A Flow-Level Simulation Framework for HSDPA-enabled UMTS Networks

Andreas Mäder
University of Wuerzburg, Dept. of Distributed Systems
Würzburg, Germany
maeder@informatik.uni-wuerzburg.de

Dirk Staehle
University of Wuerzburg, Dept. of Distributed Systems
Würzburg, Germany
staehle@informatik.uni-wuerzburg.de

ABSTRACT

The High Speed Downlink Packet Access is designed for the efficient transport of best-effort traffic in UMTS networks. HSDPA uses fast scheduling with adaptive modulation and coding for the rapid adaptation of the instantaneous channel bandwidth to the channel quality. A simulator for such a system has to model these mechanism properly, which is classically done with help of link-level simulations. Additionally, traffic-dynamics on flow-level and radio resource management have a significant influence on the system performance, but a statistical sound evaluation requires long simulation runs, which prohibits an exact simulation on a small time-scale. Our contribution is a novel flow-level simulation framework, which captures on the one hand the impact of the physical layer, and enables on the other hand the efficient simulation of large scenarios over a long time. The framework uses analytical methods to calculate user bandwidths for different scheduling types and propagation environments as well as the required transmit powers and code resources for arbitrary radio resource management schemes.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Modeling techniques

General Terms

Performance, Design

Keywords

simulation technique, flow-level simulation, UMTS, HSDPA

1. INTRODUCTION

Mobile network operators continue to deploy the High Speed Downlink Packet Access (HSDPA) service in their existing Universal Mobile Telecommunication System (UMTS) networks. From the users perspective, the HSDPA offers high bit rates (promised are up to 14.4 Mbps) and low latency. From operators perspective, the HSDPA is hoped to play a key role for the much longed for breakthrough of high quality mobile data services. From a technical perspective, HSDPA introduces a new paradigm to UMTS: Instead of adapting the transmit power to the radio channel condition in order to ensure constant link quality, HSDPA adapts the link quality to the radio channel conditions. This enables a more efficient use of scarce resources like transmit power, code resources and also hardware resources.

The basic principle of the HSDPA is to adapt the link to the radio channel condition with help of adaptive modulation and coding (AMC). For this reason, it employs a shared channel, the High Speed Downlink Shared channel (HS-DSCH), which is used by all HSDPA users. By using a shared channel, radio resources are occupied only if a transmission occurs, which enables a more efficient transport of bursty traffic. In each transport time interval (TTI), the scheduler located in the NodeB decides about the users to be scheduled and about their data rate. The scheduling decision can be either on behalf of channel quality indicator (CQI) reports from the User Equipments (UE) to enable opportunistic scheduling schemes which use the air interface more efficiently, or simple non-opportunistic schemes like Round-Robin can be used which shares the resources time-fair among the users. The rate selection in the NodeB is done according to a direct relation between CQI and the transport format resource combination (TFRC), which describes the number of information bits per TTI, the modulation scheme and following from that the code rate.

In a system where HSDPA and Dedicated Channels (DCH) coexist within one frequency spectrum, both channel types have to share the radio resources, which are transmit power and code resources. This means that both traffic types can influence each other significantly [13, 15] not only by resource preemption at the same NodeB, but also by interference in surrounding cells.

The performance evaluation of such a system is a difficult task due to its complexity and different time-scales involved. On a small time-scale, the HSDPA user bandwidth may change every 2 ms, and the DCH power control takes place every slot, i.e. 0.67 ms. On a larger time-scale, the traffic dynamics influence the radio-resource situation of the system and have to be captured properly. Also, the effects of admission control and resource reservation have to be considered. Additionally, the size of the studied network scenario must be large enough to get a realistic view on the impact of other-cell interference.

Figure 1: Schematic view of the HS-DSCH

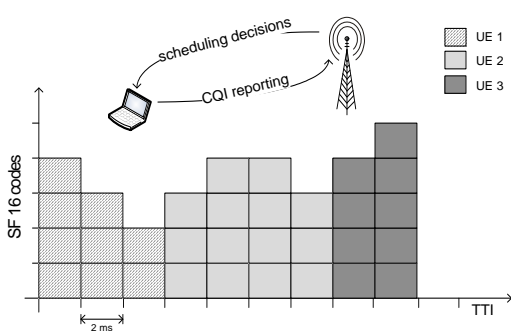**Figure 2: RRM schemes for HSDPA/DCH**

Because of this complexity, simulations are mostly the tool of choice for performance evaluation studies. However, it is difficult to build a simulation which on the one hand considers the physical layer with its effects on small time-scales properly, and on the other hand considers effects on large time-scales like the impact of traffic dynamics. In the literature, one can therefore mostly find papers which focus on a particular effect in an otherwise fixed scenario.

Our contribution is a simulation framework which models the *impact* of small time-scale mechanisms, but is on a higher abstraction level which also enables the effective simulation of large scenarios. This is achieved by analytical models for the physical layer, which are then used by a flow-level simulation covering long-term traffic dynamics.

In the next section, we motivate this work and give an overview of existing simulation tools as far as they are known to the authors. Section 3 introduces the basic principles of our simulation framework. In Sec. 4, radio resource sharing as well as the transmit power and interference model is explained. The HSDPA bandwidth model is introduced in Sec. 5, with a special focus on different scheduling disciplines including a validation by means of detailed link-trace simulations. In Sec. 6 we conduct simulation studies in order to inspect the accuracy of the approximation model for the user throughput in a static state that the system assumes for a finite period of time. Finally, we conclude our work in Sec. 7.

2. MOTIVATION AND RELATED WORK

The performance of a combined HSDPA/UMTS-system is influenced by several key factors which we describe in this section. We focus on the HSDPA, since models for the DCH are well known in the literature with several assumptions regarding power control, code limitations etc. [17, 18]. HSDPA uses the HS-DSCH as shared transport medium for all users within one sector or cell. The HS-DSCH shares two radio resources with the DCH users: Orthogonal Variable Spreading Factor (OVSF) channelization codes and transmit power. In contrast to a DCH connection, which uses normally only one code, the HS-DSCH may utilize up to 15 codes with spreading factor (SF) 16 in parallel for the transmission to one user.

Every TTI, corresponding to 2 ms, the controlling NodeB selects a UE for transmission according to the implemented scheduling discipline. The most common scheduling disciplines in this context are the simple Round-Robin scheduler, the Proportional-Fair and the MaxRate or MaxTBS sched-
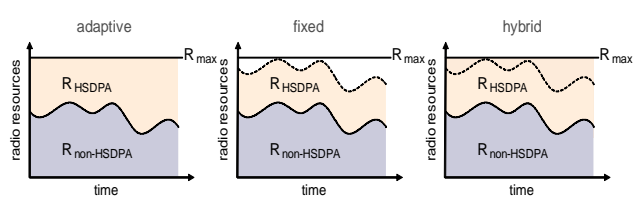
uler. The latter two are channel-aware schedulers. Their decision process considers the current channel quality, which is signaled to the NodeB in a feedback loop as CQI. The CQI is a discretization of the signal-to-interference-ratio (SIR) at the UE and defines a direct relationship to the Transport Block Size (TBS) defined in [1], which is the maximum number of bits that can be transmitted within one TTI, the number of required parallel codes as shown in Fig. 1, the modulation scheme and the coding rate, i.e. the ratio of information bits to total number of bits. Consequently, the number of available codes directly restricts the maximum possible TBS.

The performance of the HSDPA is directly or indirectly influenced by the following factors: Code resources, transmit power, scheduling discipline, number of users which share the HS-DSCH, interference situation, UE category and multipath propagation profile. The code resources and the transmit power are radio resources which can be influenced by the implemented radio resource management (RRM) scheme in the NodeB. Here, different approaches exists from fast adaptation to static reservation with and without preemption for DCH users or hybrid schemes as illustrated in Fig. 2 [8, 13]. However, what they have in common is that the performance of the HSDPA is also influenced by the amount of radio resources required by the DCH users, so the impact of these schemes make long-term simulations necessary since otherwise the traffic dynamics, e.g. if a flow with a large data volume gets only a small bandwidth, cannot be captured properly. Another example is the impact of different transmit power allocation schemes, which mainly affect the performance indirectly by the amount of generated interference: here it is important to take the traffic dynamics in surrounding cells into consideration. Also, performance metrics like blocking probabilities for HSDPA and DCH users require long simulation runs in order to get statistically valid results. So, these are factors which influence the performance on a large time scale.

The instantaneous CQI of a user depends on the current interference situation, the HS-DSCH transmit power and the multi-path channel condition. These are factors which change every 2 ms and are often modeled with link-layer simulations. However, the small time-scale makes this approach impracticable for large scenarios and long simulation runs. For power-controlled DCH-connections which are designed to keep the channel bandwidth at a certain QoS-target, well-known analytical models exist which provide good approximations of the resource requirements. For the HSDPA, such approximative models do not exist yet.

In the literature, an often used approach to speed up simulation times 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. It can be said that this is an attribute common to all 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 on the base of NS-2 [5]. The EURANE simulator uses pre-generated physical layer traces which provide block error rate (BLER) and current bandwidth in each TTI. Further examples for trace-based simulators can be found in [4, 6, 10, 14].

Other approaches [11, 15, 20] 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, 12] for Round-Robin scheduling.

Flow-level simulation tools as proposed in this work are to the best knowledge of the authors not described in the literature yet. However, in [21], 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. FLOW-LEVEL SIMULATION FRAMEWORK

The framework provides the means to build a discrete-event flow-level simulation. The term "flow-level" refers to the modeling of an ongoing data transport 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 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 [16]. An import 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.

The advantage of the flow-level concept is that it is sufficient to know the *mean* bandwidth a flow gets between two events. In the context of HSDPA simulations, this avoids the explicit calculation of TBS values every 2 ms. Summarizing, the basic idea is to find a balance between number of events which have to be processed in the total simulation time and accuracy.

An event is characterized by its type and its event time t_e. We denote the time between two events with Δt_e. In the simulation framework, events are generated if flows are arriving into or departing from the system, or if it is required to recalculate radio resource requirements, interference or HS-
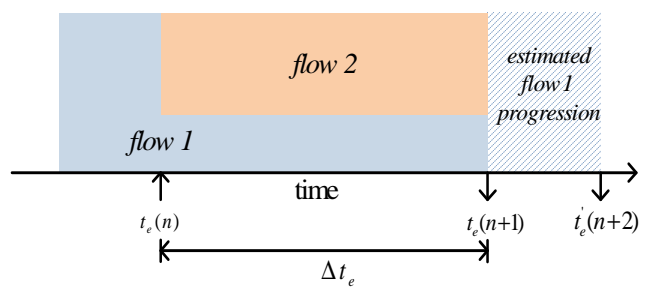

Figure 3: Flow-level simulation concept

DPA 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 changes it's location significantly, or if Radio Resource Management (RRM) 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 different in the simulation. 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 bandwidth they can utilize. So at each event, the transmitted 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 on base of the new bandwidths. Fig. 3 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 $t'_e(n+2)$.

In the following sections, we introduce the RRM and interference model and explain how the mean bandwidth of the HSDPA users is calculated.

4. HSDPA RESOURCES: CODES AND TRANSMIT POWER

The radio resources shared by DCH and HSDPA users are transmit power and channelization codes. The controlling NodeB is responsible for enforcing an RRM scheme. In the following, $C_{x,h}$ denotes the number of SF 16 codes and $T_{x,h}$ the transmit power available for the HS-DSCH. If adaptive resource allocation without reservation is used, the radio resources available for the HS-DSCH are

$$
\begin{aligned}
C_{x,h} &= C_x^* - C_{x,d} - C_{x,c} \quad \text{and} \\
T_{x,h} &= T_x^* - T_{x,d} - T_{x,c},
\end{aligned}
\tag{1}
$$

where C_x^* and T_x^* are the maximum number of codes and maximum transmit power. $C_{x,c}$ and $T_{x,c}$ are reserved for signaling and pilot channels. $C_{x,d}$ and $T_{x,d}$ are the resources used for dedicated channels.

A UMTS network is defined as a set $\mathcal{L}$ of NodeBs with associated UEs, $\mathcal{M}_x$. A DCH user k corresponds to a radio access bearer at NodeB $x \in \mathcal{L}$ that is defined by its channelization code, the information bit rate R_k, and a target bit-energy-to-noise ratio (E_b/N_0) ε_k^*. Let $C_{x,k}$ be the number of SF16 codes user k occupies, i.e. a SF128 code leads to $C_{x,k} = 1/8$ and a SF8 code to $C_{x,k} = 2$. Altogether the

DCH users occupy

$$C_{x,d} = \left\lceil \sum_{k \in \mathcal{M}_x} C_{x,k} \right\rceil \quad (2)$$

codes. The transmit power requirement from NodeB x for a DCH user k is

$$T_{x,k} = \frac{\varepsilon_k^* \cdot R_k}{W} \cdot \left(\frac{W \cdot N_0 + I_k^{oth}}{d_{x,k}} + \alpha \cdot T_x \right), \quad (3)$$

where W denotes the system bandwidth of 3.84 Mcps, N_0 denotes the thermal noise spectral density of $-174\,\text{dBm/Hz}$, α is the orthogonality factor, and $d_{x,k}$ is the average propagation gain from x to k. The other-cell interference I_k^{oth} is the total power received at mobile k from the surrounding NodeBs, and according to the most commonly used interference model [7] the own-cell interference is approximated by the orthogonality factor:

$$I_k^{oth} = \sum_{y \in \mathcal{L} \setminus x} T_y \cdot d_{y,k} \quad \text{and} \quad I_k^{own} = \alpha \cdot T_x \cdot d_{x,k} \quad (4)$$

We introduce now the boolean variable $\delta_{x,h}$ that indicates whether NodeB x serves at least one HSDPA user in order to enable off-switching of the HS-DSCH transmit power if no user is active. Furthermore, we follow [18] in defining the load of cell x with respect to cell y as

$$\eta_{x,y} = \sum_{k \in \mathcal{M}_x} \omega_{k,y}$$

$$\text{with } \omega_{k,y} = \frac{\varepsilon_k^* \cdot R_k}{W} \cdot \begin{cases} \alpha & \text{, if } \mathcal{L}(k) = y \\ \frac{d_{y,k}}{d_{\mathcal{L}(k),k}} & \text{, if } \mathcal{L}(k) \neq y. \end{cases} \quad (5)$$

Using these variables we are able to formulate a compact equation of the total NodeB transmit power:

$$T_x = \delta_{x,h} \cdot T_x^* + (1 - \delta_{x,h}) \cdot \left(T_{x,c} + \sum_{y \in \mathcal{L}} \eta_{x,y} \cdot T_y \right) \quad (6)$$

where the DCH transmit power is given as

$$T_{x,d} = \sum_{y \in \mathcal{L}} \eta_{x,y} \cdot T_y. \quad (7)$$

In these equations, we neglected the thermal noise since it is by magnitudes smaller than the multiple access interference for a reasonable cell layout. The introduction of the vector

$$V[x] = \delta_{x,h} \cdot T_x^* + (1 - \delta_{x,h}) \cdot T_{x,c} \quad (8)$$

and the matrix

$$M[x,y] = (1 - \delta_{x,h}) \cdot \eta_{x,y} \quad (9)$$

leads to the following matrix equation:

$$T = V + M \cdot T \Leftrightarrow T = (I - M)^{-1} \cdot V \quad (10)$$

The matrix I is the identity matrix, and T is the vector with the cell transmit powers T_x. The DCH and HSDPA transmit powers are then calculated with Eq. (7) and Eq. (1).

Admission control for DCHs is performed under the condition that the HSDPA power is switched on in every cell, since an idealistic admission control should accept a new user whenever the total DCH transmit powers including the new user and with active HSDPA in every cell stay below the admission control threshold. Thus, even in case that the HSDPA power is switched off, above equations with $\delta_{x,h} = 1$ 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 $\delta_{x,h}$ for computing the actual cell transmit powers.

5. HSDPA BANDWIDTH MODEL

In the following we describe a model for the calculation of the mean bandwidth or throughput of an HSDPA user depending on the other-to-own-cell interference ratio, the HS-DSCH transmit power, the number of available codes and the scheduling discipline.

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

$$\text{CQI} = \max \left(0, \min \left(30, \left\lfloor \frac{\text{SIR[dB]}}{1.02} + 16.62 \right\rfloor \right) \right). \quad (11)$$

In the following we first explain our SIR model including multi-path propagation, and afterwards we describe how to determine the resulting CQI and TBS. A validation of the model can be found in [19]. Then we describe how to consider different scheduling mechanisms to obtain the mean bandwidth over a certain time span.

5.1 CQI and TBS Distribution

The propagation channel from NodeB x to mobile k consists of a set $\mathcal{P}_{x,k}$ of paths p with associated average relative received power m_{β_p} and delay τ_p, as e.g. defined by the 3GPP [2] for evaluating the HSDPA performance. The average relative received powers are normalized, i.e. their sum equals one. Furthermore, let $d_{x,k}$ be the average propagation gain from NodeB x to mobile k. Then, the power $S_{x,k,p}$ mobile k receives on path p is

$$S_{x,k,p} = T_x \cdot d_{x,k} \cdot \beta_p \quad (12)$$

where β_p is a random variable for the instantaneous relative propagation loss of multi-path component p. If every multi-path component experiences independent Rayleigh fading, β_p is exponentially distributed with mean m_{β_p}. Assuming that the RAKE receiver has a finger on every multipath component and uses perfect Maximal Ratio Combining, the receiver achieves a SIR of

$$\gamma_k = \frac{T_{x,h}}{T_x} \cdot \sum_{f \in \mathcal{P}_{x,k}} \frac{\beta_f}{\left(\sum_{y \in \mathcal{L} \setminus x} \frac{T_y \cdot d_{y,k}}{T_x \cdot d_{x,k}} \cdot B_{y,k} \right) + B_{x,k,f}} \quad (13)$$

$$\text{with } B_{y,k} = \sum_{p \in \mathcal{P}_{y,k}} \beta_p \text{ and } B_{x,k,f} = \sum_{p \in \mathcal{P}_{x,k} \setminus f} \beta_p.$$

In Eq. (13) every finger experiences the same other-cell interference as we assume slowly varying channel conditions. Thermal noise is neglected since in well-designed networks, it is by magnitudes less than the multiple access interference.

Let us introduce the variable $\Delta_x = T_{x,h}/T_x$ for the ratio of HSDPA power to total cell power, and the variable γ_k for the SIR achieved due to the total cell power, i.e.

$$\gamma_{k,h} = \Delta_x \cdot \gamma_k. \quad (14)$$

For the rest of this paper, we refer to the variable γ_k as the normalized SIR (nSIR), and to the variable Δ_x as the HSDPA power ratio (PR) .
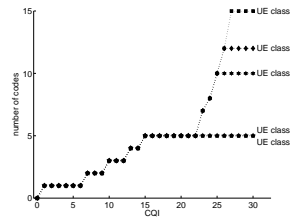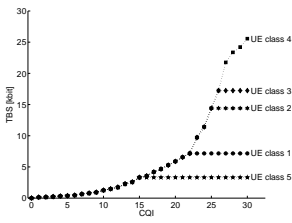
Figure 4: CQI vs. TBS   Figure 5: CQI vs. codes

The distribution function of the CQI follows from the distribution function of γ since the PR Δ means only an offset in the decibel scale. A direct calculation of the distribution function of γ, or even of its mean is numerically intractable. Accordingly, our approach is to estimate the type of distribution and approximate the mean and standard deviation. Therefore, we assume that mean and variance of γ are functions of the ratio Σ of average other-cell received power to average own-cell received power:

$$\Sigma_k = \sum_{y \in \mathcal{L} \setminus x} \sigma_{k,y} \text{ with } \sigma_{k,y} = (T_y \cdot d_{y,k})/(T_x \cdot d_{x,k}) \quad (15)$$

This is of course an approximation, since γ depends not only on Σ_k but on the received power ratio $\sigma_{k,y}$ of every non-serving NodeB.

In [19] it is shown that both the mean and the standard deviation of the normalized SIR are well approximated by a four-parametric Weibull-function as a function of Σ:

$$f_{a,b,c,d}(\Sigma) = a - b \cdot e^{-c \cdot \Sigma^d}, \quad (16)$$

such that $E[\gamma] = f_E(\Sigma)$ and $\text{STD}[\gamma] = f_{\text{STD}}(\Sigma)$. The parameters of the Weibull-function depend on the used propagation profile. The mean and the standard deviation are then used to fit the nSIR with an appropriate distribution. In [19], the normal, the inverse Gaussian and the lognormal distribution lead to good results, again depending on the propagation profile.

Discretizing the nSIR-distribution according to Eq. (11) results in the CQI distribution $p_{CQI,k}(q)$, which now depends on the other-to-own interference ratio Σ and the HSDPA power ratio Δ. While Σ determines the shape of the CQI distribution, the HSDPA power ratio Δ determines the location of the distribution on the range of possible CQIs. Finally, the TBS distribution results from mapping TBS values that are unfeasible due to a lack of codes or restrictions by the UE class to the maximal feasible TBS. We define the set of allowed TBS $\mathcal{V}_k = \{v | C(v) \leq C_{x,h} \wedge v \text{ supported by class of } k\}$ and denote the highest allowed TBS with v^*. Then, the TBS distribution follows as

$$p_{TBS,k}(v) = \begin{cases} p_{CQI,k}(q(v)), & \text{if} \quad v < v^* \\ \sum_{q=q(v^*)}^{30} p_{CQI,k}(q), & \text{else} \end{cases} \quad (17)$$

where $q(v)$ is the CQI value corresponding to v. We denote the CDF as $P_{TBS}(v) = \sum_{v' \leq v} p_{TBS}(v')$. The relation between CQI and TBS and CQI and codes is clarified by Figures 4 and 5 for different UE classes[1] .

---

[1]Here, "class" refers to the corresponding table in [1]. Class 1 comprises UE categories 1 to 6, class 2 categories 7 and 8, class 3 category 9, class 4 category 10 and class 5 categories 11 and 12.

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 leads to an optimal 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. For all three disciplines, the expected transmitted data volume per TTI is given by

$$E[V_k] = \frac{1}{1 + p_{err}} \cdot \sum_{v \in \mathcal{V}_k} v \cdot p_{TBS,k}(v) \cdot p_{s,k}(v), \quad (18)$$

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 bandwidth approximation.

5.2.1 Round-Robin Scheduling

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

$$p_{s,k} = \frac{1}{|\mathcal{M}|} \quad (19)$$

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

$$E[V_k] = \frac{E[v_k]}{|\mathcal{M}| \cdot (1 + p_{err})}, \quad (20)$$

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 speeding up the computation times by storing the possible mean TBS for all values of Σ, Δ, and $C_{x,h}$ in a database.

5.2.2 MaxTBS Scheduling

With *MaxTBS* 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. 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 [11], 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. For this reason, let us first define the Bernoulli random variable $\xi_k(v)$ as

$$\xi_k(v) = \begin{cases} 1 & \text{if } V_k = v \text{ under condition } V_k \leq v, \\ 0 & \text{else.} \end{cases} \quad (21)$$

The probability $P(\xi_k(v) = X)$ is then given by

$$P(\xi_k(v) = X) = \begin{cases} 1 - \frac{p_{TBS}(v)}{P_{TBS}(v)} & \text{if } X = 0, \\ \frac{p_{TBS}(v)}{P_{TBS}(v)} & \text{else.} \end{cases} \quad (22)$$

This enables us to calculate the probability that n users have the same TBS if user k has TBS v by defining the random variable $N_{k,v} = \sum_{m \in \mathcal{M} \setminus k} \xi_m(v)$. The distribution of $N_{k,v}$ is the convolution of the distributions of $\xi_k(v)$ and is defined as

$$p_{N_{k,v}}(n) = \bigcircledast_{m \in \mathcal{M} \setminus k} P(\xi_m(v) = n), \quad (23)$$

where $\circledast$ denotes the discrete convolution operator. $p_{N_{k,v}}(n)$ denotes the probability that n of the users except k have a TBS of v under the condition that none of these users has a higher TBS. This means that if user k also has TBS k, $1/(n+1)$ is the probability that k is actually scheduled. Expressed in a formula, the probability that k is scheduled under the condition that her TBS is v is given by

$$p_{s,k}(v) = \left(\prod_{m \in \mathcal{M} \setminus k} P_{TBS,m}(v) \right) \cdot \left(\sum_{n=0}^{|\mathcal{M}|-1} \frac{p_{N_{k,v}}(n)}{n+1} \right). \quad (24)$$

The first factor corresponds to the probability the a user with TBS v is scheduled and the second factor corresponds to the probability that this user is user k.

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 Eq. (23). 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. This can be achieved by sorting the users according some metric, e.g. to Σ_k. So, the convolution starts with the user with best average channel and the consequent users have always worse average channel conditions. Before convolving the TBS distribution from the best n users with the TBS distribution of the next best user, the chance of the user to be scheduled decides whether the user is still considered or whether this and all following users are ignored.

5.2.3 Proportional-fair Scheduling

Proportional-fair (PF) scheduling is a scheduling discipline which has been developed for the 1xEv-DO-system in the downlink, [9]. Its basic principle is to give each user the bandwidth proportional to its link quality. This is achieved by choosing the user which has the best instantaneous relative throughput over its past throughput, which is normally calculated with help of a sliding window. However, different versions of PF scheduling exist. The most fundamental difference is the way how the past throughput is calculated. One has the option to update the past throughput every scheduling period regardless whether the user has been scheduled or not, or the past throughput is updated only if the users is indeed chosen for transmission. The difference between both version is that in the first case the

mean throughput of a user is proportional to its channel quality only, while in the second case it is also related to the generated traffic. In this work, we consider the first case only since the second case is very difficult to handle analytically. However, we will show the difference between both PF variants in our simulation studies.

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

$$P_{PF,k}(\theta) = P_{TBS,k}(\theta \cdot E[v_k]). \quad (25)$$

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

$$p_{s,k}(v) = \prod_{m \in \mathcal{M} \setminus k} P_{PF,m}(\theta_m(v)). \quad (26)$$

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.

6. SIMULATION STUDIES

The following simulation study has the major goal to inspect in how far the proposed simulation model matches with a detailed simulation. The two main uncertainties in the simulation model are the derivation of the average throughput in a short-term static situation and whether the approximation of the transmitted data volume by the average throughput is justified, i.e. whether the transmitted data volume exhibits a small level of randomness. In order to investigate these two questions we conduct two simulation studies: First, we investigate the transient behavior of the HSDPA bandwidth. In particular we are interested how the coefficient of variation (c_v) of the transmitted volume develops with time depending on scheduling scheme and channel profile. Second, we evaluate the accuracy of the proposed model for approximating the long-term HSDPA bandwidth in a static situation again with different scheduling schemes and channel profiles.

6.1 Simulation design

Producing results with a validity as general as possible is quite a challenge for the design of simulation experiments. The possible parameter space includes the network layouts, the NodeB transmit power, the HSDPA powers, the available codes, the number of HSDPA users and their location within a cell, the UE class, the speed of the users, the channel profile, and many more. We focus on a two tier hexagonal network layout and consider $L=20$ possible user locations randomly located within the center cell. All NodeBs transmit with equal power, all power and code resources are available to the HSDPA, and the mobiles are also capable of using them, i.e. we have class 4 UEs. We consider three multi-path profiles, ITU Pedestrian A (Ped. A), ITU Pedestrian B (Ped. B), and ITU Vehicular A (Veh. A), and assume that the users move with a velocity of $1\,\text{m/s}$.

For every path $p \in \mathcal{P}_{y,\ell}$ between any NodeB y and any location ℓ we generate independent fading traces according to the three chosen channel profiles which yield the instantaneous relative propagation losses. We repeat this simulation $I = 50$ times and obtain the propagation losses $\beta_{p,Ch}(i,t)$, where t denotes the TTI, i the instance of the fading trace, and Ch the channel profile. The propagation losses are correlated in time domain and otherwise independent. According to formulas Eq. (13) and Eq. (11) we obtain I CQI traces for every location where $CQI_{\ell,Ch}(i,t)$ is the CQI of replication i for location ℓ in TTI t with multi-path profile Ch.

For actually applying scheduling mechanisms to the CQI traces, we focus on $K = 5$ simultaneously active users. We refer to a set of K users as a configuration and generate $C = 20$ different configurations in which the users are randomly placed on different locations. A configuration c then corresponds to K CQI traces.

According to the number of codes and UE class we translate the CQI traces to TBS traces and are then able to determine the scheduled user and the transmitted data volume for every TTI according to the scheduling schemes $Sch \in \{Round\text{-}Robin, MaxTBS, Prop.\text{-}fair\ (ideal,\ possible), Prop.\text{-}fair(ideal,\ actual)\}$. $Round\text{-}Robin$ schedules the user in arbitrary order oblivious of the CQI. $MaxTBS$ always schedules the user with largest TBS and breaks ties randomly. $Prop.\text{-}fair$ schedules the user k with highest ratio of actual TBS $v(k,t)$ to average TBS $\bar{v}(k,t)$ in the past. The average past TBS is determined by averaging over all past TTIs, i.e.

$$\bar{v}(k,t) = \begin{cases} \frac{1}{t-1} \sum_{i=1}^{t-1} \delta(k,i) \cdot v(k,i) & \text{"actual"} \\ \frac{1}{t-1} \sum_{i=1}^{t-1} v(k,i) & \text{"possible"} \end{cases}. \quad (27)$$

We refer to this scheduler as *ideal* since we do not use an exponentially weighted moving average for calculating the average as would be done in practice. The reason for using this idealized scheduler is simply not to introduce another parameter. The variable $\delta(k,i)$ indicates whether user k is scheduled in TTI i or not. The notations *actual* and *possible* are introduced in order to distinguish the scheduler computing the average bandwidth from the actually transmitted data volume and the scheduler computing the average bandwidth by the possible data volume, i.e. the data volume allowed for by the channel oblivious of the scheduling decision. Note that both implementations for *Prop.-fair* schedulers are found in the literature. The model for the proportional bandwidth in Section 5.2.3 relates to the *possible* scheduler.

Applying a certain scheduler to the K CQI/TBS traces of a configuration c yields a set of transmitted data volume traces with $v_{i,Ch,Sch}(c,k,t)$ being the data volume transmitted in TTI i to user $k \in \{1,\ldots,K\}$ of configuration c taking the ith replication of the fading traces. Now, $v_{i,Ch,Sch}(c,k,t)$ is zero if user k is not scheduled in TTI t. The throughput for user k until TTI t is defined by

$$R_{i,Ch,Sch}(c,k,t) = \frac{1}{t} \sum_{s=1}^{t} v_{i,Ch,Sch}(c,k,s). \quad (28)$$

6.2 Transient behavior of HSDPA bandwidth

We are now interested in the transient behavior of the HSDPA bandwidth, which means our simulator enters a certain static simulation scenario, namely a configuration, and we are interested in the impact that different fading traces have
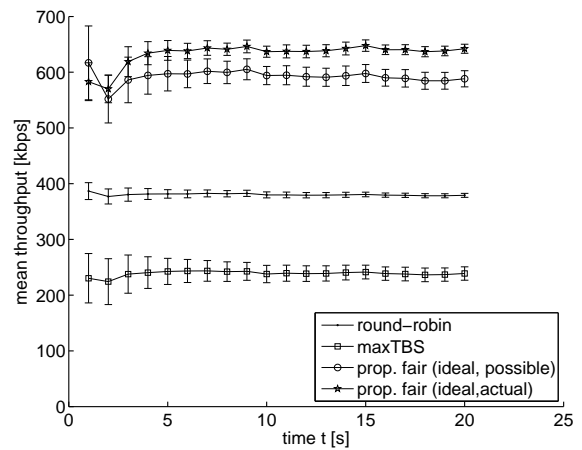

Figure 6: Confidence intervals of the mean throughput decrease with the length t of the considered period of time

on the transmitted data volume and the HSDPA bandwidth of the single users. Therefore, we consider the coefficient of variation of the user throughput after t TTIs with respect to the different replications of the fading traces

$$c_v \left[B_{Ch,Sch}(c,k,t) \right] = \frac{\sqrt{\text{VAR} \left[B_{Ch,Sch}(c,k,t) \right]}}{\text{E} \left[B_{Ch,Sch}(c,k,t) \right]} \quad (29)$$

Note that $B_{Ch,Sch}(c,k,t)$ is now a random variable for the throughput and $B_{i,Ch,Sch}(c,k,t)$ is the instance of this random variable according to trace i.

Let us first consider a single user with $\Sigma = 0.72$, i.e. the ratio of average other-cell interference and average own-cell interference is 0.72. The other four users have Σ values of 0.01, 1.53, 1.39, and 1.18. We do not show results for the other users, but their Σ-values help us to understand the results for the different scheduling disciplines. The multi-path channel profile is Ped A.

Fig. 6 shows how the mean throughput $\text{E} \left[B_{Ch,Sch}(c,k,t) \right]$ develops with the time t, here given in seconds instead of TTIs. The mean throughputs are marked with 95% confidence intervals that are derived over the different fading traces. We can see that the throughput is lowest for *MaxTBS* and highest for the two porportional fair variants with *actual* exceeding *possible*. *MaxTBS* is of course the worst since the user with $\Sigma = 0.01$ experiences by far less other-cell interference and accordingly has much larger TBSs. *Round-Robin* is worse than *Prop.-fair* since it is oblivious of the channel. These results are not surprising and already often mentioned. What is more important for us is that the confidence intervals for the mean throughput descrease with the time. We can interpret the confidence intervals as the region the throughputs of 95% of the fading traces fall inside.

Let us also consider the c_v of the throughput. A low c_v close to zero means that the throughput is rather deterministic and an approximation of the transmitted data volume by its mean bandwidth works well. A high coefficient of variation means that the data volume transmitted within the considered period of time is actually stochastic and approximating it by the mean throughput may lead to a significant error. Fig. 7 shows the c_v of the mean throughput.
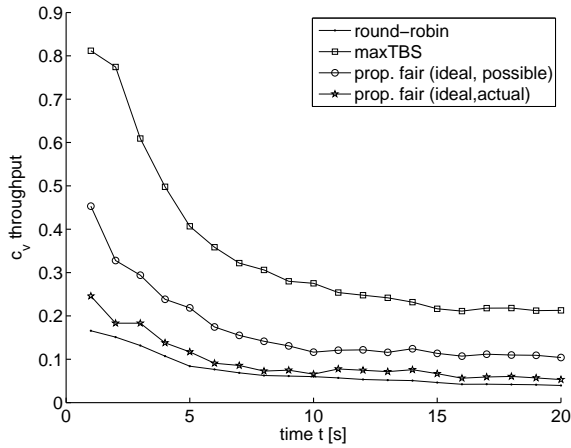
Figure 7: Coefficient of variation decreases with the length t of the considered period of time

Figure 8: Coefficient of variation decreases with the duration t of the considered period of time

As expected the c_v decreases with t. Assuming uncorrelated transmitted data volumes per TTI, the c_v of the throughput after n TTIs will be $\sqrt{1/n}$ of the c_v for a single TTI. This explains why the c_v falls of strongly for the first TTIs and then converges to a certain level. The c_v with *MaxTBS* is larger since here we observe a larger autocorrelation for the TTIs.

Let us now move from the arbitrarily selected special case to some more general results. We consider the c_v of the throughputs from the five users of all 20 configurations together, i.e. we have 100 instances of the c_v for every scheduling discipline Sch and every multi-path channel profile Ch. Please note that we used the same user configurations for all fading traces and scheduling disciplines. Fig. 8 summarizes all these results in a single scatter plot. The figure shows 24 columns of 100 symbols reflecting the c_v of the throughput on the y-axis. The 24 columns are first separated into groups of six according to the four scheduling disciplines. Then each of the groups of six columns is separated into two parts, the left one representing the c_v of the throughput after 500 TTIs or 1 s, and the right one after 10000 TTIs or 20 s. The remaining groups of three columns consist of the results for Veh. A to the left, Ped. A in the middle, and Ped. B to the right.

From the figure, we can first learn that the c_v is quite situation dependent. For *MaxTBS* the c_v covers a range of nearly two magnitudes independent of the time or the scheduling scheme. Furthermore, *MaxTBS* produces the largest c_v and is consequently the most difficult to approximate by the mean bandwidth since it is most subject to the randomness of the fading. Both variants of proportional fair scheduling lead to a c_v below 0.5 after 1 s and below 0.12 after 20 s. *Actual* achieves a smaller c_v than *possible*. The c_v is rather independent of the multi-path channel profile. The c_v for *Round-Robin* is the smallest one with values below 0.2. Ped. A with a single strong multi-path component produces a higher c_v than Veh. A or Ped. B with a second strong multi-path component. Interestingly, this effect mainly occurs for *Round-Robin*, the other channel-aware schedulers seem to be able to counter the randomness of the single strong multi-path component.
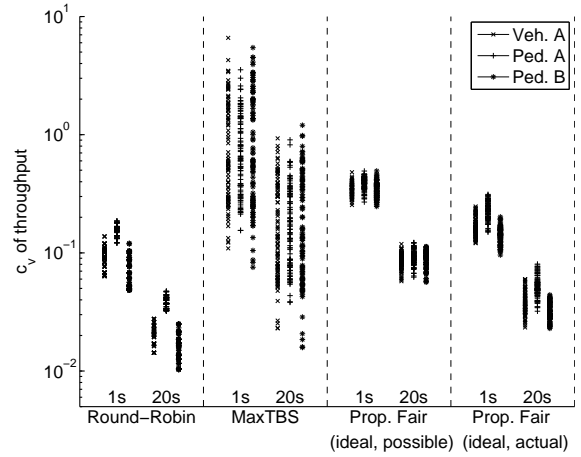
We conclude that with *Round-Robin* or *Proportional-Fair* scheduling static time periods of 1 s are sufficient to approximate the transmitted data volume. *MaxTBS* scheduling is very subject to the randomness of the channel and may have large variations even after longer periods of time though these mainly correspond to users with small data volumes. However, we have to keep in mind that approximating the data volume transmitted in a certain period of time means that the static system states in sequential periods of time differ only little since such a change is caused e.g. through a small movement of a user which leads to a change in the shadow fading or a different path loss.

6.3 Accuracy of the throughput approximation

Let us now investigate the accuracy of the throughput model proposed in Section 5.2. The CQI distributions required for the throughput model are derived according to the model introduced in [19]. This means that the approximation model is able to derive the mean throughput for the different users only based on their Σ-values which are easily derived from NodeB transmit powers and average propagation losses.

As in the previous section we arbitrarily choose a single scenario with Ped. A multi-path channel that we investigate in detail before we come to more general results. Fig. 9 is constructed similar to Fig. 7 but is different to read. There are now four groups of five columns of symbols each. The groups correspond to the scheduling discipline and the columns represent the five different users that are ordered with an increasing Σ-value from left to right. The users have Σ-values of 0.001, 0.037, 0.217, 0.251, 1.073. A symbol now corresponds to the throughput of a single replication, i.e. each symbol shows the value of $B_{i,Ped.\,A,Sch}(c, k, 1e4)$ for one replication i. The bars crossing the columns indicate the throughput obtained from the approximation model. We can see that the model matches the middle of the different traces quite well except for *Prop.-fair (ideal, actual)*. However, this is not surprising since we depicted the results obtained for the *possible* algorithm as a very rough approximation for the simple reason that no better approximation
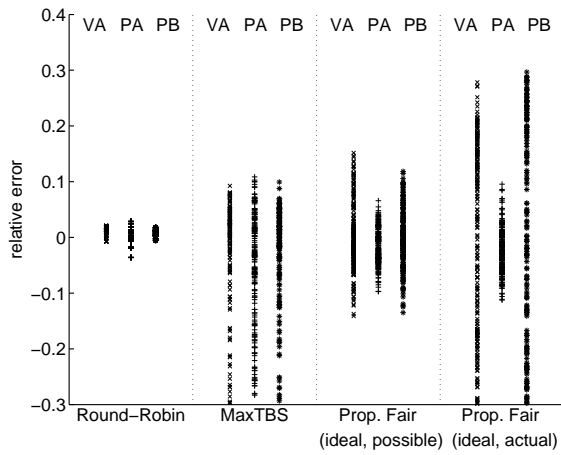
Figure 10: Good accuracy with less than 10% **error for approximation with Round-Robin and Prop.fair (ideal, possible) scheduling**
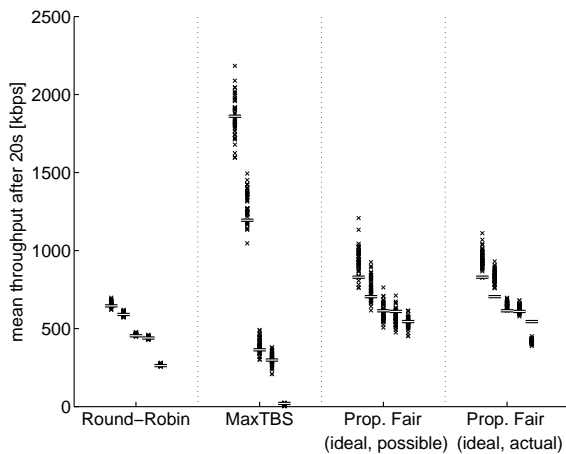

Figure 9: Spreading of the simulated throughputs around the approximated throughput

is available. Furthermore, we have to mention that the arbitrarily selected configuration is not entirely arbitrary but shows a situation with two users having rather low Σ-values that leads to rather good results of our model.

So, let us next study what happens if we consider the performance in an arbitrary situation, i.e. if we again mix up all users and all configurations. We define the relative error of the mean throughput for user k in configuration c as

$$\psi_{Ch,Sch}(c,k) = \frac{\mathrm{E}\big[B_{Ch,Sch}(c,k,1e4)\big] - \tilde{B}_{Ch,Sch}(c,k)}{\mathrm{E}\big[B_{Ch,Sch}(c,k,1e4)\big]}, \qquad (30)$$

where $\tilde{B}_{Ch,Sch}(c,k)$ is the approximated throughput. Fig. 10 shows the relative error for the three multi-path channel profiles and the four scheduling disciplines. The plot is again grouped in columns of symbols, and a single column corresponds to a multi-path channel profile. A single symbol represents the relative error for a single user in a single configuration. At first glance, we see that *Round-Robin* leads to

very small errors and both *MaxTBS* and *Prop.-fair (ideal, actual)* to rather large errors. Note that for *MaxTBS* very large relative errors up to -2 can occur for users that are very rarely scheduled. However, their bandwidth is in the order of 10 kbps and less. For *Prop.-fair (ideal, actual)* the large errors are not surprising since, as already mentioned, we only have no real solution but only a very rough approximation. In contrast, it is rather astonishing that for Ped. A our rough approximation works quite well and leads to positive and negative relative errors below 10%. Also for *Prop.-fair (ideal, possible)* we obtain a quite good accuracy with most errors below 10%. Again, the model works best for Ped. A. Remarkable is also that the model quite symmetrically under- and overestimates the actual mean bandwidth.

7. CONCLUSION

In this work, we proposed a novel simulation framework for UMTS systems with coexisting HSDPA and DCH users. The framework implements a discrete-event time-dynamic simulation on flow level. The framework uses a statistical model for the HSDPA bandwidth which considers several scheduling disciplines and takes the multi-path propagation profile, the available code resources, the UE category and the location-dependent other-cell interference into account. Furthermore, an analytical model for the DCH transmit powers is used. The abstraction of these small time-scale effects allows for time-efficient simulations of large-scale network scenarios and furthermore enables it to investigate the impact of traffic dynamics on flow-level in a statistically significant way.

A special focus was the modeling of the scheduling. Analytical models for *Round-Robin* scheduling, *MaxTBS* scheduling and *Proportional-fair* scheduling have been developed. We showed that the analytical approximation of the mean user bandwidth is in most cases of sufficient accuracy, given that the time between two events is long enough. However, the results depend on the scheduling discipline: the results for *MaxTBS* are generally more variant than for *Round-Robin* and *Proportional-fair* scheduling.

Finally we want to point out that the underlying principle is not restricted on HSDPA, but can be formulated more generally: in large networks the investigation of traffic dynamics on flow level requires simulations which combine technical accuracy with efficient simulation techniques. This also applies to future mobile networks like UMTS Long Term Evolution, WiMAX, or Mesh networks.

8. REFERENCES

[1] 3GPP. 3gpp ts 25.214 v6.7.0 physical layer procedures (fdd) (release 6). Technical report, 3GPP, Sep 2005.

[2] 3GPP. 3GPP TS 34.121-1 User Equipment (UE) conformance specification; Radio transmission and reception (FDD); Part 1. Technical report, Dec 2006.

[3] 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*, pages 844–848, Sidney, Australia, Aug 2004.

[4] H. Buddendick, G. Wölfle, S. Burger, and P. Wertz. Simulator for Performance Analysis in UMTS FDD Networks with HSDPA. In *Proc. of 15th IEEE PIMRC*, Barcelona, Spain, Sep 2004.

[5] EURANE. Eurane website. http://www.ti-wmc.nl/eurane/.

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

[7] H. Holma and A. T. (Eds.). *WCDMA for UMTS*. John Wiley & Sons, Ltd., Feb 2001.

[8] H. Holma and A. Toskala, editors. *HSDPA/HSUPA for UMTS. High Speed Radio Access for Mobile Communications*. Wiley & Sons, Jun 2006.

[9] 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*, pages 1854–1858, Tokyo, May 2000.

[10] 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*, volume 1, pages 477–481, Vancouver, Canada, Sep 2002.

[11] R. Litjens, J. van den Berg, and M. Fleuren. Spatial Traffic Heterogeneity in HSDPA Networks and its Impact on Network Planning. In *Proc. of 19th ITC*, pages 653–666, Bejing, China, Aug 2005.

[12] A. Mäder, D. Staehle, and H. Barth. A Novel Performance Model for the HSDPA with Adaptive Resource Allocation. In *Proc. of 20th ITC*, Ottawa, Canada, Jun 2007.

[13] 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.

[14] M. Necker. A comparison of scheduling mechanisms for service class differentiation in hsdpa networks. *AEÜ International Journal of Electronics and Communications*, 60(2):136–141, Feb 2006.

[15] 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*, pages 2296–4500, Milan, Italy, Sep 2004.

[16] J. W. Roberts. A survey on statistical bandwidth sharing. *Computer Networks*, 45(3):319–332, 2004.

[17] K. Sipilä, Z.-C. Honkasalo, J. Laiho-Steffens, and A. Wacker. Estimation of capacity and required transmission power of WCDMA downlink based on a downlink pole equation. In *Proc. of IEEE VTC Spring '00*, pages 1002–1005, Tokyo, Japan, 2000.

[18] 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.

[19] D. Staehle and A. Mäder. A Model for Time-Efficient HSDPA Simulations. In *Proc. of IEEE VTC Fall '07*, Baltimore, MD, USA, Oct 2007.

[20] 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*, pages 2555–2559, Milan, Italy, May 2004.

[21] 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*, pages 283–292, Venice, Italy, Oct 2004. ACM.