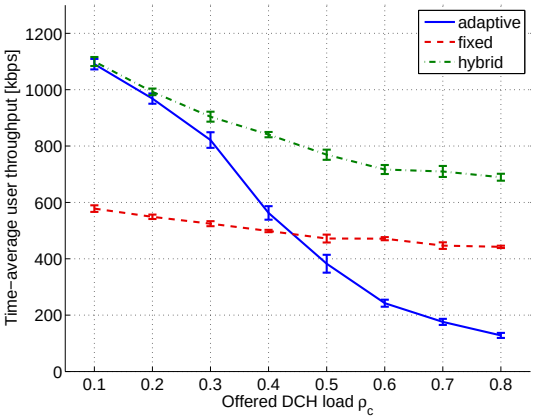
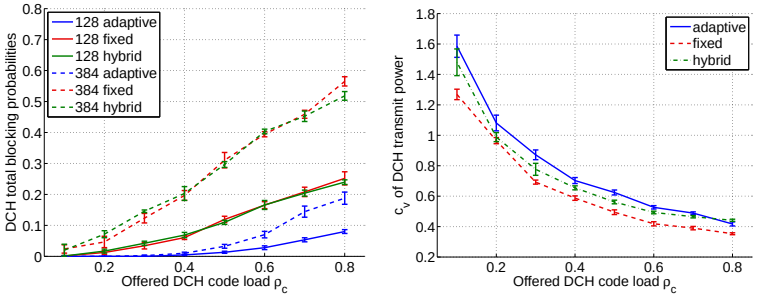


Figure 3.13: *HSDPA user throughput vs. offered DCH load.*

3.6.3 Impact of DCH Radio Bearers

We now investigate the impact of the three radio resource reservation schemes, namely *adaptive*, *hybrid* and *fixed* on the user and system performance. The results are generated for the 19-cell hexagonal network layout with homogeneous user arrivals. In the first scenario, we keep the reserved resources for the HS-DSCH constant and increase the arrival rate of the DCH users, which can either use 384 kbps or 128 kbps radio bearers with a service mix of 0.4 to 0.6. Both for the *hybrid* and for the *fixed* scenario the reserved transmit power is set to $T_h^* = 4$ W. Additionally, $N_h = 5$ codes are reserved. The impact on the HS-DPA user throughput which is calculated as the weighted time-average over all connections is shown in Fig. 3.13.

In the *adaptive* case, the user throughput decreases steeply with increasing DCH load. This trend increases with a DCH load of 0.3, since in this case the resource preemption by DCH users leads to lower peak rates in the cell center and increases the probability that HSDPA users at the cell border get CQI values



(a) Total DCH blocking probabilities vs. offered DCH load

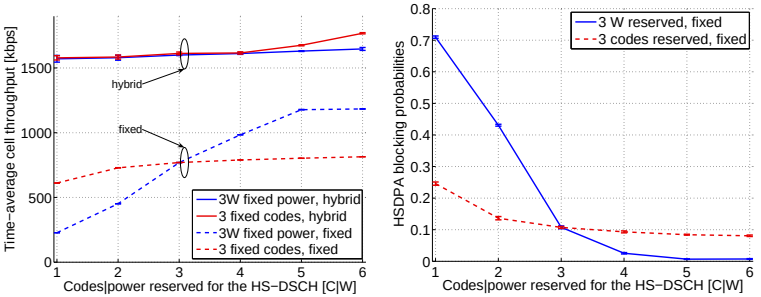
(b) Coefficient of variation of DCH transmit power vs. offered DCH load

Figure 3.14: *The trade-off between HSDPA and DCH performance is visible in the DCH blocking probabilities. The adaptive strategy leads to the highest DCH power fluctuations.*

of 0, which means that no transmission is possible at all. The resource reservation of the *hybrid* and *fixed* strategies prevents this strong decline. The *fixed* scheme shows as expected a smaller sensitivity to the DCH load, but the HS-DSCH also is not able to exploit the spare resources from the DCH connections, which leads to a significant lower throughput if compared to the *hybrid* scheme. Notable is the influence of the interference in the *fixed* case, which leads to a decrease of 100 kbps over the total range of the offered loads.

Figure 3.14(a) clarifies the trade-off between HSDPA and DCH performance: Due to resource reservation, the total blocking probabilities, which comprise code and soft blocking, reach up to 40% for the 384 kbps service class. Note that the blocking probabilities for the *fixed* and *hybrid* RRA schemes are nearly identical. The HSDPA blocking probabilities in the *fixed* and *hybrid* case are very close to zero. For the *adaptive* scheme, the blocking probability increases from 0.025 up to 0.2 for a DCH load from 0.3 to 0.8.

The impact of the allocation schemes on the coefficient of variation of the



(a) HSDPA cell throughput

(b) HSDPA blocking probabilities

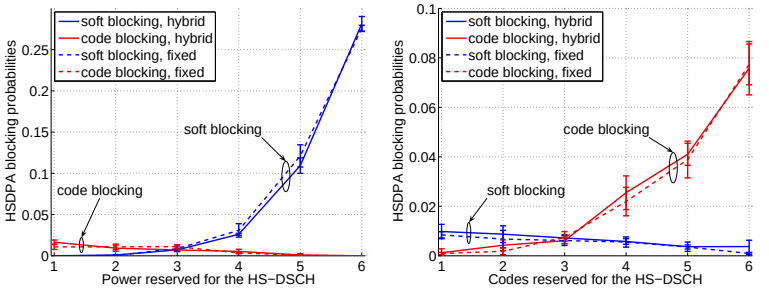
Figure 3.15: *Impact of code and power reservation on the code/power balance for hybrid and fixed reservation schemes.*

DCH transmit powers is shown in Fig. 3.14(b). Generally, a high variability of the transmit powers is malicious to the system, since large steps of the interference level (e.g. because of on-off-switching of connections) has to be compensated by the inner loop power control, which has normally a step size of only 1 dB. So, a high variability may lead to increased target-E_b/N_0 values for DCH users. From this perspective, the *fixed* scheme has advantages over the other schemes.

3.6.4 Sensitivity against Resource Reservation

In the next scenario, we investigate the sensitivity of the *fixed* and *hybrid* radio resource allocation (RRA) against code and power reservation. We keep either the number of reserved codes constant and vary the power reservation or vice versa. The range for the resource reservation is in both cases from 1 to 6, i.e. 1 code to 6 codes or 1 Watt to 6 Watt. The constant resource is set to 3 codes and 3 W power, respectively. The DCH offered load is set to 0.4, and only one service class (128 kbps) is considered.

Figure 3.15(a) shows the impact of resource reservation on the mean HSDPA



(a) Fixed power (b) Fixed codes

Figure 3.16: *DCH soft and code blocking probabilities for fixed power and fixed code reservation. In this scenario, reservation of* 3 *W transmit power and* 3 *codes leads to a balanced code/power ratio.*

cell throughput. As in the previous scenario the *hybrid* RRA has a significant performance gain due to the exploitation of spare resources. The *hybrid* scheme in this scenario is not very sensitive against reservation, only in the case of 6 W power reservation an increase is notable. This is in contrast to the *fixed* scheme, where especially code reservation up to 5 codes leads to an increased bandwidth. More than 5 codes cannot be used due to the multipath profile. The insensitivity of the *fixed* scheme against power shows that with 3 codes, 3 W transmit power is sufficient. However, we see in Fig. 3.15(b) that the resulting bandwidth with 3 codes leads to quite high HSDPA blocking probabilities, but 4 and 5 codes with 3 W power show acceptable results. The HSDPA blocking probabilities for *hybrid* RRA are very small, so we do not inlude them in this figure. The DCH blocking probabilities show in both cases a strong sensitivity against power reservation, as shown in Fig. 3.16(a). More than 4 W power reservation leads to soft blocking probabilities higher than 5%. Code reservation naturally leads to higher code blocking probabilities, however, the curve progression is not as steep as for soft blocking with power reservation. We can state here that 4 reserved codes do

still lead to an acceptable performance for both schemes. The soft blocking probabilities for both schemes are nearly identical, which indicates that interference does not have a large influence here.

3.6.5 Impact of Transmit Power Allocation Schemes

Additionally to resource reservation schemes, the transmit power allocation scheme has a significant influence on the system performance, as we will show in this section. The basic difficulty is the following: if the offered load for the HSDPA is not very high, periods without traffic (OFF-periods) alternate with periods where the HS-DSCH is switched ON. If we consider a "traffic-aware" transmit power scheme for the HS-DSCH, this means that the HS-DSCH is switched on and off possibly in a very fast pattern, since the minimum scheduling time corresponds to the transport time interval of 2 ms. This is a potential problem for the power control of DCH connections. Firstly, the power control step size is maximally ± 1 dB, but the maximum transmit power for the HS-DSCH is around 42.5 dBm (corresponds to approximately 18 W, if we assume 20 W maximum output power and 2 W pilot/common channels). Secondly, 2 ms TTI allows only for 3 power control commands (1 per slot). This means that the DCH fast power control may not be able to counter the fast fluctuations in interference fast enough, which leads to the degradation of connection quality or to increased average transmit powers due to higher target-E_b/N_0-values from outer loop power control.

An obvious solution to this problem is to leave the HS-DSCH always switched ON and to transmit padding bits if buffers are empty. We refer to this scheme as "continuous". However, this leads to additional interference in the system and to lower overall performance. A hint of the influence of interference can be seen in Figure 3.13 for the *fixed* reservation scheme. Therefore, we additionally consider the power-ramping schemes which avoids fast interference fluctuations by increasing and decreasing the HS-DSCH transmit power in small steps. Figure

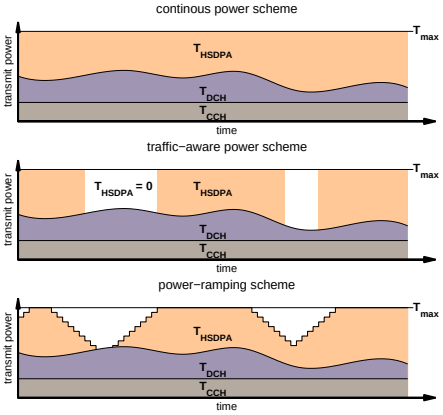


Figure 3.17: *Continuous, traffic-aware and power-ramping transmit power allocation schemes for the HS-DSCH.*

3.17 clarifies the concept of the different schemes.

However, we are not able to implement power control directly, since this would require the complete simulation of all power control commands in the whole network. Instead, we focus on the large-scale effects of the different schemes on the network-wide interference and the resulting impact on HSDPA bandwidth and blocking probabilities.

In the literature, power allocation schemes are mostly considered in the context of general radio resource management schemes. In [94] it is assumed that the HSDPA is always saturated with traffic, which then corresponds to a continuous transmit power scheme. In [92], a traffic-aware scheme has been implemented. In general, only few publications can be found on this subject.

The evaluation of the three power allocation schemes requires a small modification of the transmit power calculation. In a ramping phase, i.e. in the time

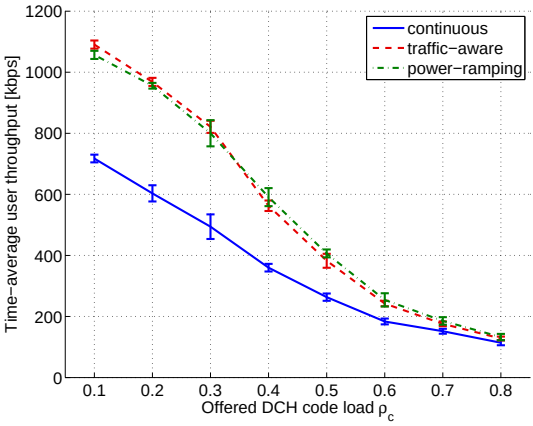


Figure 3.18: *HSDPA user throughput vs. offered DCH load for different power allocation strategies.*

between the HS-DSCH is switched ON and the transmit power reaches its maximum, the boolean variable $\delta_{x,HS}$ is set to zero. Additionally, the ramping power $T_{x,r}$ is step-wise increased (decreased) until the target-power is reached (the power is zero). The time between two ramping steps is set to 2 ms, which gives the fast power control time for 3 adjustments. The total transmit power is then calculated as

$$T_{x,tot} = \delta_{x,HS} \cdot T_x^* + (1 - \delta_{x,HS}) \cdot \left(T_{x,CCH} + T_{x,r} + \sum_{y \in \mathcal{L}} \eta_{x,y} \cdot T_{y,tot} \right), \tag{3.31}$$

and the equation system (3.13) is accordingly modified.

The evaluation is performed on the hexagonal, homogeneous network with adaptive resource reservation and Round-Robin scheduling. In the first scenario, shown in Figure 3.18, we increase the DCH offered load and see as expected the user throughput decreasing. Notable is the large difference between the continu-

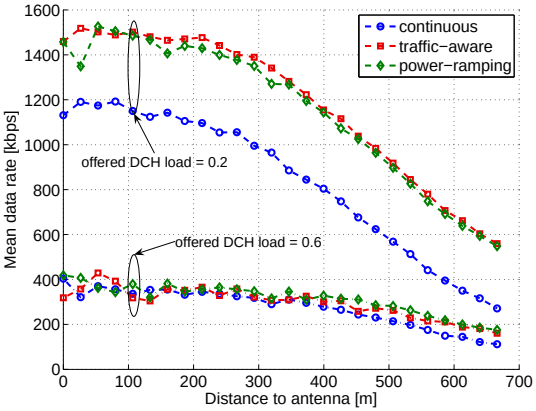


Figure 3.19: *Mean data rate vs. distance to antenna. The gain of the traffic-aware and power-ramping schemes is nearly independent of the distance.*

ous scheme on the one side and the traffic-aware and power-ramping schemes on the other side, which is around 250 kbps for lower DCH loads and is then diminishing with higher loads. The following reasons can be identified: First, the larger interference values in the continuous case generally leads to lower bit rates. Secondly, since HSDPA user behavior follows a volume-based traffic model, higher bit rates also mean shorter sojourn times, which in turn leads to lower HS-DSCH activity and therefore again to a lower average interference for the traffic-aware scheme.

In Figure 3.19, the conditional mean user throughput at a certain distance of the user to the NodeB for two scenarios with DCH loads 0.2 and 0.6 is shown. In case of a DCH load equal to 0.2, the continuous scheme leads to an almost constant performance loss of more then 200 kbps. The power-ramping scheme leads to slightly lower throughputs than the traffic-aware scheme due to the up-ramping at the beginning of a HSDPA transmission. However, in case of a DCH load of 0.6, the power-ramping scheme shows better results than the traffic-aware

scheme for larger distances, which is counter-intuitive at first sight. An explanation for this behavior is that DCH users in surrounding cells generate more other-cell interference the higher the transmit power in the own cell is, which in turn leads to lower bit rates for HSDPA users close to the cell boundary. Since this dependency is non-linear, the power-ramping scheme causes less other-cell interference such that especially users at the cell border can benefit. In case of the traffic-aware scheme, the transmission with maximum power leads therefore effectively to lower bit rates for users close to the cell boundary. It is also an interesting fact that the performance loss for the continuous scheme is nearly independent of the distance, although the additional interference is higher for users which are closer to the cell edge. The reason for this behavior is the employed Round-Robin scheduling, which leads to time-fair resource assignments between all users. This means that if users stay longer in the system on the cell edges, they also affect users in the inner area by taking away their resources.

3.7 Analytical Model

In the previous sections, we used time-dynamic flow-level simulations for evaluation. In this section, we introduce an analytic model based on a queuing model approach. The need for an analytical model is motivated by their lower computing time requirements if compared to simulations. This is especially useful for the planning or optimization of large networks. The trade-off, as we will see later, is in flexibility and level of detail.

Our model integrates DCH and HSDPA connections in a common state space, similar to the approach proposed in [108, 109]. Since the resource requirements of DCH connections depend on their service class, this approach theoretically requires at least an $(S + 1)$-dimensional state space where S is the number of supported service classes. This "state-space explosion" leads inevitably to computational problems, which we counter with a state-space reduction technique based on the Kaufman-Roberts [110, 111] (or Fortet-Grandjean, [112]) recursion.

We assume that DCH and HSDPA users arrive according to an Poisson arrival process with rates λ_s and λ_H, respectively. As in the previous sections, DCH users follow a time-based traffic model with exponentially distributed sojourn times with mean $1/\mu_s$. HSDPA users transmit a exponentially distributed data volume with mean $E[V_H]$. We evaluate the hexagonal network scenario with homogeneous user distribution.

We consider adaptive resource allocation, so theoretically it would also have been possible to implement a time-decomposition approach as in [113]. We decided against this option in order to preserve the flexibility to implement resource reservation schemes or more sophisticated admission control schemes.

3.7.1 Transmit Powers

The calculation of downlink transmit powers, or more specifically, the power requirements of the DCH connections, is performed similarly as in Section 3.4, but with some modifications and simplifications. An important difference is that we are now interested in the mean transmit power over the whole cell. Additionally, we consider continuous transmit power allocation for the HS-DSCH (see Section 3.6.5), which means that the NodeBs always transmit with their target power T^*. This assumption enables us to calculate the mean DCH transmit power requirements independently of the situation in other NodeBs, since the other-cell interference is constant. It would be alternatively possible to use a similar approach as in [4] or [40] to calculate the average NodeB transmit power.

We define the average power load that one DCH user with service class s inflicts at its controlling NodeB as the DCH transmit power divided by the total transmit power, which is equivalent to the target transmit power:

$$\omega_s = \nu_s \cdot \frac{T_{k,x}}{T_{x,tot}}. \tag{3.32}$$

The class-specific activity factor ν_s indicates the ratio between the time the user is active and the total lifetime of the connection, given that at an observer perceives the probability to see an active connection as Bernoulli random variable. Note

that the average load is equal for all mobiles of the same service class. Also, due to the constant other-cell interference, we omit the specifier indicating the controlling NodeB. Neglecting thermal noise and with equations (3.5) and (3.6) we can formulate the single user DCH load as:

$$\omega_s = \nu_s \cdot \frac{\varepsilon_s R_s}{W} \cdot \left(\sum_{y \in \mathcal{L} \setminus x} E\left[\frac{d_{y,k}}{d_{x,k}}\right] + \alpha \right). \tag{3.33}$$

The total DCH load η_{DCH} and the mean HSDPA transmit power $T_{x,HS}$ are then given by

$$\eta_{DCH} = \sum\nolimits_{s \in \mathcal{S}} n_s \omega_s, \quad \text{and} \quad T_{HS} = T_{\max} - T_{CCH} - \eta_{DCH} \cdot T^*, \tag{3.34}$$

where T_{CCH} is the power required for common channels. Note since we only consider the mean transmit power, an error is induced in the resulting calculation of the HSDPA data rates. The reason is that since the DCH transmit powers depend on the other-cell interference, DCH users on the cell edge influence the HSDPA performance much more than DCH users in the cell center. A possible way to increase the accuracy is to partition the cell area into one or more tiers. However, this requires one additional state space dimension per tier and the accuracy is only moderately increased for two tiers, a partition which would keep the number of states in a decent extent.

3.7.2 State Space Description

We consider the number of occupied code units as state space description. We have learned in Section 3.3 that each DCH service connection requires a number c_s of SF 512 code equivalents. The code resources form a shared resource which can be used in the Kaufman-Roberts recursion [110] to form an one-dimensional state space. However, since we also want to include the HSDPA users in the state space, we construct a two-dimensional state space with the number of occupied code resources by DCH users as the first dimension, and the number of active

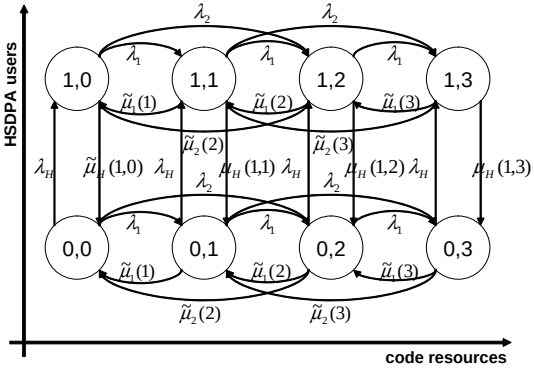


Figure 3.20: *Structure of the two-dimensional state space.*

HSDPA flows as the second dimension. A state j in the DCH dimension corresponds to the number of occupied code units $c_u = \min\{c_s\}$, i.e. in state j the DCH users occupy an equivalent of $j \cdot c_u$ codes with SF 512. The size of the state space is $C/c_u \times n_{H,max}$, where $n_{H,max}$ is the maximum number of HSDPA users. As in the simulations, we assume a simple count-based admission control for the HSDPA.

Figure 3.20 shows a part of the state space. The departure rates for the DCH dimension are calculated according to the following equation (see [110]):

$$\tilde{\mu}_s(j) = \mu_s \cdot E[n_s|j], \tag{3.35}$$

where $E[n_s|j]$ is the mean number of service class s connections in state j:

$$E[n_s|j] = \frac{\lambda_s}{\mu_s} \cdot \frac{\tilde{p}_{kr}(j - \frac{c_s}{c_u})}{\tilde{p}_{kr}(j)}. \tag{3.36}$$

The variable $\tilde{p}_{kr}(j)$ denotes the un-normalized probability for the DCH state j which is calculated recursively as

$$\tilde{p}_{kr}(j) = \frac{c_u}{C} \sum_{s \in \mathcal{S}} \frac{\lambda_s}{\mu_s} \cdot c_s \cdot \tilde{p}_{kr}(j - \frac{c_s}{c_u}). \tag{3.37}$$

Note that this model only allow to calculate the mean DCH power load per state. This prevents the computation of soft blocking probabilities, i.e. blocking due to transmit power limitation, for DCH users. However, in [37] it has been shown that the code capacity is the dominating factor for the system capacity except for nearly full DCH activity and concurrently bad orthogonality conditions, i.e. high orthogonality factors.

The HSDPA user throughput depends on the current DCH power load and on the number of active HSDPA flows. While the available code resources are directly available through the state, the current transmit power for the HSDPA requires knowledge of the mean DCH power load, which is calculated as in Equation (3.34) using the mean number of DCH users $E[n_s|j]$. Consequently, the number of usable HSDPA codes is $C_{hs}(j, n_H) = \lfloor (480 - j \cdot c_u)/32 \rfloor$ and the mean ratio of HSDPA power to total cell power is

$$\Delta_T(j, n_H) = 1 - \tfrac{T_c}{T_{\max}} - \sum\nolimits_s E[n_s|j] \cdot \omega_s. \tag{3.38}$$

The HSDPA bandwidth for a certain location f follows then under consideration of the mean HSDPA power ratio and the average number of available codes according to the model defined in Section 3.5:

$$\begin{aligned} R_f(j, n_H) &= R_f(C_{hs}(j, n_h), \Delta_T(j, n_H)) \\ &= \frac{1\,\mathrm{s}}{2\,\mathrm{ms}} \cdot E[V_{k,TTI}]. \end{aligned} \tag{3.39}$$

The average other-to-own interference ratio Σ_f follows directly from the constant other-cell interference assumption.

A straightforward method for computing the HSDPA departure rate would be to determine the mean bandwidth for an HSDPA user by averaging over the cell area. We will later refer to this approach as the “naïve” approach. However, we have to consider the following: With a volume-based user model, the lifetime of an HSDPA connections depends on its data volume V_H and its data rate. Even with Round-Robin scheduling, users at the cell border receive a smaller bandwidth than users in the cell center, and accordingly, they stay in the system for a

longer time. Consequently, as we have observed in Section 3.6.1, the probability p_f to meet a user at location f when looking into the system at a random instance of time is larger for a location close to the cell border than for one close to the cell center. More precisely, the probability p_f is proportional to the reciprocal bandwidth available at this position:

$$p_f \sim \frac{1}{R_f(j, n_H)}. \tag{3.40}$$

This effect is also mentioned by Litjens et al. [93] regarding Monte Carlo simulations. We approximate the average time $E[T|(j, n_H)]$ by summing over all positions in the cell and, after some algebraic operations, obtain the following formulation:

$$E[T|(j, n_H)] = E[V_H] \cdot E\left[\frac{1}{R_f(j,n_H)^2}\right] \cdot E\left[\frac{1}{R_f(j,n_H)}\right]^{-1} \tag{3.41}$$

For a more detailed derivation, see Appendix A. In the following, we will refer to this method as the "location-aware" approach. A corresponding formula for other, channel-aware scheduling disciplines is difficult to find because the user throughputs depend on each other through their TBS distributions.

In order to calculate the steady-state user distribution, we arrange the transition rate matrix Q with help of an index function $\phi(j, n_H) \rightarrow \mathbb{N}$ according to the following rules for all valid states:

$$\begin{aligned}
&Q(\phi(j, n_H), \phi(j + \tfrac{c_s}{g_c}, n_H)) = \lambda_s \\
&Q(\phi(j, n_H), \phi(j - \tfrac{c_s}{g_c}, n_H)) = \tilde{\mu}_s(j) \\
&Q(\phi(j, n_H), \phi(j, n_H + 1)) = \lambda_H \\
&Q(\phi(j, n_H), \phi(j, n_H - 1)) = \tfrac{1}{E[T|(j,n_H)]}
\end{aligned} \tag{3.42}$$

In all other cases $Q(i, j)$ is set to zero and $Q(i, i)$ is set to the negative row-sum of all entries to keep the state equations balanced. The steady-state distribution is then obtained by solving $Q \cdot \bar{\pi} = 0$ s.t. $\sum \pi = 1$ for the state vector $\bar{\pi}$. Performance measures like blocking probabilities or moments of the user throughput

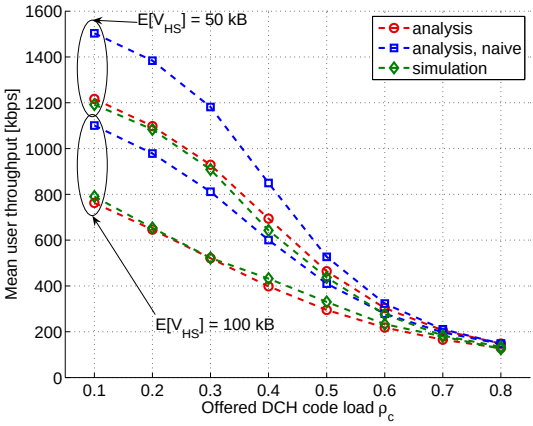


Figure 3.21: *Mean HSDPA user throughput vs. offered DCH load.*

are then calculated with help of the steady-state distribution and under assumption of PASTA [114]. For example, the mean HSDPA user throughput at a random time instance is

$$E[R_U] = \sum_{(j,n_H)|n_H>0} R_U(j,n_H) \cdot \frac{n_H \cdot \bar{\pi}(\phi(j,n_H))}{\sum_{(j',n'_H)|n'_H>0} n'_H \cdot \bar{\pi}(\phi(j',n'_H))}, \tag{3.43}$$

where the normalization term in the denominator ensures that only states with at least one active HSDPA user are considered.

3.7.3 Numerical Example

Let us now define an example scenario with the following parameters: we consider two DCH service classes with 128 kbps and 384 kbps. The service mix is 0.6 to 0.4. The activity factor is 0.55 for both service classes. HSDPA flows arrive with rate $\lambda_H = 1$. The orthogonality factor α for the DCH part of the air inter-

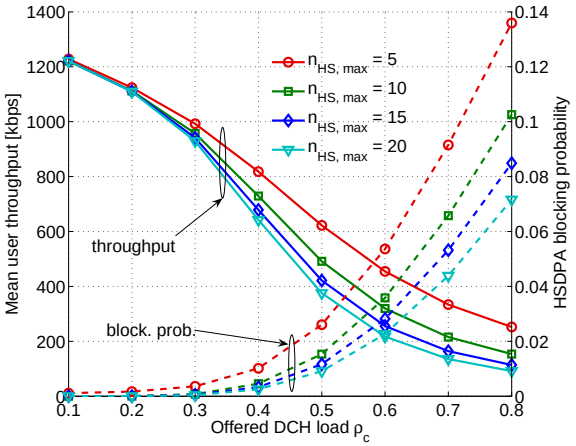


Figure 3.22: *Impact of $n_{H,\max}$ on throughput and blocking probabilities.*

face model is set to 0.35 corresponding to the HSDPA physical layer abstraction model. We validate our analytical results with the event-based simulation we also used for the results in Section 3.6.

Figure 3.21 shows the mean user throughput versus the offered DCH code load as defined in Equation (3.30). The figure shows two scenarios, one with a mean HSDPA data volume of 50 kbyte and one with 100 kbyte. The influence of the spatial user distribution can be clearly seen on the large difference between the naïve approach (squares) and the location-aware approach (circles). Especially for a low DCH load the difference is nearly 50%. With location-awareness, the analytical and the simulation results match well. The curves for 50 kbyte and 100 kbyte converge with a higher offered load for the DCH users since in this case the HSDPA is in an overload situation due to insufficient power and code resources.

In Figure 3.22 we show the impact of the HSDPA admission control on the

trade-off between user throughput and HSDPA blocking probabilities. The data volume is $E[V_H] = 50$ kbyte. The solid lines indicate the user throughput, while the dashed lines indicate the HSDPA blocking probabilities. The maximum number of allowed HSDPA users, $n_{H,max}$ is set to 5, 10, 15 and 20, respectively. We see that with higher DCH loads the increasing blocking probabilities for low values of $n_{H,max}$ leads to a significant improvement of the user throughput. The difference between the curves becomes smaller with higher values of $n_{H,max}$ which indicates that they will converge if $n_{H,max}$ would be increased further.

3.8 Concluding Remarks

In this chapter, we presented flow-level models for HSDPA-enabled UMTS networks. The models rely on a physical layer abstraction model for the HSDPA user throughput which gave us insights into the relationship between multi-path propagation profile, available code resources, interference and transmit power. With a model for the NodeB transmit power we implemented a flow-level simulation, which we used to investigate the HSDPA and DCH performance with different radio resource sharing and scheduling schemes. We also developed an analytical model which is suitable for the application in radio network planning tools.

The numerical results unveiled an interesting interrelationship between traffic model, scheduling disciplines and throughput. Volume-based user behavior leads to spatial inhomogeneity which has a significant impact on user and cell throughput. Channel-aware scheduling amplifies this effect, and leads to an unexpected result: Due to starvation of users at cell edges, MaxTBS scheduling performed worse than Proportional-fair scheduling, although the former achieves optimal throughput in static scenarios and in scenarios with time-based traffic. The volume-based model itself reflects a user with infinite patience, a behavior which could be realistic for some P2P file sharing applications. For other applications like web surfing, there may be a correlation between the willingness to stay in the system and the received throughput (the so called fun-factor effect, [115]), or a mixture between time and volume-based behavior as proposed in some publications (e.g. in [91]). This again may lead to better performance results for the MaxTBS scheduler. Another aspect is that the system operator may put unsatisfied HSDPA users on DCH connections, leads to the question whether this procedure is beneficial for the overall performance or not.

We conclude from the results presented in this chapter that HSDPA performance modeling has to take flow-level dynamics into account, i.e. the arrival and departure of users with different traffic demands and behaviors.

4 Performance of the Enhanced Uplink

As we elaborated in Chapter 2, the *Enhanced Uplink* or *High Speed Uplink Packet Access* (HSUPA) is the evolutionary pendant to the HSDPA in uplink direction. In the following chapter, we introduce models for evaluating the performance of the Enhanced Uplink. Analogously to the downlink we also considerate the still existing QoS Dedicated Channel users. The focus of this chapter are stochastic capacity models which capture the traffic-dynamics on the flow-level, but for a thorough understanding of the UMTS enhanced uplink it is also important to address the fundamental problem of system feasibility. In the next section we give a short overview of the different aspects we consider in the remainder of this chapter.

4.1 Overview and Related Work

The UMTS enhanced uplink provides power-controlled radio bearers with fast rate control. The capacity limiting radio resource in such a system is the multiple access interference at the NodeB antenna, which must not exceed a certain threshold in order to keep the system stable. This follows from the well known pole-capacity formula in [116] that gives a relation between the *noise rise*, defined as interference over thermal noise, the cell load η, which must not exceed 1, and the number of supported users. With rate controlled radio bearers as provided by the enhanced uplink, the amount of required resources depends on the

instantaneous data rate, which means that if a certain maximum load is reached, the data rate of the radio bearers can be reduced ("slow-down") in order to avoid dropping of connections. This may also happen if QoS users with DCH radio bearers are in the system that do not have this possibility. In Section 4.3 we describe a framework for the calculation of the interferences, considering DCH QoS user and E-DCH users, at all NodeBs in the system.

A fundamental question in such a system is how radio resource sharing is performed. The basic principle is similar to that in the downlink we discussed in the previous chapter: E-DCH connections use the remaining resources left by DCH users up to a certain threshold. We formulate two different radio resource management schemes based on this approach in Section 4.2. The *other-cell aware* scheme tries to reach a certain operator-defined total cell load, i.e. the instantaneous E-DCH capacity is affected by the amount of other-cell interference. In contrast, the *other-cell unaware* scheme tries to reach a certain own-cell load, ignoring other-cell interference.

In the UMTS uplink, the spatial configuration of the users may lead to situations where a NodeB is "overflooded" by other-cell interference, for example if a high data rate user is close to the cell border. In such a case it may be impossible to reach a system state where the resources at all NodeBs are fully utilized, i.e. the problem of radio resource assignment is infeasible. The feasibility region defines therefore the space spanned by assignments within the allowed resource constraints. A condition based on the global link gain matrix and the SIR requirements for feasibility has been first introduced in [117].

The rate assignment problem is connected to the practical implementation of the rate assignment procedure: in a system with global knowledge of all relevant parameters, optimal assignment can be achieved. However, this would require a network-wide centralized rate assignment which is difficult to achieve with the UTRAN architecture. Another possibility is to use distributed algorithms using feedback control loops which converge toward the (utility-) optimal solution, as it has been proposed in [118]. An alternative approach is described in [119]: a semi-distributed algorithm controls the other-cell interference between the NodeBs,

while the rate assignment is done by the NodeBs itself according to these constraints. A solution that requires no interaction between NodeBs is also proposed earlier in [120], but for the price of a reduced feasibility region due to linearized constraints. The underlying optimization problem has been mentioned in [121] and has been further investigated in [122], where it has been shown that the problem can be solved with convex optimization also for effective bandwidth utilities, which correspond to the concept of service load factors defined in Section 4.3, where an uplink interference model is described. Our contribution in Section 4.4 is a characterization of the feasibility region specifically for the enhanced uplink, i.e. with consideration of transmit power constraints, other-cell DOWN grants, and the presence of QoS DCH users.

The impact of traffic dynamics on average performance measures like blocking probabilities or throughput requires flow-level capacity models. In Section 4.5, we develop a single-cell capacity model with a queuing model approach. Similar to the models in Chapter 3 for the HSDPA, it is assumed that E-DCH best effort users follow a volume-based traffic model. In principle, the model follows the classical queuing approaches by Viterbi & Viterbi [123] or Evans & Everitt [124]. However, the system model includes similarly to HSDPA a physical layer abstraction model which takes the features and impairments of the E-DPDCH into account. Data rates are calculated according to the instantaneous transmission capacity of the E-DPDCH, and the impact of power control errors is considered with the assumption of lognormal distributed E_b/N_0-values. Additionally, the system state space is extended by a dimension for the best-effort users. Related work can be found in [125] for rate controlled radio bearers only and with perfect power control. An extension for a system with QoS radio bearers is presented in [126]. A slightly different model is proposed in [127] with minimum and maximum allowed data rates. These works concentrate on performance measures like average throughput and blocking probabilities, but with admission control for QoS calls only.

In contrast to the HSDPA, power control enables minimum guaranteed data rates for E-DCH connections. This means that admission control has to guarantee

that the minimum data rate is maintained. However, it is commonly assumed that QoS traffic is more valuable for service providers then best-effort traffic which implicates a certain hierarchy in case of exhausted radio resources. Specifically, best-effort connections (i.e. E-DCH connections) may be dropped if radio resources are insufficient for an incoming DCH connection. The impact of an preemptive admission control in the context of uplink CDMA systems was subject in [128] and [5]. In Section 4.5.5 the performance of preserving and preemptive admission control is compared.

So far, we presumed that the UEs are scheduled in parallel since the UMTS uplink is not synchronized. However, it has been shown in [76] that one-by-one scheduling leads to increased capacity in terms of throughput. This reasonable result can be also conjectured by the simple fact that with one-by-one scheduling, users within the same cell do not generate any interference to each other. This insight is used with the assumption of synchronized uplink transmissions in [78] to design a one-by-one scheduler which takes the current channel qualities into account. But also channel-blind strategies like Round-Robin lead to significant performance gains, as we show in Section 4.5.6 and in [7].

Especially in network planning, an accurate approximation of the other-cell interference is crucial, so in Section 4.6, a stochastic capacity model for a multi-cell environment is proposed. The model is based on the multi-cell model in [10, 40] and [129] with the queuing approach for the single-cell case in Section 4.5.

4.2 Radio Resource Sharing for the E-DCH Best Effort Service

The scheduling of the E-DCH users is done in the NodeBs which control the maximum allowed transmit power of the mobiles and therefore also the maximum user data rate. Throughout this chapter we assume saturated traffic which means that the maximum user data rate corresponds to the chosen data rate. The

NodeBs send scheduling grants on the absolute or relative grant channel (AGCH and RGCH) which either set the transmit power to an absolute value or relative to the current value. The mobiles choose then the transport block size (TBS) which is most suitable to the current traffic situation and which does not exceed the maximum transmit power. The grants can be sent every TTI, i.e. every 2 ms, which enables a very fast reaction to changes of the traffic or radio conditions. Grants can be received from the serving NodeB and from non-serving NodeBs. However the latter may just send relative DOWN grants to reduce the other-cell interference in their cells.

Generally, the WCDMA uplink is interference limited [116]. Therefore, following [30] we define the uplink load in a cell as

$$\eta = \frac{I_D + I_E + I_{oc}}{I_0 + WN_0}, \tag{4.1}$$

with I_D and I_E as received powers from the DCH and E-DCH users within the cell, I_{oc} as other-cell interference from mobiles in adjacent cells, W as system chip rate, N_0 as thermal noise power spectral density and $I_0 = I_D + I_E + I_{oc}$. It can be readily seen that this load definition allows the decomposition of the cell load according to its origin, hence we define

$$\begin{aligned} \eta &= \tfrac{I_D}{I_0+WN_0} + \tfrac{I_E}{I_0+WN_0} + \tfrac{I_{oc}}{I_0+WN_0} \\ &= \eta_D + \eta_E + \eta_{oc} \end{aligned} \tag{4.2}$$

subject to $\eta < 1$. The goal of the RRM is now twofold: First, the cell load should be below a certain maximum load in order to prevent outage. Second, the RRM tries to maximize the resource utilization in the cell to provide high service qualities to the users. The second goal allows also the interpretation of the maximum load as a target load, which should be met as close as possible. Since the DCH-load and the other-cell load cannot be influenced, the E-DCH load can be used to reach the target cell load. The fast scheduling gives operators the means to use the E-DCH best-effort users for "waterfilling" the cell[1] load at the NodeBs up to a desired target.

[1] corresponding to a sector in case of multiple sectors per NodeB

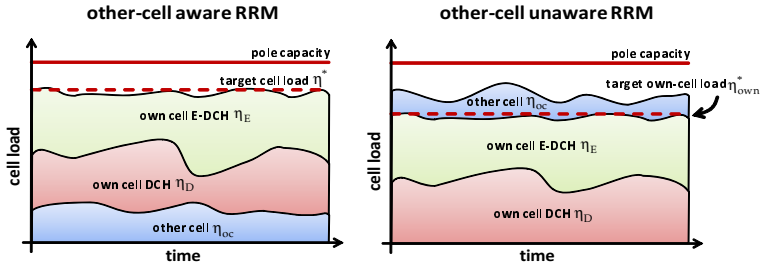


Figure 4.1: *Illustration of the RRM principles for the E-DCH best-effort service. The scheme on the left side denoted "other-cell aware RRM" tries to maintain the total cell load below a certain target load η^*. In contrast, "other-cell unaware RRM" only keeps the own-cell load below a (lower) threshold η^*_{own}, leaving the other-cell load to the Grant processing.*

This radio resource management strategy is illustrated in two variants in Figure 4.1. Generally, the total cell load comprises the varying other-cell load, the load generated by DCH users and the E-DCH load. The first variant, denoted "other-cell-aware", adapts the available load for the E-DCH users to a *total* target cell load η^*, including the load generated by other-cell interference, such that under perfect conditions

$$\eta^* = \eta_D + \eta_E + \eta_{oc}. \tag{4.3}$$

The second variant, "other-cell unaware", does not consider the other-cell load but only adapts the E-DCH load to an own-cell target load η^*_{own} and the own-cell load η_D of the dedicated channel users:

$$\eta^*_{own} = \eta_D + \eta_E. \tag{4.4}$$

In both cases the received power for the E-DCH users is adapted such that the considered fraction of the cell load is close or equal to the target load, but in

the case of “other-cell unaware” RRM, the other-cell interference is handled by other-cell DOWN grants in order to avoid outage.

In the rest of this chapter, we address both types of radio resource sharing. The problem of system feasibility investigated in Section 4.4 does only occur with the other-cell aware scheme. The single-cell capacity models in Section 4.5 also consider the other-cell aware scheme, however, the model can be easily adapted to the other-cell unaware scheme without much effort. Finally, the multi-cell capacity model in Section 4.6 considers the other-cell unware scheme.

4.3 Basic Interference Model

We consider an arbitrary UMTS network with a set of NodeBs $\mathcal{L}$ and a set of user equipments $\mathcal{M}$. The UEs are further divided into the mutual exclusive sets $\mathcal{E}_x$ and $\mathcal{D}_x$, the first containing the E-DCH users and the second the DCH users controlled by a NodeB x. We write $k \in x$ with $x \in \mathcal{L}$ to denote a UE controlled by NodeB x, regardless of its bearer.

The received power $S_{k,x}$ of a mobile k at its controlling NodeB x depends on the target-E_b/N_0 requirement and the data rate of the connection. If we assume perfect fast power control the following must hold:

$$\varepsilon_k^* = \frac{W}{R_k} \frac{S_{k,x}}{W \cdot N_0 + \sum_{j \in \mathcal{M} \setminus k} S_{j,y}}, \tag{4.5}$$

where ε_k^* is the target-E_b/N_0-value, W is the system chiprate ($3.84\,\text{Mcps}$ in UMTS FDD), R_k is the instantaneous data rate, and N_0 is the one-sided thermal power density. Solving Equation (4.5) for the received power $S_{k,x}$ yields [130]

$$S_{k,x} = \omega_k \cdot \left(W \cdot N_0 + \sum_{j \in \mathcal{M}} S_{j,y} \right). \tag{4.6}$$

The term ω_k is an effective bandwidth measure of the load this mobile generates at its controlling NodeB. We will denote it as *service load factor* (SLF) in the rest of this monograph. It is defined as

$$\omega_k = \frac{\varepsilon_k^* \cdot R_k}{\varepsilon_k^* \cdot R_k + W}, \tag{4.7}$$

and depends only on the target-E_b/N_0-value and the data rate R_k. For E-DCH radio bearers, the total received power of the E-DPDCH is relative to the power controlled control channel. This means that the combined service load factor for E-DCH radio bearers consists of the control channel SLF and the data channel SLF. The relation between both SLFs is defined by the power offset Δ_T:

$$\omega_{k,E} = \omega_{k,C} \cdot (1 + \Delta_{k,T}). \tag{4.8}$$

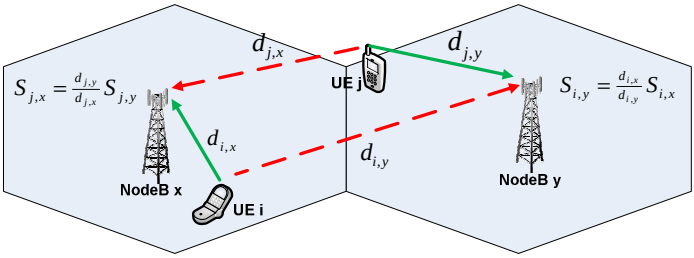


Figure 4.2: *Simple example scenario with two mobiles and two NodeBs.*

The magnitude of $\omega_{k,C}$ depends on the target-E_b/N_0 for the DPCCH control channel, however, due to the small information data rate (5 kbps with a spreading factor of SF $= 256$) and for sake of simplicity, we neglect the control channel SLF if not stated otherwise in the rest of this work.

The cell interference in a unsynchronized CDMA system depends generally not only on the number of transmitting mobiles in the own cell, but also on the interference generated in surrounding cells and sectors, respectively. This interference is called other-cell (or inter-cell) interference, since it originates from cells other than the cell we are currently looking on. With the same argument, the interference which is generated in other cells also depends on the interference of the own-cell. Figure 4.2 clarifies that with an example scenario with two mobiles. In this scenario, mobile i is close to its controlling NodeB x and requires therefore only a small amount of transmit power to reach its target-E_b/N_0 value. The second mobile j, controlled by NodeB y, is close to the cell edge such that both NodeB x and NodeB y nearly receive the same power from this mobile. The interference power $S_{k,y}$ from one mobile to an none-controlling NodeB is given by the ratio between the link gains between the mobile and the two NodeBs:

$$S_{k,y} = \frac{d_{k,y}}{d_{k,x}} S_{k,x} = \Delta_{k,y}^{x} S_{k,x}, \tag{4.9}$$

where x is the controlling NodeB, y is a non-controlling NodeB, $d_{k,x}$ is the link

gain between mobile k and NodeB x, and $S_{k,x}$ is the received power at NodeB x. Note that the link gain ratio $\Delta^x_{k,y} = 1$ if $x = y$.

We define the total interference at a NodeB x as the sum of all received signal powers from all mobiles in the network. With Equation (4.9) the interferences over all NodeBs form a linear equation system with

$$I_x = \sum_{l \in \mathcal{L}} \sum_{k \in l} \Delta^x_{k,l} \cdot \omega_k \cdot (W \cdot N_0 + I_l). \tag{4.10}$$

With matrices, we formulate this equation as

$$\bar{I} = \widetilde{G} \cdot \left(\bar{N}_0 + \bar{I}\right), \tag{4.11}$$

where $\bar{I}$ is the $|\mathcal{L}| \times 1$-vector of interferences, $\bar{N}_0$ is a $|\mathcal{L}| \times 1$-vector with $(\bar{N}_0)_j = W \cdot N_0$, and $\widetilde{G}$ is an $|\mathcal{L}| \times |\mathcal{L}|$ matrix with the sum of the link-gain ratios multiplied with the corresponding SLF as elements, such that

$$(\widetilde{G})_{ij} = \sum_{k \in j} \Delta^i_{k,j} \cdot \omega_k. \tag{4.12}$$

Note that in our notation j is the set of mobiles which are connected to NodeB j. The interference at each NodeB is the result of solving Eq. (4.11) for $\bar{I}$:

$$\bar{I} = \left(\widetilde{E} - \widetilde{G}\right)^{-1} \cdot \left(\bar{N}_0 \cdot \widetilde{G}\right), \tag{4.13}$$

where $\widetilde{E}$ is the identity matrix.

Up to now, our model does not make any distinction between DCH and E-DCH users. Both user types are characterized through their service load factor ω. However, in Section 4.2 we defined an RRM strategy for the E-DCH users which tries to maximize the resource utilization in each cell. Since the resource in our case is the interference and corresponding to that, the cell load, the remaining resources are distributed to the E-DCH users according to the other-cell aware scheme as in Equation (4.3). Essentially this means that DCH users have fixed SLFs, while E-DCH users get the remaining load in a typical best-effort manner.

We can express this by splitting the interference equation further up after the signal source:

$$I_x = I_{x,D}^{own} + I_{x,D}^{oc} + I_{x,E}^{own} + I_{x,E}^{oc}, \tag{4.14}$$

which corresponds to the matrix form

$$\begin{aligned} \bar{I} = \tilde{G}_D^{own}(\bar{N}_0 + \bar{I}) + \tilde{G}_D^{oc}(\bar{N}_0 + \bar{I}) \\ + \tilde{G}_E^{own}(\bar{N}_0 + \bar{I}) + \tilde{G}_E^{oc}(\bar{N}_0 + \bar{I}). \end{aligned} \tag{4.15}$$

Here, the elements of the load matrices correspond to the set of users which generate interference. The matrices $\tilde{G}_D^{own}$ and G_E^{own} are diagonal matrices with elements $(\widetilde{G}_D^{own})_{ii} = \sum_{k \in \mathcal{D}_x} \omega_k$ and $(\widetilde{G}_E^{own})_{ii} = \sum_{k \in \mathcal{E}_x} \omega_k$. The matrizes for the other-cell interference contain zeros at the diagonal, and on the remaing entries the sum of SLFs multiplied with their link gain ratios, i.e. $(\widetilde{G}_D^{oc})_{ii} = 0$ and $(\widetilde{G}_D^{oc})_{ij} = \sum_{k \in \mathcal{D}_j} \Delta_{k,j}^x \cdot \omega_k$ for all $i \neq j$.

4.4 System Feasibility

In the previous section, we defined a general framework for the calculation of the interferences and loads at a NodeB. We now use this framework for obtaining the actual resources that can be assigned to an E-DCH user. Let us for this reason first define some common constraints for resource assignment:

1. The maximum load or interference should not be exceeded. The purpose is to guarantee a stable system, since if the cell loads get too high, the required transmit powers for the mobiles tend to infinity, which makes it impossible for them to reach their required target-E_b/N_0. Hence we define the constraint

$$\mathrm{C}_{\text{load}}: \quad 0 \leq \eta_x \leq \eta_x^*. \tag{4.16}$$

2. All E-DCH users have a certain minimum bandwidth guarantee which corresponds to a minimum TBS and thus to a minimum SLF ω_{min}. This

condition avoids quasi-outage of users. Further, the maximum SLF ω_{max} is defined by the highest TBS, which corresponds to 5.74 Mbps. So it is mandatory that

$$\mathrm{C_{SLF}}: \quad \omega_{min} \leq \omega \leq \omega_{max}. \tag{4.17}$$

3. The mobiles have a maximum transmit power T_{max} which is normally either 125 mW (21 dBm) or 250 mW (24 dBm), so

$$\mathrm{C_{pow}}: \quad T_m \leq T_{max}. \tag{4.18}$$

Note that transmit powers can be easily calculated from the interference at the serving NodeB x and the pathloss as

$$T_m = {d_{m,x}}^{-1} \cdot \omega_m \cdot (WN_0 + I_x). \tag{4.19}$$

4. In [131] it is stated that DOWN grants are sent to mobiles in adjacent cells if the ratio between the E-DCH other-cell interference and the total interference from E-DCH users exceeds a certain operator-defined threshold. This reduces flooding of cells from adjacent sites due to high-bitrate mobiles near the cell borders. Let $\mathcal{H}_x$ be the set of UEs which are in the soft handover area but not controlled by NodeB x. The condition can then be expressed as

$$\mathrm{C_{grant}}: \quad \frac{\sum_{h \in \mathcal{H}_x} S_{h,x}}{I_E} \leq t_{\mathrm{SHO}}, \tag{4.20}$$

where t_{SHO} is an operator-defined threshold.

The goal of the resource assignment procedure is that all this conditions are fulfilled. Under certain circumstances, this may not always be possible, which may lead to a load overshoot event. A load overshoot does not necessarily mean that a UE experiences outage, however it may affect the connection or system stability negatively, so it should be avoided if possible.

In our model, load overshoots corresponds to a resource assignment which is not in the feasibility region, which is defined by the constraints above. Depending on the RRM strategy and the degree of knowledge that the executing entity

has on the global load situation, the feasibility regions significantly differ from each other. We distinguish between three kinds of RRM implementations: One with global knowledge of the system load which constitutes the optimal case, one with global knowledge but with a distributed implementation such that it has a reduced feasibility region, and a totally decentralized one with only local knowledge of the load, which corresponds to the single cell resource assignment scheme. Generally, load overshoots can occur because of two reasons: First, the load generated by the DCH users is so high that the target load is exceeded. Normally, the admission control prevents such events. The second case is due to an RRM implementation that allows cells to be flooded with interference from adjacent cells. This may occur with the local RRM implementation. Apart from load overshoots also the contrary it may happen, i.e. that the target load is not reached. This occurs if the RRM implementation decides to lower the load in some cells to prevent load overshoots, i.e. for the global RRM implementation.

4.4.1 Global Resource Assignments

From Equation (4.10) we see that the SLF and interference calculation can be interpreted as an optimization problem. In our model we try to optimize the cell load with a utility function $U(\cdot)$. In the literature, several options are mentioned to optimize for different fairness goals. The most straightforward utility function is to sum over all individual loads of the E-DCH users. However, this approach leads to unfair assignments in the sense that UEs close to the NodeB get as much load as possible, while the more distant UEs may only get the minimum SLF. An often mentioned generic fairness criterion is that of α-fairness, where the optimization converges to different fairness goals according to the setting of a parameter α, [132]:

$$U(\omega_m) = \frac{\omega_m^{1-\alpha}}{1-\alpha} \tag{4.21}$$

With this utility function, proportional fairness [58] can be achieved with $\alpha \to 1$ and max-min-fairness can be achieved in the limit $\alpha \to \infty$. The optimization

problem is then formulated as:

$$\text{OPT}_{\text{nlin}}: \quad \max. \quad \sum_{m \in \mathcal{M}} U(\omega_m) \tag{4.22}$$

$$\text{s.t.} \quad \text{C}_{\text{load}}: \quad 0 \leq \eta_x \leq \eta_x^* \tag{4.23}$$

$$\text{C}_{\text{SLF}}: \quad \omega_{min} \leq \omega_m \leq \omega_{max} \tag{4.24}$$

We consider the load and SLF as the basic set of constraints. Later throughout the paper we additionally take the power and the DOWN-grant constraints into account.

The above formulation leads to an optimal load factor assignment if the RRM entity has knowledge of the load situation in all cells. In practice, however, this is very difficult to implement since it would need a very high amount of signaling to a central point which should be avoided. In [120] and [119] the authors therefore propose an RRM implementation which can be implemented in a distributed way. One way to achieve that is to base the constraints only on locally available information. In this case this constitutes a row in the link-gain ratio matrix $\widetilde{G}$. The optimization problem is therefore in our model complemented with a linear constraint on the row sums:

$$\text{OPT}_{\text{lin}}: \quad \max. \quad \sum_{m \in \mathcal{M}} U(\omega_m) \tag{4.25}$$

$$\text{s.t.} \quad \text{C}_{\text{lin}}: \quad \sum_{x} \sum_{k \in x} \Delta_{k,j}^{i} \cdot \omega_k \leq \eta_x^* \tag{4.26}$$

$$\text{C}_{\text{SLF}}: \quad \omega_{min} \leq \omega_m \leq \omega_{max} \tag{4.27}$$

Note that with condition C_{lin} also condition C_{load} is fulfilled (see e.g. [120]).

4.4.2 Local Resource Assignments

The target load relates to an equivalent target interference by $I^* = \frac{\eta^*}{1-\eta^*}(WN_0)$. Let us now assume that the target interference is reached in all cells, i.e. $I_x = I_x^*$ for all NodeBs. The total interference term in Equation (4.10) is then independent

of the actual spatial user configuration. If we divide by the constant term $(WN_0 + I_x)$, the left hand side is per definition the target load, and the right hand side is the sum of all SLFs times their link gain ratios, if we assume that the target load is equal for all NodeBs:

$$\eta_x^* = \sum_{l \in \mathcal{L}} \sum_{k \in l} \Delta_{k,l}^x \omega_k. \tag{4.28}$$

Under this assumption, we can calculate with the cell load directly. If we split up the total load according to its sources, Equation (4.28) becomes

$$\eta_x^* = \eta_{x,D} + \eta_{x,E}^{own} + \sum_{y \in \mathcal{L} \setminus x} \sum_{j \in \mathcal{E}_y} \Delta_{j,y}^x \omega_j. \tag{4.29}$$

The most straightforward way to calculate the SLFs for the E-DCH users is to solve the load equation system for the E-DCH own cell load η_E^{own}. This means, we assume that the load at each NodeB is constant and corresponds to the target load and solve for the own-cell load for the E-DCH users. This requires that if we have more than one user per cell[2], we have to fix the partitioning of the E-DCH load to the individual SLFs with a *policy* factor g such that

$$\sum_{j \in \mathcal{E}_x} g_j \cdot \eta_{x,E}^{own} = 1. \tag{4.30}$$

The policy factor can rely just on the number of E-DCH mobiles such that $g_j = \frac{1}{|\mathcal{E}_x|}$ or can include distances or path gains to prioritize mobiles which are close to the NodeB. Following Equation (4.3), we calculate the own-cell E-DCH load directly in matrix formulation:

$$\bar{\eta}_E^{own} = \bar{\eta}^* - \bar{\eta}_D - \tilde{F}_E^{'oc} \cdot \bar{\eta}_E^{own}, \tag{4.31}$$

where $\tilde{F}_E^{'oc}$ contains the link gain ratios as well as the policy factor g_j:

$$(\widetilde{F}_E^{'oc})_{ij} = \begin{cases} \sum_{k \in \mathcal{E}_j} \Delta_{k,j}^i \cdot g_k, & \text{if} \quad i \neq j \\ 0 & \text{else} \end{cases} \tag{4.32}$$

[2]Note that we assume at least one E-DCH user in each cell

Solving for $\bar{\eta}_E^{own}$ yields the own-cell E-DCH load at each NodeB and with the policy factor also the resource assignment for each individual E-DCH user:

$$\bar{\eta}_E^{own} = (\tilde{E} + \tilde{F}_E^{'oc})^{-1} \cdot (\bar{\eta}^* - \bar{\eta}_D). \tag{4.33}$$

This approach, which we will call "local" or "direct" in the reminder, leads to negative results for $\bar{\eta}_E^{own}$ in two cases. Either is $\bar{\eta}_D > \bar{\eta}^*$ for one element which means that the DCH load is higher than the target load, or the spatial configuration is such that the other-cell E-DCH load is higher than $\eta^* - \eta_D$. In this case, we assume that the SLFs for the E-DCH users in the specific cell is set to its minimum, leading to a load overshoot.

4.4.3 Feasible Load Region and Boundaries

Table 4.1: *Example scenario*

	E-DCH 1	E-DCH 2	DCH 1	DCH 2
S-NodeB	A	B	A	B
ω			0.1	0.05
Δ	0.9	$6 \cdot 10^{-4}$	$3 \cdot 10^{-4}$	$1 \cdot 10^{-4}$

Let us now consider a simple example with two cells two E-DCH users (one per cell) and two DCH users. The values for Δ correspond to the path gain ratio between the non-serving and the serving NodeB, see Table 4.1. The first E-DCH user is close to the cell edge, which leads to a high Δ of 0.9. The second E-DCH user in the second cell is close to its serving NodeB. The DCH user in the first cell has moderate distance to NodeB A. As fairness criterion for the global RRM schemes we chose max-min-fairness since it is closest to the behavior of the local RRM scheme with equal load assignments for all E-DCH users in a cell.

The resulting feasible SLF regions for the two E-DCH users are shown in Figure 4.3. For the global RRM strategies we considered the power, the linear

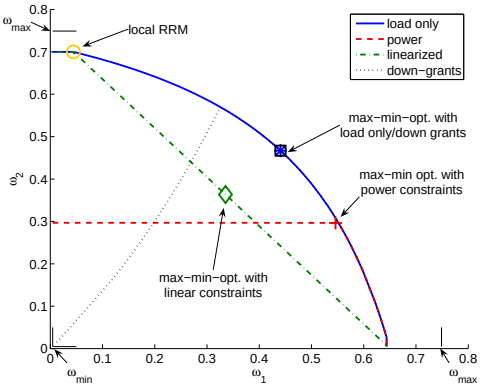


Figure 4.3: *Feasible SLF region for the two-cell scenario for different constraints. The markers indicate the max-min optimal solution for the system with exception of the local RRM case.*

and the DOWN-grants constraints individually, i.e. we consider only one constraint additionally to the load and SLF constraint. The max-min-optimal points for the global RRM differ significantly from the direct approach, and yields a very unbalanced result between the two E-DCH users but still is within the feasible region. The power constraints in this scenario leads to an SLF configuration which favors the first E-DCH user, while for the load-only and the DOWN-grant constraint as well as for the linear constraint the SLF values are balanced. The direct approach for the local RRM corresponds to the linear constrained RRM with sum-optimal utility function. The feasible region does not reach the maximum possible SLF ω_{max} due to the load from the DCH users. The optimal solution for the DOWN-grant constraint corresponds in this case to the solution with load constraints only, however, this changes if the maximum allowed ratio between own-cell to total E-EDCH load was set to a lower value.

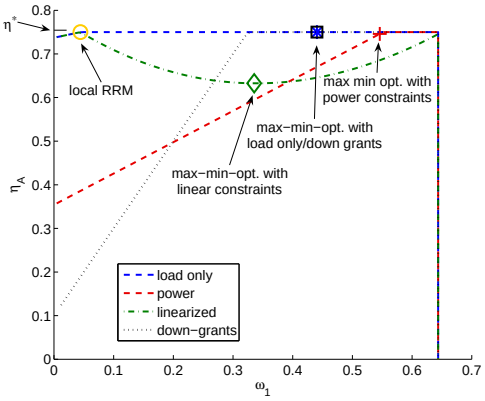


Figure 4.4: *Cell loads at NodeB A. With excepetion of the linear row-sum constraint, all assignments reach the target load.*

The corresponding load η_A at the first NodeB is shown in Figure 4.4. The loads for the non-linear and linear case begin to diverge on the solution point for the direct approach. The effect of the row-sum constraint on the load is that the target load is not reached for a large range of the feasible SLF region. Further, the max-min-optimal point in this case is significantly lower than for the non-linear case. The direct approach naturally reaches the target load at both NodeBs, but at the expense of a very low SLF for the first E-DCH UE. The non-linear approaches both reach the target-load at NodeB A. It should be mentioned that this scenario is quite extreme, which is the reason for the different results of the approaches.

As we see in Figure 4.5, the probability that the system is not feasible with local RRM depends on the number of E-DCH users in the system. The result has been generated from 1000 simulation runs where in total 10 users (i.e. DCH and E-DCH users) are randomly placed in every cell. The fraction between E-DCH and DCH users is increased from 2 to 8. Equal-rate scheduling is assumed for the

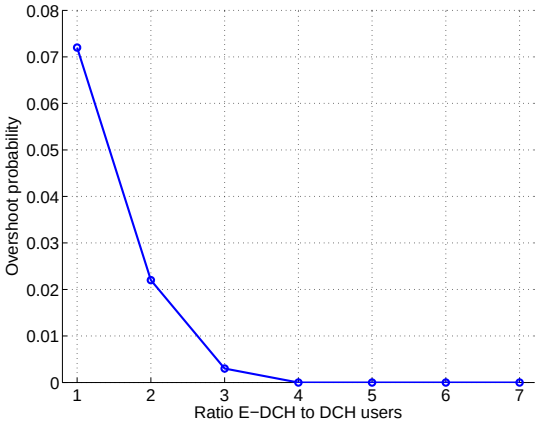


Figure 4.5: *Probability of a load overshoot depending on the fraction of E-DCH users in each cell.*

E-DCH users. In this scenario, the probability of a load overshoot (which corresponds to an infeasible system) is largest with only one E-DCH user in each cell and reaches 7 %. With an increasing number of E-DCH users, the system gets more stable, i.e. the probability that overshoots occur gets rapidly lower and is nearly zero with 4 E-DCH users. The reason is that with a higher number of E-DCH users the assigned service load factors and correspondingly the transmit powers are lower such that the probability that a NodeB is overflooded with interference decreases. Overflooding may still happen, but it requires that several users are nearly at the same location in a cell. That indicates that for example cluster arrivals (like at hot spots) may again lead to higher overshoot probabilities.

4.5 Single-Cell Capacity Model

In this section, we develop a capacity model which considers imperfect power control and lognormal-distributed other-cell interference. Imperfect power control means that the received E_b/N_0-values at the NodeB fluctuate around the target-E_b/N_0-value due to the granularity of the power-control steps in the dimension of time and power (three ± 1 dB commands per slot) or signaling errors. Measurements and link-level simulations have shown that the received E_b/N_0-values are approximately lognormal distributed with the target-E_b/N_0 value as mean (see for example [123, 133]).

Although the assumption of independence between other-cell and own-cell interference is not realistic for scenarios with an unequal load distribution over the network, it enables us to propose an analytical model for the "other-cell aware" RRM scheme without the need to consider system feasibility explicitely. That is, we assume firstly that the system prevents outage due to infeasibility, and secondly that the effect on the overall-performance is negligible. This is true for scenarios where the number of users is not too low as Figure 4.5 shows.

If we consider power control errors and independent random other-cell interference there is always the possibility of a load "overshoot". The probability for such an event should be kept low. We define that the goal of the RRM is to keep the probability of a load overshoot below a maximum tolerable probability p_t:

$$P\{\eta \geq \eta^*\} \leq p_t. \tag{4.34}$$

This means that the received signal power (i.e. the E-DCH interference) of the E-DCH users depends on the amount of dedicated channel and other-cell interference. More precisely, the E-DCH users are slowed down if the DCH or the other-cell load is growing, or are speed up, if more radio resources are available for the E-DCH users. If we now assume that the buffers in the mobiles of the E-DCH users are always saturated, we can use this relation to calculate the grade-of-service the E-DCH users receive depending on the scheduling strategy.

4.5.1 Single Cell Load Model with Imperfect Power Control

We assume imperfect power control, so the received E_b/N_0 is a lognormally distributed random variable with the target-E_b/N_0-value ε_k^* as mean value and parameters $\mu = \varepsilon_k^* \cdot \frac{\ln(10)}{10}$ and $\sigma = \text{Std}[\varepsilon_k] \cdot \frac{\ln(10)}{10}$. The received power of each mobile is calculated according to Equation (4.6) as

$$S_k = \omega_k \cdot (WN_0 + I_0) \quad \text{with} \quad \omega_k = \frac{\varepsilon_k R_k}{W + \varepsilon_k R_k}. \tag{4.35}$$

We further assume that the DCH users are distinguished by their service class $s \in \mathcal{S}$ which corresponds to a certain data rate, target-E_b/N_0 value and service load factor. The state vector $\bar{n}$ comprises the users per DCH service class, n_s, and the E-DCH users n_E:

$$\bar{n} = (n_1, \ldots, n_{|S|}, n_E). \tag{4.36}$$

The sum of all concurrently received powers constitutes the received own-cell interference, i.e.

$$I_D(\bar{n}) = \sum_{s \in \mathcal{S}} \sum_{k \in n_s} S_k \quad \text{and} \quad I_E(\bar{n}) = \sum_{j \in n_E^a} S_j. \tag{4.37}$$

I_D is the total received power of the DCH users and I_E of the E-DCH users. Note that the set of currently active E-DCH users n_E^a depends on the scheduling discipline. For parallel scheduling, $n_E^a = n_E$, since we assume that all users are concurrently active. For one-by-one scheduling, $|n_E^a| = 1$ since in this case only one E-DCH user is transmitting at the same time.

The substitution of I_D and I_E in Equation (4.2) with Equation (4.37) gives us the load definitions depending on $\bar{n}$:

$$\eta_D(\bar{n}) = \sum_{s \in \mathcal{S}} \sum_{k \in n_s} \omega_k \quad \text{and} \quad \eta_E(\bar{n}) = \sum_{j \in n_E^a} \omega_j, \tag{4.38}$$

and the total load as

$$\eta(\bar{n}) = \eta_D(\bar{n}) + \eta_E(\bar{n}) + \eta_{oc}. \tag{4.39}$$

We assume that the service load factors are lognormal r.v.'s which follows from the power control error of the received E_b/N_0 around the target-E_b/N_0 (see e.g. [123]). The parameters μ, σ are then derived from the mean and variance of the E_b/N_0 distributions. These parameters depend on the service class of the users, but are equal for all users within one class. So we can write $E[\hat{\omega}_k] = E[\hat{\omega}_s]$ for all mobiles k with the same service class s. The other-cell load η_{oc} is modeled as a lognormal random variable with constant mean and variance.

Since the total load η is a sum of independent lognormal distributed random variables, we assume that η also follows a lognormal distribution [134]. We get the distribution parameters from the first moment and variance of the cell load which can be calculated directly from the moments of the SLFs:

$$E[\eta(\bar{n})] = \sum\nolimits_{s \in \mathcal{S}} n_s \cdot E[\omega_s] + n_E^a \cdot E[\omega_E] + E[\eta_{oc}]. \tag{4.40}$$

The variance is calculated analogously. The accuracy of this approach is validated e.g. in [40]. The effect of Hybrid ARQ can be modeled as a constant gain which is included in the target-E_b/N_0 of the E-DCH and with an additional overhead on the mean data volumes of the E-DCH.

4.5.2 Rate Assignement

With other-cell aware radio resource sharing, the available E-DCH load depends on the DCH and other-cell load. The task of the RRM is to assign each E-DCH mobile a service load factor ω such that the E-DCH load is completely utilized if possible. Due to the very flexible scheduling mechanism of the E-DCH, this can be reached in several ways. We consider two fundamentally different scheduling disciplines: The first one is parallel equal-rate scheduling, which means that every E-DCH user gets the same service load factor in every TTI. The second is equal-rate one-by-one-scheduling, where each E-DCH user gets the maximum possible

service load factor in a round-robin-fashion. Note that we assume for the latter discipline that the E-DPDCH physical data channels are perfectly synchronized, hence do not generate any interference to each other. Although this is a strong assumption for current specifications of the Enhanced Uplink, the results can be seen as an upper bound for a more realistic system. We further assume that the network is dimensioned such that the transmit powers of the mobiles are sufficient to reach the maximum bitrate.

Generally, the user bit rate depends on the magnitude of the E-DCH cell load which may be generated without violating the RRM target in Eq. (4.34). The channel bit rate of the E-DCH is defined by the amount of information bits which can be transported within one TTI. This quantity is defined in [73] by the set of transport block sizes $\mathcal{TBS}$. With a TTI of 2 ms, the information bit rate per second is $R^I_{i,E} = TBS_i \cdot \frac{1\,\mathrm{s}}{2\,\mathrm{ms}}$, where $i = 1, \ldots, |\mathcal{TBS}|$ indicates the index of the TBS. We further define $R^I_{0,E} = 0$. With this interpretation we can map the E-DCH bit rate to a service load factor according to Equation (4.35) as

$$\omega_{i,E} = \frac{\varepsilon_E \cdot R^I_{i,E}}{\varepsilon_E \cdot R^I_{i,E} + W}, \tag{4.41}$$

where ε_E is the E_b/N_0-value of the E-DCH radio access bearer. Note that we assume here that the target-E_b/N_0-values are equal for all rates. However, this restriction can be easily avoided by introducing individual target-E_b/N_0-values for each TBS.

The next step is to select the information bit rate such that condition (4.34) is fulfilled:

$$R^I_E(\bar{n}) = \max\{R_{i,E} | P(\eta_D(\bar{n}) + |n^a_E| \cdot \omega_{i,E} + \eta_{oc} \geq \eta^*) \leq p_t\} \tag{4.42}$$

The actual user data rates are now calculated according to the scheduling mechanism under the condition that the rate is higher than a certain minimum bit rate $R_{\min,E}$: In case of parallel scheduling, the user data rate is simply the information data rate. In case of one-by-one scheduling, the user data rate is approximated by

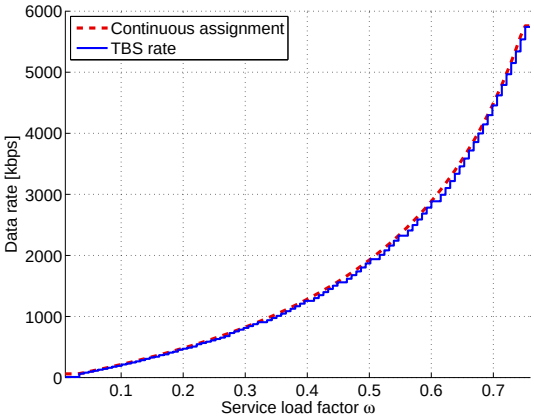


Figure 4.6: *Service load factors vs. data rates. The dashed line corresponds to the continuous data rate obtained from the SLF equation, the solid line to the transport block sizes defined in the specifications.*

dividing the information data rate by the number of E-DCH users:

$$R_E(\bar{n}) = \begin{cases} R_E^I(\bar{n}), & \text{if} \quad R_E^I(\bar{n}) \geq R_{\text{min},E} \quad \text{and parallel scheduling} \\ \frac{R_E^I(\bar{n})}{|n_E|}, & \text{if} \quad \frac{R_E^I(\bar{n})}{|n_E|} \geq R_{\text{min},E} \quad \text{and one-by-one scheduling} \\ 0 & \text{else.} \end{cases} \tag{4.43}$$

Figure 4.6 shows the mapping of the service load factors to information bit rates in case of a target-E_b/N_0 of 3 dB. The continuous case indicated by the dashed line is calculated from the definition of the service load factors as $R_{\text{opt}} = \frac{\omega \cdot W}{\hat{\varepsilon}^*(1-\omega)}$. The solid line corresponds to the SLFs calculated from the TBS-table in [73]. Both curves nearly match perfectly, and we see that for high SLFs, a small change means a large change on the data rate. The non-linear dependency between data rate and SLF leads to the argument that a slow-down (in terms of data rate) of the users increases the system capacity in terms of admissible ses-

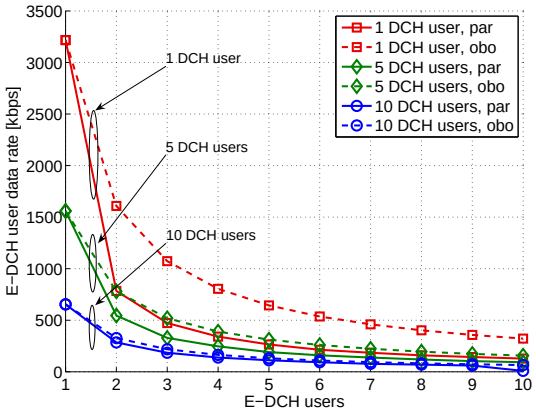


Figure 4.7: *Rate selection with one-by-one and parallel scheduling for different numbers of DCH and E-DCH connections.*

sions if an admission control based on the cell load is used ([125] and [127]). However, if we define capacity as the cumulated data rate per cell, the capacity decreases with the number of parallel transmitting users due to the increased interference, as we can see in Figure 4.7 and in Section 4.5.6.

Figure 4.7 shows the E-DCH user data rate for different numbers of DCH and E-DCH users. The solid lines indicate parallel scheduling and the dashed one-by-one scheduling. Different marker shapes indicate different numbers of concurrently active DCH users. We see that for only one E-DCH user, parallel and one-by-one scheduling have the same throughput. However, with more E-DCH users the gain of one-by-one scheduling over parallel scheduling increases since the users do not interfere with each other and thus are able to utilize the radio resources more efficiently, i.e. they get high SLFs and correspondingly also high data rates. This gain depends on the number of DCH users in the system: With more DCH users, the gain is lower. With 10 DCH users, there is nearly no difference between one-by-one and parallel scheduling anymore. In this case the

available resources for the E-DCH are already quite low and, thus, only SLFs with low transmission rates are possible.

4.5.3 Preserving and Preemptive Admission Control

The admission control (AC) is responsible for keeping the cell load below the maximum load. Generally, we model the AC based on the RRM target condition. We distinguish between two RRM policies for incoming QoS users: The first, which we call *preserving*, treats E-DCH and QoS equally. An incoming connection of either class is blocked if there are not enough resources available. The second, which we call *preemptive*, gives priority to QoS users. Active best effort connections may be dropped from the system in order to make room for the incoming QoS user. In both policies existing E-DCH connections are slowed-down if the number of QoS-connections increases. However, with the preserving strategy incoming QoS-calls are blocked if the RRM cannot slow-down the E-DCH connections any more. With the preemptive strategy, one or more E-DCH connections are dropped from the system in this case. Hence, blocking for the QoS users occurs only if nearly all resources are occupied by QoS connections (cf. Figure 4.8).

If a new connection arrives to the network, the AC performs two steps: At first, the amount of resources ω the incoming connection will occupy is identified. In case of a QoS-connection, this is simply ω_s. In case of an E-DCH connection, incoming connections are admitted if a minimum bit rate $R_{min,E}$ can be guaranteed. The corresponding SLF is denoted with $\omega_{min,E}$. Let us further denote with $\bar{n}^+$ the state vector $\bar{n}$ plus the incoming connection with service class s or with an additional E-DCH connection. The second step is then to estimate the probability for exceeding the maximum load with the new connection included. This step depends on the implemented policy:

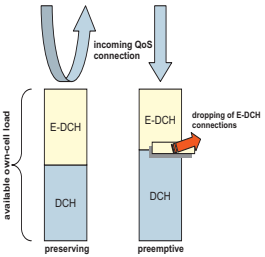


Figure 4.8: *Principle of the preserving and preemptive policy.*

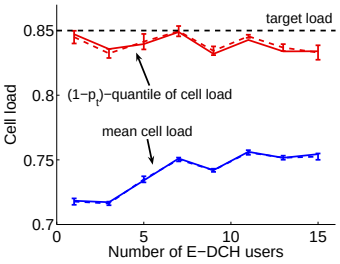


Figure 4.9: *Mean cell load and 95%-quantiles.*

Preserving Policy: In the preserving case, we calculate the parameters for the distribution of the expected cell load η_{AC} as in Equation (4.40), but with $\omega_{\min,E}$ for the E-DCH users:

$$\eta_{AC}(\bar{n}^+) = \eta_D(\bar{n}^+) + n_E^+ \cdot \omega_{\min,E} + \eta_{oc}, \tag{4.44}$$

where n_E^+ is the number of E-DCH mobiles with the incoming mobile included, if any. So, if the probability $P(\eta_{AC} \geq \eta^*)$ is higher than the target probability p_t, the connection is rejected. Otherwise the connection is admitted.

Preemptive Policy: With preemption, the incoming call is admitted if enough resources are available such that $P(\eta_{AC} \geq \eta^*) \leq p_t$ as in the preserving case. However, if the resources are insufficient, we distinguish two cases: If the incoming call belongs to an E-DCH user, the call is blocked. If the incoming call belongs to a QoS user, the RRM calculates from the service requirement ω_s the number of E-DCH connections with minimum rate $R_{E,min}$ which must be dropped from the system such that the incoming call can be admitted. The number of E-DCH connections $n_d(\bar{n}, s)$ which must be dropped depends on the current state and on the SLF of the incoming QoS-connection. It is given by the

following rule:

$$n_d(\bar{n}, s) = \min\{n | P(\eta_D(\bar{n}^+) + (n_E - n) \cdot \omega_{\min,E} + \eta_{oc} \geq \eta^*) \leq p_t\}. \quad (4.45)$$

Note that $0 \leq n_d \leq \lceil \frac{\omega_s}{\omega_{\min,E}} \rceil$. Blocking for QoS-users occurs if the number of E-DCH connections is too low to meet the requirements of the service class, i.e. if $n_d(\bar{n}, s) > n_E$. Blocking for E-DCH users occurs if the existing connections cannot be slowed down any further, due to the constraint on the minimum bit rate.

After admission control, the RRM executes the rate assignment as in Eq. (4.42) to adjust the bit rate of the E-DCH users to the new situation. Figure 4.9 illustrates the principle of admission control and rate selection. It shows the mean and the $(1-p_t)$-quantile (here $p_t = 5\%$) of the cell load distribution for 5 DCH users and an increasing number of E-DCH users. The target load is $\eta^* = 0.85$. Note that the results from a Monte-Carlo-simulation with random E_b/N_0-values (denoted by dashed lines) are very close to the analytical results, showing the accuracy of the lognormal approximation. Due to the discretization of the available rates, the $(1 - p_t)$-quantile does not exactly meet the target-load, but stays slightly below. Since the coefficient of variation of the cell load is decreasing with the number of users in the system, the mean load comes closer to the target load with an increasing number of E-DCH users.

4.5.4 Performance Evaluation

We assume that all calls arrive with exponentially distributed interarrival times with mean $\frac{1}{\lambda}$. The users choose a DCH service class or the E-DCH with probability p_s and p_E, hence the arrival rates per class are $\lambda_s = p_s \cdot \lambda$ and $\lambda_E = p_E \cdot \lambda$. The holding times for the DCH calls are also exponentially distributed with mean $\frac{1}{\mu_s}$. For the E-DCH users we assume a volume based user traffic model [18]. With exponentially distributed data volumes, the state-dependent departure rates of the E-DCH users are then given as

$$\mu_E(\bar{n}) = n_E \cdot \frac{R_E(\bar{n})}{E[V_E]}, \quad (4.46)$$

where $E[V_E]$ is the mean traffic volume of the E-DCH users.

The resulting system is a multi-service $M/M/n - 0$ loss system with state dependent departure rates for the E-DCH users. We are now interested in calculating the steady-state distribution of the number of users in the system. Since the joint Markov process is not time-reversible which can be instantly verified with Kolmogorov's reversibility criterion, no product form solution exists. The steady state probabilities follow by solving

$$Q \cdot \bar{\pi} = 0 \quad \text{s.t.} \quad \sum \pi = 1 \tag{4.47}$$

for $\bar{\pi}$, where Q is the transition rate matrix. The rate matrix Q is defined with help of the bijective index function $\phi(\bar{n}) : \Omega \rightarrow \mathbb{N}$, which maps the state vector $\bar{n}$ to a single index number. The transition rate $q(\phi(\bar{n}), \phi(\bar{n} \pm \bar{1}))$ in the rate matrix between states $\bar{n}$ and $\bar{n} \pm \bar{1}$ is then

$$\begin{aligned} q(\phi(\bar{n}), \phi(\bar{n} + \bar{1}_s)) &= \lambda_s \\ q(\phi(\bar{n}), \phi(\bar{n} + \bar{1}_E)) &= \lambda_E \\ q(\phi(\bar{n}), \phi(\bar{n} - \bar{1}_s)) &= n_s \cdot \mu_s \\ q(\phi(\bar{n}), \phi(\bar{n} - \bar{1}_E)) &= \mu_E(\bar{n}) \end{aligned} \tag{4.48}$$

for all valid states in the state space Ω. Additionally, $Q_{ii} = -\sum_j Q_{ij}$ for local balance and all other entries are set to zero. The sets of $\Omega^{+}_{ps,b}$ states where blocking occurs in the preserving case are defined by the condition $P(\eta(\bar{n}^{+}) \geq \eta^{*}) > p_t$, i.e. they form the 'edges' of the state space. With preemption, an E-DCH connection is dropped if $P(\eta(\bar{n}^{+s}) \geq \eta^{*}) > p_t$ and $n_d(\bar{n}, s) \geq \lceil \frac{\omega_s}{\omega_{\min,E}} \rceil$, i.e. in case of an incoming QoS connection. We define this set as $\Omega^{+s}_{pe,d}$. Blocking occurs then in the set $\Omega^{+s}_{pe,b} = \Omega^{+s}_{ps,b} \setminus \Omega^{+s}_{pe,d}$. The set of blocking states for E-DCH connections is the same for both policies. For the preemptive policy, an additional entry in the transition rate matrix is generated for states where preemption may occur:

$$q(\phi(\bar{n}), \phi(\bar{n} + \bar{1}_s - \bar{n}_d(\bar{n}, s))) = \lambda_s. \tag{4.49}$$

As performance measures we choose the service-dependent call blocking probabilities P_s, the call dropping probability P_d which applies only in the case of the

preemptive strategy, and the mean user bit rate $E[R_U]$ achieved by the E-DCH users. The call blocking probabilities are easily calculated as the sum of all states probabilities in which blocking may occur:

$$P_s = \sum_{\bar{n}|\bar{n}\in\Omega_b^{+s}} \pi(\bar{n}). \tag{4.50}$$

Note that we omit the qualifier for the admission control policy. We define the call dropping probability in our analysis as the probability that an E-DCH connection is dropped if a QoS-call is arriving in the system. This probability is given by

$$P_d = \sum_{\bar{n}|\Omega_{pe,d}^{+s}} \frac{\pi(\bar{n}) \cdot \sum_{s'\in\mathcal{S}} P_{s'}^a \cdot P_{s'}^{\text{sel}}}{\sum_{\bar{n}'|n_E>0} \pi(\bar{n}')}, \tag{4.51}$$

where P_s^a is the probability that the incoming connection is of class s and P_s^{sel} is the probability that an active E-DCH connection is selected for dropping:

$$P_s^a = \frac{\lambda_s}{\sum_{s'\in\mathcal{S}} \lambda_{s'}}, \quad \text{and} \quad P_s^{\text{sel}} = \frac{n_d(\bar{n}, s)}{n_E}. \tag{4.52}$$

We further define the mean throughput per user at a random time instance as

$$E[R_U] = \sum_{R_E>0} R_E \cdot \frac{\sum_{\bar{n}|R_E(\bar{n})=R_E} n_E \cdot \pi(\bar{n})}{\sum_{\bar{n}'|n_E>0} n'_E \cdot \pi(\bar{n}')}, \tag{4.53}$$

which is conditioned with the probability that at least one E-DCH user is in the system. Finally, the mean total cell throughput (or system throughput) follows according to Little as

$$E[R_{T,E}] = \lambda_E \cdot (1 - P_{b,E}) \cdot E[V_E]. \tag{4.54}$$

Note that the mean cell throughput includes cases where no E-DCH users are active or in the system.

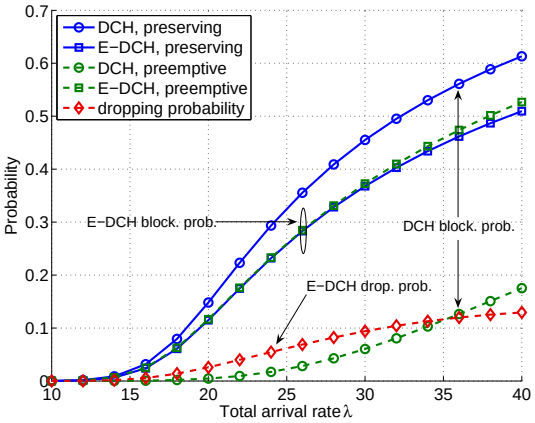


Figure 4.10: *Blocking and dropping probabilities for DCH and E-DCH users.*

4.5.5 Impact of User Preemption

In this section we give some numerical examples for the impact of user preemption. The scenarios, if not stated otherwise, consist of two service classes: 64 kbps QoS-users (i.e. DCH users) with a target-E_b/N_0 of 4 dB and the E-DCH best effort users with a target-E_b/N_0 of 3 dB. The service probabilities are $p_D = 0.4$ and $p_E = 0.6$ for DCH and E-DCH connections, respectively. The mean volume size for E-DCH connections is $E[V_E] = 72$ kbit. The minimum guaranteed data rate for E-DCH users is 60 kbps.

Figure 4.10 shows blocking and dropping probabilities for both admission policies. The curves with circles indicate the blocking probabilities for the 64 kbps DCH QoS users, while the curves with square markers show the blocking probabilities for the E-DCH users. The dashed line with diamond markers shows the dropping probabilities in case of preemption. In this scenario, the total arrival rate λ is increased from $10\,s^{-1}$ to $40\,s^{-1}$. Although a system with such high blocking probabilities can be considered as heavily overloaded, we show

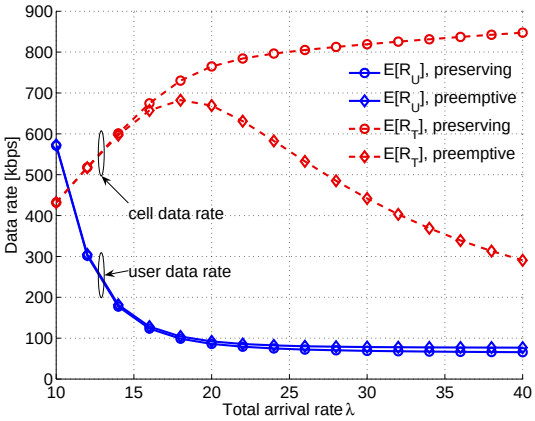


Figure 4.11: *Mean user and cell bit rates. With preemptive AC, DCH connections force out E-DCH connections with increasing load, leading to shrinking cell throughputs.*

these results for a better understanding of the effect of preemption. The comparison of the DCH blocking probabilities for both schemes reveals an enormous performance gain for DCH QoS users with preemption, which is caused by the substantially smaller set of states where blocking can occur at all. The blocking probabilities for the E-DCH users, however, are nearly identical and only begin to diverge under high load conditions, where the preserving scheme has a slight advantage over the preemptive scheme. The dropping probabilities for E-DCH users increase with the arrival rate, but do not exceed approximately 10 %. With a higher load, the system is nearly fully occupied with QoS users and the remaining E-DCH users cannot be dropped due to their lower resource requirements.

The impact of preemption on the average user and cell data rates (defined as the cumulated data rates of all users) is shown in Figure 4.11 for the same scenario. The user data rates are indicated with solid lines, the cell data rates with dashed

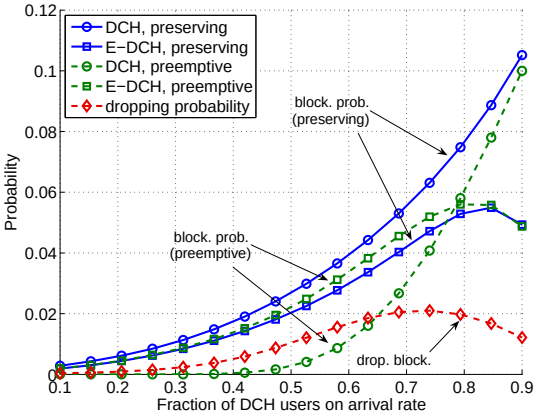


Figure 4.12: *The impact of preemption depends on the ratio between DCH and E-DCH users.*

lines. Note that the average cell data rates include the case where no E-DCH user is in the system. This explains why they are indeed lower than the average user data rates for very small arrival rates. The expected user data rates $E[R_U]$ are in both cases nearly identical with a slight advantage for the preemptive case for higher loads, reflecting the slightly higher blocking probabilities shown in Figure 4.10. However, due to the dropping of E-DCH users the average cell data rates $E[R_T]$ in the preemptive case becomes significantly lower than in the preserving case with increasing load. The small increase for the preserving scheme with arrival rates higher than $20\,s^{-1}$ indicates that the system is saturated above this point.

In the next scenario we fix the total arrival rate to $15\,s^{-1}$ and vary the ratio between DCH and E-DCH arrivals from $10\%/90\%$ to $90\%/10\%$. The results are shown in Figure 4.12. We see that an increasing number of DCH arrivals leads to a higher system load due to the slightly lower minimum guaranteed data rate for E-DCH users and the time-based traffic model of the DCH users, leading to

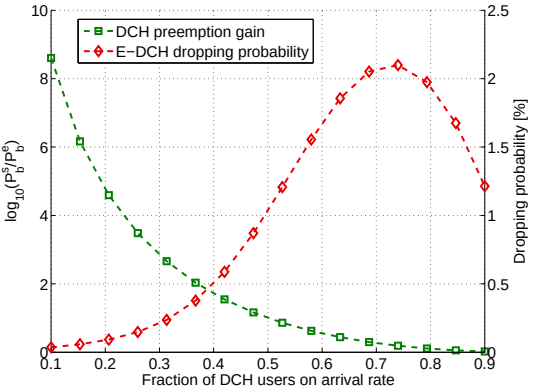


Figure 4.13: *Logarithmic gain of DCH blocking probabilities with preemption over preserving admission control and corresponding E-DCH dropping probabilities.*

overall higher blocking probabilities for DCH connections. Further, in situations with a high fraction of best-effort traffic, preemption leads to a substantial decrease of the blocking probabilities for the QoS users with still acceptable dropping probabilities. However, if the ratio is shifted to the QoS side, the decreasing load available to the E-DCH users leads to increased dropping probabilities. With preemption, the blocking probabilities for E-DCH users are slightly higher than without preemption if more DCH users are in the system, since then, incoming E-DCH connections see more DCH connections and a higher system load due to E-DCH connection dropping. For low fractions of E-DCH users, the blocking probability shrinks again.

Figure 4.13 clarifies the trade-off between blocking and dropping probabilities. The values indicated with square markers show the DCH blocking preemption gain, defined as blocking probability for preserving AC over blocking probability for preemptive AC, P_b^s / P_b^e. The preemption gain is logarithmically scaled with

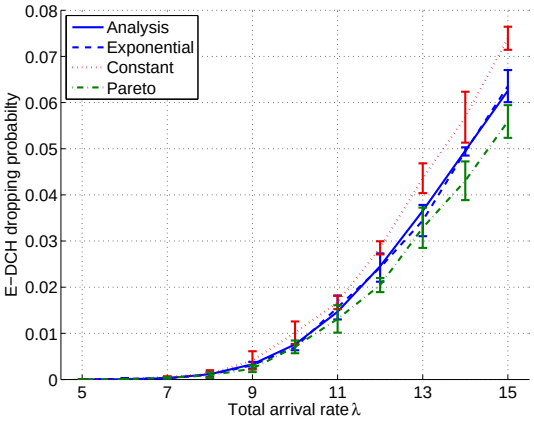
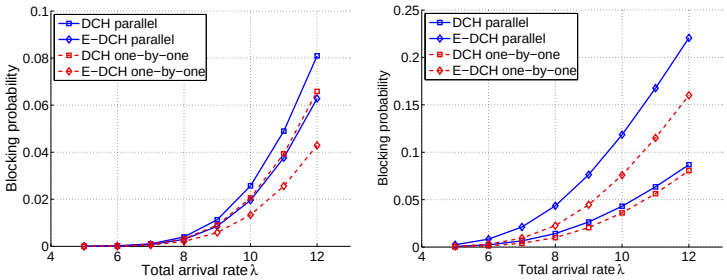


Figure 4.14: *Sensitivity of dropping probabilities against the volume size distribution: higher variability leads to lower dropping probabilities.*

base 10. The corresponding E-DCH dropping probability is indicated by diamond markers. As an example let us consider a balanced configuration, i.e. 50 % DCH and 50 % E-DCH users. The E-DCH dropping probability in that case is 1 %, but the blocking probability for DCH users with preemption is 10 times lower than in the preserving case.

Figure 4.14 shows the sensitivity of the system to different volume size distributions for the E-DCH users. The results are calculated with an event-discrete flow-level simulation which was also used for the validation of the analytical results. Three cases are presented: Constant volume size, exponentially, and Pareto-distributed volume sizes (with parameters $k = 1.5$ and $x_{min} = 2.4 \cdot 10^4$), all with the same mean. The solid line represents the analytical results, which fit very well to the results of the simulation with exponentially distributed transfer volumes. We further see that a higher coefficient of variation leads to lower dropping probabilities due to the resulting higher occurrence of smaller sojourn times



(a) 60 kbps guaranteed E-DCH data rate (b) 200 kbps guaranteed E-DCH data rate

Figure 4.15: *Comparison of blocking probabilities for different minimum guaranteed E-DCH data rates.*

(see e.g. [109]). In case of lower load, the differences are quite small, which may lead to the conclusion that then the exponential assumption may be a sufficient approximation.

4.5.6 Parallel vs. One-by-One Scheduling

In this section we investigate the impact of parallel and one-by-one scheduling as well as the minimum guaranteed E-DCH data rate on the system performance. Remember from Section 4.5.2 that one-by-one scheduling, although in the current UMTS specification not feasible due to lacking synchronization of the uplink physical data channels, has the inherent advantage that a transmission does not suffer from any interference from other mobiles in the same cell. A disadvantage is that a single mobile may not be able to completely occupy the available resources due to transmit power restrictions. However, we assume that the cells are dimensioned such that the transmit power is no limiting factor.

The evaluation scenario is the same as in the previous section: 64 kbps QoS-users (i.e. DCH users) with a target-E_b/N_0 of 4 dB and E-DCH best effort users with a target-E_b/N_0 of 3 dB are considered. The service probabilities are $p_D =$

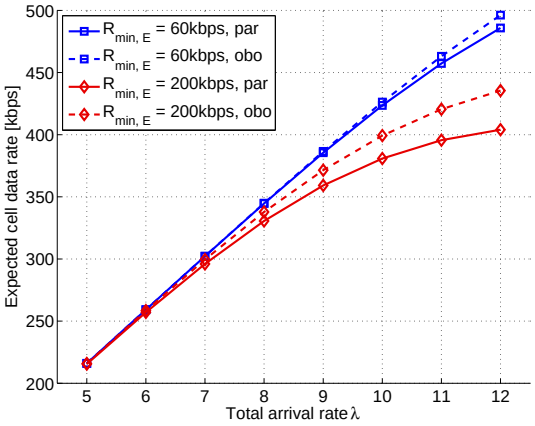


Figure 4.16: *Mean total E-DCH throughput. The performance gain of one-by-on scheduling is larger for high minimum data rates.*

0.4 and $p_E = 0.6$. We consider only preserving admission control.

In the first scenario we compare the blocking probabilities for parallel and one-by-one scheduling. In Figure 4.15(a), the minimum guaranteed data rate for E-DCH users, $R_{\min,E}$, is 60 kbps. The mean volume size $E[V_E]$ is 72 kbit. Diamond markers indicate the blocking probabilities for E-DCH users, square markers for DCH users. We see that in this scenario the blocking probabilities for the DCH users are higher than for the E-DCH users. Due to the low minimum E-DCH bit rate and the resulting low minimal service load factor, E-DCH users may still connect to the system even if DCH users are already blocked. The comparison of the parallel (solid lines) with the one-by-one scheduling case (dashed lines) shows that the throughput gain of the one-by-one users leads to lower blocking probabilities, and also to a larger difference between DCH and E-DCH users.

In Figure 4.15(b), the scenario is equal to the previous one with the exception of the minimum guaranteed data rate $R_{\min,E}$, which is now 200 kbps. In this case,

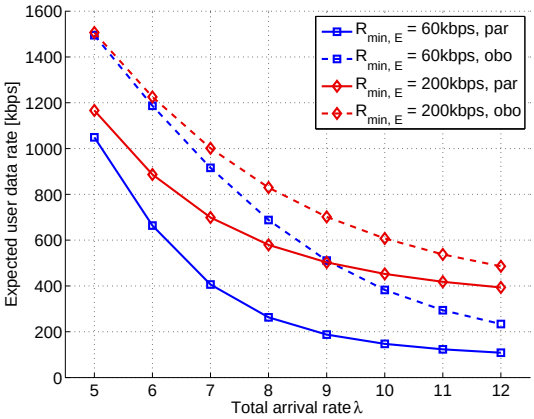


Figure 4.17: *Mean E-DCH data rates per user. The performance gain for one-by-one scheduling shrinks with increasing load due to the interference generated by DCH connections.*

the service load factor for the minimum data rate of the E-DCH connections is higher than the load requirements of the DCH users. Consequently, the E-DCH blocking probabilities are now higher than the DCH blocking probabilities. We further see that the DCH blocking probabilities for both scheduling disciplines are nearly identical. The reason is that with both schemes, DCH users are still able to connect even if E-DCH connections are blocked, leveling the advantage of one-by-one scheduling due to less interference.

In Figure 4.16, the total cell data rates are compared, dashed lines again indicate one-by-one scheduling. As expected, the data rate for one-by-one scheduling is always higher than for parallel scheduling, although this becomes visible only if a certain system load is reached since otherwise the average time with no E-DCH user is so high that the performance differences are diminished. The gain becomes larger with increasing load due to lower blocking probabilities for one-

by-one scheduling, (cf. Figure 4.15). This is an interesting fact since in Section 4.5.2 we have seen that the performance gain for one-by-one scheduling becomes smaller with an increasing number of DCH users. However, the performance gain becomes visible only if the probability that no E-DCH user is in the system is sufficiently small. Finally, the impact of the minimum cell data rate is more visible than the impact of the scheduling due to the significant higher blocking probabilities for $R_{\min,E} = 200\,\text{kbps}$.

The higher blocking probabilities for the scenario with a more generously guaranteed data rate are to the benefit of the users admitted to the system, as it can be concluded from Figure 4.17. Here, dashed lines indicate one-by-one scheduling and solid lines parallel scheduling, while square markers denote a minimum data rate of $R_{\min,E} = 60\,\text{kbps}$ and diamond markers of $R_{\min,E} = 200\,\text{kbps}$. Corresponding to the cell data rates we see that the impact of the minimum data rates on the average user data rates increases with the total arrival rate. The close values for low loads reflect the low blocking probabilities (e.g. at $\lambda = 5\,s^{-1}$). For the scheduling discipline, the opposite can be observed. The performance gain of one-by-one scheduling decreases with higher loads. The reason is that the increasing number of DCH users leads to more interference which decreases the advantage of the one-by-one scheduling.

4.6 Multi-Cell Capacity Model

The single cell capacity model described in the previous Section assumes that the influence of mobiles in surrounding cells is independent of the load at the considered NodeB. More specifically, we assumed that the parameters of the interference and load distribution are constant. In this section, we propose a more accurate model that takes the mutual dependency between other-cell interference and transmit power of the mobiles into account. The model uses a similar approach as in [10, 40, 135] by calculating the first and second moments of the other-cell interference [23].

The general formulation in Section 4.3 for the uplink received powers in a UMTS network enables us to calculate the interference for a deterministic situation, i.e. in a situation where all parameters (distance, target-E_b/N_0, etc.) are fixed. We now extend this model by a stochastic component: we assume that the number of users in the cells and their position is random. This means that we are now interested in the *distribution* (and its parameters) of the received interference at every NodeB.

In contrast to the previous sections, we consider the other-cell unaware radio resource management scheme, i.e. we assume that the NodeB tries to keep the own-cell load below a certain target own-cell load such that $\eta_{own} \leq \eta^*_{own}$ (see Figure 4.1). This implicates that the own-cell load η_E available for the E-DCH users is now independent from the other-cell interference, which means in turn that that an "infeasible" situation as we investigated in Section 4.4 cannot occur anymore. Outage, caused by insufficient transmit power at the UE, may still occur, but only because of the direct influence of the other-cell interference. We further assume perfect power control and equal-rate assignment for all E-DCH users, i.e. that the received E_b/N_0-values are deterministic and equal to the target-E_b/N_0, and that the data rate of all E-DCH users is equal.

4.6.1 Steady State User Distribution

According to the other-cell unaware RRM scheme and the employed equal-rate parallel scheduling, the service load factor of an E-DCH user k controlled by an arbitrary NodeB with state vector $\bar{n}$ is

$$\omega_{k,E}(\bar{n}) = \frac{\eta^*_{own} - \eta_D(\bar{n})}{n_E}, \tag{4.55}$$

This enables us to calculate the steady state distribution of the users at that NodeB independently from the other-cell interference with a $S+1$-dimensional Markov chain model similar as in the single-cell case, where S is the number of DCH service classes. The resulting state space Ω is restricted by admission control based on the own-cell load and a minimum guaranteed bit rate R^E_{min} for the E-DCH users which corresponds to a minimum service load factor ω^E_{min}. An incoming user is accepted if the total own-cell load plus the service load factor of the incoming user does not exceed the own-cell target load such that the following condition must is fulfilled:

$$n_E \cdot \omega_E^{min} + \eta(\bar{n}) \leq \eta^*_{own}. \tag{4.56}$$

So in contrast to the probabilistic admission control proposed for the single-cell model due to the influence of lognormal other-cell interference and imperfect power control, here a deterministic policy is enforced.

The computation of the steady-state distribution is equal to the calculation described in Section 4.5.4 in the single-cell case. So we will not describe it in detail again. An important distinction is the calculation of the user data rates. Since we assume perfect power control and perfect equal-rate assignment, the data rates are directly given by the E-DCH service load factor and target-E_b/N_0-value:

$$R_E(\bar{n}) = \frac{W}{\varepsilon^*_E} \cdot \frac{\omega_E(\bar{n})}{1 - \omega_E(\bar{n})}, \tag{4.57}$$

where ε^*_E is the target-E_b/N_0 value of an E-DCH connection.

4.6.2 Other-Cell Interference Revisited

Recalling the interference model in Section 4.3, the other-cell interference at NodeB x can also be characterized by the following two equations in a static scenario, i.e. in a fixed system state $\bar{n}$:

$$I_x^{oc} = \sum_{l \in \mathcal{L} \setminus x} I_{y \to x}^{out} \tag{4.58}$$

$$I_{y \to x}^{out} = \eta_{y,x} \cdot (W \cdot N_0 + I_y), \tag{4.59}$$

where $\eta_{y,x}$ is defined as the load inflicted on NodeB x by NodeB y:

$$\eta_{y,x} = \sum_{k \in \mathcal{M}_y} \Delta_{k,x}^{y} \cdot \omega_k. \tag{4.60}$$

Note that we follow the notation established in [40] for compact representation, and that the total own-cell load η_x comprises DCH and E-DCH users:

$$\eta_{y,x} = \eta_{y,x}^{D} + \eta_{y,x}^{E}. \tag{4.61}$$

Recalling furthermore that I_x consists of other-cell and own-cell interference, whereas the latter can be written depending on the other-cell interference as

$$I_x^{own} = \frac{\eta_{x,x}}{1 - \eta_{x,x}} \cdot (W \cdot N_0 + I_x^{oc}), \tag{4.62}$$

the "outgoing" interference becomes

$$I_{y \to x}^{out} = \frac{\eta_{y,x}}{1 - \eta_{x,x}} \cdot (W \cdot N_0 + I_x^{oc}), \tag{4.63}$$

as it can be easily shown with some algebraic manipulations. Since both equations for other-cell and "outgoing" interference now depend on each other only, the other-cell interference can be calculated directly by solving the matrix equation as in [40, 10]:

$$\bar{I}^{oc} = \tilde{\zeta} \cdot (\bar{N}_0 + \bar{I}^{oc}), \tag{4.64}$$

where the elements of the $|\mathcal{L}| \times |\mathcal{L}|$-matrix are set to $(\tilde{\zeta})_{i,j} = \frac{\eta_{j,i}}{1-\eta_{i,i}}$, and the $|\mathcal{L}| \times 1$ column vectors $\bar{N}_0$ and $\bar{I}^{oc}$ contain the thermal noise and the other-cell interference, respectively. Another way is to use a fixed-point iteration scheme as it has been proposed in [22, 23].

4.6.3 A Stochastic Other-Cell Interference Model

An exact calculation of the other-cell interference distribution would require to consider all state permutations over all cells in the network, which is computationally intractable. Instead, we follow the same approach as in [10, 40, 129] and assume that the other-cell interference is lognormal distributed, such that we need to calculate only the first two moments in order to get approximative statistics. The difference to a system with only DCH users is that we have to distinguish between the case that E-DCH users are active, which means that the own-cell target load is reached, or no E-DCH users are active, which means that the own-cell load is in most cases much lower than the target-load. With this approach, we can calculate the moments of the other-cell interference as well as the probability that a certain maximum sustainable load is exceeded with the theorem of total probability.

We are therefore interested in the first and second moment of the random variable ζ, which is given by the theorem of total probability as follows:

$$E[\zeta_{y,x}] = \sum_{\bar{n}|n_E=0} \pi(\bar{n}) \cdot \zeta(\bar{n}) \cdot E[\Delta_y^x] + \sum_{\bar{n}|n_E>0} \pi(\bar{n}) \cdot \zeta^* \cdot E[\Delta_y^x], \quad (4.65)$$

where $\zeta(\bar{n})$ and ζ^* are defined as

$$\zeta(\bar{n}) = \frac{\eta_D(\bar{n})}{1-\eta_D(\bar{n})} \qquad \text{and} \qquad \zeta^* = \frac{\eta^*_{own}}{1-\eta^*_{own}}, \quad (4.66)$$

where the latter reflects the case of fully used own-cell target-load if E-DCH users are active. Correspondingly, the second moment follows as

$$E[\zeta^2_{y,x}] = \sum_{\bar{n}} \pi(\bar{n}) \cdot \left(\text{VAR}[\zeta_{y,x}(\bar{n})] + E[\zeta_{y,x}]^2\right), \quad (4.67)$$

where the conditional variance is defined as

$$\mathrm{VAR}[\zeta_{y,x}(\bar{n})] = \begin{cases} \frac{\sum_{s\in\mathcal{S}} n_s\cdot(\omega^D_{y,s})^2}{(1-\eta^D_{y,own})^2}\cdot \mathrm{VAR}[\Delta^x_y], & \text{if } n_E = 0 \\ \frac{\sum_{s\in\mathcal{S}} n_s\cdot(\omega^D_{y,s})^2 + \frac{(\eta^E_{y,own})^2}{n_E}}{(1-\eta^*_{own})^2}\cdot \mathrm{VAR}[\Delta^x_y], & \text{else.} \end{cases} \tag{4.68}$$

We have now the means to calculate the moments of the other-cell interference over the whole network, either by solving the matrix equations similar to the previous section or with iterative methods. For a more detailed description the reader is referred to Appendix B and [40].

Under assumption of lognormal distributed other-cell interference, the probability that a certain interference threshold corresponding to a maximum sustainable load is exceeded can now be computed. We define the maximum allowed total load as η_{max} and following from that the maximum interference as

$$I_{max} = \frac{\eta_{max}}{1-\eta_{max}}\cdot W\cdot N_0. \tag{4.69}$$

The probability that the maximum interference is exceeded follows again from the theorem of total probability:

$$P^{out}_x = \sum_{\bar{n}|n_E=0} \pi(\bar{n})\cdot P^{out}_x(\bar{n}), \tag{4.70}$$

where $P^{out}_x(\bar{n})$ is the conditional probability that the maximum interference is exceeded, which we assume as lognormal distributed. We have to distinguish between the two cases that E-DCH users are in the cell or not, consequently the probabilities are defined as

$$P^{out}_x(\bar{n}) = \begin{cases} 1-\mathrm{LN}_{\mu_x,\sigma_x}(I_{max}-I^D_x(\bar{n})), & \text{if } n_E = 0 \\ 1-\mathrm{LN}_{\mu_x,\sigma_x}(I_{max}-I^*_{own}), & \text{otherwise.} \end{cases} \tag{4.71}$$

The involved interference values can be calculated as $I = \frac{\eta}{1-\eta}\cdot W\cdot N_0$. The parameters μ_x, σ_x for the lognormal distribution are calculated directly from the moments of the other-cell interference as $\mu_x = \ln(E[I^{oc}_x]) - \frac{\sigma^2_x}{2}$ and $\sigma^2_x = \ln\left(\frac{\mathrm{VAR}[I^{oc}_x]}{E[I^{oc}_x]^2}+1\right)$.

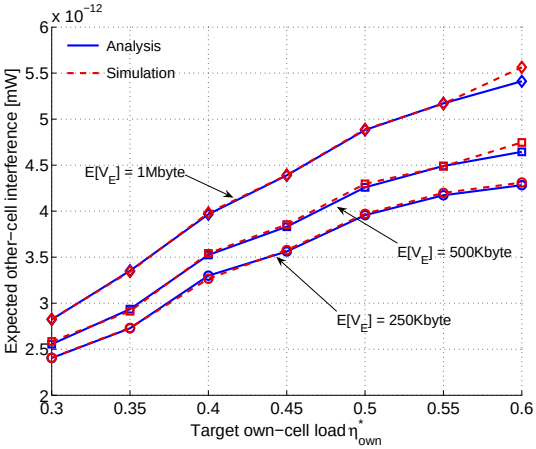


Figure 4.18: *Mean other-cell interference for different mean volume sizes. The small spikes in the curves result from the minimum data rates for E-DCH users and admission control.*

4.6.4 Numerical Results

We now present some numerical results. The scenario uses the hexagonal 19-cell layout with homogeneous user arrivals. The data rate of the DCH users is assumed to be 64 kbps with a target-E_b/N_0-value of 3 dB. The minimum guaranteed data rate of the E-DCH connection is 9 kbps with a fixed target-E_b/N_0 of also 3 dB. All other system parameters (like chiprate, thermal noise density) are equal to the standard assumptions we use throughout this monograph. Only the center cell is considered for the results. The analytical results are verified with a time-dynamic flow-level simulation that calculates the complete interference situation at every arrival and departure event. Users arrive with exponentially distributed inter-arrival times with rate $\lambda_E = 40^{-1}s^{-1}$ and $\lambda_D = 10^{-1}s^{-1}$. The call times $E[T_D]$ for the DCH connections as well as the volume sizes for the E-

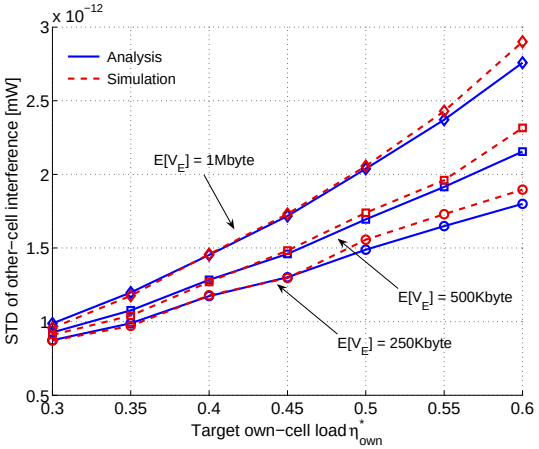


Figure 4.19: *Standard deviation of the other-cell interference for different mean volume sizes. For high loads, the results for analysis and simulation begin to diverge.*

DCH connections are also exponentially distributed. We assume $E[T_D] = 100s$.

In Figure 4.18, the first moment of the other-cell interference versus an increasing target own-cell load is shown for different values of the mean E-DCH volume size $E[V_E]$. Solid lines indicate analytical results, dashed lines the corresponding simulation results. We observe that both results match very well, with small errors for higher own-cell target loads. An interesting phenomenon are the small spikes (for example at $\eta^*_{own} = 0.4$), which originate from a discretization effect due to the minimum data rate of the E-DCH connections. We further observe that the mean other-cell interference increases both with the target own-cell load (which constitutes the upper limit of the mean own-cell load) and the E-DCH mean volume size, the latter due to the longer time periods without active E-DCH users.

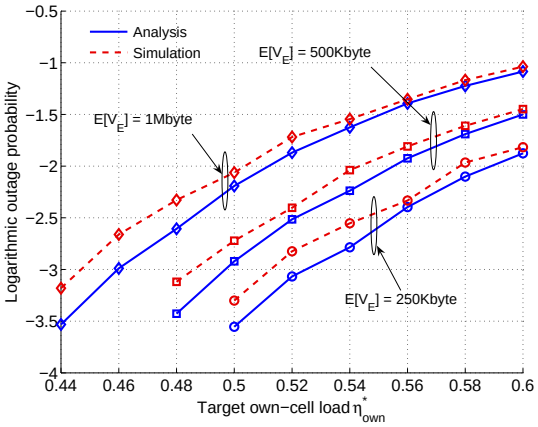


Figure 4.20: *Logarithmic outage probabilities versus increasing own-cell target load.*

The corresponding standard deviations are shown in Figure 4.19. Again, the simulation results and the analytical results match quite well, however, the results begin to diverge with higher cell loads. The reason is that we assume independency between the random variables ζ (reflecting the other-cell load) and the other-cell interference which leads to an increasing error correlated to the system load. Also, the curves, especially for $E[V_E] = 1$ MByte, are convex in contrast to the concave progression of the first moment, leading to an increasing coefficient of variation with increasing own-cell target load.

Finally, the probability that the maximum sustainable interference I_{max} is exceeded ("outage probability") is shown in Figure 4.20. For a better visibility of very small values the log-probability is plotted on the y-axis. We see that the analytical results slightly underestimate the outage probability. The reason is that the lognormal assumption for the other-cell interference becomes more valid with an increasing number of users in the system.

4.7 Concluding Remarks

The flexibility of the rate control mechanism implemented by the UMTS enhanced uplink is a challenge for modeling. In this chapter, we presented new approaches how to calculate flow-level performance measures like blocking probabilities and average cell and user data rates. The comparison between one-by-one and parallel scheduling clarified the inherent advantage of one-by-one scheduling and showed the relation to the number of concurrently active DCH users. Radio resource sharing and admission control have a large influence on the overall performance. Preemption of E-DCH users leads to drastically increased QoS for DCH connections in terms of blocking probability. However, operators must trade off the DCH QoS with the dropping probability for E-DCH users. We also showed that the impact of preemption on the perceived average user data rate is insignificant. Finally, the multi-cell capacity model provides network planners a tool to include the enhanced uplink in their considerations for proper network dimensioning.

The generic nature of the models allow for inclusion of different radio resource management strategies which have not been addressed in this monograph, for example resource reservation schemes. Some problems, however, would require major modifications, like channel-aware scheduling schemes or the inclusion of other-cell DOWN grants into the stochastic other-cell interference model.

Another aspect is the interaction between HSDPA and the enhanced uplink from the perspective of resource efficiency. The enhanced uplink requires several signaling channels on the downlink with substantially large data rates and also correspondingly low spreading factors. Since high data rates for the HSDPA require a large fraction of the code resources, the use of E-DCH radio bearers could lead to decreasing HSDPA data rates.

Like for every shared resource system with variable data rate, the user behavior has a large impact on the system performance. This applies to the difference between volume-based and time-based traffic as well as for applications which purposefully adapt their traffic volume to the current quality of service like Skype.

Interactions between such applications and QoE-aware schedulers are therefore a point for further research.

5 Summary

The complexity of integrated, 3.5G-enabled UMTS networks is a challenge for system and performance modeling. However, operators and other interested parties need tools for evaluation, planning and optimization of such networks. In this monograph, we proposed and validated modeling frameworks for such tasks. Numerical studies confirmed the importance of an integrative approach and revealed the interactions between user behavior, radio resource management and physical layer effects. Our approach for all models in this work is based on the concept of *flows*, which constitute a coherent stream of packets with the same source or destination address. This approach made it possible to find a compromise between an accurate inclusion of lower layer effects on the one hand, and computational efficient performance evaluation tools for time-average performance measures on the other hand.

We introduced an HSDPA model that consists of two parts: A physical layer abstraction model that approximates the stochastic properties of the user throughput in an inter-event time, and a radio resource model that describes the consumption of transmit power, interference and channelization codes. The physical layer abstraction model alone gave us insights into the dominating factors on HSDPA data rates. Radio channels with a high multi-path diversity constrain the HSDPA data rate effectively even in locations close to the transmitting antenna. With less severe propagation conditions, the number of available codes for the HS-DSCH transport channel is the restricting factor.

These insights gave us the instruments to interpret the trade-off between HSDPA and DCH user performance with different radio resource sharing schemes. We considered the fixed scheme, which is a pure reservation scheme with a con-

stant amount of radio resource reserved for HSDPA, an adaptive scheme, where the HS-DSCH instantaneously utilizes all resources which are not used by DCH connections and the hybrid scheme, which is a combination of the two previous schemes. Apart from an unexpected high performance gain of the hybrid scheme over the fixed scheme in high load situations, we showed that the code/power balance is an important factor for the system performance. Furthermore, we observed that the transmit power allocation scheme has a significant influence on performance. Based on the assumption that rapid on and off-switching of the HS-DSCH decreases the effectiveness of fast power control, an always-on scheme and a power-ramping scheme was compared to the on-off scheme. Interestingly, the performance gain of the power-ramping and on-off scheme proved to be nearly independent of the user distance to the base station antenna. All these results have been gained from a discrete-event flow-level simulation with Round-Robin scheduling in order to exclude side effects due to channel-aware scheduling.

To complete our analysis we therefore compared Round-Robin, Proportional-fair and MaxTBS scheduling, constituting the three most common scheduling disciplines for HSDPA. MaxTBS scheduling is a “greedy” scheduling discipline which always favors the best user and is therefore often referred to as cell throughput optimal in the literature. However, we showed that with volume-based user traffic, i.e. with users that leave the system only if a certain data volume has been completely transmitted, Proportional-fair performs better than MaxTBS scheduling. The reason are the very long sojourn times of “starving”, low-bandwidth users at cell edges. This observation was confirmed with a comparison of the spatial arrival probability with the residence probability, i.e. the probability to meet an active user at a certain location. With MaxTBS scheduling the difference between arrival and residence probability was largest, followed by Proportional-fair and channel-blind Round-Robin scheduling. In the further course of this work we proposed an analytical queuing model with state-space reduction based on the Kaufman-Roberts equation. The model comprises an explicit expression of the conditional HSDPA sojourn time with Round-Robin

scheduling, however, for channel-aware scheduling schemes no such expressions have been found.

The second major focus of our work was the Enhanced Uplink. Although being the uplink counterpart to HSDPA, it is based on a different concept: The NodeB adjusts the maximum allowed transmit power of the mobile stations, which in turn adjust their instantaneous data rate accordingly. Fast power control is still in effect. We introduced mathematical expressions for such a rate and power controlled system with QoS DCH users and Enhanced Uplink user in co-existence. Then we investigated the feasibility region of such a system under different constraints and degrees of rate control distribution. We showed that the probability of outage events decreases with the number of Enhanced Uplink users in the system. In the further course of our work, we developed a capacity model which takes other-cell interference as independent lognormal random variable into account. Soft capacity in this model is considered with a probabilistic admission control scheme that keeps the load overshoot probability below a certain threshold. This model gave us insights into the impact of admission control and scheduling schemes. Due to fast power control, Enhanced Uplink radio bearers can provide bandwidth guarantees. A minimal user bandwidth is chosen as admission criterion for incoming connections. Preemptive admission control drops Enhanced Uplink radio bearers if a QoS DCH connections arrives into the system. The comparison with a preserving, non-dropping scheme unveiled the trade-off between E-DCH dropping probability and DCH blocking probability. Numerical results suggested that a substantial performance gain in terms of DCH blocking probabilities can be achieved with tolerable E-DCH dropping probabilities. However, these results are sensitive to the service mix between E-EDCH and DCH connections. Nevertheless, preemptive admission control proved to be a useful option for operators to ensure quality of service for DCH connections.

A further point that draw our interest was the impact of own-cell interference on the E-DCH performance. We investigated two antipodal scheduling schemes: With one-by-one scheduling, scheduling grants are assigned to the users in a Round-Robin manner, such that at every time instance only one user is transmit-

ting. With parallel scheduling, all users transmit at the same time. In the first case, the absence of any own-cell interference from other Enhanced Uplink users lead to a significant performance gain; however, the gain diminishes rapidly with an increasing number of DCH users.

Finally, we developed a model which takes other-cell interference explicitly into account. The model is based on the assumption that interference is approximately lognormal distributed, such that an explicit calculation of the interference distribution is avoided. A comparison with flow-level simulation results confirmed this assumption, with an increasing accuracy of the analytical model in case of higher loaded systems. The model therefore enables the inclusion of the Enhanced Uplink into the network planning and optimization process.

Summarizing, both the results for HSDPA as well as for Enhanced Uplink confirmed that it is essential to take a holistic approach for the evaluation of todays and future mobile communication networks. Physical layer effects, radio resource management and sharing for QoS and best-effort users, and user behavior together determine the system performance.

The complexity of future communication systems will not diminish, although one design goal of UTRAN LTE is to decrease the technical complexity on the fixed part of the radio access network. However, it is reasonable to anticipate that the *behavior* of the system will be more complex: A partly self-organizational RAN and frequency-selective scheduling in a common frequency spectrum require highly sophisticated algorithms for radio resource management. The methods we developed in the course of this work are also applicable to future systems like LTE.

A Conditional Mean Sojourn Time of an HSDPA User

The assumption is that the probability p_f to meet a user at a certain location f is proportional to the reciprocal of the location dependent data rate $R_f(j, n_H)$, i.e.

$$p_f \sim \frac{1}{R_f(j, n_H)}, \qquad \text{such that} \qquad p_f = \frac{\frac{1}{R_f(j,n_H)}}{\sum_f \frac{1}{R_f(j,n_H)}}. \tag{A.1}$$

The occupied code resources are denoted by j, the number of HSDPA users is n_H. The mean conditional sojourn time is

$$\begin{aligned} E[T|(j, n_H)] &= E\left[\frac{V_H}{R(j, n_H)}\right] \\ &= E[V_H] \cdot E\left[\frac{1}{R(j, n_H)}\right] \\ &= E[V_H] \cdot \sum_f p_f \cdot \frac{1}{R_f(j, n_H)} \\ &= E[V_H] \cdot \sum_f \frac{\frac{1}{R_f(j,n_H)}}{\sum_f \frac{1}{R_f(j,n_H)}} \cdot \frac{1}{R_f(j, n_H)} \\ &= E[V_H] \cdot E\left[\frac{1}{R_f(j, n_H)^2}\right] \cdot E\left[\frac{1}{R_f(j, n_H)}\right]^{-1}. \end{aligned} \tag{A.2}$$

Note that the second line of the equation above is possible due to the assumption of a constant data rate during every inter-event time.

B Moments of the Enhanced Uplink Other-Cell Interference

With Equations (4.58), (4.63), and the assumption of independence between the other-cell interference and the load variable ζ, the first moments and variance of the other and outgoing interferences are

$$E[I_x^{oc}] = \sum_{l \in \mathcal{L} \setminus x} E[I_{y \to x}^{out}], \tag{B.3}$$

$$E[I_{y \to x}^{out}] = E[\zeta_{y,x}] \cdot \left(W \cdot N_0 + E[I_y^{oc}] \right), \tag{B.4}$$

$$\mathrm{VAR}[I_x^{oc}] = \sum_{l \in \mathcal{L} \setminus x} \mathrm{VAR}[I_{y \to x}^{out}], \tag{B.5}$$

$$\begin{aligned} \mathrm{VAR}[I_{y \to x}^{out}] = \mathrm{VAR}[\zeta_{y,x}] \cdot \left(W^2 \cdot N_0^2 + 2 \cdot N_0 \cdot E[I_y^{oc}] \right) \\ + E[(\zeta_{y,x})^2] \cdot E[(I_y^{oc})^2] - E[\zeta_{y,x}]^2 \cdot E[I_y^{oc}]^2. \end{aligned} \tag{B.6}$$

The first moment of the other-cell interference is calculated by solving the following equation system:

$$E[\bar{I}_{oc}] = E[\tilde{\zeta}] \cdot (\bar{N}_0 + E[\bar{I}_{oc}]), \tag{B.7}$$

where $E[\bar{I}_{oc}]$ is a $|\mathcal{L}| \times 1$ column vector containing the first moment of the other-cell interference at all NodeBs, N_0 is a $|\mathcal{L}| \times 1$ column vector with elements $W \cdot N_0$ and $E[\tilde{\zeta}]$ is a $|\mathcal{L}| \times |\mathcal{L}|$ matrix with entries $(E[\tilde{\zeta}])_{ij} = E[\zeta_{i,j}]$ if $i \neq j$ and $(E[\tilde{\zeta}])_{ij} = 0$ otherwise. Solving for $E[\bar{I}_{oc}]$ yields

$$E[\bar{I}_{oc}] = (\tilde{I} - E[\tilde{\zeta}])^{-1} \cdot (E[\tilde{\zeta}] \cdot \bar{N}_0), \tag{B.8}$$

with $\tilde{I}$ as $|\mathcal{L}| \times |\mathcal{L}|$ identity matrix. The second moment of the other-cell interference is formulated as follows using Equation (B.6):

$$E[(I_x^{oc})^2] = \mathrm{H}_x + \sum_{l \in \mathcal{L} \setminus x} E[(\zeta_{l,x})^2] \cdot E[(I_l^{oc})^2], \tag{B.9}$$

$$\begin{aligned} \mathrm{H}_x = E[I_x^{oc}]^2 &+ \sum_{l \in \mathcal{L} \setminus x} \mathrm{VAR}[\zeta_{l,x}] \cdot \left(W^2 \cdot N_0^2 + 2 \cdot N_0 \cdot E[I_l^{oc}]\right) \\ &- \sum_{l \in \mathcal{L} \setminus x} E[\zeta_{l,x}]^2 \cdot E[I_l^{oc}]^2 \end{aligned} \tag{B.10}$$

This equation system in matrix form is then

$$E[(\bar{I}^{oc})^2] = \bar{\mathrm{H}} + E[\tilde{\zeta}^2] \cdot E[(\bar{I}^{oc})^2], \tag{B.11}$$

where $E[(\bar{I}^{oc})^2]$ is the $|\mathcal{L}| \times 1$ column vector containing the second moment of the other-cell interference, $\bar{\mathrm{H}}$ is a $|\mathcal{L}| \times 1$ column vector with entries H_x, and $E[\tilde{\zeta}^2]$ is a $|\mathcal{L}| \times |\mathcal{L}|$ matrix with entries $(E[\tilde{\zeta}^2])_{ij} = E[(\zeta_{i,j})^2]$ if $i \neq j$, and $(E[\tilde{\zeta}])_{ij} = 0$ otherwise. Finally, solving for $E[(\bar{I}^{oc})^2]$ yields

$$E[(\bar{I}^{oc})^2] = (\tilde{I} - E[(\tilde{\zeta})^2])^{-1} \cdot \bar{\mathrm{H}}. \tag{B.12}$$

C Acronyms

3GPP	3rd Generation Partnership Program
AAA	Authentication, Authorization and Accounting
AC	Admission Control
BPSK	Binary Phase Shift Keying
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CQI	Channel Quality Indicator
CPICH	Common Pilot Channel
CS	Circuit-Switched
DPCCH	Dedicated Physical Control Channel
DPDCH	Dedicated Physical Data Channel
DCH	Dedicated Channel
E-AGCH	Enhanced Absolute Grant Channel
E-DCH	Enhanced Dedicated Channel
E-DCCH	Enhanced Dedicated Control Channel
E-DPCCH	Enhanced Dedicated Physical Control Channel
E-DPDCH	Enhanced Dedicated Physical Data Channel
E-HICH	Enhanced Hybrid ARQ Indicator Channel
E-RGCH	Enhanced Relative Grant Channel
E-TFC	Enhanced Transport Format Combination
FACH	Forward Access Channel
FDD	Frequency Division Duplex
FTP	File Transfer Protocol
GGSN	General GPRS Support Node
GMSC	Gateway Mobile Switching Center
GPRS	General Packet Radio Service
HARQ	Hybrid Automatic Repeat Request
HLR	Home Location Register
HS-DPCCH	High Speed Dedicated Physical Control Channel

HS-DSCH	High Speed Downlink Shared Channel
HS-PDSCH	High Speed Physical Downlink Shared Channel
HS-SCCH	High Speed Signaling Control Channel
HSDPA	High Speed Downlink Packet Access
HSPA	High Speed Packet Access
HSUPA	High Speed Uplink Packet Access
IMS	IP Multimedia Subsystem
IMT	International Mobile Telecommunications
IP	Internet Protocol
ITU	International Telecommunication Union
LTE	Long Term Evolution
MAC	Medium Access Control
MIMO	Multiple Input Multiple Output
MSC	Mobile Switching Center
NRT	Non-Realtime
OVSF	Orthogonal Variable Spreading Factor
PDF	Probability Density Function
PDU	Packet Data Unit
PF	Proportional Fair
PS	Packet-Switched
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAB	Radio Access Bearer
RAN	Radio Access Network
RLC	Radio Link Control
RNC	Radio Network Controller
RRM	Radio Resource Management
RT	Realtime
RTP	Realtime Transport Protocol
TBS	Transport Block Size

TCP	Transmission Control Protocol
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TTI	Transport Time Interval
SF	Spreading Factor
SGSN	Supporting GPRS Support Node
SIR	Signal-to-Interference Ratio
SINR	Signal-to-Interference-and-Noise Ratio
SNR	Signal-to-Noise Ratio
SPI	Service Priotity Indicator
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
UPH	Uplink Power Headroom
UTRAN	UMTS Terrestrial Radio Access Network
VLR	Visiting Location Register
WCDMA	Wideband Code Division Multiple Access

Bibliography of the Author

— Book Chapters —

[1] D. Staehle and A. Mäder, "Radio Resource Management Strategies for HSDPA-enhanced UMTS Networks," in *Wireless Multimedia: Quality of Service and Solutions*, N. Cranley and L. Murphy, Eds. Hershey, USA: IGI Global, 2008.

— Journal Articles —

[2] A. Mäder and D. Staehle, "Radio Resource Management for the UMTS Enhanced Uplink in Presence of QoS Radio Bearers," *Annals of Operation Research, Special issue on Stochastic Performance Models for Resource Allocation in Communication Systems*, 2008.

— Conference Articles —

[3] R. Pries, A. Mäder, D. Staehle, and M. Wiesen, "On the Performance of Mobile IP in Wireless LAN Environments," in *Proc. of the 3rd EuroNGI Workshop on Wireless Systems and Mobility in Next Generation Internet*, ser. LNCS, J. García-Vidal and L. Cerdà-Alabern, Eds., vol. 4369. Sitges, Spain: Springer, Jun. 2006, pp. 155–170.

[4] A. Mäder and D. Staehle, "Interference Estimation for the HSDPA Service in Heterogeneous UMTS Networks," in *Proc. of the 2nd EuroNGI Workshop on Wireless Systems and Mobility in Next Generation Internet*, ser.

LNCS, M. Cesana and L. Fratta, Eds., vol. 3883. Como, Italy: Springer, Jul. 2005, pp. 145–157.

[5] A. Mäder and D. Staehle, "Comparison of Preemptive and Preserving Admission Control for the UMTS Enhanced Uplink," in *Proc. of the 15th KiVS*, Bern, Switzerland, Feb. 2007, pp. 201–212.

[6] A. Mäder and D. Staehle, "A Flow-Level Simulation Framework for HSDPA-enabled UMTS Networks," in *Proc. of IEEE/ACM MSWiM '07*, Crete, Greece, Oct. 2007, pp. 269–278.

[7] A. Mäder and D. Staehle, "An Analytical Model for Best-Effort Traffic over the UMTS Enhanced Uplink," in *Proc. of IEEE VTC Fall '06*, Montreál, QC, Canada, Sep. 2006, pp. 1–5.

[8] A. Mäder and D. Staehle, "Dimensioning of Hardware Components in UMTS Networks with Sectorized NodeBs," in *Proc. of the 19th International Teletraffic Congress*, Bejing, China, Sep. 2005.

[9] A. Mäder and D. Staehle, "Analytic Modelling of the WCDMA Downlink Capacity in Multi-Service Environments," in *Proc. of the 16th ITC Specialist Seminar*, Antwerp, Belgium, Aug. 2004, pp. 229–238.

[10] A. Mäder and D. Staehle, "Uplink Blocking Probabilities in Heterogeneous WCDMA Networks considering Other-Cell Interference," in *Proc. of the Southern African Telecommunication Networks & Applications Conference*, South Western Cape, South Africa, Sep. 2004.

[11] A. Mäder, D. Staehle, and H. Barth, "A Novel Performance Model for the HSDPA with Adaptive Resource Allocation," in *Proc. of the 20th International Teletraffic Congress*, ser. LNCS, L. Mason, T. Drwiega, and J. Yan, Eds., vol. 4516. Ottawa, Canada: Springer, Jun. 2007, pp. 950–961.

[12] A. Mäder, D. Staehle, and C. Gößwein, “Performance of Internet Services over the UMTS Enhanced Uplink,” in *Proc. of NGMAST ’07*, Cardiff, Wales, UK, Sep. 2007, pp. 298–303.

[13] A. Mäder, D. Staehle, T. Liu, and H. Barth, “Feasible Load Regions for different RRM Strategies for the Enhanced Uplink in UMTS Networks,” in *Proc. of the 3rd EuroNGI Workshop on Wireless Systems and Mobility in Next Generation Internet*, ser. LNCS, J. García-Vidal and L. Cerdà-Alabern, Eds., vol. 4369. Sitges, Spain: Springer, Jun. 2006, pp. 213–228.

[14] A. Mäder, D. Staehle, and M. Spahn, “Impact of HSDPA Radio Resource Allocation Schemes on the System Performance of UMTS Networks,” in *Proc. of IEEE VTC Fall ’07*, Baltimore, USA, Oct. 2007.

[15] A. Mäder, D. Staehle, and M. Spahn, “Impact of HSDPA Transmit Power Allocation Schemes on the Performance of UMTS Networks,” in *Proc. of ATNAC ’07*, Christchurch, New Zealand, Dec. 2007, pp. 315–319.

[16] A. Mäder, B. Wagner, T. Hoßfeld, D. Staehle, and H. Barth, “Measurements in a Laboratory UMTS Network with time-varying Loads and different Admission Control Strategies,” in *Proc. of the 4th International Workshop on Internet Performance, Simulation, Monitoring and Measurement*, Salzburg, Austria, Feb. 2006.

[17] F.-U. Andersen, H. de Meer, I. Dedinski, C. Kappler, A. Mäder, J. Oberender, and K. Tutschku, “An Architecture Concept for Mobile P2P File Sharing Services,” in *Proc. of Workshop at Informatik 2004 - Algorithms and Protocols for Efficient Peer-to-Peer Applications*, Ulm, Sep. 2004.

[18] T. Hoßfeld, A. Mäder, K. Tutschku, and F.-U. Andersen, “Time-Discrete Analysis of the Crawling Strategy in an Optimized Mobile P2P Architecture,” in *Proc. of the 2nd EuroNGI Workshop on Wireless and Mobil-*

ity, ser. LNCS, M. Cesana and L. Fratta, Eds., vol. 3883. Como, Italy: Springer, Jul. 2006, pp. 211–225.

[19] T. Hoßfeld, A. Mäder, and D. Staehle, "When Do We Need Rate Control for Dedicated Channels in UMTS?" in *Proc. of IEEE VTC Fall '06*, Melbourne, Australia, May 2006, pp. 425–429.

[20] T. Hoßfeld, A. Mäder, K. Tutschku, P. Tran-Gia, F.-U. Andersen, H. de Meer, and I. Dedinski, "Comparison of Crawling Strategies for an Optimized Mobile P2P Architecture," in *Proc. of the 19th International Teletraffic Congress*, Beijing, China, Sep. 2005.

[21] T. Liu, A. Mäder, and D. Staehle, "A Novel Linear-Programming Based Approach for Near-Optimal Rate Allocation in the UMTS Enhanced Uplink," in *Proc. of IEEE WiMob*, White Plains, New York, USA, Oct. 2007, pp. 1–4.

[22] T. Liu, A. Mäder, and D. Staehle, "Analytical Other-Cell Interference Characterization over HSUPA-Enabled Multi-cell UMTS Networks," in *Proc. of IEEE VTC Fall '07*, Baltimore, MD, USA, Oct. 2007, pp. 96–100.

[23] T. Liu, A. Mäder, D. Staehle, and D. Everitt, "Analytic Modeling of the UMTS Enhanced Uplink in Multi-Cell Environments with Volume-Based Best-Effort Traffic," in *Proc. of IEEE ISCIT '07*, Sydney, Australia, Oct. 2007, pp. 439–444.

[24] J. Oberender, F.-U. Andersen, H. de Meer, I. Dedinski, T. Hoßfeld, C. Kappler, A. Mäder, and K. Tutschku, "Enabling Mobile Peer-to-Peer Networking," in *Proc .of the 1st EuroNGI Workshop on Wireless Systems and Mobility in Next Generation Internet*, ser. LNCS, G. Kotsis and O. Spaniol, Eds., vol. 3427, Dagstuhl, Germany, Jan. 2005, pp. 219–234.

[25] R. Pries, A. Mäder, , and D. Staehle, “A Network Architecture for a Policy-Based Handover Across Heterogeneous Networks,” in *Proc. of OPNETWORK ’06*, Washington D.C., USA, Aug. 2006.

[26] R. Pries, A. Mäder, , and D. Staehle, “Do we need Header Compression for VoIP in Wireless LANs?” in *Proc. of the 12th EUNICE Open European Summer School 2006*, Stuttgart, Germany, Sep. 2006.

[27] D. Staehle and A. Mäder, “An Analytic Approximation of the Uplink Capacity in a UMTS Network with Heterogeneous Traffic,” in *Proc. of the 18th International Teletraffic Congress*, Berlin, Sep. 2003, pp. 81–90.

[28] D. Staehle and A. Mäder, “A Model for Time-Efficient HSDPA Simulations,” in *Proc. of IEEE VTC Fall ’07*, Baltimore, MD, USA, Oct. 2007, pp. 819–823.

[29] D. Staehle and A. Mäder, “An Analytic Model for Deriving the Node-B Transmit Power in Heterogeneous UMTS Networks,” in *Proc. of IEEE VTC Spring ’04*, Milano, Italy, May 2004, pp. 2399–2403.

General References

[30] H. Holma and A. Toskala, Eds., *WCDMA for UMTS*, 3rd ed. John Wiley & Sons, Ltd., Jul. 2004.

[31] 3GPP, “TS 22.340 V7.0.0 IP Multimedia System (IMS) messaging; Stage 1,” Tech. Rep., Dec. 2005.

[32] M. C. Necker and A. Weber, “Impact of Iub Flow Control on HSDPA System Performance,” in *Proc of IEEE PIMRC ’05*, Berlin, Germany, Sep. 2005, pp. 1703–1707.

[33] 3GPP, “TS 25.213 V6.4.0 Spreading and modulation (FDD),” Tech. Rep., Sep. 2005.

[34] Y.-H. Jung and Y. H. Lee, “Scrambling code planning for 3GPP W-CDMA systems,” in *Proc. of IEEE VTC Spring ’01*, Rhodos, Greece, May 2001, pp. 2431–2434.

[35] A. Goldsmith, *Wireless Communications.* Cambridge University Press (UK), 2005.

[36] 3GPP, “TS 23.107 V6.1.0 Quality of Service (QoS) concept and architecture,” 3GPP, Tech. Rep., Mar. 2004.

[37] D. Staehle, “On the Code and Soft Capacity of the UMTS FDD Downlink and the Capacity Increase by using a Secondary Scrambling Code,” in *IEEE PIMRC ’05*, vol. 3, Berlin, Germany, Sep. 2005, pp. 2099–2103.

[38] V. V. Veeravalli and A. Sendonaris, “The Coverage-Capacity Tradeoff in Cellular CDMA Systems,” *IEEE Transactions on Vehicular Technology*, vol. 48, no. 5, pp. 1443–1450, Sep. 1999.

[39] M. H. Ahmed, “Call admission control in wireless networks: A comprehensive survey,” *IEEE Communications Surveys & Tutorials*, vol. 7, no. 1, pp. 50–59, Jul. 2005.

[40] D. Staehle, “Analytical Models for UMTS Radio Network Planning,” Ph.D. dissertation, University of Würzburg, Würzburg, Germany, Nov. 2004.

[41] D. Staehle, K. Leibnitz, K. Heck, B. Schröder, A. Weller, and P. Tran-Gia, “Analytical Characterization of the Soft Handover Gain in UMTS,” in *Proc. of IEEE VTC Fall ’01*, Atlantic City, NJ, USA, Oct. 2001, pp. 291–295.

[42] K. Heck, D. Staehle, and K. Leibnitz, "Diversity Effects on the Soft Handover Gain in UMTS networks," in *Proc. of IEEE VTC Fall '02*, Vancouver, Canada, Sep. 2002, pp. 1269–1273.

[43] 3GPP, "TS 23.009 V6.4.0 Handover procedures," Tech. Rep., Mar. 2006.

[44] J. W. Roberts, "Traffic Theory and the Internet," *IEEE Communications Magazine*, vol. 39, no. 1, pp. 94–99, Jan. 2001.

[45] H.-K. Choi and J. O. Limb, "A Behavioral Model of Web Traffic," in *Proc. of the 7th International Conference on Network Protocols*, Toronto, Canada, Oct. 1999, pp. 327–334.

[46] P. Tran-Gia, D. Staehl, , and K. Leibnitz, "Source Traffic Modeling of Wireless Applications," *AEÜ International Journal of Electronics and Communications*, vol. 55, no. 1, pp. 27–36, Nov. 2001.

[47] R. Kwan, P. H. J. Chong, E. Poutiainen, and M. Rinne, "The effect of code-multiplexing on the high speed downlink packet access (HSDPA) in a WCDMA network," in *Proc. of IEEE WCNC '03*, New Orleans, LA, USA, Mar. 2003, pp. 1728–1732.

[48] A. J. Goldsmith, "Adaptive Coded Modulation for Fading Channels," *IEEE Transactions on Communications*, vol. 46, no. 5, pp. 595–602, May 1998.

[49] T. E. Kolding, F. Frederiksen, and P. E. Mogensen, "Performance Evaluation of Modulation and Coding Schemes Proposed for HSDPA in 3.5G UMTS Networks," in *Proc. of the International Symposium on Wireless personal Multimedia Communications*, Aalborg, Denmark, Sep. 2001, pp. 307–312.

[50] M. Döttling, J. Michael, and B. Raaf, "Hybrid ARQ and Adaptive Modulation and Coding Schemes for High Speed Downlink Packet Access," in

Proc. of IEEE PIMRC '02, vol. 3, Lisboa, Protugal, Sep. 2002, pp. 1073–1077.

[51] 3GPP, "3GPP TS 25.214 V6.7.0 Physical layer procedures (FDD) (Release 6)," 3GPP, Tech. Rep., Sep. 2005.

[52] D. Chase, "Code Combining – A Maximum-Likelihood Decoding Approach for Combining an Arbitrary Number of Noisy Packets," *IEEE Transactions on Communications*, vol. 33, no. 5, pp. 385–393, May 1985.

[53] J.-F. Cheng, "On the coding gain of incremental redundancy over chase combining," in *Proc. of IEEE GLOBECOM '03*, San Francisco, CA, USA, Dec. 2003, pp. 107–112.

[54] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with chase combining and incremental redundancy for HSDPA," in *Proc. of IEEE VTC Fall '01*, Vancouver, Canada, Sep. 2001, pp. 1829–1833.

[55] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. of IEEE ICC '95*, Seattle, WA, USA, Jun. 1995, pp. 331–335.

[56] P. Visnawath, D. N. C. Tse, and R. Laroia, "Opportunistic Beamforming Using Dumb Antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[57] F. Berggren and R. Jäntti, "Multiuser Scheduling over Rayleigh Fading Channels," in *Proc. of IEEE GLOBECOM '03*, San Francisco, CA, USA, Dec. 2003, pp. 158–162.

[58] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, Mar. 1998.

[59] A. Jalali, R. Padovani, and R. Pankaj, “Data throughput of CDMA-HDR: a high efficiency-high data rate personal communication wireless system,” in *Proc. of IEEE VTC Spring ’00*, Tokyo, Japan, May 2000, pp. 1854–1858.

[60] M. Andrews, “Instability of the proportional fair scheduling algorithm for HDR,” *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1422–1426, Sep. 2004.

[61] T. E. Kolding, “Link and system performance aspects of proportional fair scheduling in WCDMA/HSDPA,” in *Proc. of IEEE VTC Fall ’03*, Orlando, FL, USA, Oct 2003, pp. 1717–1722.

[62] T. E. Klein, K. K. Leung, and H. Zheng, “Enhanced Scheduling Algorithms for Improved TCP Performance in Wireless IP Networks,” in *Proc. of IEEE GLOBECOM ’04*, Dallas, TX, USA, Nov. 2004, pp. 2744–2748.

[63] T. E. Kolding, “QoS-Aware Proportional Fair Packet Scheduling with Required Activity Detection,” in *Proc. of VTC Fall ’06*, Montreal, CA, Oct 2006.

[64] G. Aniba and S. Aïssa, “Adaptive proportional fairness for packet scheduling in HSDPA,” in *In Proc. of IEEE GLOBECOM 2004*, Montreál, QC, Canada, Dec. 2004, pp. 7413–7417.

[65] M. Lundevall, B. Olin, J. Olsson, N. Wiberg, S. Wänstedt, J. Eriksson, and F. Eng, “Streaming Applications over HSDPA in Mixed Service Scenarios,” in *Proc. of IEEE VTC Fall ’04*, vol. 2, Los Angeles, CA, USA, Sep 2004, pp. 841–845.

[66] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, “Providing Quality of Service over a Shared Wireless Link,” *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, Feb. 2001.

[67] P. Ameigeiras, J. Wigard, and P. Morgensen, "Performance of the M-LWDF Scheduling Algorithm for Streaming Services in HSDPA," in *Proc. of IEEE VTC Fall '04*, Los Angeles, CA, USA, Sep 2004, pp. 999–1003.

[68] A. D. F. Khattab and K. M. F. Elsayed, "Channel-Quality Dependent Earliest Deadline Due Fair Scheduling Schemes for Wireless Multimedia Networks," in *Proc. of IEEE/ACM MSWiM '04*, Venice, Italy, Oct. 2004, pp. 31–38.

[69] S. Shakkottai and A. L. Stolyar, "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR," in *Proc. of 17th International Teletraffic Congress*, Salvador da Bahia, Brazil, Dec. 2001, pp. 793–804.

[70] K. I. Pedersen, P. E. Mogensen, and T. E. Kolding, "Overview of QoS options for HSDPA," *IEEE Communications Magazine*, vol. 44, no. 7, pp. 100–105, Jul. 2006.

[71] J. Li and S. Sampalli, "QoS-guaranteed wireless packet scheduling for mixed services in HSDPA," in *Proc. of ACM/IEEE MSWiM '06*, Terromolinos, Spain, Oct. 2006, pp. 126–129.

[72] M. Necker, "A comparison of scheduling mechanisms for service class differentiation in HSDPA networks," *AEÜ International Journal of Electronics and Communications*, vol. 60, no. 2, pp. 136–141, Feb. 2006.

[73] 3GPP, "3GPP TS 25.321 V6.6.0 Medium Access Control (MAC) protocol specification," 3GPP, Tech. Rep., Sep. 2005.

[74] T. Henttonen and T. Chen, "Power Headroom Measurements for E-TFC Elimination/Selection in HSUPA," in *Proc. of IEEE VTC Spring '06*, Melbourne, Australia, May 2006, pp. 415–419.

[75] W. Xiao, R. Ratasuk, A. Ghosh, and R. Love, “Scheduling and Resource Allocation of Enhanced Uplink for 3GPP W-CDMA,” in *Proc. of IEEE PIMRC ’05*, Berlin, Germany, Sep. 2005, pp. 1905–1909.

[76] K. Kumaran and L. Qian, “Uplink Scheduling in CDMA Packet-Data Systems,” in *Proc. of IEEE INFOCOM ’03*, San Francisco, CA, USA, Mar. 2003, pp. 292–300.

[77] C. Rosa, J. Outes, K. Dimou, T. B. Sørensen, J. Wigard, F. Fredrikson, and P. E. Mogensen, “Performance of fast node B scheduling and L1 HARQ schemes in WCDMA uplink packet access,” in *Proc. of IEEE VTC Spring ’04*, Milan, Italy, May 2004, pp. 1635–1639.

[78] C. Rosa, J. Outes, T. B. Sørensen, J. Wigard, and P. E. Mogensen, “Combined Time and Code Division Scheduling for Enhanced Uplink Packet Access in WCDMA,” in *Proc. of IEEE VTC Fall ’04*, Los Angeles, USA, Sep. 2004, pp. 851–855.

[79] J. Voigt and K. Pannhorst, “Optimizations on Scheduling Strategies for Enhanced Uplink on WCDMA,” in *Proc. of IEEE VTC Spring ’07*, Dublin, Ireland, Apr. 2007, pp. 1172–1176.

[80] E. Damosso and L. M. Correia, Eds., *Digital Mobile Radio Towards Future Generation Systems; COST 231 Final Report*, 1998.

[81] 3GPP, “3GPP TS 34.121-1 User Equipment (UE) conformance specification; Radio transmission and reception (FDD); Part 1,” 3GPP, Tech. Rep., Dec. 2006.

[82] A. Das, F. Khan, A. Sampath, and H.-J. Su, “Performance of hybrid ARQ for high speed downlink packet access in UMTS,” in *Proc. of IEEE VTC Fall ’01*, Atlantic City, NJ, USA, Oct. 2001, pp. 2133–2137.

[83] F. Frederiksen and T. E. Kolding, "Performance and modeling of WCDMA/HSDPA transmission/H-ARQ schemes," in *Proc. of IEEE VTC Fall '02*, Vancouver, Canada, Sep. 2002, pp. 472–476.

[84] S. Bliudze, N. Billy, and D. Krob, "On optimal Hybrid ARQ control schemes for HSDPA with 16QAM," in *Proc. of IEEE WiMob '05*, vol. 1, Montréal, QC, Canada, Aug. 2005, pp. 121–127.

[85] M. Döttling, T. Grundler, and A. Seeger, "Incremental Redundancy and Bit-Mapping Techniques for High Speed Downlink Packet Access," in *Proc. of IEEE GLOBECOM '03*, vol. 2, San Francisco, CA, USA, Dec. 2003, pp. 908–912.

[86] F. Brouwer, I. de Bruin, J. de Bruin, N. Souto, F. Cercas, and A. Correia, "Usage of Link-Level Performance Indicators for HSDPA Network-Level Simulations in E-UMTS," in *Proc. of IEEE ISSSTA '04*, Sidney, Australia, Aug. 2004, pp. 844–848.

[87] P. Ameigeiras, J. Wigard, and P. Morgensen, "Performance of Packet Scheduling Methods with Different Degree of Fairness in HSDPA," in *Proc. of IEEE VTC Fall '04*, Los Angeles, CA, USA, Sep 2004, pp. 860–864.

[88] EURANE, "EURANE website," http://www.ti-wmc.nl/eurane/. [Online]. Available: http://www.ti-wmc.nl/eurane/

[89] H. Buddendick, G. Wölfle, S. Burger, and P. Wertz, "Simulator for Performance Analysis in UMTS FDD Networks with HSDPA," in *Proc. of IEEE PIMRC '04*, vol. 3, Barcelona, Spain, Sep. 2004, pp. 2251–2255.

[90] A. Furuskär, S. Parkvall, M. Persson, and M. Samuelsson, "Performance of WCDMA high speed packet data," in *Proc of IEEE VTC Spring '02*, Birmingham, UK, May 2002, pp. 1116–1120.

[91] T. E. Kolding, F. Frederiksen, and P. E. Mogensen, "Performance Aspects of WCDMA Systems with High Speed Downlink Packet Access (HSDPA)," in *Proc. of IEEE VTC Fall 02*, vol. 1, Vancouver, Canada, Sep. 2002, pp. 477–481.

[92] K. Hiltunen, M. Lundevall, and S. Magnusson, "Performance of Link Admission Control in a WCDMA System with HS-DSCH and Mixed Services," in *Proc. of IEEE PIMRC '04*, vol. 2, Barcelona, Spain, Sep. 2004, pp. 1178–1182.

[93] R. Litjens, J. van den Berg, and M. Fleuren, "Spatial Traffic Heterogeneity in HSDPA Networks and its Impact on Network Planning," in *Proc. of the 19th International Teletraffic Congress*, Bejing, China, Aug. 2005, pp. 653–666.

[94] K. I. Pedersen, T. F. Lootsma, M. Støttrup, F. Frederiksen, T. E. Kolding, and P. E. Mogensen, "Network Performance of Mixed Traffic on High Speed Downlink Packet Access and Dedicated Channels in WCDMA," in *Proc. of IEEE VTC Fall '04*, Milan, Italy, Sep. 2004, pp. 2296–4500.

[95] U. Türke, M. Koonert, R. Schelb, and C. Görg, "HSDPA performance analysis in UMTS radio network planning simulations," in *Proc. of IEEE VTC Spring '04*, Milan, Italy, May 2004, pp. 2555–2559.

[96] H. van den Berg, R. Litjens, and J. Laverman, "HSDPA flow level performance: the impact of key system and traffic aspects," in *Proc. of IEEE/ACM MSWiM '04*. Venice, Italy: ACM, Oct. 2004, pp. 283–292.

[97] J. W. Roberts, "A survey on statistical bandwidth sharing," *Computer Networks*, vol. 45, no. 3, pp. 319–332, 2004.

[98] K. I. Pedersen and P. H. Michaelsen, "Algorithms and Performance Results for Dynamic HSDPA Resource Allocation," in *Proc. of IEEE VTC Fall '06*, Montréal, QC, Canada, Sep. 2006, pp. 1–5.

[99] S. Brueck, E. Jugl, H.-J. Kettschau, M. Link, J. Mueckenheim, and A. Zaporozhets, "Radio Resource Management in HSDPA and HSUPA," *Bell Labs Technical Journal*, vol. 11, no. 4, pp. 151–167, Mar. 2007.

[100] 3GPP, "TS 25.433 V7.5.0 UTRAN Iub interface Node B application part (NBAP) signalling," 3GPP, Tech. Rep., 2007.

[101] 3GPP, "TS 25.331 V7.5.0 Radio Resource Control (RRC) Protocol Specification," 3GPP, Tech. Rep., 2007.

[102] H. Holma and A. Toskala, Eds., *HSDPA/HSUPA for UMTS. High Speed Radio Access for Mobile Communications*, 1st ed. Wiley & Sons, Jun. 2006.

[103] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 3, pp. 636–647, Jun. 2005.

[104] H. J. Kushner and P. A. Whiting, "Convergence of Proportional-Fair Sharing Algorithms Under General Conditions," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1250–1259, Jul. 2004, konvergenz von PF, gezeigt mittels ODE.

[105] J. D. Little, "A proof of the queueing formula $L = \lambda W$," *Operations Research*, vol. 9, no. 3, pp. 383–387, 1961.

[106] J. M. Holtzman, "CDMA Forward Link Waterfilling Power Control," in *Proc. of IEEE VTC Spring '00*, Tokyo, Japan, May 2000, pp. 1663–1667.

[107] A. Haider, R. Harris, and H. Sirisena, "Simulation-Based Performance Analysis of HSDPA for UMTS Networks," in *Proc. of ATNAC '06*, Melbourne, Australia, Dec. 2006.

[108] R. Litjens and R. J. Boucherie, "Performance Analysis of Fair Channel Sharing Policies in an Integrated Cellular Voice/Data Network," *Telecommunication Systems*, vol. 19, no. 2, pp. 147–186, Feb. 2002.

[109] R. Litjens and R. J. Boucherie, “Elastic calls in an integrated services network: the greater the call size variability the better the QoS,” *Performance Evaluation*, vol. 52, no. 4, pp. 193–220, May 2003.

[110] J. Kaufman, “Blocking in a Shared Resource Environment,” *IEEE Transactions on Communications*, vol. 29, no. 10, pp. 1474–1481, Oct. 1981.

[111] G. Pujolle, Ed., *Performance of Data Communication Systems and their Applications*. Amsterdam: Elsevier Science/North Holland, 1981.

[112] R. Fortet and C. Grandjean, “Congestion in a loss system when some calls want several devices simultaneously,” *Electrical Communication*, vol. 39, no. 4, pp. 513–526, 1964.

[113] R. Núñez-Queija and H.-P. Tan, “Location-based admission control for differentiated services in 3G cellular networks,” in *Proc. of IEEE/ACM MSWiM ’06*, Terromolinos, Spain, Oct. 2006, pp. 322–329.

[114] R. W. Wolff, “Poisson Arrivals See Time Averages,” *Operations Research*, vol. 30, no. 2, pp. 223–231, Mar. 1982.

[115] J. Charzinski, “Fun Factor Dimensioning for Elastic Traffic,” in *Proc. of the ITC Specialist Seminar on Internet Traffic Measurement, Modeling and Management*, Sep. 2000.

[116] H. Holma and J. Laakso, “Uplink Admission Control and Soft Capacity with MUD in CDMA,” in *Proc. of IEEE VTC Fall ’99*, vol. 1, Amsterdam, The Netherlands, Sep. 1999, pp. 431–435.

[117] J. Zander, “Performance of Optimum Transmitter Power Control in Cellular Radio Systems,” *IEEE Transactions on Vehicular Technologies*, vol. 41, no. 1, pp. 57–62, Feb. 1992.

[118] P. Hande, S. Rangan, and M. Chiang, “Distributed Uplink Power Control for Optimal SIR Assignment in Cellular Data Networks,” in *Proc. of IEEE INFOCOM ’06*, Barcelona, Spain, Apr. 2006, pp. 1–13.

[119] E. G. Lundin, F. Gunnarsson, and F. Gustafsson, "Robust Uplink Resource Allocation in CDMA Cellular Radio Systems," in *Proc. of Joint IEEE Conference on decision and control (CDC) and European Control Conference (ECC)*, Sevilla, Spain, Dec. 2005, pp. 1848–1853.

[120] T. Javidi, "Decentralized rate assignments in a multi-sector CDMA network," in *Proc. of IEEE GLOBECOM '03*, San Francisco, CA, USA, Dec. 2003, pp. 65–69.

[121] K. Leibnitz, "Analytical Modelling of Power Control and its Impact on Wideband CDMA Capacity Planning," Ph.D. dissertation, University of Würzburg, Jan. 2003.

[122] H. Boche and M. Wiczanowski, "Optimal Scheduling for High Speed Uplink Packet Access," in *Proc. of IEEE VTC Spring '04*, Milan, Italy, May 2004.

[123] A. M. Viterbi and A. J. Viterbi, "Erlang Capacity of a Power Controlled CDMA System," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 6, pp. 892–900, Aug. 1993.

[124] J. Evans and D. Everitt, "On the Teletraffic Capacity of CDMA Cellular Networks," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, pp. 153–165, Jan. 1999.

[125] E. Altman, "Capacity of Multi-Service Cellular Networks with Transmission-Rate Control: A Queueing Analysis," in *Proc. of ACM MobiCom '02*, Atlanta, GA, USA, Sep. 2002, pp. 205–214.

[126] N. Hedge and E. Altman, "Capacity of multiservice WCDMA Networks with variable GoS," *ACM Wireless Networks*, vol. 12, no. 2, pp. 241–253, Mar. 2006.

[127] G. Fodor and M. Telek, “Performance Analysis of the Uplink of a CDMA Cell Supporting Elastic Services,” in *Proc. of IFIP NETWORKING ’05*, Waterloo, Canada, May 2005, pp. 205–216.

[128] G. Fodor, M. Telek, and L. Badia, “On the Tradeoff Between Blocking and Dropping Probabilities in CDMA Networks Supporting Elastic Services,” in *Proc. of IFIP Networking ’06*, Coimbra, Portugal, May 2006, pp. 954–965.

[129] T. Liu and D. Everitt, “Analytical Approximation of Other-cell Interference in the Uplink of CDMA Cellular Systems,” in *Proc. of IEEE VTC Spring ’06*, Melbourne, Australia, May 2006, pp. 693–697.

[130] V. V. Veeravalli, A. Sendonaris, and N. Jain, “CDMA coverage, capacity and pole capacity,” in *Proc. of IEEE VTC Spring ’97*, Phoenix, AZ, USA, May 1997, pp. 1450–1454.

[131] 3GPP, “TS 25.309 V6.4.0 FDD enhanced uplink; Overall description,” 3GPP, Tech. Rep., Jun. 2005.

[132] J. Mo and J. Walrand, “Fair end-to-end window-based congestion control,” *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, 2000.

[133] K. Sipilä, J. Laiho-Steffens, A. Wacker, and M. Jäsberg, “Modeling the Impact of the Fast Power Control on the WCDMA Uplink,” in *Proc. of IEEE VTC Spring ’99*, Housten, TX, USA, May 1999, pp. 1266–1270.

[134] L. F. Fenton, “The Sum of Log-Normal Probability Distributions in Scatter Transmission Systems,” *IRE Transactions on Communication Systems*, vol. 8, no. 1, pp. 57–67, Mar. 1960.

[135] D. Staehle, K. Leibnitz, K. Heck, B. Schröder, A. Weller, and P. Tran-Gia, “Approximating the Othercell Interference Distribution in Inhomogeneous UMTS Networks,” in *Proc. of IEEE VTC Spring ’02*, vol. 4, Birmingham, AL, USA, May 2002, pp. 1640–1644.

ISSN 1432-8801