Optimizing the ARQ Performance in Downlink Packet Data Systems With Scheduling

Haitao Zheng, *Member, IEEE* and Harish Viswanathan, *Senior Member, IEEE*

***Abstract*—Third generation wireless systems typically employ adaptive coding and modulation, scheduling, and Hybrid Automatic Repeat reQuest (HARQ) techniques to provide high-speed packet data service on the downlink. Two main considerations in designing such a system are algorithms for the selection of coding and modulation schemes based on the channel quality of the link and algorithms for the selection of the user to whom a particular slot is assigned. We propose a systematic approach to optimize the mapping between signal-to-interference-and-noise ratio (SINR) and modulation and coding scheme (MCS) to maximize the throughput by taking into account the type of HARQ scheme employed. We also propose to incorporate frame error rate (FER) and retransmission information as a part of the scheduling decision. The proposed scheduler ranking methods based on using an effective rate rather than the instantaneous rate provide natural priority to retransmissions over new transmissions, and priority to users with better channel quality. Extensive simulation results comparing performance of the proposed methods to conventional methods are presented.**

***Index Terms*—Automatic repeat request, hybrid automatic repeat reQuest (HARQ), link adaptation, packet radio networks, packet scheduling algorithm, quality of service, scheduling, wireless communications.**

I. Introduction

THIRD-GENERATION cellular systems are designed to support high-speed packet data service on the downlink from the base station to the mobile device. The high-speed packet service is provided through a shared downlink channel, where time division multiplexing is used within a code division multiple access (CDMA) system. Time slots are assigned to users at specific data rates through a scheduling algorithm based on the user data backlog and the channel quality from a feedback channel.[1] Such a transmission scheme allows multiple users to share time and Walsh code resources. The sharing of resources permits adaptation to network and channel variations and also reduces the channelization code limitation that might occur if each user is allocated a dedicated channel. Therefore, it has the potential to improve the capacity for bursty services. Examples of such a system include the CDMA 2000 1X-EV-DO, 1X EV-DV, and the UMTS high-speed downlink packet access (HSDPA) [1]–[3]. Several advanced technologies are used in high-speed downlink transmission to improve link throughput or reduce packet delay by adapting to the time varying channel conditions, traffic statistics, and quality of service requirements. Some of these adaptive techniques relevant to this paper are summarized below.

Dynamic Link Adaptation or Adaptive Modulation and Coding: Typically, each transmission in the shared downlink channel is at maximum power available and no power control is employed. Therefore, link adaptation, which continuously adjusts the modulation and coding scheme (MCS), provides an efficient way of maximizing the instantaneous usage of the wireless channel [4]–[6]. For example, it enables the use of spectral-efficient higher-order MCSs when channel conditions are favorable while reverting to the MCSs that are more robust but with lower transmission rates when channel conditions degrade. The transmitter selects an appropriate MCS, based on the user's channel quality feedback, for the user that is to be served in the current time slot. Henceforth, we refer to the MCS selection as the mapping design.

Automatic Repeat reQuest (ARQ) or Hybrid ARQ: The performance of link adaptation largely depends on the accuracy of channel quality measurement, which is difficult to maintain as user mobility increases. That packet data is delay-tolerant makes it feasible to use retransmission schemes to recover erroneous packets. Recently, Hybrid Automatic Repeat reQuest (HARQ) techniques have been adopted by several wireless standards, e.g., 1X EV-DV and HSDPA [1]–[3]. HARQ can compensate for link adaptation errors and provide a finer granularity of coding rate [7], [8] and improve throughput performance. Upon detecting a transmission failure, typically through cyclic redundancy check (CRC), the mobile user sends a request to the base station for retransmission. The packet decoder at the mobile combines the soft information of original transmission with that of the subsequent retransmissions, and the combined signal has higher probability of successful decoding. In general, there are two ways of performing soft combining. Using Chase combining based HARQ mechanism, the base station repeatedly sends the same packet and the receiver aggregates the energy from the (re)transmissions to improve signal to noise ratio (SNR) [9], [10] and thus, the combined packet has higher probability to be decoded correctly. A more sophisticated HARQ mechanism, named incremental redundancy (IR) [11]–[16], transmits additional redundant information in each retransmission and gradually refines coding rate and SNR till a successful decoding is achieved. Transmission starts with the highest rate code and additional redundancy bits are transmitted whenever necessary. Rate compatible punctured convolutional (RCPC) codes [17] and Turbo-codes [18] can be used to produce redundancy bits. Compared to Chase combining,

Manuscript received December 9, 2002; revised June 6, 2003 and October 21, 2003; accepted December 15, 2003. The editor coordinating the review of this paper and approving it for publication is M. Zorzi.

H. Zheng is with Microsoft Research Asia, Beijing 100080, China.

H. Viswanathan is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: harishv@lucent.com).

Digital Object Identifier 10.1109/TWC.2004.843012

[1]Each mobile user measures the channel condition and translates it into a metric which is fed back to the serving base station.

IR has superior performance but also requires larger receiver buffer [16], [19]. It also provides finer granularity of coding rate and allows better adaptation to channel variations. A channel capacity based analysis of different HARQ mechanisms can be found in [20].

Scheduler: For a multiuser system where user channel conditions change over time, a scheduler can take advantage of channel variations by giving certain priority to the users with transitorily better channel conditions. Hence, the choice of the scheduling algorithm critically impacts the system performance. Several scheduling algorithms have been proposed in the literature [3], [21]–[26] to maximize the packet data throughput, subject to various fairness conditions. For example, the proportional fair algorithm [3], [22] utilizes asynchronous channel variations by selecting the user with the maximum value of $R/R_{\rm avg}$. Here, R represents the instantaneous transmission rate chosen from the set of available MCS schemes based on the signal-to-noise-and-interference ratio (SINR) feedback from the receiver, and $R_{\rm avg}$ represents the moving average of the rate at which the user was served in the previous slots. This scheduler gives approximately the same number of time slots to all the users, but assigns the transmission to each user when its channel condition is at its best.

The above three technologies are tightly coupled. One of the main design issues to be addressed in link adaptation is the mapping between the channel quality feedback, e.g., SINR, and the MCS to be used to transmit user packets [27]–[30]. The mapping would determine the success probability of transmissions and hence the achieved throughput. When some level of channel uncertainty exists and the system supports HARQ, it may be beneficial to pick an aggressive MCS that has a high packet error rate at the reported SINR value. However, an overly aggressive mapping could produce too many unsuccessful packet transmissions that diminish the overall throughput while an overly conservative mapping fails to fully utilize channel resources. To date, the mapping is typically chosen based on maximizing data throughput subject to the frame error rate being below a certain threshold [29]–[31]. The data throughput is computed as the instantaneous data rate which is the ratio of the 1) number of information bits transmitted in a frame; and 2) frame duration. However, the instantaneous rate does not fully represent the true data throughput since it fails to account for transmission errors and the impact of the resulting retransmissions. Sometimes an aggressiveness factor is introduced through trial and error, when HARQ is used as the retransmission scheme [16]. Overall, the conventional mapping design fails to take into account the performance improvement by HARQ and, therefore, is not optimized for new systems. At the same time, the mapping design and the associated throughput computation are crucial to the scheduler since the decision of which user to transmit in each scheduling interval is closely related to the throughput that the user can achieve. Most scheduling algorithms proposed to date rely on the instantaneous rate of a particular mobile user to determine the priority of the user. Examples are the proportional fair scheduler [22], the maximum weight scheduler [21], [24], and the exponential weight scheduler [23]. However, these algorithms do not include the fact that the packet transmission is error-prone and that link layer retransmission techniques such as HARQ are used to recover the lost packets.

In this paper, we address some of the key design issues associated with the choice of the MCS scheme to be used for transmission, given that an ARQ scheme is being used. We present a systematic approach to selecting the mapping between the SINR and MCS schemes, which performs well as shown in simulations. The proposed mapping design is based on maximizing a reasonable approximation to the user throughput while taking into account the HARQ scheme being used. Simulation results show that the sensitivity of the performance to the choice of the mapping depends on the granularity of the MCS set. Thus, we present results for the cases with 4 MCS levels (referred to as *4MCS set* hereafter) where the improvement is marked and for the cases with 6 MCS (referred to as *6MCS set* hereafter) levels where there is limited improvement using the scheme presented in the paper. In addition, we propose modifications to the scheduling algorithm so as to use the frame error rate (FER) and retransmission information in the scheduling process. When a scheduler is choosing between multiple users, one would expect that using information about retransmissions versus new transmissions in addition to information related to backlog and channel condition should improve the accuracy of the scheduling decision. We investigate schemes in which the scheduler is designed to provide priority to retransmissions over new transmissions.

The paper is organized as follows. In Section II, we present the mapping design in the presence of HARQ. Along with the general formulation, a number mapping criteria related to special HARQ mechanisms are discussed. Modifications to the scheduling algorithm to include frame error rate (FER) and retransmission information in the scheduling process are discussed in Section III. We compare the performance of different mapping designs and scheduler ranking methods and illustrate TCP level results in Section IV. Conclusions are drawn in Section V.

II. Optimizing the Mapping Between SINR and MCS

In this paper, we assume that the channel quality feedback carries SINR. The mapping is between the SINR and the MCS so that it can be applied to any user regardless of its distance to the base station. In addition, the mapping only depends on the instantaneous channel SINR rather than channel statistics. Before we proceed, we will first review the mapping criteria that have been widely used in the existing systems. Traditionally, the mapping is determined so as to maximize the instantaneous rate while maintaining certain target FER.[2] For voice transmission, the mapping is selected to maintain a 1% FER, while for data transmission, the mapping is usually generated to maximize the user throughput. For a general user k, the traditional mapping selection criteria between the SINR η and the chosen MCS are the following.

***FER* x**: which is bounded by the maximum allowable FER of $x\%$, i.e.,

$$\mathrm{MCS}_{\mathrm{FER}\,x}(\eta) = \arg\max_{i \in M}\{R_i \mid F_i(\eta) < x\} \quad (1)$$

[2]This target FER is the measured frame error rate prior to HARQ operation.

RFER: which is to maximize the estimated short term per packet throughput assuming no HARQ operation, i.e.,

$$\mathrm{MCS_{RFER}}(\eta) = \arg\max_{i \in M} R_i \cdot (1 - F_i(\eta)) \tag{2}$$

where R_i represents the instantaneous rate of MCS $i, F_i(\eta)$ is associated FER of MCS i at SINR η, and M represents the MCS set.

However, when applying HARQ as the retransmission scheme, the mapping should be selected to maximize the throughput after HARQ operation. Therefore, the average throughput cannot be simply represented by $R_{\mathrm{MCS}(\eta)} \cdot (1 - F_{\mathrm{MCS}(\eta)}(\eta))$. In this paper, we consider two HARQ mechanisms: simple ARQ and Chase combining. Upon detecting a packet error, simple ARQ based transmitter retransmits the same packet and repeats the procedure until the packet is received successfully. Hence, the associated receiver always uses the last received packet to decode the information and there is no combining. Chase combining based receiver, on the other hand, combines the soft decision of the original and the retransmitted signal for a higher probability of successful decoding [9], [10].[3] Therefore, Chase combining outperforms simple ARQ, especially at high FER for new transmissions. On the other hand, despite the performance gain, some systems could still favor simple ARQ since it does not require any advanced processing or extra buffer to save the erroneous packets. It should be noted that for both Chase combining and simple ARQ, the MCS used in retransmissions is the same as that in the original transmission.

Motivated by the above discussions, we propose a single unified mapping criterion that maximizes the average user throughput [33]. For simplicity, we assume the channel condition stays constant during retransmissions as that of the initial transmission since the retransmissions are likely to occur soon after the initial transmission for slowly varying channels. Furthermore, in practice it is difficult to predict the future channel conditions accurately and hence assuming constant channel condition can be viewed as a simplified predictor of future channel. Therefore, for a general user, during each transmission period, the channel can be approximated as an AWGN channel, and maximizing the average throughput is equivalent to maximizing the average throughput for each AWGN channel. In principle, if the distribution of the SINR is known then one can determine the mapping so as to maximize the average throughput over the time-varying SINR. However, such an approach will result in a mapping that depends on the SINR distribution, which in itself depends on the user location within the cell. It is impractical to have a separate mapping for each user. Thus, a mapping derived by assuming a constant SINR is a reasonable approach.

The average throughput of an AWGN channel using MCS i is defined as the ratio of the average number of successfully received information bits $\mathrm{E}(S \mid \eta, i)$ and the average time taken by the packet $\mathrm{E}(T \mid \eta, i)$ [32], i.e.,

$$\mu(\eta, i) = \frac{E(S \mid \eta, i)}{E(T \mid \eta, i)} \tag{3}$$

where η represents the SINR. This metric includes the contribution from not only the current transmission but also that of the future retransmission(s) and thus depends on the retransmission scheme. The corresponding criteria for Chase combining and simple ARQ are as follows:

TRPT-Chase:

$$\begin{aligned}\mathrm{MCS_{thrpt}^{CHASE}}(\eta) &= \arg\max_{i \in M} \mu(\eta, i) \\ &= \arg\max_{i \in M} \frac{S_i}{\tau \cdot \sum_{k=1}^{\infty} k(1 - F_i(k\eta)) \prod_{q=1}^{k-1} F_i(q\eta)} \\ &= \arg\max_{i \in M} \frac{R_i}{\sum_{k=1}^{\infty} k(1 - F_i(k\eta)) \prod_{q=1}^{k-1} F_i(q\eta)}\end{aligned} \tag{4}$$

where the probability of successful reception after k transmissions can be computed as $(1 - F_i(k\eta)) \prod_{q=1}^{k-1} F_i(q\eta)$. It should be pointed out that (3) provides a valid throughput estimation only when the number of retransmissions is unlimited. For a given bound on the maximum number of retransmissions $T_{\max}$, we modify the criterion to (see (5) at the bottom of the page). For simple ARQ we have the following.

TRPT-SARQ: (See (6) at the bottom of the page.) The second equality above is established in the Appendix. It should be noted that the mapping design described above could be extended for systems that employ IR based HARQ by changing the FER function to include the coding rate information.

III. Optimizing the Scheduler Ranking for HARQ

The scheduling algorithms for downlink wireless transmission are designed to exploit channel variations due to fading by using feedback from the mobile device. Most scheduling

[3]The soft combining increases the SINR after a retransmission to the sum of the SINR corresponding to the original transmission and the retransmission.

$$\mathrm{MCS_{thrpt}^{CHASE}}(\eta) = \arg\max_{i \in M} \frac{R_i \cdot \left(1 - \prod_{q=1}^{T_{\max}+1} F_i(q\eta)\right)}{\sum_{k=1}^{T_{\max}+1} k(1 - F_i(k\eta)) \prod_{q=1}^{k-1} F_i(q\eta) + (T_{\max}+1) \cdot \prod_{q=1}^{T_{\max}+1} F_i(q\eta)} \tag{5}$$

$$\begin{aligned}\mathrm{MCS_{thrpt}^{SARQ}}(\eta) &= \arg\max_{i \in M} \frac{R_i \cdot \left(1 - F_i(\eta)^{T_{\max}+1}\right)}{\sum_{k=1}^{T_{\max}+1} k(1 - F_i(\eta)) F_i(\eta)^{k-1} + (T_{\max}+1) \cdot F_i(\eta)^{T_{\max}+1}} \\ &= \arg\max_{i \in M} R_i(1 - F_i(\eta)) = \mathrm{MCS_{RFER}}(\eta)\end{aligned} \tag{6}$$

algorithms in the literature use the instantaneous rate to determine the user to be served in each scheduling interval, but ignore the link level transmission errors and retransmission mechanisms such as HARQ to recover transmission errors. Given that packet queues contain new packets waiting to be transmitted and packets to be retransmitted, one could expect that giving higher priority to retransmissions could enhance the scheduler performance.

Before we proceed, we will briefly summarize the operations by the traditional schedulers that are unaware of the existence of HARQ. After selecting a MCS for each user, the scheduler determines the user to be served in the future based on the instantaneous rate of each user. For example, assuming $\mathrm{MCS}(k)$ represents the chosen MCS for user k, and $R_{\mathrm{MCS}(k)}$ represents the instantaneous rate of use k that uses $\mathrm{MCS}(k)$, then the Proportional Fair scheduler [22] selects to serve user U according to

$$U = \arg\max_{k \in \Pi} \frac{R_{\mathrm{MCS}(k)}}{R_{\mathrm{avg}}(k)} \tag{7}$$

where $R_{\mathrm{avg}}(k)$ is the average rate received by user k in the past over a certain time horizon and Π represents the set of active users in the system. Another example of scheduler method, corresponding to the maximum rate scheduler, selects the user to be served according to

$$U = \arg\max_{k \in \Pi} R_{\mathrm{MCS}(k)}. \tag{8}$$

However, for a system that employs HARQ, the instantaneous rate $R_{\mathrm{MCS}(k)}$ cannot fully represent the true transmission rate since it fails to recognize the following two important characteristics:

Frame Error Rate Information: The instantaneous rate only represents the transmission rate when the packet is successfully received. Upon a packet error, the actual transmission rate is zero and the frame should be retransmitted or discarded. This miscalculation is especially harmful in a multiuser system. For example, Mobile A has a SINR of 4 dB and selects MCS 1 of 2.4 Mbps, i.e., the instantaneous rate $R = 2.4$ Mbps, and $F_1(4\ \mathrm{dB}) = 90\%$; Mobile B reports a 3 dB SINR that maps to MCS 0 of 0.6 Mbps which results in a FER of 10%. Based on the instantaneous rate, Mobile A is scheduled. However, it is very likely that Mobile A's transmission fails due to high FER. On the other hand, Mobile B has a higher success possibility and thus should be scheduled first according to the FER information.

Retransmission Information: For Chase combining, retransmission has higher success probability since the SINR after soft combining is the sum of the current SINR and that of the previous transmission(s).[4] Therefore, the rate calculation should take into account the retransmission information.

Unlike in the mapping design case, it is not straightforward to derive an expression for the average throughput for a multiuser scheduled system since it intricately depends on the scheduling algorithm. This is especially the case for a scheduling algorithm such as the Proportional Fair algorithm that guarantees a certain fairness criterion. Thus, determining an *optimum* scheduling policy in the presence of frame errors and HARQ appears to be intractable. Accordingly, we present a heuristic approach that relies on the fact that if in the scheduling algorithm, it is possible to substitute the assumed error free rate by a more accurate rate that can be achieved through multiple transmissions, then the essential nature of the scheduling algorithm will be relatively unchanged and the same fairness criterion will be retained while at the same time, HARQ information is also taken into account. It should be noted that the scheduler design and the mapping selection can be conducted jointly to optimize the system performance. For simplicity, we assume that they are performed independently. We thus propose and investigate a set of *effective rates* $R_{\mathrm{eff}}(k)$ that include both the FER information and the retransmission information [34]. For any scheduling algorithm, replacing the instantaneous rate $R_{\mathrm{MCS}(k)}$ in (7) and (8) with $R_{\mathrm{eff}}(k)$ provides a natural way of tuning the scheduling algorithm to the particular retransmission scheme implemented in a system.

In the following, we assume that for a general user $k, F_{\mathrm{MCS}(k)}(\eta)$ represents the estimated frame error rate for an input SINR η using $\mathrm{MCS}(k)$, which could take into account the impact of channel uncertainty including mobile speed and channel estimation errors. As mentioned before, $R_{\mathrm{MCS}(k)}$ represents the instantaneous data rate using $\mathrm{MCS}(k)$ as the modulation and coding scheme. There are several possible ways to compute the *effective rate* $R_{\mathrm{eff}}(k)$. We present a list of methods below and compare them through detailed simulations in the next section. The simulation results show that the methods of computing effective rate taking into account the FER and retransmission information have superior performance.

Ranking A: Instantaneous Rate $[R_{\mathrm{inst}}(k)]$*:* This is the conventional approach where the scheduler does not take past and future (re)transmissions into account, i.e.,

$$R_{\mathrm{inst}}(k) = R_{\mathrm{MCS}(k)} \tag{9}$$

Ranking B: ARQ Success Probability Weighted Instantaneous Rate $[R_{AS}(k)]$*:* In this approach, the instantaneous rate obtained from the mapping is weighted by the success probability of the current transmission. Note that the success probability depends on whether it is a new transmission or a retransmission. In the case where it is a retransmission, the success probability depends on the type of hybrid ARQ scheme employed. Formally, the equivalent rate of the current SINR at the Lth transmission of the packet is given by

$$R_{AS}(k) = R_{\mathrm{MCS}(k)} \cdot \left(1 - F_{\mathrm{MCS}(k)}(\eta + \hat{\eta}_L)\right) \tag{10}$$

where η is the current SINR and $\hat{\eta}_L$ is the accumulated SINR of the last $L-1$ transmission(s). In particular, for simple ARQ $\hat{\eta}_L = 0$; and for Chase combining $\hat{\eta}_L$ is the accumulated SINR of the previous $L-1$ transmission(s), i.e., $\hat{\eta}_L = \sum_{k=1}^{L-1} \eta(t_k)$, and $\hat{\eta}_1 = 0$. Note that this scheme provides priority to retransmissions naturally for Chase combining since the success probability is higher for retransmissions.

Ranking C: Average Packet Throughput Based Effective Throughput $[R_{\mathrm{AT}}(k)]$*:* In this approach, the effective

[4] The success probability increases with the value of SINR.

throughput for the current SINR is computed as the ratio of the: 1) average amount of information that the future transmission(s) could carry and 2) average amount of time or resource that will take to carry such information, i.e., (3). It can be shown that this approach achieves the largest set of arrival rates for which the queues remain stable [35]. This ranking can be viewed as a refinement of the $R_{AS}(k)$, given that not only the current transmission but also the effect of possible future retransmission(s) of the packet are taken into account. Clearly, the throughput computation depends on the type of HARQ scheme employed and the channel statistics. As in Section II, we assume that the SINR remains constant for possible future retransmission(s) and derive the equivalent rates at the Lth transmission of the current packet, for both Chase combining and simple ARQ, i.e., (see (11) and (12) at the bottom of the page). where $T_{\max}$ represents the maximum number of transmissions. As $T_{\max} \to \infty$, (11) becomes (see (13) at the bottom of the page.) It should be noted that in (11) the summation in the denominator has to reach the kth term where $F_{\mathrm{MCS}(k)}(k\eta + \hat{\eta}_L)$ is fairly close to zero, which could result in certain computational complexity for aggressive transmissions.

Ranking D: Success Probability Weighted Instantaneous Rate $[R_S(k)]$*:* The instantaneous rate obtained from the SINR to MCS mapping is weighted by the success probability of the current transmission regardless of whether it is a new transmission or a retransmission, i.e.,

$$R_S(k) = R_{\mathrm{MCS}(k)} \cdot \left(1 - F_{\mathrm{MCS}(k)}(\eta)\right). \tag{14}$$

It can be shown that this ranking corresponds to the effective rate in (12) when simple ARQ is used as retransmission scheme.

Ranking E: Approximated Average Packet Throughput Based Effective Throughput $[R_{\mathrm{AA}}(k)]$*:* In this approach, the average packet throughput in (11) is approximated by

$$R_{\mathrm{AA}}(k) = R_{\mathrm{MCS}(k)} \cdot \left(\sum_{m=1}^{T_{\max}-L} \frac{1}{m} \cdot \left(\prod_{q=1}^{k-1} F_{\mathrm{MCS}(k)}(q\eta + \hat{\eta}_L) \right) \times \left(1 - F_{\mathrm{MCS}(k)}(m\eta + \hat{\eta}_L)\right) \right). \tag{15}$$

It can be seen that (10) is an approximation of (15) by using the first term of the summation.

IV. Simulation and Discussion

In this section, we compare the performance of different mapping criteria and scheduling algorithms in the context of a high-data rate downlink packet access system similar (but not standards compliant) to the UMTS HSDPA system [36] that is specifically developed for delay-tolerant data. We have developed a simulation tool that captures the dynamic processes in a radio network based on the Opnet network simulation tool.[5] The simulated radio network consists of a radio network controller (RNC), a base station (Node B), and mobile terminals. Packet flow between the application server and the base station experiences a total networking delay of 50 ms (30 ms between server and RNC, 20 ms between RNC and base station).[6] The mobile terminals share the downlink data channel by time multiplexing, and a scheduler at MAC layer determines the user to be served in each frame. Each scheduling interval or frame lasts 2 ms. We assume that the base station transmits to only one user in each frame at full power. The HARQ functionality resides between the base station and the mobile terminal to permit soft combining and fast NACK/ACK feedback. We do not simulate any kind of handoff. Fig. 1 illustrates the simulated protocol stack from RNC to mobile terminal. We apply the Opnet built-in module to IP, TCP, UDP, HTTP, and FTP, and contribute to the following modules:

Radio Link Protocol (RLP): It performs data block segmentation and reassembly. The ARQ functionality at this layer is enabled. The RLP PDU size is chosen to be 40 bytes.

Medium Access Control $(\mathrm{MAC}_{\mathrm{HSDPA}})$*:* It performs scheduling, MCS selection and HARQ functionality, based on the SINR feedback and the NACK/ACK signaling from each mobile device. For HARQ, the retransmission scheme can be either simple retransmission or Chase combining. The downlink HARQ operates asynchronously where the retransmissions can take place anytime after the NACK/ACK signaling is received

[5]Opnet Technologies Inc., Bethesda, MD; http://www.opnet.com/

[6]These numbers are selected arbitrarily as an example.

$$R_{\mathrm{AT}}^{\mathrm{CHASE}}(k) = \frac{R_{\mathrm{MCS}(k)} \cdot \left(1 - \prod_{q=1}^{T_{\max}-L} F_{\mathrm{MCS}(k)}(q\eta + \hat{\eta}_L)\right)}{\sum_{m=1}^{T_{\max}-L} m \left(\prod_{q=1}^{m-1} F_{\mathrm{MCS}(k)}(q\eta + \hat{\eta}_L)\right) \left(1 - F_{\mathrm{MCS}(k)}(m\eta + \hat{\eta}_L)\right) + (T_{\max} - L) \prod_{q=1}^{T_{\max}-L} F_{\mathrm{MCS}(k)}(q\eta + \hat{\eta}_L)} \tag{11}$$

and

$$\begin{aligned} R_{\mathrm{AT}}^{\mathrm{SARQ}}(k) &= \frac{R_{\mathrm{MCS}(k)} \cdot \left(1 - \prod_{q=1}^{T_{\max}-L} F_{\mathrm{MCS}(k)}(\eta)\right)}{\sum_{m=1}^{T_{\max}-L} m \left(\prod_{q=1}^{k-1} F_{\mathrm{MCS}(k)}(\eta)\right) \left(1 - F_{\mathrm{MCS}(k)}(\eta)\right) + (T_{\max} - L) \prod_{q=1}^{T_{\max}-L} F_{\mathrm{MCS}(k)}(\eta)} \\ &= R_{\mathrm{MCS}(k)} \cdot \left(1 - F_{\mathrm{MCS}(k)}(\eta)\right) \end{aligned} \tag{12}$$

$$R_{\mathrm{AT}}^{\mathrm{CHASE}}(k) = \frac{R_{\mathrm{MCS}(k)}}{\sum_{m=1}^{\infty} m \cdot \left(\prod_{q=1}^{k-1} F_{\mathrm{MCS}(k)}(q\eta + \hat{\eta}_L)\right) \left(1 - F_{\mathrm{MCS}(k)}(k\eta + \hat{\eta}_L)\right)} \tag{13}$$

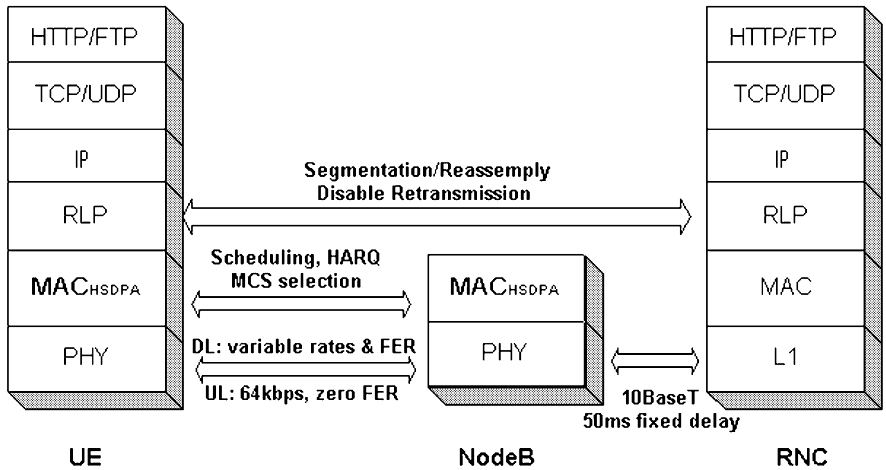


Fig. 1. System architecture.

and the exact time is to be determined by the base station scheduler. To compensate for NACK/ACK feedback delay, i.e., 2 frames or 4 ms delay, HARQ operates in terms of three Stop And Wait (SAW)[37] processes. The multiple process structure allows HARQ to transmit continuously without waiting for NACK/ACK signal. The maximum number of retransmissions is 4. This layer also provides in-sequence delivery to RLP by reordering the received data blocks according to the transmission sequence number (TSN). Therefore, an error-free data block cannot be forwarded to the upper layer if the blocks with smaller TSN are missing. To avoid buffer exhaustion and protocol stall, a timer of 1s is invoked upon receiving an error-free data block that cannot be forwarded [36]. MAC scheduler makes scheduling decision about 1 ms prior to the actual transmission, based on the estimated SINR value at this time. As such, the SINR at the time of scheduling will be different from that during transmission. The backlog includes the packets that already enter the MAC buffer, and exclude the packets that enter the MAC buffer at the time of the scheduling. When choosing the MCS, MAC takes into account the size of backlog, i.e., if the code block size corresponding to the selected MCS is larger than the backlog, MAC will select the next MCS with the smaller code block size that can carry the backlog.

PHY: The physical layer simulation consists of a sequence of events such as transmission and reception of signals, SINR evaluation and estimation. It employs 5-MHz spectrum and 2-ms frame. For simplicity, we assume that the uplink channel (from the mobile terminal to the base station) operates at 64 kbps and 0% FER. For the downlink channel, the frame error is generated by relating the SINR at each mobile terminal to a link level performance curve, i.e., a FER versus SINR curve for each MCS, shown in Fig. 2. We define user geometry as the ratio of the total power received from the signaling base station power to the sum of total power received from the interfering base stations and the noise power. In Rayleigh fading environments, geometry also represents the average received SINR at the user. We simulate two speeds: pedestrian speed of 3 km/h and a typical driving speed of 30 km/h. The channel encounters frequency-flat fading and the channel estimation noise is modeled as AWGN with zero mean and 0.5 dB variance. Jakes model with Doppler corresponding to the speed at 2-Ghz carrier frequency was used to simulate the Rayleigh fading. Physical mobility of the user has not been considered since it is not necessary to consider path loss variations and shadowing variations in addition to Rayleigh fading for the purpose of illustrating the advantage from the proposed schemes. This is especially the case since the scheduling interval is only 2 ms and hence a very large number of time slots are required for physical mobility of the users to take effect.

The system can employ up to 6 MCS schemes. The *4MCS set* involves four MCSs of 640 kbps, 1.28 Mbps, 1.92 Mbps, and 2.56 Mbps while the *6MCS set* consists of 320 kbps, 480 kbps, 640 kbps, 1.28 Mbps, 1.92 Mbps, and 2.56 Mbps. Based on the FER versus SINR performance, we calculate the mapping $\mathrm{MCS}(\eta)$ in terms of a SINR threshold set $\{S_0, S_1, S_2, S_{,3}, S_4, S_5\}$ where $\mathrm{MCS}(\eta) = i$, if $S_i \leq \eta < S_{i+1}$, with $S_0 = -\infty$ and $S_5 = \infty$. We introduce a MCS 0 as no transmission (see Fig. 3). Tables I and II list the SINR thresholds in dB associated with different mapping criteria assuming the *4MCS set* and *6MCS set*, respectively. FER10, FER50, and FER90 represent the mapping design criteria, which limit the FER to 10%, 50%, and 90%, respectively. The RFER (2) and TRPT-sarq mappings are equivalent, so we use TRPT-sarq (6) to represent the two. We also introduce an AGG criterion, which is 3 dB more aggressive than TRPT-chase. We observe that for the given MCS rates, the TRPT-chase criterion recommends aggressive transmissions in low and medium SINRs, i.e., less than 6 dB, but falls back to conservative transmissions for high SINRs. Among all, the FER10 criterion is the most conservative and the AGG criterion is the most aggressive.

Scenario A: Single User With TCP: In this subsection, we study the performance of mapping design in the context of HTTP/TCP performance, since most data applications use TCP as underlying transport protocol. An important consideration here is the tradeoff that although being aggressive could help to improve the MAC layer throughput, retransmissions create delay variations that TCP could falsely interpret as congestion and induce congestion control to scale down the transmission rate. Each mobile user runs HTTP 1.1 application and requests

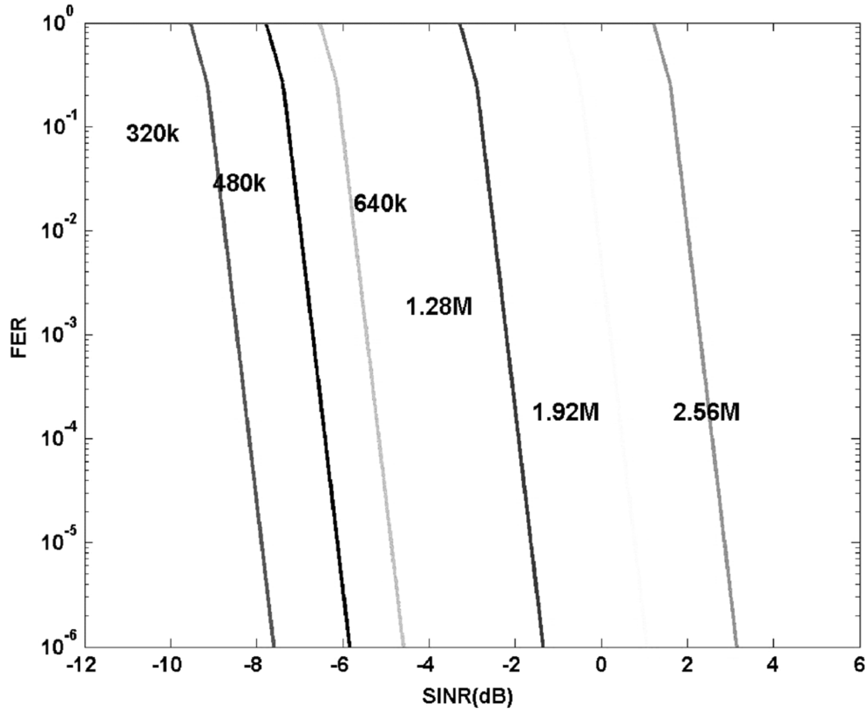


Fig. 2. Link-level FER.

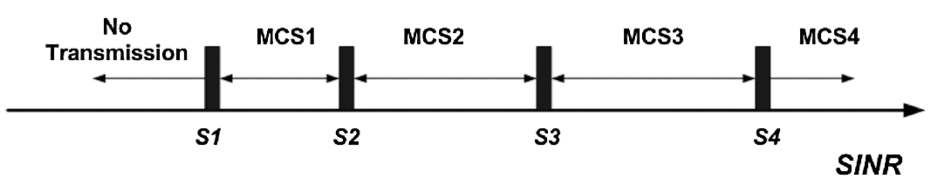


Fig. 3. Link adaptation and mapping thresholds.

TABLE I
SINR THRESHOLD FOR the *4MCS SET* IN Decibels OF VARIOUS MAPPING CRITERIA

	S1	S2	S3	S4
FER01	-5.370	-2.49	-0.08	2.00
FER10	-5.65	-2.77	-0.36	1.72
FER50	-5.97	-3.09	-0.68	1.4
TRPT_sarq	-6.595	-3.09	-0.565	1.6050
FER90	-6.15	-3.27	-0.86	1.22
TRPT_chase	13.595	-3.295	-0.685	1.525
AGG	16.595	-6.295	-3.685	-1.475

a web page of 400 000 bytes every 20 s. We use TCP Reno with 1460 bytes per TCP segment. The performance is measured in terms of the HTTP throughput, $(S(i)/T(i))$, where $S(i)$ represents the byte size of the ith HTTP page and $T(i)$ represents the download time of the corresponding page. Figs. 4 and 5 show the CDF of the HTTP throughput using Chase combining. The proposed TRPT-chase mapping is superior for both 4MCS and 6MCS rate sets at pedestrian speed of 3 km/h. In particular, the average improvement over FER01 is about 50% for the *4MCS set* and 26% for the *6MCS set*, while the improvement over TRPT-sarq and FER50 is about 25% for the *4MCS set* and 10% for the *6MCS set*. Hence, finer rate granularity could reduce the sensitivity to the mapping criterion. This is mainly due to the fact that finer rate granularity reduces the average FER and thus the average number of retransmissions. As such, the contribution to throughput from retransmissions becomes less important. We also observe that in Table II, the difference between TRPT-sarq, TRPT-chase, FER 50, and FER 90 are fairly small.

As speed increases to 30 km/h, the assumption of constant channel SINR during first transmission and following retransmissions becomes less appropriate. In this case, we see that a more aggressive AGG mapping performs slightly better, i.e., 5% than TRPT-chase. For TRPT-chase, these are still 5%–15% improvement over TRPT-sarq, FER10, and FER01 for the *6MCS set*. Once the difference between the feedback SINR and the actual SINR during transmission can be correctly modeled, the link level FER curve can be reshaped accordingly, e.g.,

$$f(\eta) = \int_{\varepsilon=-\infty}^{\infty} f(\eta+\varepsilon)\cdot pdf(\eta,\varepsilon)\,d\varepsilon \tag{16}$$

to take into account the impact of SINR estimation and prediction noise. We also observe that the throughput fluctuation reduces compared to that of 3 km/h. This is mainly due to the averaging effect of the fast channel fading during each HTTP page

TABLE II
SINR THRESHOLD IN Decibels FOR *6MCS SET* OF VARIOUS MAPPING CRITERIA

	S0	S1	S2	S3	S4	S5
FER01	-8.740	-6.985	-5.370	-2.490	-0.08	2.005
FER10	-9.02	-7.265	-5.65	-2.77	-0.36	1.725
FER50	-9.34	-7.585	-5.97	-3.09	-0.68	1.405
TRPT_sarq	-9.595	-7.465	-6.150	-3.09	-0.565	1.6050
FER90	-9.52	-7.765	-6.15	-3.27	-0.86	1.225
TRPT_chase	16.595	-7.585	-6.215	-3.295	-0.685	1.525
AGG	19.595	10.585	-9.215	-6.295	-3.68	-1.475

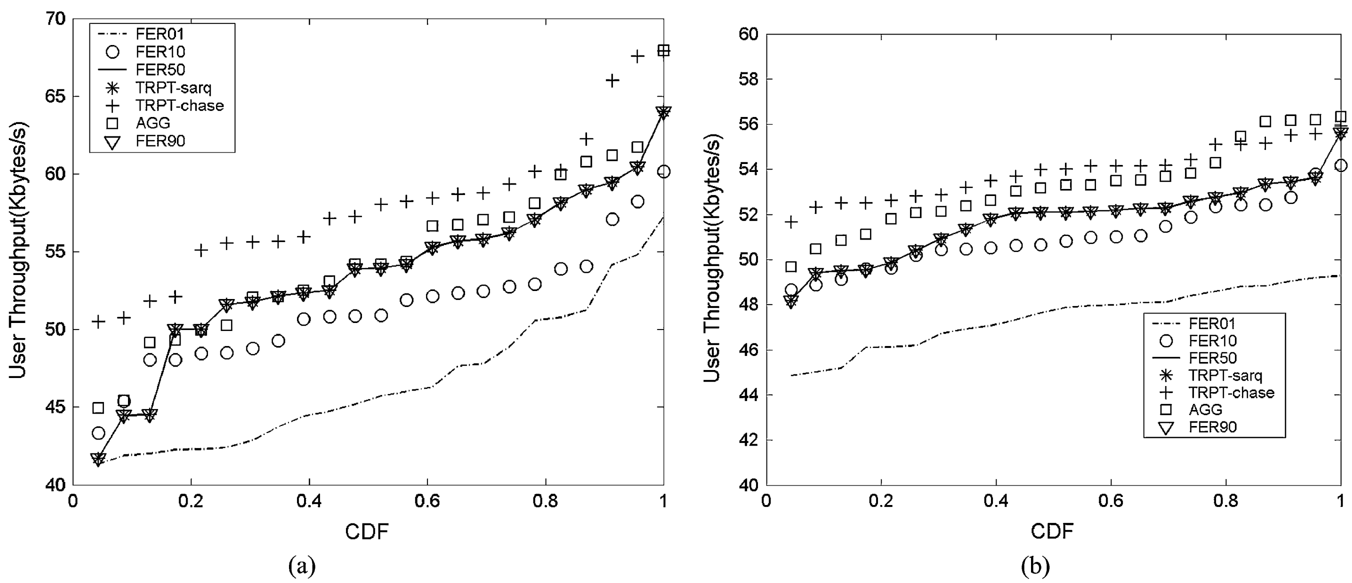


Fig. 4. Single user HTTP performance using Chase combining and *6MCS set*. (a) 3 km/h and (b) 30 km/h.

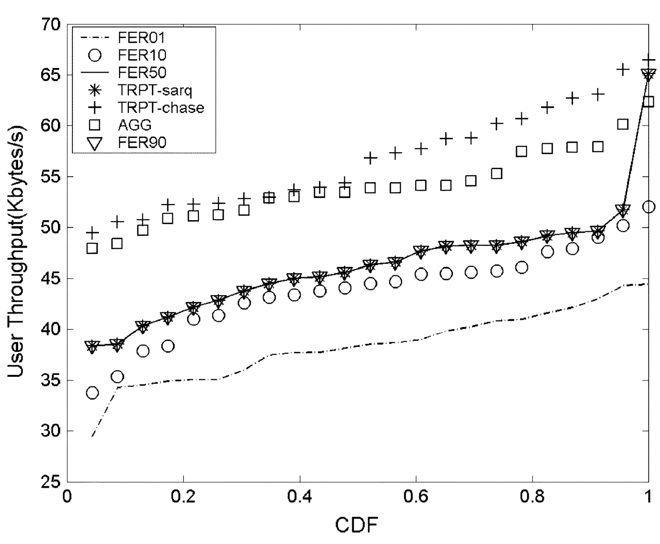


Fig. 5. Single-user HTTP performance using Chase combining and 4 MCS rate set at 3 km/h.

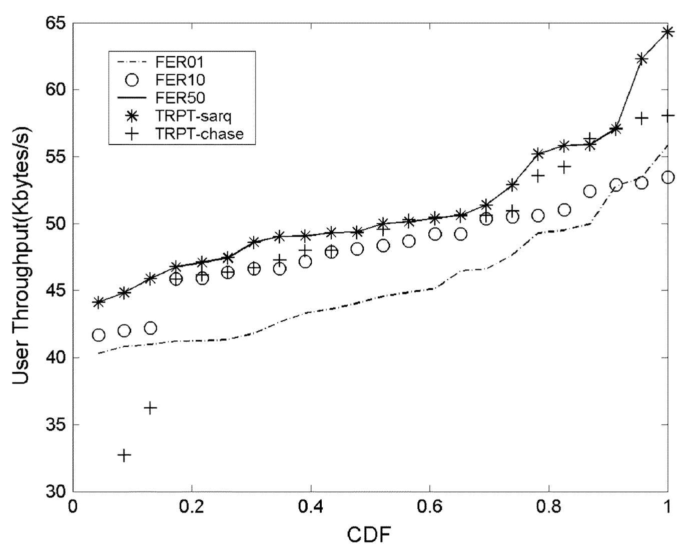


Fig. 6. Single-user HTTP performance using Simple ARQ and *6MCS set* at 3 km/h.

download. Overall, we can conclude that TRPT-chase mapping achieves superior performance regardless of rate profile and user speed. Since the mapping is in general derived off-line, the computational complexity of TRPT-chase mapping is negligible.

In Fig. 6, we investigate the HTTP/TCP performance when simple ARQ is used as MAC layer retransmission scheme. We observe that in this case the retransmissions at RLP layer is necessary since sometimes MAC layer retransmissions cannot fully

TABLE III
USER GEOMETRY VALUES

UE	0	1	2	3	4	5	6	7	8	9
geometry (dB)	1.55	2.83	3.66	4.42	4.54	5.23	6.65	6.67	6.81	7.11

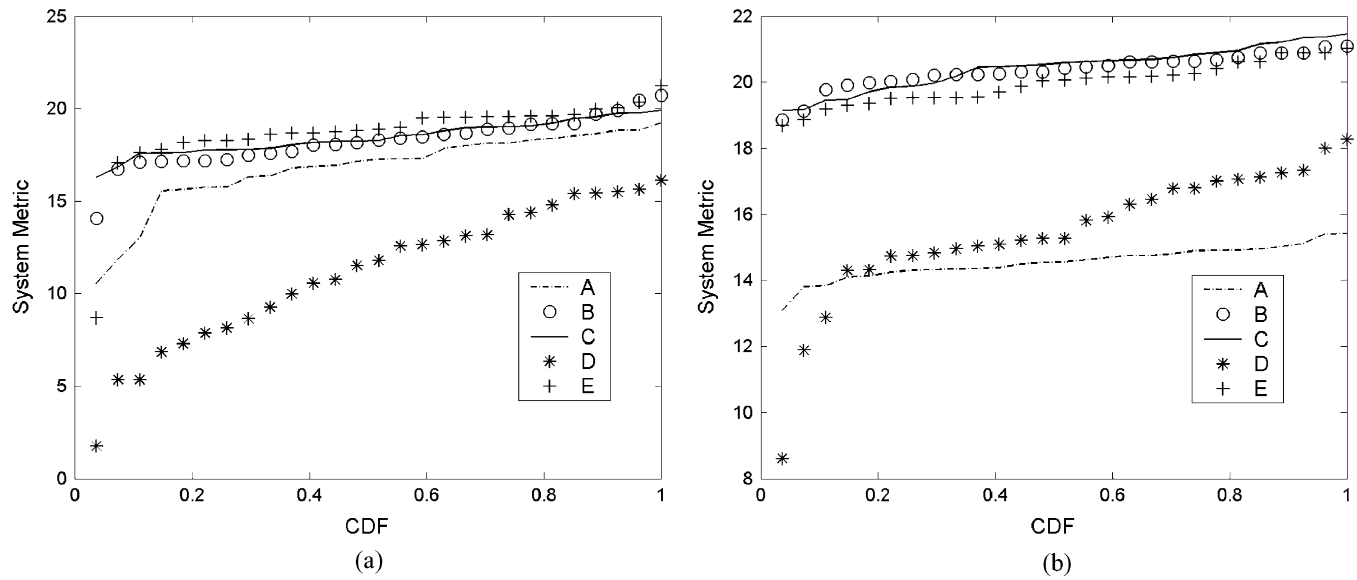


Fig. 7. Multiple-user HTTP performance of various scheduler rankings using Chase combining and TRPT-chase mapping criterion, *6MCS set*. (a) 3 km/h and (b) 30 km/h.

recover transmission errors. Overall, the TRPT-sarq mapping has the superior performance.

Scenario B: Multiple Users With TCP and Modified Scheduler Rankings: Next, we evaluate the performance of a multiuser system using the modified proportional fair scheduler. We replace the $R_{\mathrm{MCS}(k)}$ in (7) with the effective rates $R_{eff(k)}$ in Section 4. The system simulated includes 10 users, each associated with a geometry value which measures the average signal to noise ratio, descending with the user index, as listed in Table III.[7] Each user runs HTTP1.1 over TCP/IP, requesting a web page of 100 000 bytes every 40s. The proportional fair scheduler maximizes the following fairness [22], [38]:

$$\log\left(\prod_{n=1}^{10} R_{MAC}(n)\right) \tag{17}$$

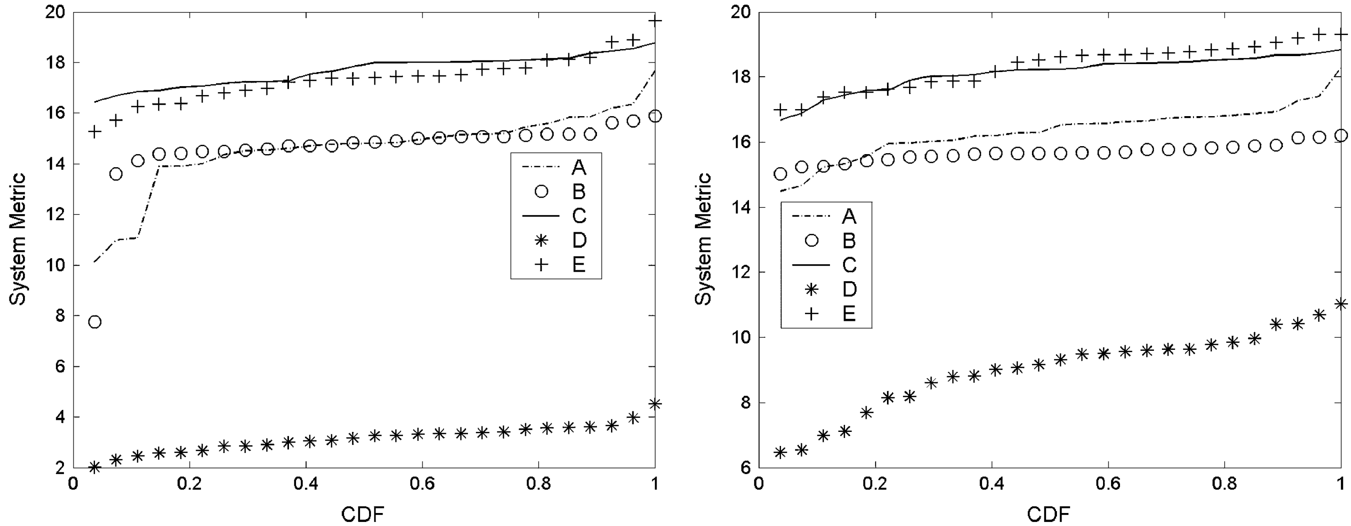
where $R_{MAC}(n)$ represents the MAC layer throughput. To reflect the HTTP/TCP level performance, we modified the above metric to

$$\text{system metric}(i) = \left(\prod_{n=1}^{10} \frac{S_n(i)}{T_n(i)}\right)^{\frac{1}{10}} \tag{18}$$

where $S_n(i)$ represents the size in bytes of the ith HTTP page for user n and $T_n(i)$ represents the corresponding download time. This metric allows us to compare the performance of the different scheduler rankings using a single scalar. Fig. 7 shows the performance of various rankings using TRPT-chase mapping using the *6MCS set*. We plot the CDF of the system metric over different instances of page downloads. Since the channel is time varying and the page download times are not large compared to the coherence time, the system metric is different at different download times. In this experiment, the calculation of Ranking E includes up to 3 retransmissions, i.e., $N = 4$ in (15). We observe that Rankings B, C, and E perform well, and, therefore, Rankings B and E can be utilized in practical systems from a complexity consideration. The essence of proportional fair scheduler is to schedule a user when its channel quality is relatively better than the average. Using TRPT-chase mapping, the average number of retransmissions is below 0.2. Hence, the contribution of future retransmissions to the packet throughput is fairly small, and so is the difference among Rankings B, C, and E. On the other hand, Ranking A performs poorly by ignoring the FER and retransmission information and failing to give appropriate priority to retransmissions. Ranking D estimates throughput incorrectly based on wrong ARQ algorithm, and, therefore, has the worst performance. This again proves the importance of choosing the right scheduler ranking. Quantitatively, using the *6MCS set*, the proposed Ranking C outperforms Ranking A by 10% and Ranking E by more than 30% at pedestrian speed of 3 km/h, and the improvement becomes more than 40%, at 30 km/h.

It is obvious that the sensitivity to scheduler ranking method depends on the level of aggressiveness in link adaptation. To validate this claim, we compare in Fig. 8 the performance of various scheduler rankings, when AGG mapping criterion is used to perform link adaptation. The corresponding average number of retransmissions is in general greater than 0.3. In this case, the contribution of future retransmissions cannot be ignored when comparing packet throughput. Therefore, the performance of

[7]We use a prefixed grometry value for each use to permit a fair comparison of various scheduler rankings.



(a) (b)

Fig. 8. Multiple-user HTTP performance of various scheduler rankings using Chase combining and AGG mapping criterion, *6MCS set*. (a) 3 km/h and (b) 30 km/h.

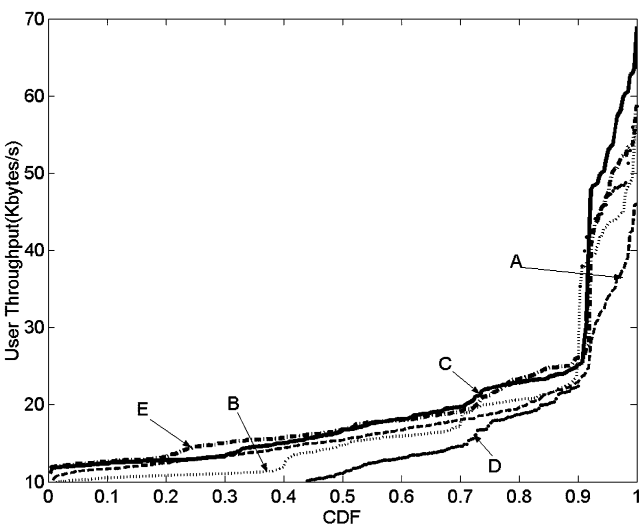


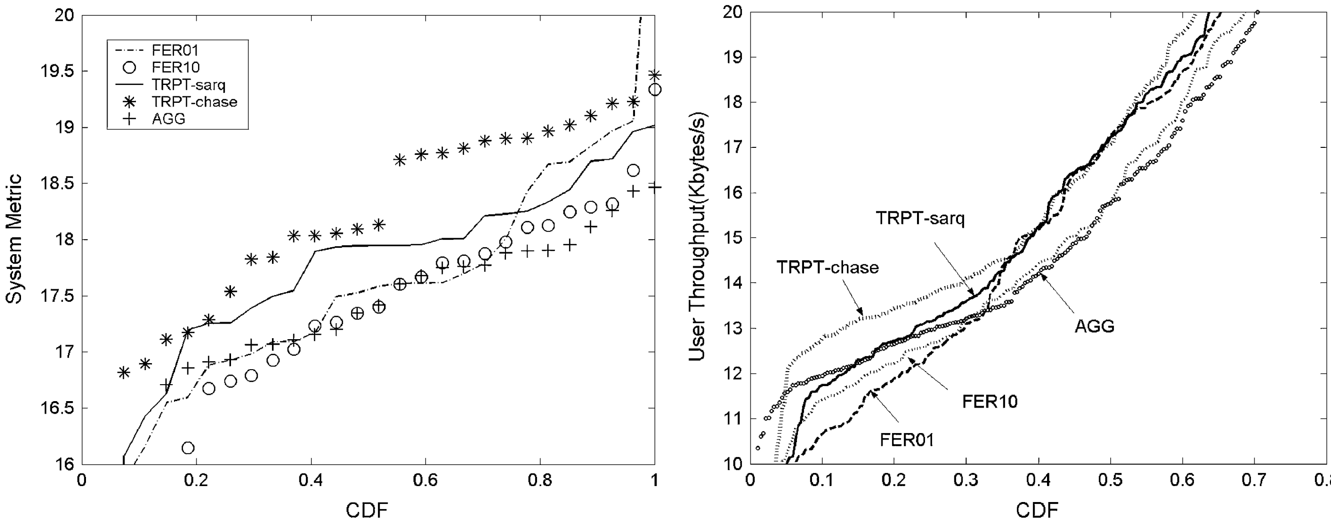
Fig. 9 CDF of user HTTP performance of various scheduler rankings using Chase combining and AGG mapping criterion, *6MCS set*, 30 km/h

Ranking B is about 15% lower compared to those of Rankings C and E. The individual user performance is shown in Fig. 9 in terms of the CDF of user HTTP throughput. We observe that, relatively, Rankings B and D favor users that are in good location over those in bad location or far away from the base station, while Rankings C and E balance among users and allow fast adaptation to channel variations.

Scenario C: Multiple Users With TCP Using Ranking C for Various Mapping Criterion: We evaluate the performance of mapping design in a multiuser environment using the modified proportional fair scheduler with Ranking C. Fig. 10(a) depicts the system metric using different mapping criteria where TRPT-chase mapping achieves the best performance. The user throughput CDF is shown in Fig. 10(b). An overly aggressive mapping like AGG is beneficial to low geometry users but is detrimental to other users since low geometry users consume extra channel resources. On the other hand, conservative mappings like FER01 and FER10 only favor high geometry users.

V. Summary and Future Work

In this paper, we have addressed some of the key design issues associated with the choice of the MCS scheme used for transmission given that an ARQ scheme is being used. We have presented a systematic approach to selecting the mapping between the SINR and MCS schemes, which performs well based on simulation results. Our mapping design criterion is to maximize the throughput, taking into account the ARQ scheme being used. It achieves 5%–50% throughput improvement compared to that of the traditional mapping design. Our results also show that the sensitivity of the performance to the mapping depends on the granularity of the MCS set with decreasing sensitivity for larger MCS sets. The channel estimation and prediction error can also be included in the mapping design by modifying the frame error rate accordingly (16). We have described modifications to scheduling algorithm to utilize the FER and retransmission information in the scheduling process by using the effective rate instead of the instantaneous rate in making the scheduling decision. Several alternatives to approximating the effective rate were considered. The modified proportional fair scheduler achieves 10%–30% performance improvement compared to that of the traditional proportional fair scheduler. The gain from the proposed modification to the scheduler increases with the aggressiveness level of the link adaptation and the user speed. It should be noted that the proposed effective rate computation methods were based on the assumption that the SINR remains constant during the multiple transmissions of the packet. If this assumption is severely violated, as in the case of high Dopplers, the proposed effective rate approach may not result in enhanced performance.



(a) (b)

Fig. 10. User HTTP performance of various mapping criterion using Chase combining, proportional fair scheduler with ranking C, *4MCS set*, 3 km/h. (a) System metric and (b) user throughput.

APPENDIX

Proof of (6):

$$\sum_{k=1}^{N} kx^{k-1} = \frac{\partial\left(\sum_{k=1}^{N} x^k\right)}{\partial x} = \frac{\partial\left(\frac{x(1-x^N)}{(1-x)}\right)}{\partial x}$$
$$= \frac{1-(N+1)x^N + Nx^{N+1}}{(1-x)^2}. \quad \text{(a)}$$

Take (a) into the left hand side of (6), (to simplify notation we replace $F_i(\eta)$ with x on the right hand side of the first equation)

$$\frac{R_i \cdot \left(1-F_i(\eta)^{T_{\max}+1}\right)}{\sum_{k=1}^{T_{\max}+1} k(1-F_i(\eta))F_i(\eta)^{k-1} + (T_{\max}+1)\cdot F_i(\eta)^{T_{\max}+1}}$$
$$= \frac{R_i\left(1-x^{T_{\max}+1}\right)}{\frac{1-(T_{\max}+2)x^{T_{\max}+1}+(T_{\max}+1)x^{T_{\max}+2}}{1-x} + (T_{\max}+1)x^{T_{\max}+1}}$$
$$= R_i(1-x) = R_i(1-F_i(\eta)).$$

ACKNOWLEDGMENT

The authors would like to thank A. Stolyar, G. Rittenhouse, and D. Morgan for several useful discussions. Equation (11) was proposed by A. Stolyar. The authors would also like to thank the reviewers for their detailed comments that have helped us significantly improve the quality of the presentation.

REFERENCES

[1] "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Physical Layer Aspects of UTRA High Speed Downlink Packet Access," 3G, Tech. Rep. 25.848, 2000.

[2] "CDMA2000 High Rate Packet Data Air-Interface Specification," 3GPP, Document C. S00024, 3GPP2, Tech. Specifications, 2002.

[3] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR: A high efficiency high data rate personal communication wireless system," in *Proc. IEEE Veh. Technol. Conf.*, vol. 3, Tokyo, Japan, May 15–18, 2000, pp. 1854–1858.

[4] A. J. Goldsmith and S. G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595–602, May 1998.

[5] K. Balachandran, S. R. Kadaba, and S. Nanda, "Channel quality estimation and rate adaptation for cellular mobile radio," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 7, pp. 1244–1256, Jul. 1999.

[6] B. Vucetic, "An adaptive coding scheme for time varying channels," *IEEE Commun. Mag.*, vol. 9, no. 5, pp. 653–663, May 1991.

[7] D. Costello, J. Hagenauer, H. Imai, and S. Wicker, "Applications of error control coding," *IEEE Trans. Inf. Theory*, vol. 44, no. 10, pp. 2531–2560, Oct. 1998.

[8] S. Lin and D. Costello, *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[9] G. Benelli, "An ARQ scheme with memory and soft error detectors," *IEEE Trans. Commun.*, vol. COM-33, no. 3, pp. 285–288, Mar. 1985.

[10] D. Chase, "Code combining—A maximum likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. Commun.*, vol. 33, no. 5, pp. 385–393, May 1985.

[11] S. Kallel, "Analysis of a type-II hybrid ARQ scheme with code combining," *IEEE Trans. Commun.*, vol. 38, no. 8, pp. 1133–1137, Aug. 1990.

[12] A. Shiozaki, "Adaptive type-II hybrid broadcast ARQ system," *IEEE Trans. Commun.*, vol. 44, no. 4, pp. 420–422, Apr. 1996.

[13] K. R. Narayanan and G. L. Stuber, "A novel ARQ technique using the turbo coding principle," *IEEE Commun. Lett.*, vol. 1, no. 3, pp. 49–51, Mar. 1997.

[14] D. N. Rowitch and L. B. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes," *IEEE Trans. Commun.*, vol. 48, no. 6, pp. 948–959, Jun. 2000.

[15] R. Love, B. Classon, A. Ghosh, and M. Cudak, "Incremental redundancy for evolutions of 3G CDMA systems," in *Proc. IEEE 55th Veh. Technol. Conf.*, vol. 1, May 6–9, 2002, pp. 454–458.

[16] A. Das, F. Khan, A. Sampath, and H. J. Su, "Performance of hybrid ARQ in high speed downlink packet access in UMTS," in *Proc. IEEE Veh. Tech. Conf.*, vol. 4, 2001, pp. 2133–2137.

[17] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, Apr. 1998.

[18] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[19] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with chase combining and incremental redundancy for HSDPA," in *Proc. IEEE 54th Veh. Technol. Conf.*, vol. 3, Oct. 7–11, 2001, pp. 1829–1833.

[20] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 1971–1988, Jul. 2001.

[21] M. Andrews *et al.*, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 150–154, Feb. 2001.

[22] D. Tse, "Forward link multiuser diversity through rate adaptation and scheduling," presented at the Bell Laboratories Presentation, 1999.

[23] A. Stolyar and S. Shakkottai, “Scheduling for multiple flows sharing a time-varying channel: The exponential rule,” in *Translations of AMS*. Providence, RI: American Mathematical Society, 2001, A Volume in Memory of F. Karpelvich.

[24] H. Viswanathan and K. Kumaran, “Rate scheduling in multiple antenna downlink,” in *Proc. 39th Annu. Allerton Conf. Commun. Control*, Oct. 2001, pp. 747–756.

[25] X. Liu, E. Chong, and N. Shroff, “Transmission scheduling for efficient network utilization,” in *Proc. IEEE INFOCOM*, vol. 2, Anchorage, AK, Apr. 2001, pp. 776–785.

[26] T. Ng, I. Stoica, and H. Zhang, “Packet fair queuing algorithms for wireless networks with location dependent errors,” in *Proc. IEEE INFOCOM*, vol. 3, San Francisco, CA, Mar. 1998, pp. 1103–1111.

[27] X. Qiu, J. Chuang, K. Chawla, and J. Whitehead, “Performance enhancement of incremental redundancy in wireless data networks by measurement-based transmission control,” in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, Amsterdam, The Netherlands, Sep., 19–22, 1999, pp. 517–522.

[28] R. Van Nobelen, N. Seshadri, J. Whitehead, and S. Timiri, “An adaptive radio link protocol with enhanced data rates for GSM evolution,” *IEEE Pers. Commun.*, vol. 6, no. 1, pp. 54–64, Feb. 1999.

[29] F. Tang, L. Deneire, M. Engels, and M. Moonen, “A general optimal switching scheme for link adaptation,” in *Proc. IEEE 54th Veh. Technol. Conf.*, vol. 3, Oct. 2001, pp. 1598–1602.

[30] W. Featherstone and D. Molkdar, “Performance analysis of varying the link adaptation BLER switching point in EGPRS,” in *Proc. IEEE 12th Int. Symp. Pers., Indoor, Mobile Radio Commun.*, vol. 2, Sep.-Oct., 2001, pp. F116–F120.

[31] J. Gozalvez and J. Dunlop, “GPRS link adaptation switching thresholds and intervals,” *Electron. Lett.*, vol. 36, no. 15, pp. 1311–1312, Jul. 2000.

[32] M. Zorzi and R. R. Rao, “On the use of renewal theory in the analysis of ARQ protocols,” *IEEE Trans. Commun.*, vol. 44, no. 9, pp. 1077–1081, Sep. 1996.

[33] H. Zheng and H. Viswanathan, “Optimizing TCP performance with hybrid ARQ and scheduler,” in *Proc. IEEE 14th Int. Symp. Pers., Indoor, Mobile Radio Commun.*, vol. 2, Beijing, China, Sep. 7–10, 2003, pp. 1785–1789.

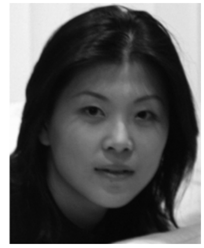
[34] ——, “HARQ aware scheduler in downlink packet data systems,” in *Proc. IEEE Veh. Technol. Conf.* , Oct. 2003, pp. 1578–1582.

[35] A. Stolyar and H. Viswanathan, “Scheduling with effective rates for lossy transmission,” manuscript in progress.

[36] “UTRA High Speed Downlink Packet Access: UTRAN Overall Description,” 3GPP, Doc. No., 3GPP TS 25.308, 2002. 3rd Generation Partnership Project; Technical Specification Group Radio Access Network.

[37] M. Moeneclaey and H. Bruneel, “Efficient ARQ scheme for high error rate channels,” *Electron. Lett.*, vol. 20, pp. 986–987, 1984.

[38] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Trans. Inf. Theory*, no. 6, pp. 1277–1294, Jun. 2002.



Haitao Zheng (S’98–A’99–M’00) received the B.S. degree (highest honors) in electrical engineering from Xian Jiaotong University, Xian, China, in 1995, and the M.S. and Ph.D. degrees from the University of Maryland, College Park, MD, in 1998 and 1999, respectively, all in electrical engineering.

From 1995 to 1998, she was an Institute for System Research Fellow at the University of Maryland. From August 1999 to March 2004, she was with the Wireless Research Laboratory, Bell Labs, Lucent Technologies, Holmdel, NJ. Since March, 2004, she has been with Microsoft Research Asia, Beijing, China. Her research interests include wireless communications and networking, multimedia communications, and signal processing.

Dr. Zheng currently serves as the Technical Programme Committee Member of the IEEE Multimedia Signal Processing Technical Committee, Technical Programme Committee of ICME 2003, Technical Programme Committee of Globecom 2003, Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, Special Issue on Advanced Mobility Management and QoS “Protocols for Wireless Internet” and Guest Editor of the EURASIP *Journal on Application Signal Processing*,Special Issue on Cross Layer Design for Communications and Signal Processing. She has served as a Technical Programme Committee Member of ICME 2002, ICC 2003, and ICASSP 2002, etc. She received the 1998–1999 George Harhalakis Outstanding Systems Engineering Graduate Student Award in recognition of outstanding contributions in cross-disciplinary research from the University of Maryland, she received Bell Laboratories 2002 President’s Gold Award, in recognition of outstanding level of innovation, technical excellence, and business impact.

Harish Viswanathan (S’93–M’97–SM’03) received the B.Tech. degree from the Department of Electrical Engineering, Indian Institute of Technology, Chennai, India, in 1992, and the M.S. and Ph.D. degrees from the School of Electrical Engineering, Cornell University, Ithaca, NY, in 1995 and 1997, respectively.

He is presently with Bell Laboratories, Lucent Technologies, Murray Hill, NJ. His research interests include information theory, communication theory, wireless networks, and signal processing.

Dr. Viswanathan was a recipient of the Cornell Sage Fellowship.