ELSEVIER

Computer Networks 38 (2002) 43–59

COMPUTER NETWORKS

www.elsevier.com/locate/comnet

Packet service in UMTS: delay-throughput performance of the downlink shared channel

Flaminio Borgonovo *, Antonio Capone, Matteo Cesana, Luigi Fratta

Dipartimento di Elettronica e Informazione, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

Received 20 June 2001; accepted 28 June 2001

Responsible Editor: I.F. Akyildiz

Abstract

The Universal Mobile Telecommunications System, wide band code-division multiple access (W-CDMA) radio interface is characterized by great flexibility and a variety of different physical and logical channel types: for example, on the downlink the dedicated channels (DCH) offers circuit switching, the downlink shared channel (DSCH) and forward access channel use packet switching, the former with closed loop power control. Furthermore, several user rates and protections are possible, by choosing suitable parameters, such as spreading factors (SFs), code rates and automatic repeat request (ARQ) schemes. In this paper we present the results, obtained by a detailed simulation, about the effect of several parameters and system alternatives on the capacity of the downlink segment of the W-CDMA interface with packet service. In particular, we investigate the effect of the SF, the code rate, the ARQ and the power control on the DSCH capacity and delay-throughput performance.

Keywords: UMTS; WCDMA; Packet service; Third generation wireless

1. Introduction

In the story of the extraordinary evolution of telecommunications systems and their impact on the mass market in the last ten years, two stars can be easily recognized: mobile communications systems and data networks based on IP technology. The use of IP applications such as web and e-mail has pushed the data traffic to grow quickly and to become now of the same volume of the voice traffic. At the same time, the second generation cellular systems, such as GSM, have heavily changed the way in which users access the telecommunication services. However, in this scenario a strange discrepancy is that the market of mobile communication is still dominated by the old telephone service even if second generation cellular systems can also provide low rate data services. For this reason the challenge of third generation mobile communication systems is to provide access for a wide range of multimedia applications and services.

Universal Mobile Telecommunications System (UMTS) [1–3] is the third generation mobile communication system developed by ETSI, the European Telecommunications Standard Institute,

* Corresponding author. Tel.: +39-2-23993637; fax: +39-2-23993413.

E-mail addresses: borgonov@elet.polimi.it (F. Borgonovo), capone@elet.polimi.it (A. Capone), cesana@elet.polimi.it (M. Cesana), fratta@elet.polimi.it (L. Fratta).

1389-1286/02/$ - see front matter
PII: S1389-1286(01)00239-0

which will use a new frequency spectrum [4] and will extend the present GSM service to include multimedia. UMTS is also in the family of standards considered by International Telecommunications Union for IMT-2000 [5].

The Third Generation Partnership Project, which is now in charge of the standardization process, has defined two schemes for the radio access [6], UMTS Terrestrial Radio Access (UTRA), a wide band code-division multiple access (W-CDMA) which adopts frequency division duplexing, and time division CDMA (TD-CDMA) which is based on time division duplexing.

Due to the effort of the standardization bodies, the radio interface is characterized by great flexibility and a variety of different physical and logical channel types. For instance, several user rates and protections are possible by choosing suitable parameters, such as spreading factors (SFs), code rates and automatic repeat request (ARQ) schemes [7]. This approach is quite different from that adopted by second generation systems where a small set of services can be implemented by vendors and provided by operators. If from one side this added flexibility is an advantage of UMTS, from the other it makes the task of real services implementation more complex. In fact, the complexity of detailed services definition and system parameters optimization has been moved out of specifications and let to UMTS vendors and operators.

Among the new services offered by UMTS, the packet data service is one of the most critical mainly because of the characteristics of the code division access scheme adopted at the radio interface. Up to now, no study that thoroughly investigates the effects of the different parameters settings on the performance of UMTS data services has been published yet.

In this paper we study the performance of the downlink segment of the W-CDMA interface with packet service and present the results obtained through detailed simulation of the UMTS system. After a brief overview of UMTS radio interface, aimed to provide the basics to the readers that are not familiar with the system (Section 2), we present in Section 3 the problems related to the implementation of packet data services. In Section 4 we describe the system model adopted for simulations and in Section 5 we discuss the results obtained. Conclusions are given in Section 6.

2. UTRA basics

In January 1998 ETSI, the European Telecommunications Standard Institution, has selected two access schemes for its radio interface, the W-CDMA scheme and the TD-CDMA, to be used in the paired part of the spectrum assigned to UMTS, 60 MHz from 1920 to 1980 MHz (uplink) and 60 MHz from 2110 to 2170 MHz (downlink), and in the unpaired part, 35 MHz from 1900 to 1920 MHz and from 2010 to 2025 MHz, respectively.

The W-CDMA scheme adopts a QPSK modulation and a chip rate of 3840 Mchip/s for the in-phase and quadrature channels. It presents a carrier separation of 5 MHz, so that up to 12 carriers can be defined in the available bandwidth. For the downlink direction a QPSK modulation is adopted after spreading, while for the uplink direction the in-phase and quadrature channels are used to transmit two BPSK flows with different spreading codes [8].

The spreading process is based on two codes, namely the spreading code and the scrambling code. The spreading code increases the flow bit rate to the chip rate of the air interface according to the SF. Different values of SF ranging from 4 to 512 are available and they are obtained using a tree of orthogonal codes. The tree has the characteristic that two codes, even with different SF, are orthogonal if they are located in different branch of the tree. Multiple trees can be generated using a scrambling code which varies the order of chips. The channels transmitted by the same station (base or mobile) can use codes in the same tree so that the mutual interference is greatly reduced, while channels transmitted by different stations should use different scrambling codes so that the mutual interference is almost independent of the delay offset at the receiver due to different propagation paths.

Physical channels are defined by the associated spreading and scrambling codes. The bit flow is divided into time slots (625 μs). During a time-slot both physical control bits and data bits can be

transmitted, and while the number of chips is fixed, the total number of bits depends on the SF. The minimum transmission unit offered by the physical layer to the upper layers is the transmission time interval (TTI), also called frame in the following, and is composed of multiples of 16 slots.

The transport services provided by the physical layer to the upper layers are based on transport channels which are mapped into the physical channels [9]. Transport channels are divided into dedicated and common channels.

Dedicated channels (DCH) are used to transmit user and control information in the uplink and downlink direction and are devoted to the connection between a single mobile station and the UTRA Network. A DCH is mapped into two physical channels, namely the dedicated physical data channel (DPDCH) and the dedicated physical control channel (DPCCH). The DPDCH carries user data while the DPCCH carries physical signaling used to control the channel. In particular, in the DPCCH the pilot symbols are transmitted for channel response and interference estimation, the transport format combination indicator symbols describe the format adopted for the channel (mainly SF and error protection code), the transmit power control (TPC) symbols are used to transmit the commands for the closed loop power control algorithm, and the feedback information symbols are used to perform a closed-loop signal-quality control. In the downlink direction the DPDCH and the DPCCH are time multiplexed into each slot, while in the uplink direction they are transmitted into the in-phase and quadrature channels.

Common transport channels are used to transmit both control and user information. Among the channels mainly used for control are the broadcast channel, used to broadcast system information, the paging channel, used to transmit downlink control information into a location area, and the synchronization channel used for mobile synchronization control. The random access channel and the forward access channel (FACH) can be used in each cell to transmit control information or packets in uplink and downlink direction, respectively. Finally, the common packet channel in the uplink and the downlink shared channel (DSCH) in the downlink are used for packet only.

The physical layer directly controls the transmission power of physical channels. The power control exerted on the DCH is based on a closed-loop signaling, outlined in the following. The TPC symbols in each slot carry a command for increasing or decreasing the transmission power in each direction. The power step is fixed and usually is equal to 1 dB. On the receiving side, if the estimated signal-to-interference ratio (SIR) after despreading is lower than a SIR target value, an increase command is sent in the subsequent slot. A decrease command is sent otherwise. The SIR target value is controlled by an outer control loop which is based on the quality of the decoded bit flow. Other channels adopt different power control mechanisms. As an example, the DSCH is not directly power controlled, but its transmission power is computed on the basis of the power of the DCH associated to the user actually transmitting on the DSCH. This implies that each mobile station involved in the DSCH packet transmission has an active DCH.

The information received by higher layers can be protected by the physical layer using forward error correction (FEC) codes [10]. The basic coding schemes use convolutional codes with rate $1/3$ or $1/2$, or a turbo code with rate $1/3$. Different rates can be obtained using the rate matching process which can increase the code rate by means of puncturing.

On top of the physical layer in the user plane, the link layer is split into the medium access control (MAC), the radio link control (RLC) and packet data convergence protocol (PDCP) [11]. The MAC layer [12] provides logical channels to the RLC and maps logical channels into transport channels. On common channels, the MAC provides addressing of user equipments and scheduling of packet data units (PDUs). The RLC layer [13] can offer a transparent or non-transparent data transfer mode. With the non-transparent mode it adds control information on each transmitted PDU and an error check on each received PDU. If the unacknowledged mode is selected, erred PDU are simply discarded, while with the acknowledged mode an ARQ mechanism is adopted. Finally, the PDCP maps each network layer instance into one RLC entity and performs

higher layers header compression, if required (for example TCP/IP header compression).

3. Downlink packet data services

As mentioned in the previous section, the provision of transport services at the radio interface of UMTS systems is a complex task since many possible configurations are available and a large number of parameters are involved.

In the downlink direction, data packets can be transmitted on three different channel types, namely the DCH, the DSCH and the FACH.

The DCHs are assigned to single users through set-up and tear down procedures and are subject to closed loop power control that, if used for circuit service such as voice, stabilizes the bit error rate (BER) and optimizes CDMA performance.

In alternative to the use of dedicated channels, data transmissions of many users can be time multiplexed on the DSCH. No set-up and tear down procedures are required and the physical channel on which the DSCH is mapped does not carry power control signaling. However, since the closed loop power control is still required, users must have an associated active DCH to access DSCH services.

The FACH is shared by many users to transmit short bursts of data, but, unlike DSCH, no closed-loop power control is exerted and no DCH must be activated to access this channel.

Since for each one of the above channels different combinations of SF and code rate can provide the bandwidth and the protection required for different services and environments, the problem to select the best choice arises.

For real-time circuit traffic, well known results show that CDMA with closed-loop power control is very effective in spectrum exploitation [14]. Efficiency can be further enhanced by using powerful FEC codes, which have been proved more effective than spreading codes [15].

The scenario with packet service traffic is different. In fact, due to its burstiness and depending on the number of interfering channels and their power levels, errors can be more efficiently obviated by ARQ techniques than by forward error correcting codes [16,17]. For the same reason, the protection obtained with high SFs is questionable. Also the effectiveness of the closed loop power control is to be verified, since the frequent changes in the SIR due to bursty transmissions could be tracked with difficulty by the power control mechanism.

Finally, as data service can be delivered also with circuit switching technology, it is interesting to investigate whether UMTS achieves the highest data throughput with circuit or packet switching technology. For example, the capacity, assuming equal rate sources and circuit switching technology, is defined by the maximum number of channels that can be accommodated by the system without exceeding a given BER. The addition of any further channel beyond the capacity cannot be accepted since the BER will increase beyond tolerable values, due to power limitations.

On the contrary, with packet switching, occasional increases in BER over its target value can be tolerated, because of the use of ARQ techniques. For this reason, some traffic, and interference, can be added without impairment even beyond the point at which the power control can no further guarantee the BER target. Therefore, throughput gain with respect to the circuit switching case is expected.

The aim of our investigation, presented in the remaining part of this paper, is to provide a quantitative evaluation of the performance of the different alternatives so far outlined and to support the UMTS design in optimizing the parameters setting.

4. Simulator description

The UMTS system considered is composed of 49 hexagonal cells laying on a torus surface to avoid border effects. The base stations (BS) are located at the center of each cell and irradiate with omni-directional antennas with unit gain.

4.1. Propagation model

In the propagation model assumed in this work, according to the guidelines of ETSI [18], the received power P_r is given by

$$P_r = P_t \alpha^2 10^{\epsilon/10} L \quad (1)$$

where P_t is the transmitted power, L is the path loss, $10^{\epsilon/10}$ accounts for the loss due to slow shadowing, being ϵ a normal variate with zero mean and σ^2 variance, and α^2 represents the gain, with a negative exponential distribution of unit mean, due to fast fading.

In the following we refer to a macro-cellular environment, for which the cell radius is 300 m, and the path loss L is expressed as

$$10 \log L = -(128.1 + 37.6 \log r) \text{ (dB)}$$

where r (in kilometers) represents the distance between the mobile and the base. Furthermore, we assume no fading and shadowing standard deviation equal to 5 dB.

When a new user is generated, its position is chosen at random over the torus surface and the path losses for the radio links toward all BSs are determined. The user is assigned to the BS with the minimum attenuation. No user mobility is considered at this stage.

Each cell is assigned a single tree of orthogonal variable spreading factors, so that channels in the same cell are always orthogonal. The loss of orthogonality of the received signals due to the multipath effect is accounted for in the receiver model as specified in Section 4.4.

4.2. Traffic model

As we shall see in the following, the performance of the UMTS downlink heavily depends on the input traffic characteristics. Here we have adopted a basic traffic model where users become active according to a Poisson point process of intensity λ, as described in Ref. [18]. Each user, upon activation, generates a flow of packets whose length is negative exponentially distributed with mean 3840 bits. The packet flow is composed of a number of packets, geometrically distributed with mean $N_p = 25$ and packets arrive according to a Poisson point process whose intensity is chosen to match a given source speed.

In the present investigation we have chosen a source speed equal to 929.4 kb/s, for the following three reasons. First, as no user multiplexing is allowed within a 10 ms frame, the packet arrival rate must be high enough to keep the average frame utilization sufficiently high. Second, the user must be able to generate packets at a speed comparable with the maximum throughput, which is, as we will find later, around 1000 kb/s. Third, the chosen speed reflects the rate at which present and future Web servers can transmit files.

A user leaves the system as soon as the last packet has been transmitted.

4.3. Transmission model

The packets generated by each user are delivered to the RLC layer where they are subdivided into transmission blocks before being queued for transmission. Each transmission block includes an RLC header of 16 bits that also account for an ARQ mechanism.

At each frame, the MAC layer chooses an user queue according to the scheduling mechanism and, after adding the MAC header, sends to the physical layer a number of blocks up to filling the space available in the frame. Before transmission, the physical layer adds the redundancy bits according to the coding scheme adopted.

Our simulator adds the parity bits required by convolutional codes, with 256 states, constraint

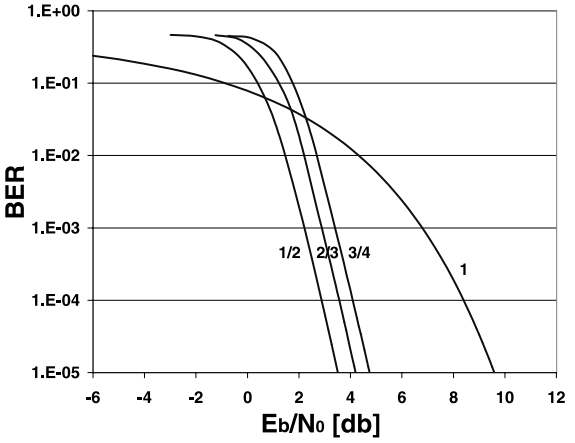

Fig. 1. BER of the convolutional codes adopted in UMTS as function of the bit normalized energy.

Table 1
Physical parameters adopted in the physical layer of the downlink physical shared channel

Rate	SF	Frame bits		Physical speed (Kb/s)	Number of TB in a frame	Length of TB (bits)	Expansion factor
		Gross	Net				
$R = 1$	SF = 4	19200	1879.2	18792	24	749	1.02
	SF = 8	9600	9396	939.6	12	749	1.02
$R = 1/2$	SF = 4	19200	9492	949.2	12	757	2.02
	SF = 8	9600	4746	474.6	6	757	2.02
$R = 2/3$	SF = 4	19200	12608	1260.8	16	754	1.52
	SF = 8	9600	6304	630.4	8	754	1.52
$R = 3/4$	SF = 4	19200	14166	1416.6	18	753	1.355
	SF = 8	9600	7083	708.3	9	753	1.355

length $K = 9$ and optimal puncturing, whose BER, obtained through link level simulations [19], is shown in Fig. 1. In particular we have considered code rates, SFs and block sizes such that the bits introduced by rate matching, which add overhead without increasing error protection, are very few. Table 1 shows the physical parameters that have been used in our simulations. Furthermore, we have used, with different codes and SFs, almost the same transport-blocks size to avoid throughput differences due to different mappings of bit from packets to blocks. A block length of about 750 bits has been found optimal with respect to the maximum throughput. In fact, with the chosen traffic profile, a longer block size is inefficient because blocks can hardly be filled, whereas a smaller size increases block overheads.

4.4. Receiver model

At the receiving side and for each transmission block the carrier to interference ratio is evaluated as

$$\frac{C}{I} = \frac{P_{\mathrm{r}}}{\alpha I_{\mathrm{intra}} + I_{\mathrm{inter}} + P_N} \tag{2}$$

where P_N is the thermal noise assumed equal to -99 dBm, I_{inter} is the sum of the signal powers received from the other cells, I_{intra} is the sum of the signal powers received from other users in the same cell, and α is the loss-of-orthogonality factor that, according to Ref. [20], is assumed equal to 0.4. All the received powers are obtained by Eq. (1).

From the C/I evaluated as in Eq. (2) the normalized energy per information bit is obtained as

$$\frac{E_b}{N_0} = \frac{1}{2R} \times \mathrm{SF} \times \frac{C}{I}. \tag{3}$$

where R is the coding rate. For each value of the ratio E_b/N_0 the curves in Fig. 1 give the corresponding BER values which allow to obtain the block error rate (BLER) as

$$\mathrm{BLER} = 1 - (1 - \mathrm{BER})^l, \tag{4}$$

l being the transmission block length.

The correctness of the transmission is then decided by testing the value of a normalized random variable against the BLER.

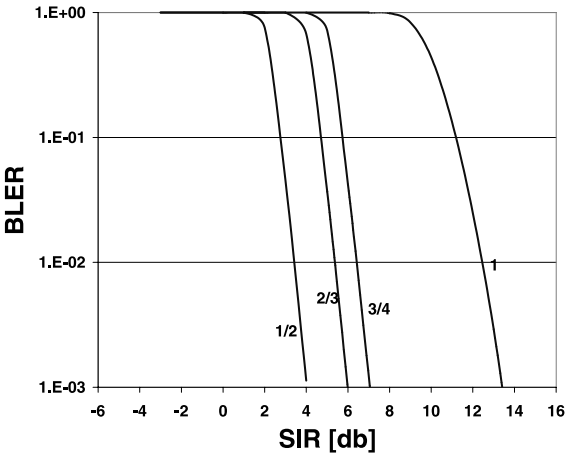

Fig. 2. BLER of the convolutional codes adopted in UMTS as function of the signal to interference ratio after despreading.

Fig. 2 shows the BLER of blocks of 750 bits versus the SIR after despreading, which is defined as $\mathrm{SIR} = \mathrm{SF} \times C/I$.

Our simulator assumes an ideal ARQ procedure, i.e. the transmitted block is kept in the transmitting queue in case of error and is canceled otherwise. After 10 failed transmissions the block is dropped and the user is declared in outage.

4.5. Power control model

The closed loop power control mechanism adopted for DCHs uses two loops. The inner loop controls the transmitted power to maintain the SIR at the target value, whereas the outer loop controls the SIR target to provide a target BLER. This last control mechanism has been envisaged to provide different qualities to different services. As in our simulation we investigate a service at a time, the corresponding BLER requirement can be assumed constant and therefore we have implemented the inner loop only, treating the BLER as an input.

Though the UMTS specifications state that power-update requests of ± 1 dB are transmitted every time slot (0.625 ms), we have assumed to transmit power updates every frame (10 ms) to simplify and speed up the simulator. To meet the dynamic of the real mechanism the power updates are in the range of ± 16 dB. This simplification let all blocks in a frame, which belong to the same user, to be transmitted at the same level. However no impact is expected on the convergence of the mechanism though some slight differences in the number of retransmitted blocks are possible. The value of power updates, with the constraints stated above, are determined at each frame by the difference between the SIR target and the SIR evaluated on the last frame.

Physical power constraints are also added as specified in Ref. [20]. Each channel cannot exceed a transmitted power of 30 dBm, whereas the overall power transmitted by a BS is limited to 43 dBm. The constraint on the channel power is first enforced by setting at 30 dBm the transmission power of channels exceeding the limit, and then the BS power constraint is checked. If it is not satisfied, the transmission power of all channels is proportionally reduced to obtain the global transmitted power equal to 43 dBm.

5. Simulation results

The simulation model described in the previous section has been implemented in C++. Due to the complexity of the overall system and the interaction among system parameters and performance variables, a simple and straightforward analysis is impossible. We have been forced to split the study in several sub-problems and take simplifying assumptions.

First, as it will be made clearer later, the use of DCH control channels beside the PDSCH (physical DSCH), increases the global interference suffered at the receiving end and, consequently, reduces the available capacity, even when the user is queued and not transmitting on the DSCH. In turn, the reduced capacity increases the number of queued users, therefore causing a positive feedback which leads to instability and zero throughput as the queue size becomes infinite. As this dependence adds complexity to the analysis, in order to understand the effect of the basic mechanism, we have at first investigated the simplified system in which the interference generated by DCHs is not taken into account into the SIR that determines the performance of the data channel and the power control updates.

With this assumption we have investigated the effect of spreading factor and code rate on the single PDSCH (Section 5.1), the effect of user traffic (Section 5.2), and the performance with multiple physical channels (Section 5.3). Then, once acquired a clearer understanding of the complex mechanisms which affect the system behavior, we have considered the DCH interference and have investigated its effect on the DSCH throughput (Section 5.4). The trade off between power control benefit and control channel (DCH) interference is investigated by comparing FACH and DSCH throughputs (Section 5.5).

Finally, we must note that the operation with ARQ makes the system intrinsically unstable for several values of system parameters. In fact, when the system operates far from capacity, an

interference increase on a channel is counteracted by a power increase on the same channel, to maintain the SIR target. However, when the maximum power allowed on a channel is reached an increase in the interference causes errors that are counteracted by retransmitting the packets. Retransmissions increase the channel traffic G and consequently the interference. In these conditions, it may happen that the maximum interference tolerable by the system is attained with a traffic channel G less than 1, and that a further increase in G, due to statistical fluctuations in the number of retransmissions, causes a strong decrease in throughput due to the excessive number of errors.

To avoid such an occurrence, we have introduced a linear back-off (BO) mechanism, which tries to limit the instability by refraining from transmitting. More precisely, when a channel has reached its maximum transmitting power and an error occurs, a block is transmitted with a delay that increases linearly from 1 to $K + 1$, where K is the number of failed transmissions of the same block.

5.1. Effect of codes

Fig. 3 shows the average delay suffered by a packet versus the throughput when one PDSCH is adopted with SF = 4 and for codes with $R = 1$, $3/4$, $2/3$ and $1/2$. In all cases we have adopted the linear back-off mechanism described above. The best performance is obtained with light codes ($R = 3/4$ and $2/3$). Heavier codes ($R = 1/2$) achieve a poorer performance since the added redundancy bits reduce the throughput and practically no benefit is obtained by the excess protection. The low throughput measured in the case of no error correction ($R = 1$) is due to the high interference and shows that the protection of the spreading process with SF = 4 is not sufficient to fight interference.

To get a deeper understanding of the effect of the different parameters we analyze the simulation results on the behavior of the throughput, the average fraction of transmissions at the maximum power (saturation fraction) and the BLER versus the channel traffic, G, shown in Figs. 4–6, respectively. Let us consider the case of codes with $R = 1$. To achieve a BLER low enough (10^{-2}) to be suitable for correct ARQ operation a SIR target of 13 dB must be adopted (see Fig. 2). To guarantee such an high value and due to the limited transmission power, the power control drives an high fraction of stations into saturation (Fig. 5) as soon as G increases. Correspondently, the BLER increases (Fig. 6) and the BO mechanism limits the channel traffic to a maximum value, $G = 0.38$ and the throughput is bounded to 420 kb/s. In this case the system is wrongly designed: the no error protection ($R = 1$) requires to keep the interference

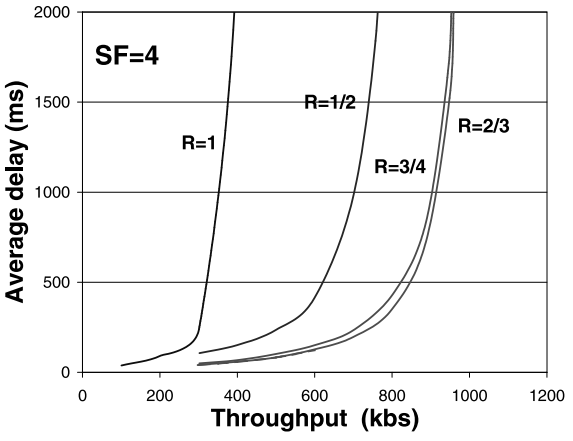

Fig. 3. Average delay versus throughput for SF = 4 and different code rates with the basic traffic model.

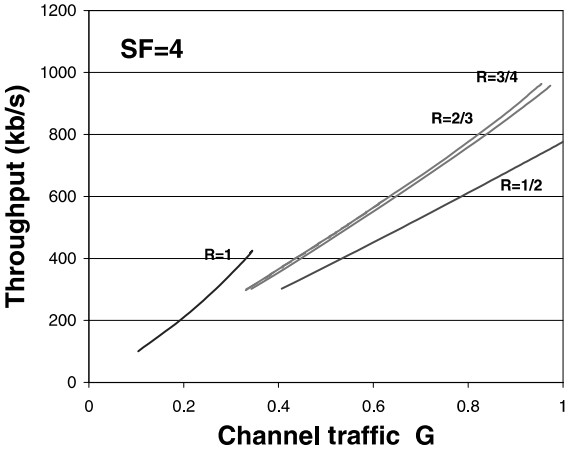

Fig. 4. Throughput versus the channel traffic G for the cases reported in Fig. 3.

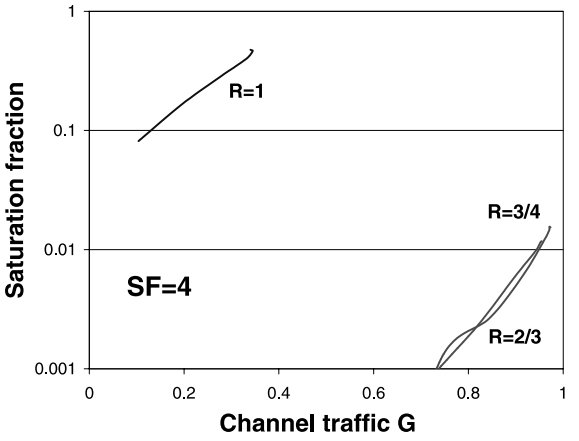

Fig. 5. Fraction of user in saturation versus the channel traffic G for the cases reported in Fig. 3.

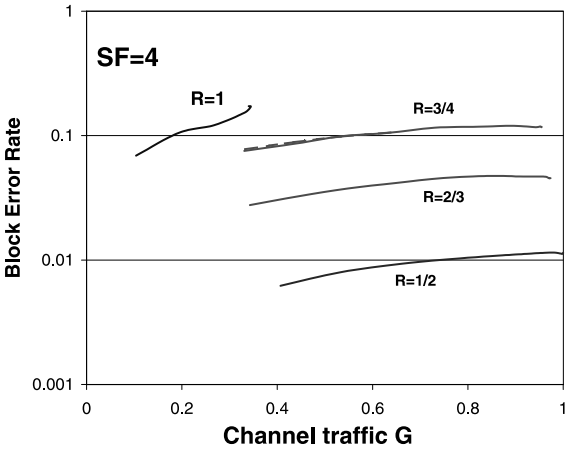

Fig. 6. BLER versus the channel traffic G for the cases reported in Fig. 3.

very low (high SIR) and excessively throttles the traffic and the throughput. To increase G we must lower the SIR target and consequently introduce some correcting codes. As an extreme case we considered, let us assume a code with $R = 1/2$. From Fig. 2 the SIR target is reduced to 4 dB. In this case the power control provides the SIR target without driving any station into saturation. The BLER is below 0.01 and the traffic limiting effect of the BO is negligible: G reaches 1. However, the system throughput is limited to 800 kb/s since it is throttled by the excessive code redundancy. Again the system parameters setting is far from optimum. The adoption of codes with intermediate rates, $R = 3/4$, $2/3$, provides much better results. For both rates the SIR target from Fig. 2 is around 7 dB and it is guaranteed by the power control with a limited (below 1%) fraction of saturated stations. The BO is activated and limits G to 0.955 for $R = 3/4$ and to 0.97 for $R = 2/3$. Both cases provide almost the same maximum throughput equal to 980 kb/s.

Note that, even if the traffic limitation is very small, the measured BLER (Fig. 6) is much higher than the expected value (Fig. 2). This is because the power control mechanism is not able to keep the SIR close to its target value when the traffic and interference are bursty. In our simulations we have measured SIR standard deviation values in the 3.7–4.3 dB (note that all interference and SIR statistics are evaluated considering logarithmic values throughout the paper). As a consequence of this behavior, the errors introduced by the radio channel are not independent and occur in burst when the SIR is too low.

In Figs. 7 and 8 we show the effect of the SIR target on the maximum throughput and the BLER, respectively. If a too small SIR target is chosen too many errors occur since the SIR fluctuations around the target value quite often drive the system in a condition where the code protection is useless. On the other side, with a high SIR

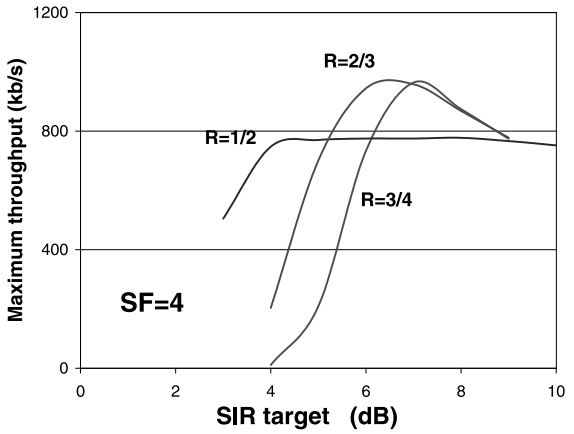

Fig. 7. Maximum throughput as function of the SIR target for the cases $R = 3/4$, $2/3$ and $1/2$.

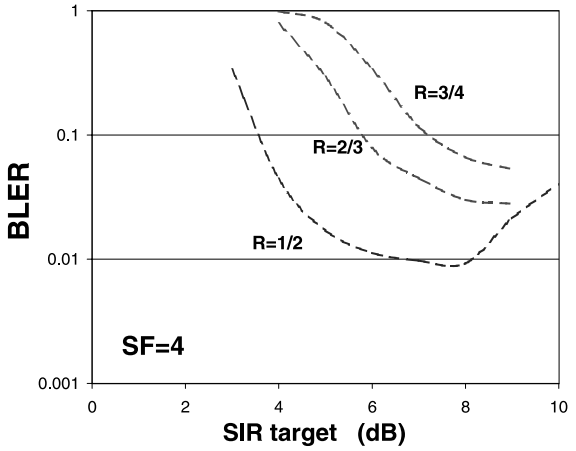

Fig. 8. BLER as function of the SIR target for the cases $R = 3/4$, $2/3$ and $1/2$.

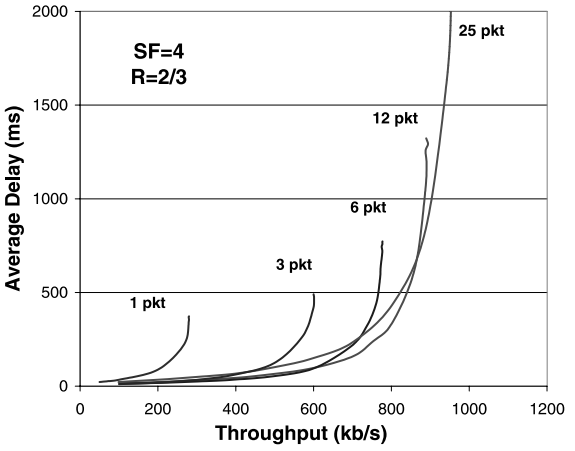

Fig. 9. Average delay versus throughput for SF = 4 and $R = 2/3$ as the average number of packet per user changes.

target the power requirement increases and too many users are driven into saturation with a consequent increase of the BLER. For each code an optimum value of SIR target exists. In our simulation we have assumed these optimal values, namely 7 dB for codes with $R = 3/4$ and $2/3$ and 6 dB for the code with $R = 1/2$.

Finally, from the comparison between Figs. 8 and 2 we observe that the SIR target needed for packet operation is at least 3 dB higher than in the case of continuous transmission.

5.2. Effect of user traffic on downlink shared channel

In UMTS, transmission blocks have a fixed length and only blocks belonging to the same user can be transmitted during each TTI. Consequently, if the amount of information is lower than the space available in the physical layer container, the efficiency is reduced due to the unused space. This effect depends on source dynamics in terms of speed, size and number of packets generated, and becomes more critical as the channel rate increases.

This behavior is quite different from the one of classical packet networks, where a better multiplexing gain and a reduced packet transmission time is always achieved by increasing the link rate.

To investigate this effect we have considered sources that generate an average number of packets N_p in the range from 1 to 25. Fig. 9 shows the average delay versus throughput with SF = 4 $R = 2/3$ and $N_p = 1$, 3, 6, 12 and 25.

As expected, the maximum throughput decreases as N_p decreases since the average frame filling fraction, as shown in Fig. 10, significantly reduces down to about 30% for $N_p = 1$. The filling fraction also changes with the code rate and, referring to cases considered in the previous section with $N_p = 25$, it reaches 0.91, 0.94, 0.95 and 0.96 for $R = 1$, $3/4$, $2/3$ and $1/2$, respectively.

The increase in the average delay, observed in Fig. 9 in the range of throughput below 900 kb/s, as N_p increases is due to the increase in queuing delay because of the increasing user burst size.

For a given throughput, G increases (Fig. 11) as N_p decreases, since each frame carries less blocks and more frames must be deserved to the transmissions of packets.

We have also observed that the achieved average SIR is always greater than the SIR target, as shown in Fig. 12. The reason for such SIR polarization is intrinsic in the power control mechanism and originates when the DSCH becomes active. In fact, the interference estimate, which is the interference measured in the preceding frame, includes also the power of the DSCH used to transmit to another user in the cell, and it is clearly higher, in

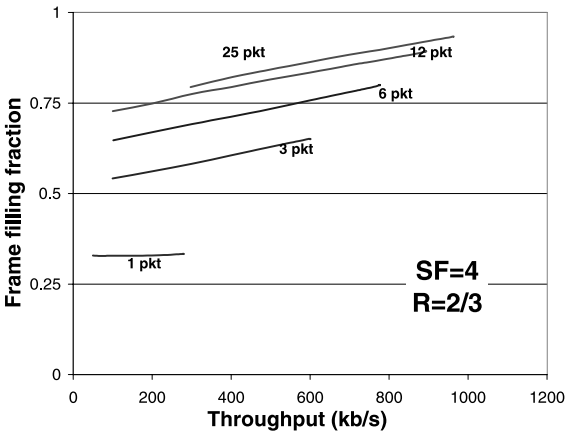

Fig. 10. Frame average filling fraction versus throughput for the cases reported in Fig. 9.

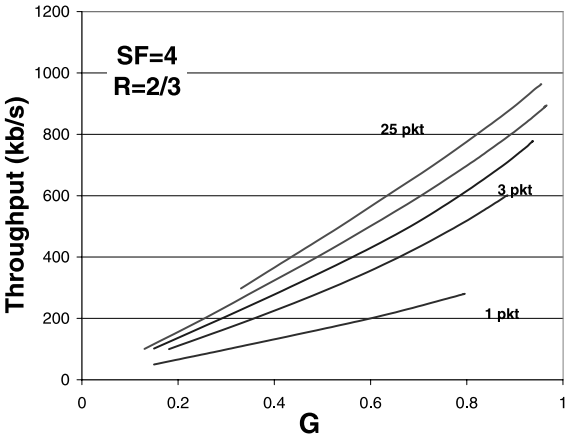

Fig. 11. Throughput versus the channel traffic G for the cases reported in Fig. 9.

the average, than the actual interference, which does not include the DSCH received power.

This SIR polarization effect increases when N_p decreases since the rate of new transmissions increases. Furthermore the polarization decreases as the traffic G increases because the overall interference increases with G, while the absolute estimation error remains unchanged. The anomalous behavior of the case $N_p = 1$ is due to the high saturation fraction experienced in this case (Fig. 5), which severely limits the SIR delivered.

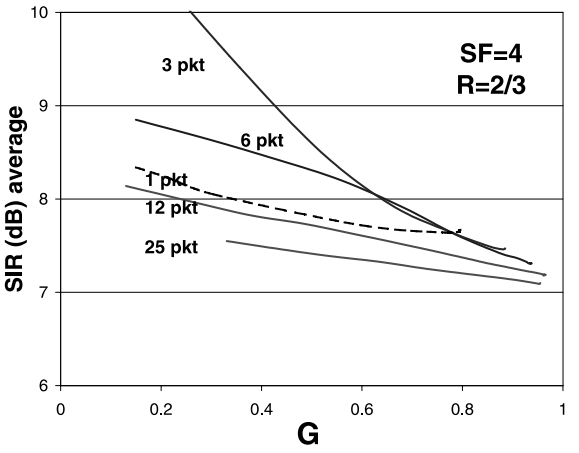

Fig. 12. Average SIR versus the channel traffic G for the cases reported in Fig. 9.

5.3. Multiple physical channels

While, in the previous sections we have considered a DSCH mapped into a single PDSCH, we now investigate the performance of the system when using several PDSCHs in parallel. First we consider the case in which the same overall channel rate is obtained by doubling both the number of channels and the SF. If the two channels with SF = 8 were orthogonal, the system would achieve the same capacity as the single channel with SF = 4 and should show an increased average delay because of the lower speed of the channels. In addition, since in our model we adopt an orthogonality loss, $\alpha = 0.4$, as suggested in Ref. [18], one would expect also a capacity decrease due to the added intra-cell interference.

On the contrary, we have measured that two channels with SF = 8 always provide a higher throughput than the single channel with SF = 4, as shown in Fig. 13 for $R = 2/3$, $1/2$ (the case $R = 3/4$ overlaps with $R = 2/3$). This unexpected behavior is due to two main reasons. First, the use of slower channels achieves higher frame filling degree (0.98 instead of 0.95 for $R = 1/2$). Second, the power control works better because the traffic burstiness is reduced, having doubled the packet transmission time. This is proved by the reduced standard deviation of interference variation that

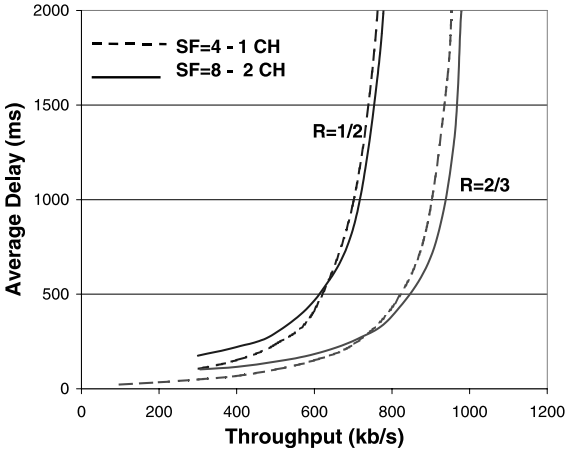

Fig. 13. Comparison of the average delays achieved with two channels and doubled SF for different code rates.

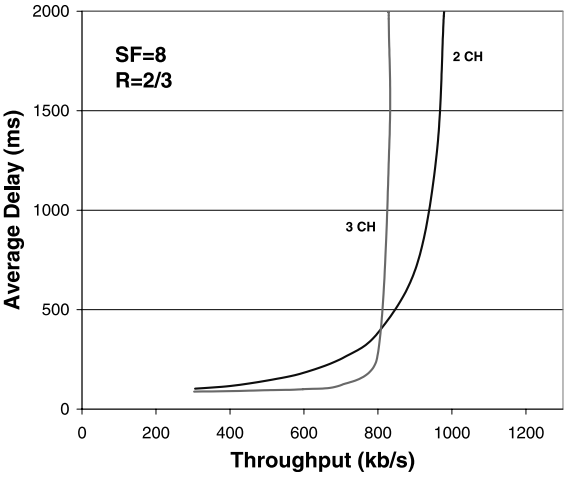

Fig. 15. Average delay versus throughput when a different number of physical channels are used in parallel for SF = 8, $R = 2/3$.

occurs in adjacent frames measured as shown in Fig. 14. However, we have observed that a further splitting of the channel no longer improves the maximum throughput since the de-orthogonalization effect prevails.

We now investigate the maximum throughput as function of the number of physical channels.

Fig. 15 shows that, for the code $R = 2/3$, no further improvement is achieved by increasing to three PDSCHs. The system is at its capacity and does not tolerate new interference sources. Therefore, the introduction of a new source of interference is counteracted by the BO which further limits the traffic channels down to 0.57. Note that an ideal BO mechanism should limit the traffic G to values corresponding to the maximum interference tolerated by the system, therefore yielding the same maximum throughput regardless of the number of channels. This does not happen, i.e., with three channels the throughput is lowered, because when the BO intervenes it increases the channel burstiness. As with three channels it must intervene more often, the channel burstiness is increased more than in the two channel case, where the BO limits the traffic to 0.96.

On the contrary, with $R = 1/2$ the performance is significantly improved (Fig. 16) by using more PDSCHs. The case of four channels, provides the capacity of the system, i.e., the maximum throughput (1240 kb/s). Capacity is reached with an optimal SIR target equal to 4 dB, which yields a BLER equal to 9%. Even in this case the BO is crucial for the correct operation of the system and limits G to 0.84 and 0.62 for four and five channels respectively.

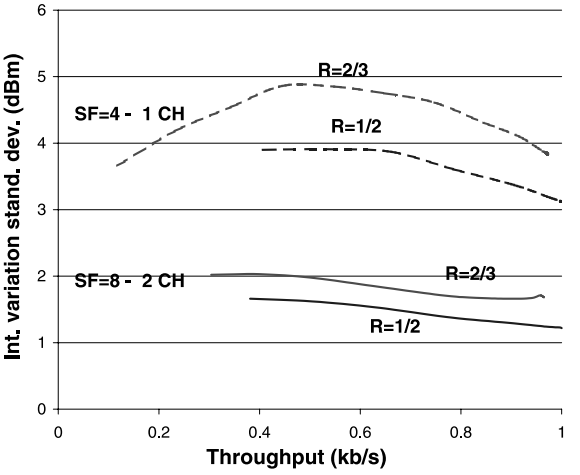

Fig. 14. Standard deviation of the interference changes in two consecutive frames for the cases reported in Fig. 15.

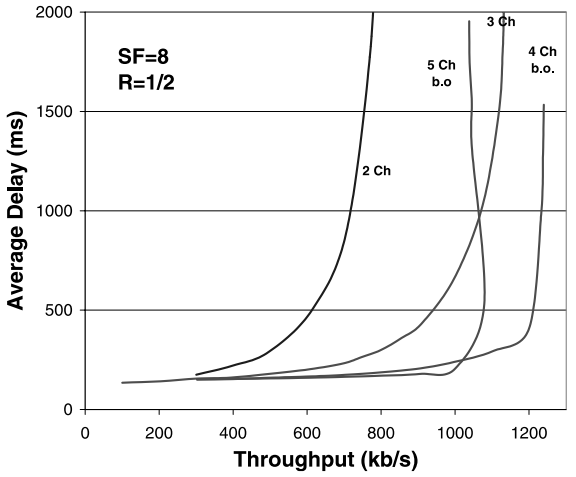

Fig. 16. Average delay versus throughput when a different number of physical channels are used in parallel for SF = 8, $R = 1/2$.

5.4. Effect of control channels

In this section we discuss the system performance when the effect of the DCH control channels is considered in the interference and SIR evaluations.

We have already observed that the positive feedback between the number of queued users and the suffered interference causes instability and reduces the throughput to zero if the queue is infinite. To control this instability we have limited to N the number of DCHs. Users that arrive beyond this limit are queued and wait for DCH availability.

To obtain numerical values of the interference of DCHs we need to determine the DCH power level. As already mentioned, since power control commands only indicate changes in the transmitted power level, both the power P_{DCH} received on the DCH and the power P_{DSCH} received on the DSCH must change in the same way. If their ratio is denoted by R, the SIR achieved after despreading on the two channels are related as

$$\frac{\mathrm{SIR}_{\mathrm{DSCH}}}{\mathrm{SIR}_{\mathrm{DCH}}} = R\frac{\mathrm{SF}_{\mathrm{DSCH}}}{\mathrm{SF}_{\mathrm{DCH}}}\frac{I_{\mathrm{DCH}}}{I_{\mathrm{DSCH}}} \tag{5}$$

where the term I_X indicates the interference received on channel X. Due to the de-orthogonalization factor, I_{DCH} includes the interference generated by the related DSCH, $P_{\mathrm{DSCH}} \times 0.4$, in addition to the "background" interference I_{B} generated by all the other channels.

If we disregard the interference generated on the DSCH by the related DCH, which is usually small, we have $I_{\mathrm{DSCH}} \simeq I_{\mathrm{B}}$, while the interference generated by the DSCH on the related DCH is given by

$$P_{\mathrm{DSCH}} \times 0.4 = \frac{\mathrm{SIR}_{\mathrm{DSCH}}}{\mathrm{SF}_{\mathrm{DSCH}}} I_{\mathrm{DSCH}} \times 0.4. \tag{6}$$

Therefore, the total interference suffered by the DCH is

$$\begin{aligned} I_{\mathrm{DCH}} &= I_{\mathrm{B}} + \frac{\mathrm{SIR}_{\mathrm{DSCH}}}{\mathrm{SF}_{\mathrm{DSCH}}} I_{\mathrm{DSCH}} \\ &\simeq \left(1 + 0.4 \times \frac{\mathrm{SIR}_{\mathrm{DSCH}}}{\mathrm{SF}_{\mathrm{DSCH}}}\right) I_{\mathrm{DSCH}}. \end{aligned} \tag{7}$$

The ratio $I_{\mathrm{DCH}}/I_{\mathrm{DSCH}}$ from Eq. (7) replaced into Eq. (5) yields

$$\frac{\mathrm{SIR}_{\mathrm{DSCH}}}{\mathrm{SIR}_{\mathrm{DCH}}} = R\frac{\mathrm{SF}_{\mathrm{DSCH}} + 0.4 \times \mathrm{SIR}_{\mathrm{DSCH}}}{\mathrm{SF}_{\mathrm{DCH}}}. \tag{8}$$

By Eq. (8), a SIR target on the DCH corresponds to a given SIR target on the DSCH. Therefore, given the SIR target on the DSCH, the target on the DCH can be achieved by suitably setting the power ratio R.

In our model we have assumed that the DCH control channels use a SF equals to 512 and require a SIR target equal to 8 dB. Furthermore, we have disregarded the term $0.4 \times \mathrm{SIR}_{\mathrm{DSCH}}$ in Eq. (8).

Fig. 17 shows the delay vs. throughput curves for the case SF = 4, $R = 2/3$ and $N = 10$, 15. We see that the effect of the DCH interference reduces the maximum throughput with respect to the case of no DCH interference (NI). The increased interference increases the BLER and the BO operates drastically reducing the maximum G to values equal 0.84 and 0.75 for $N = 10$ and $N = 15$ respectively. A similar behavior has been observed in the case of four channels, SF = 8 and $R = 1/2$ as shown in Fig. 18. Here the limiting effect of BO on the channel traffic is even more drastic as the maximum G is 0.77 and 0.69 for $N = 10$ and 20, respectively.

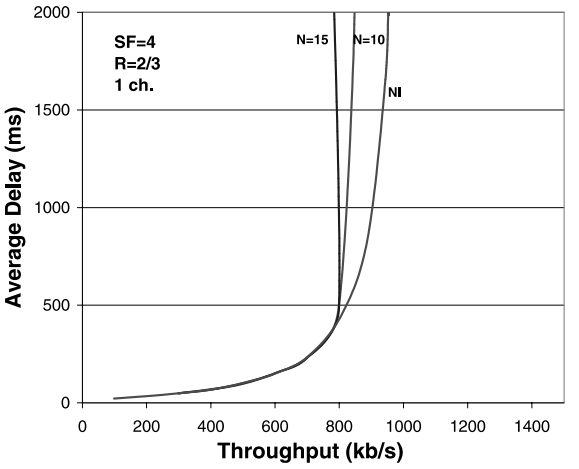

Fig. 17. Average delay versus throughput when a different number of active users is adopted in the case SF = 4, $R = 2/3$ and one physical channel.

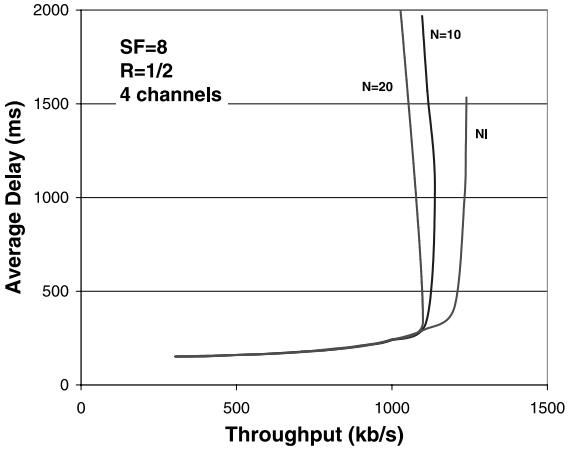

Fig. 18. Average delay versus throughput when a different number of active users is adopted in the case with four physical channels, SF = 8 and $R = 1/2$.

For the optimal choice of N a trade off exists between interference and multiplexing effect. In fact, by reducing N the system throughput from one side increases because the interference decreases, but from the other side it decreases because of the increasing multiplexing inefficiency. The latter effect becomes relevant when the N users are not able to fill the channel. In the cases we have considered such effect has never been observed since the source traffic model adopted assumes an high speed (929.4 kb/s).

5.5. Effect of power control

The closed-loop power control in the DSCH has been introduced to increase the capacity of the system. However, we have already observed in Section 3 that the burstiness of data transmission may jeopardize the gain achieved. Furthermore, the use of the control channels introduces additional interference. To evaluate the effect of this trade off, we have analyzed two alternative ways of using the FACH channel, which is very similar to the DSCH as far as the processing and the transmission of information is concerned, but do not use a closed-loop power control.

In the first alternative, the "NO-PC" case, no power control is exerted and all users transmit at the fixed level of 30 dBm. In such a channel, users experience different average SIR values, due to their different path loss and shadowing, and, therefore different BLER values. To avoid persistent error conditions of users with a bad channel, a Random Order scheduling discipline has been adopted.

In the second alternative, the "OL-PC", open-loop power control mechanism that compensates the different link losses has been implemented. In this case, the transmitted power is adjusted, within the allowed range, to reach a target value for the received power level.

Fig. 19 shows the delay vs. throughput curves for these two different alternatives. In the NO-PC case the maximum throughput, equal to 450 kb/s is obtained using the lowest spreading factor SF = 4, $R = 1/2$ and one physical channel. Furthermore, in this case, a non-negligible outage probability has been observed. In fact, 8% of users experience packet dropping due to the limit in the number of consecutive retransmissions.

The performance of the OL-PC case depends also on the target value of the received power and the best results, shown in the figure, have been obtained using the same channel parameters as in the NO-PC case, and a received power target value of -66 dBm. The maximum throughput in this case is 670 kb/s. For comparison purposes we have

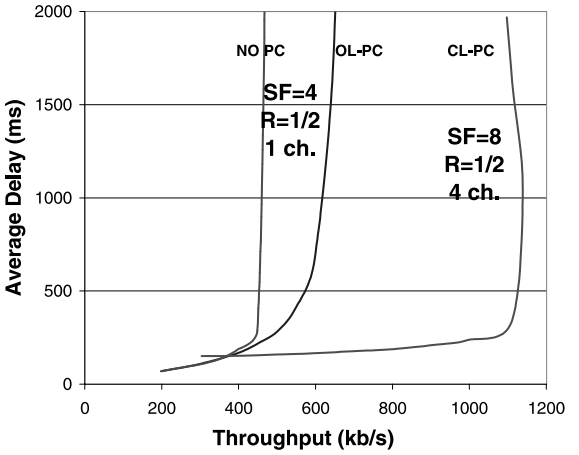

Fig. 19. Best delay throughput curves for cases NO-PC, OL-PC and CL-PC.

also reported in the figure the best performance achieved using CL-PC previously shown in Fig. 18. The comparison shows that in the UMTS system the CL-PC is very effective and almost doubles the throughput with respect to the case of OL-PC.

6. Conclusions

In this paper we have presented some preliminary results obtained by simulation on the delay-throughput curves of the packet service of the UMTS radio interface in a homogeneous cellular system. In particular, we have focused on the effect that the many system parameters and channel configurations have on the UMTS capacity. Although at present we have considered a specific scenario, which ignores fading and assume a shadowing standard deviation of 5 dB, we think that some general remarks and some warnings can be inferred by our results.

Primarily, we have verified that the close-loop power control mechanism is very effective even with packets service, where the burstiness causes abrupt changes in the overall interference. However, due to this changes, the power-control mechanism tracks the SIR target with errors and achieves a SIR whose standard deviation increases as the speed of the physical channel increases, because of the reduced packet transmission time. With the traffic model adopted we have verified that the mechanism is very efficient even with low interference protection. In particular, with a single physical channel, the maximum throughput, equal to 980 kb/s, is attained with the smallest available SF, $SF = 4$, and a light code, $R = 2/3$, corresponding to a net physical speed of 1260.8 kb/s. If a higher channel protection is chosen (lower R or higher SF) the throughput is reduced due to the increased overhead.

If the use of multiple physical channels is allowed, the maximum throughput is attained by using up to four channels with $SF = 8$ and $R = 1/2$, despite the new intra-cell interference introduced. This is mainly due to the improved efficiency of the closed-loop power control with the reduced interference burstiness. With greater SF and/or more powerful codes more channels, of lower speed and higher delay, must be used, but the capacity is not increased. To assess the generality of this results we have also investigated the capacity with shadowing increased from 5 to 10 dB, although the related results have not been reported in the paper. As expected the capacity decreases and is now achieved with 3 rather than 4 physical channels, but the optimal channel parameter choice remains unchanged ($SF = 8$, $R = 1/2$). We have also verified that these results have significance in all the cases in which fading can be counteracted by other means, such as power control (in this case the introduction of fading has the same effect as increasing the shadowing standard deviation), or a Rake receiver with optimal signal combining.

A further general result is throughput instability at capacity, i.e., when the system operates with the highest bearable interference. In this case a back off mechanism, which is able to reduce the interference by reducing the channel traffic, is crucial to let the system operate close to capacity.

Also the use of DCH for power control with DSCH may lead to instability if the number of DCH is not limited; therefore, users must be queued not only for the DSCH use but also for accessing power control channels.

Finally, we must note that a great reduction of capacity may be observed if low speed users are

served because of the excessive length of the frame, which is the minimum unit that a single user may use. In fact, at capacity, the frame length varies from 4746 bit, with four SF = 8, $R = 1/2$ channels, to 12 608 bit with SF = 4, $R = 2/3$. In these conditions slow users might not be able to fill the entire frame. As all frames are synchronized, the reduced frame filling degree does not reduce the interference, yielding a net decrease in throughput. Indeed, we expect that capacity is further reduced by the fact that the great burstiness introduced in this way further affects the power control, although we could not verify this effect because, to simplify the simulator, we did not consider power updates within the frame.

The simulator we have implemented is rather complete and allows to investigate the performance of other channels such as DCH and FACH. From the DCH analysis, we have observed that, adopting circuit switching operation with an average measured BLER not greater than 0.001, averaged on all channels, and assuming the scenario adopted in this paper, only 18 channels with SF = 64 and $R = 1/2$ can be served. Each channel provides a physical throughput of 42 kb/s. This corresponds to an aggregate maximum throughput of 756 kb/s, which should be compared with the maximum throughput of 1200 kb/s obtained with the packet service over DSCH. This confirms that, in spite of the several limitations of UMTS pointed out in this paper, packet switching, due to its intrinsic flexibility, better adapts to interference limited systems than circuit switching.

References

[1] A. Samukic, UMTS universal mobile telecommunications system: development of standards for the third generation, IEEE Transactions on Vehicular Technology 47 (4) (1998) 1099–1104.
[2] K.W. Richardson, UMTS overview, Electronics and Communication Engineering Journal 12 (3) (2000) 93–100.
[3] M. Gallagher, W. Webb, UMTS the next generation of mobile radio, IEE Review 45 (2) (1999) 59–63.
[4] D. O'Mahony, UMTS: the fusion of fixed and mobile networking, IEEE Internet Computing 2 (1) (1998) 49–56.
[5] T. Ojanpera, R. Prasad, An overview of air interface multiple access for IMT-2000/UMTS, IEEE Communications Magazine 36 (9) (1998) 82–95.
[6] 3rd Generation Partnership Project, Physical layer – General description, 3G TS 25.201, June 2000.
[7] E. Berruto, M. Gudmundson, R. Menolascino, W. Mohr, M. Pizarroso, Research activities on UMTS radio interface, network architectures, and planning, IEEE Communications Magazine 36 (2) (1998) 82–95.
[8] 3rd Generation Partnership Project, Spreading and modulation (FDD), 3G TS 25.213, June 2000.
[9] 3rd Generation Partnership Project, Physical channels and mapping of transport channels onto physical channels (FDD), 3G TS 25.211, June 2000.
[10] 3rd Generation Partnership Project, Multiplexing and channel coding (FDD), 3G TS 25.212, June 2000.
[11] 3rd Generation Partnership Project, Radio interface protocol architecture, 3G TS 25.301, June 2000.
[12] 3rd Generation Partnership Project, MAC protocol specification, 3G TS 25.321, June 2000.
[13] 3rd Generation Partnership Project, RLC protocol specification, 3G TS 25.322, June 2000.
[14] A.M. Viterbi, A.J. Viterbi, Erlang capacity of a power controlled CDMA system, IEEE Journal on Selected Areas in Communications 11 (6) (1993) 892–900.
[15] J.Y.N. Hui, Throughput analysis for Code Division Multiple Access of the Spread Spectrum Channel, IEEE Journal on Selected Areas in Communications 2 (4) (1984).
[16] R.J. McEliece, W.E. Stark, Channels with block interference, IEEE Transactions on Information Theory 30 (1) 1984.
[17] F. Borgonovo, A. Capone, L. Fratta, Retransmissions versus FEC plus interleaving for real-time applications: a comparison between CDPA and MC-TDMA cellular systems, IEEE Journal on Selected Areas in Communications 17 (11) 1999.
[18] UMTS 30.03, Annex B: Test environments and deployment models, TR 101 1112 v.3.2.0, April 1998.
[19] A. Bellini, M. Ferrari, Personal communication, Politecnico di Milano, 2000.
[20] 3rd Generation Partnership Project, RF system scenarios, 3G TR 25.942, Dic. 1999.

Flaminio Borgonovo received the Doctorate in Electronic Engineering from the Politecnico di Milano in 1971. In 1973, after a two-years period as Research Assistant at the Laboratory of Electrical Communications of the Politecnico, he reached the Italian National Research Council (CNR), where he started research activities in the multiple access and Local Area Networks field. In 1979 he became Associate Professor at the Electronic Department of Politecnico di Milano, where he was active in proposing and prototyping new Local-Area access schemes. In 1990 he got a Full-Professor position at the Università di Catania. He is currently Full Professor of Electrical Communications at the Politecnico di Milano. His current research interests include data networks, optical networks, and wireless communications systems. Dr. Borgonovo is member of IEEE.

Antonio Capone was born in Lecce, Italy in 1969. He received the "Laurea" degree (MS equivalent) and the Ph.D. degree in Telecommunication Engineering from the Politecnico di Milano in July 1994 and June 1998 respectively. From November 1997 to June 1998 he was a Teaching Assistant at the Engineering Faculty of the University of Lecce. He is now an Assistant Professor at the Dipartimento di Elettronica e Informazione of the Politecnico di Milano. His current research activities mainly involve packet access in wireless cellular networks, radio planning of cellular systems, QoS guaranteed services over IP networks. He is a member of the IEEE Communications and Vehicular Technology Societies.

Matteo Cesana (Student Member, IEEE) received the "Laurea" degree (MS equivalent) in Telecommunication Engineering from the Politecnico di Milano in July 2000. He is currently a Ph.D. student and works in the Telecommunications Networks Group of the Electronics and Information Department (DEI) of the Politecnico di Milano. His research interests are in mobile communications systems.

Luigi Fratta received the Doctorate in Electrical Engineering from the Politecnico di Milano, Milano, Italy, in 1966. From 1967 to 1970 he worked at the Laboratory of Electrical Communications, Politecnico di Milano. As a Research Assistant at the Department of Computer Science, University of California, Los Angeles, he participated in data network design under the ARPA project from 1970 to 1971. From November 1975 to September 1976 he was at the Computer Science Department of the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, working on modeling analysis and optimization techniques for teleprocessing systems. In 1979 he was a Visiting Associate Professor in the Department of Computer Science at the University of Hawaii. In the summer of 1981 he was at the Computer Science Department, IBM Research Center, San José, CA, working on local area networks. During the summers of 1983, 1989 and 1992 he was with the Research in Distributed Processing Group, Department of Computer Science, U.C.L.A., working on fiber optic local area networks. During the summer of 1986 he was with Bell Communication Research working on metropolitan area networks. In 1994 he has been Visiting Scientist at NEC Network Research Lab, Japan. Since 1980 he is a Full Professor at the Dipartimento di Elettronica e Informazione of the Politecnico di Milano. His current research interests include computer communication networks, packet switching networks, multiple access systems, modeling and performance evaluation of communication systems, local area networks, wireless cellular systems and integrated services over IP networks. Dr. Fratta is Fellow of IEEE.