Cross-layer Optimization for Multimedia Traffic in CDMA Cellular Networks

Daniele Veronesi, Stefano Tomasin, *Member, IEEE,* and Nevio Benvenuto, *Senior Member, IEEE*

***Abstract*— We consider the uplink transmission of multimedia services in a cellular network where multiple access is implemented by code division (CDMA) and the base station performs successive interference cancellation (SIC) to enhance performance. We propose a cross-layer optimization technique that operates both at the physical layer, by selecting the detection order at the SIC receiver, and at the medium access control layer, by selecting the power/rate for each mobile terminal. The optimization objective is the maximization of the overall weighted network throughput with the satisfaction of the quality of service criteria for multimedia communications. The resulting problem turns out to be $\mathcal{NP}$-complete and we resort to a discrete stochastic approximation (DSA) approach for its solution. Concerning DSA, an efficient implementation is proposed in order to reduce memory occupation and improve the convergence of the algorithm. In a UMTS cellular environment, numerical results show that the optimization provides a significant performance advantage over existing techniques at the cost of an increase of computational complexity and memory occupation.**

***Index Terms*— CDMA, cross-layer design and optimization, mobile multimedia technology, resource management and QoS provisioning, wireless personal communication systems.**

I. Introduction

In the recent past, a significant effort has been devoted to the optimization of resource management in the downlink of cellular communication systems, with the aim of reaching the required quality of service (QoS) for multimedia content distribution [1]. Most studies assume that QoS requirements in the uplink are less demanding, therefore leading to asymmetric communications with an high-rate downlink and a medium-rate uplink. However, a consensus is recently gathering on enabling the uplink of future cellular systems to support high-rate differentiated traffic [2]–[6]. This will make possible for example video streaming from the mobile terminal (MT), interactive multimedia applications and uploads of large files. On the other hand, due to the limited availability of bandwidth, efficient techniques will be required at both the physical (PHY) and medium access control (MAC) layers. For multimedia communications with different classes of traffic, efficiency is achieved also through resource allocation (RA) that optimizes rates and transmit powers according to the QoS requirements [4].

At PHY and MAC layers, code division multiple access (CDMA) is considered a good solution for uplink communications [7], [8], especially in conjunction with successive interference cancellation (SIC) [12], [13] at the base station (BS). In order to account for time-varying QoS requirements and channel conditions, various transmission parameters may be adapted, including spreading factor [5], power [2] and data rate [9]. When no QoS is considered, it has been shown that in a downlink transmission the network throughput is maximized when the BS transmits to at most one user at a time with the maximum power [10], [11].

In most of the existing literature, the optimization objective is the minimization of transmit power for a given signal to noise plus interference ratio (SNIR) of the received signal [14]–[16]. In [17], the capacity is computed for a system with power control (PC) and multiple rates, i.e., multiple target SNIRs. When all MTs transmit at the same rate, PC can be optimized to maximize the number of active users, [18]. Beyond PC, also the order of user detection in SIC (user order, UO), plays an important role in system performance [19], but studies have been focused on the minimization of power consumption rather than throughput enhancement [20], [21]. In [3], the problem of RA for uplink CDMA with SIC was analyzed for a general weighted-sum of user rates. It was shown that, when all received user signals have the same cross-correlation, users are optimally scheduled by always transmitting at the maximum power and selecting the UO according to the weights. However, we will show that for a broadband asynchronous uplink, due to different broadband channels, attenuations and spreading codes, the correlation among received user signals changes significantly and transmitting at the maximum power is not the optimal solution.

In this paper, we propose a new technique for optimizing PC and UO in SIC for an uplink CDMA system supporting multimedia communications with differentiated QoS requirements. In other words, we integrate PC and UO. The proposed optimization operates across PHY and MAC layers, since MAC RA requires information on the interference level from the PHY layer and in turns power, rate and UO at the PHY layer are determined by RA. Our objective is the maximization of a weighted sum of the MT instantaneous rates, where the weights are selected to account for any QoS constraint. In turn, the optimization is performed by dynamically selecting UO and setting the transmit power/rate of each user. The resulting solution includes a constraint on the maximum transmit power per user, which is a typical physical limitation of MT. Since the

Manuscript received November 20, 2006; revised April 12, 2007 and July 16, 2007; accepted November 11, 2007. The associate editor coordinating the review of this paper and approving it for publication was Q. Zhang. This work has been supported in part by the National Project on Fundamental Research (FIRB) "reconfigurable platforms for wide-band wireless communications."

S. Tomasin and N. Benvenuto are with the Department of Information Engineering (DEI), University of Padova, Via Gradenigo 6/B, I–35131 Padova, Italy (e-mail: {tomasin, nevio.benvenuto}@dei.unipd.it).

D. Veronesi was with the University of Padova. He is now with MgTech S.R.L., Via Verdi 14, I–24100 Bergamo, Italy (e-mail: daniele-mgtech.veronesi@st.com).

Digital Object Identifier 10.1109/TWC.2008.060960.

considered optimization problem is $\mathcal{NP}$-complete, we resort to a suboptimal algorithm, based on the discrete stochastic approximation (DSA) algorithm. The DSA algorithm was first studied in [22] and [23] for the optimization of a function defined on a discrete set, when only a noisy estimate of the objective function is available. This approach has been successfully applied in various communication problems, for spreading code optimization in CDMA systems [24], scheduling [25] and synchronization [26]. An immediate application of DSA to our optimization problem turns out to be memory expensive and slow in converge, due to two iterative loops for UO and PC. Hence, we resort to a suboptimal approach where UO and PC are jointly adapted. Moreover, memory requirement is optimized by a suitable statistical modeling of the throughput as a function of UO.

This paper is organized as follows. In Section II we describe the system model of a multimedia uplink CDMA system and derive the expression of the network throughput when the BS is equipped with SIC. In Section III we formalize the RA and PC problem. The DSA algorithm is presented in Section IV together with an efficient implementation. Numerical results are shown in Section V and conclusions are outlined in Section VI.

II. System Model

We consider an uplink CDMA transmission, where K MTs are transmitting from different locations to a common BS. Each MT may transmit a variety of multimedia traffic, with both real time (RT) and non real-time (NRT) requirements. Transmission time is divided into slots, each of duration T_{S}. The BS knows the channel conditions of all users at each time slot. According to channel conditions and transmission requests from the MTs, the BS determines the power and the transmission rate for each MT, according to the RA algorithm. In order to simplify the data exchange and the RA algorithm, we assume that the transmit power and rate are unique for the entire slot. Once RA has been performed, the BS broadcasts the transmission decisions to MTs [6].

A. CDMA transmission

MTs are identified by index k, with $k = 1, 2, \ldots, K$, associated to the spreading code $\{c_{k,\ell}\}$, $\ell = 0, 1, \ldots, N_S - 1$, having spreading factor N_S and chip period T_c. In this paper we consider short spreading codes, but the extension to systems using long spreading codes is straightforward. The signal of MT k is transmitted with power P_k. Each slot contains M data symbols $\{d_k(m)\}$, $m = 0, 1, \ldots, M - 1$, each spread by the MT spreading code. We assume that each MT transmits with a different coding/modulation technique and the power of the data signal is unitary for normalization.

The signal propagates through a wideband wireless channel that is assumed to be time-invariant for the duration of a slot and includes both path-loss and shadowing. Let us indicate with $h_k(t)$ the impulse response of the *composite* channel given by the convolution of the transmit filter, the channel impulse response and the receive filter.

By considering spreading as a filtering operation, we define the *effective* channel impulse response, which comprises spreading and filtering with $h_k(t)$, as $g_k(t) = \sum_{\ell=0}^{N_s-1} c_{k,\ell} h_k(t - \ell T_c)$. The received signal at the BS is the sum of contributions from all the MTs plus noise, i.e.,

$$\begin{aligned} r(t) &= \sum_{k=1}^{K} z_k(t) + w(t) \\ &= \sum_{k=1}^{K} \sum_{m=0}^{M-1} g_k(t - mN_sT_c)\sqrt{P_k}\, d_k(m) + w(t)\,, \end{aligned} \tag{1}$$

where $w(t)$ is a complex Gaussian noise with zero mean and power spectral density $N_0/2$ per dimension. An asynchronous transmission fits into the model (1) by randomly delaying the channel impulse responses.

B. SIC receiver

The SIC receiver performs the detection of signals in succession and interference is removed before detection of the next signal. Signals from MTs are detected according to the ordered vector $\mathcal{K} = [k_1, k_2, \ldots, k_K]$, with

$$k_p \in \{1, 2, ..., K\} \text{ and } k_p \neq k_q \text{ for } p \neq q. \tag{2}$$

Step p in SIC corresponds to the detection of the signal coming from the MT with index k_p and the removal of its contribution from the received signal.

In particular, let us indicate with $r_{k_p}(t)$ the signal obtained from $r(t)$ after the removal of the signals of MTs $k_1, k_2, \ldots, k_{p-1}$. Initially, $r_{k_1}(t) = r(t)$. The detection of MT signal k_p is performed by first applying a rake receiver to $r_{k_p}(t)$ followed by a de-spreader that provides

$$\tilde{d}_{k_p}(m) = \frac{1}{\sqrt{P_{k_p}}} \int g^*_{k_p}(-\tau) r_{k_p}(mN_ST_c - \tau)d\tau\,. \tag{3}$$

The signal $\{\tilde{d}_{k_p}(m)\}$ is detected, decoded and re-encoded to obtain an estimate of the transmitted data signal $\{\hat{d}_{k_p}(m)\}$. The received signal relative to MT k_p is then estimated as

$$\hat{z}_{k_p}(t) = \sum_{m=0}^{M-1} g_{k_p}(t - mN_sT_c)\sqrt{P_{k_p}}\,\hat{d}_{k_p}(m) \tag{4}$$

and subtracted from $r_{k_p}(t)$ to provide the signal for detection of the next MT, i.e.,

$$r_{k_{p+1}}(t) = r_{k_p}(t) - \hat{z}_{k_p}(t)\,, \quad p = 1, 2, \ldots, K - 1. \tag{5}$$

The forthcoming analysis assesses the achievable throughput of a system with SIC using coding and operating close to capacity (see (13)). In these conditions, for sufficiently long slots, errors in detection and interference cancellation are negligible, which is a realistic assumption for a SIC receiver using decoding [29], contrary to the linear SIC receiver [16]. Moreover, under the assumption of perfect interference cancellation, $r_{k_p}(t)$ is affected by noise, inter-symbol interference (ISI) and interference of undetected MTs. Provided that the spreading factor N_S is much larger than the normalized (to T_c) root mean square delay spread of the channel (rms_{ds}), ISI is negligible, as will be assumed in the following analysis. Hence, the dominant limiting factor is multiuser interference (MUI), which is strongly related to the particular propagation channel and spreading sequence. Indeed, most works in literature on SIC assume an equal correlation among all the received user signals. Although useful for a first analysis, this assumption is not realistic for channels characterized by a significant multipath, as shown in [19] and [20], [21].

We define the average correlation among signals received from the MTs as

$$\tilde{R}_{i,j}(\Delta) = \mathrm{E}\left[\tilde{d}_i(m-\Delta)d_j^*(m)\right], \quad \Delta = -1, 0, 1, \tag{6}$$

with $i, j = 1, 2, \ldots, K$, where we considered three values of Δ to take into account asynchronous transmissions. Assuming perfect cancellation we obtain

$$\tilde{R}_{i,j}(\Delta) = \int g_i^*(\tau - \Delta T) g_j(\tau) d\tau. \tag{7}$$

We define the power of the total correlation as $|R_{i,j}|^2 = |\tilde{R}_{i,j}(-1)|^2 + |\tilde{R}_{i,j}(0)|^2 + |\tilde{R}_{i,j}(1)|^2$, where we denote $R_{i,j}$ as the *equivalent correlation* among users. Moreover, we define the interference plus noise power for MT k_p as

$$I_{k_p} = \sum_{i=p+1}^{K} P_{k_i} |R_{k_i,k_p}|^2 + N_0 R_{k_p,k_p}. \tag{8}$$

Then, from (5), the SNIR of MT k_p can be written as

$$SNIR_{k_p} = \frac{P_{k_p} |R_{k_p,k_p}|^2}{I_{k_p}}, \tag{9}$$

under the assumption that data signals of different MTs are uncorrelated and statistically independent of noise. From (9) we observe that CDMA is limited by MUI, and the SIC receiver only mitigates the problem by performing partial cancellation of interference.

Usually, correlation $\{R_{i,j}\}$ and interference power I_{k_p} are not readily available at the BS and they must be estimated by either using a training sequence or a decision-directed method. This yields uncertainties on the estimate of $SNIR_{k_p}$.

III. Resource Allocation Problem

As RA objective, we consider the maximization of the sum of the MT throughputs weighted by coefficients that account for the packet priority. This approach has been widely adopted in RA algorithms [3], [30], [31] and it has been proven to be stable. In particular, let $T(SNIR_{k_p})$ be the instantaneous throughput of MT k_p, as a function of its SNIR, and let $w_{k_p}(s)$ be the associated weight at time slot s. By defining the power vector $\boldsymbol{P} = [P_1, P_2, \ldots, P_K]$ and the weighted sum-throughput as

$$T(\boldsymbol{P}, \mathcal{K}, s) = \sum_{p=1}^{K} w_{k_p}(s) T(SNIR_{k_p}), \tag{10}$$

the RA aims at solving the problem

$$\max_{\mathcal{K}, \boldsymbol{P}} T(\boldsymbol{P}, \mathcal{K}, s), \tag{11}$$

with respect to the PC vector $\boldsymbol{P}$ and the UO $\mathcal{K}$, under the constraint of a maximum transmit power per user, i.e., $P_k \leq P_{\max}$, $k = 1, 2, \ldots, K$, and UO satisfying (2).

Each user throughput is defined as the number of information bits that can be successfully received by the BS and an upper bound is provided by the Shannon capacity. In practice, the capacity is achieved up to a gap, provided that a large number of modulation and coding formats are available for transmission. Let Γ_{gap} be the signal to noise ratio (SNR) gap to achieve capacity. We define the normalized SNIR as

$$\Gamma_{k_p} = \frac{SNIR_{k_p}}{\Gamma_{\text{gap}}}. \tag{12}$$

Then the throughput relative to user k_p can be written as

$$T(SNIR_{k_p}) = \log_2\left(1 + \Gamma_{k_p}\right), \quad \text{[bit/s/Hz]} \tag{13}$$

Hence, we can rewrite (10) as

$$T(\boldsymbol{P}, \mathcal{K}, s) = \sum_{p=1}^{K} w_{k_p}(s) \log_2\left(1 + \frac{P_{k_p} |R_{k_p,k_p}|^2}{\Gamma_{\text{gap}} I_{k_p}}\right). \tag{14}$$

From (8), (9) and (14) we observe that the SNIR and, consequently the throughput are determined by the transmit power of each user. Indeed, the power determines both the useful signal level and the interference on other users' signals. As a consequence, the optimization of the power that maximizes the total weighted throughput is not a trivial problem.

Similarly to the problem in [20], [21], it can be shown that problem (11) can be modeled by a graph where the node set comprises all active users and two additional nodes, source and termination. We assign to each user node $\mathtt{i}$ the power coefficient $P_{\mathtt{i}}$ while each arc $(\mathtt{i}, \mathtt{j})$ is weighted by $R_{\mathtt{i},\mathtt{j}}$. Now, the problem of finding UO and PC that maximize the weighted throughput becomes the problem of finding an Hamiltonian path [36] on the graph (i.e., a sequence of nodes from the source to the termination node in which each node of the graph appears exactly once) that satisfies (11). Since the objective function is a non-linear function of powers $\{P_{\mathtt{i}}\}$ while the power constraints as well as the order constraints can be formulated as linear constraints [20], [21], we then conclude that the maximum weighted throughput problem is the search of an Hamiltonian path on a graph by a non-linear objective function, under linear constraints. Unfortunately, from [37] we conclude that the problem is $\mathcal{NP}$-complete and its solution requires an exhaustive search of the optimum UO over the entire set of $K!$ orderings. Hence, this approach is unfeasible for real systems with a possibly large number of active users. Therefore, to solve (11) we resort to a stochastic approach, based on the DSA algorithm [22], whose general framework is: *i*) to iteratively select the best UO and *ii*) to determine the required PC for a given UO that maximizes the weighted throughput.

For the choice of the weights in (10), various approaches may be considered. For example, by setting $w_{k_p}(s) = 1$ the max-rate RA is obtained. In general, rate and delay will be balanced according to requirements and we refer to literature for various approaches. As an example of application of optimization methods to multimedia traffic, we follow the proposal of [32], which aims at satisfying the QoS requirements in terms of minimum delay while exploiting channel conditions. Details on the choice of the weights are reported in Appendix I.

A. Power control for a given user order

We start performing PC that maximizes the weighted throughput for a given UO $\mathcal{K}$, with a constraint on the maximum available power per user. A closed-form solution for this problem is not available and we resort instead to an iterative algorithm similar to the technique proposed in [34] for PC with a single user receiver and here extended to a SIC receiver.

Let $\mathcal{P} = \{\boldsymbol{P} : \boldsymbol{0} \leq \boldsymbol{P} \leq \boldsymbol{P}_{\max}\}$ be the set of feasible power allocations, with $\boldsymbol{P}_{\max}$ a K-size vector with all entries $P_{\max}$. For a *given UO* $\mathcal{K}$, the following properties hold in $\mathcal{P}$:

- $T(\boldsymbol{P}, \mathcal{K}, s) \geq 0$. In fact, SNIR is always greater than one and consequently the logarithm is not negative.
- $T(\boldsymbol{P}, \mathcal{K}, s)$ is continuous and differentiable on $\boldsymbol{P} \in \mathcal{P}$.
- For all $\beta > 1$, $T(\beta\boldsymbol{P}, \mathcal{K}, s) > T(\boldsymbol{P}, \mathcal{K}, s)$.

From the maximum-minimum theorem [35, p. 89] we conclude that $T(\boldsymbol{P}, \mathcal{K}, s)$ has a maximum on set $\mathcal{P}$. Unfortunately, finding it is a hard task and proposed algorithms in the literature usually achieve only local maxima. This is the case for example of the iterative algorithms of [34], based on the gradient of the weighted throughput. Anyway, we also propose this strategy for the SIC receiver.

From (14) and (8), the gradient of the weighted throughput turns out to be

$$\frac{\partial T(\boldsymbol{P}, \mathcal{K}, s)}{\partial P_{k_p}} = \frac{1}{\ln(2)} \left[w_{k_p}(s) \frac{\Gamma_{k_p}}{(1+\Gamma_{k_p})} \cdot \frac{1}{P_{k_p}} - \sum_{i=1}^{p-1} w_{k_i}(s) \frac{\Gamma_{k_i}}{(1+\Gamma_{k_i})} \cdot \frac{|R_{k_i,k_p}|^2}{I_{k_i}} \right]. \tag{15}$$

In order to optimize PC, we set (15) to zero for $p = 1, 2, \ldots, K$ and obtain a non-linear system of equations that can be solved iteratively.

Let us define: *i)* $P_{k_p}^{(\nu)}$, the power allocated to user k_p at iteration ν; *ii)* $I_{k_p}^{(\nu)}$, the interference on user k_p when using the allocated power vector $\boldsymbol{P}^{(\nu)}$; *iii)* $\Gamma_{k_p}^{(\nu)}$, the normalized SNIR of user k_p computed from (12) and (9) using $\{I_k^{(\nu)}\}$; *iv)* N_{PC}, the maximum number of PC iterations. By setting (15) to zero, the tentative allocated power that maximizes the weighted throughput is

$$\tilde{P}_{k_p}^{(\nu+1)} = \left[\frac{(1+\Gamma_{k_p}^{(\nu)})}{w_{k_p}(s)\Gamma_{k_p}^{(\nu)}} \sum_{i=1}^{p-1} \frac{w_{k_i}(s)\Gamma_{k_i}^{(\nu)}}{(1+\Gamma_{k_i}^{(\nu)})} \cdot \frac{|R_{k_i,k_p}|^2}{I_{k_i}^{(\nu)}} \right]^{-1}, \tag{16}$$

for $p = 1, 2, \ldots, K$. Considering the user power constraint, the allocated power for a given UO that maximizes the weighted throughput is as follows

$$P_{k_p}^{(\nu+1)} = \begin{cases} P_{\max} & \text{if } \tilde{P}_{k_p}^{(\nu+1)} > P_{\max}, \\ \tilde{P}_{k_p}^{(\nu+1)} & \text{otherwise.} \end{cases} \tag{17}$$

The functional (14) may have many local maxima. By simulations, we verified that the initialization $P_{k_p}^{(1)} = P_{\max}/2$ provides a good performance in conjunction with our UO technique.

IV. The Discrete Stochastic Approximation Algorithm

The DSA algorithm is a technique to optimize (maximize) a function $f(\mathcal{K})$ on a discrete set $\mathcal{S}$, based upon its noisy estimate $\tilde{f}(\mathcal{K})$, [22]. The basic idea of DSA is to build a Markov chain having as states the elements of set $\mathcal{S}$ and converging toward the maximum value of $f(\mathcal{K})$ by successive evaluations of $\tilde{f}(\mathcal{K})$. Indeed, due to the noisy estimate, each state $\mathcal{K}$ has a non-null probability of yielding the maximum value of $f(\mathcal{K})$ for a given set of estimates, [23].

The DSA algorithm proceeds iteratively by updating the estimate of all the state probabilities. At the end of the iterative process, the state having the highest estimated probability is selected as the solution. We indicate with n the generic DSA iteration, in order to distinguish it from the PC iteration ν of Section III-A.

In particular, let $\pi^{(n)}(\mathcal{K})$ be the estimate, up to iteration n, of the probability that state $\mathcal{K}$ is a maximizer of f, i.e.,

$$\pi^{(n)}(\mathcal{K}) = \mathrm{P}[\mathcal{K} \text{ is a maximizer of } f]. \tag{18}$$

At iteration n, the DSA chooses the following three states:

- $\mathcal{K}_{\mathrm{rand}}$, a state selected randomly in set $\mathcal{S}$;
- $\mathcal{K}_{\mathrm{prob}}^{(n)}$, the state having the highest probability up to the current iteration, i.e.,

$$\mathcal{K}_{\mathrm{prob}}^{(n)} = \arg\max_{\mathcal{K}\in\mathcal{S}} \pi^{(n-1)}(\mathcal{K})\,; \tag{19}$$

- $\mathcal{K}_{\mathrm{curr}}^{(n-1)}$, the state selected at the previous iteration.

Among the three states, the new selected state, $\mathcal{K}_{\mathrm{curr}}^{(n)}$, maximizes the noisy objective function $\tilde{f}(\mathcal{K})$, i.e.,

$$\mathcal{K}_{\mathrm{curr}}^{(n)} = \arg \max_{\mathcal{K}\in\{\mathcal{K}_{\mathrm{rand}}, \mathcal{K}_{\mathrm{curr}}^{(n-1)}, \mathcal{K}_{\mathrm{prob}}^{(n)}\}} \tilde{f}(\mathcal{K})\,. \tag{20}$$

Next, the state probabilities are updated by increasing the probability of state $\mathcal{K}_{\mathrm{curr}}^{(n)}$ and decreasing the probabilities of all the other states,

$$\pi^{(n)}(\mathcal{K}) = \begin{cases} (1-\mu^{(n)})\pi^{(n-1)}(\mathcal{K}) + \mu^{(n)}, & \mathcal{K} = \mathcal{K}_{\mathrm{curr}}^{(n)} \\ (1-\mu^{(n)})\pi^{(n-1)}(\mathcal{K}), & \mathcal{K} \neq \mathcal{K}_{\mathrm{curr}}^{(n)}\,, \end{cases} \tag{21}$$

where $\mu^{(n)}$ is a suitable decreasing function of n. It is seen that, as the number of iterations increases, the state probabilities converge to the average value, and the coefficient $\mu^{(n)}$, together with the randomness of $\tilde{f}(\mathcal{K})$, determines the speed of convergence, [22].

A. Resource allocation by DSA

Our cross-layer RA requires that we optimize both PC and UO. By a direct application of DSA to the RA problem, the states are all the UOs

$$\mathcal{S} = \{\mathcal{K} = [k_1, k_2, \ldots, k_K] : k_i \neq k_j \text{ for } i \neq j, 1 \leq k_i \leq K\}, \tag{22}$$

and the function to be maximized is the weighted throughput

$$f(\mathcal{K}) = \max_{\boldsymbol{P}, P_i < P_{\max}} T(\boldsymbol{P}, \mathcal{K}, s)\,, \quad \mathcal{K} \in \mathcal{S}\,. \tag{23}$$

We observe that once the UO has been set, the maximum of $T(\boldsymbol{P}, \mathcal{K}, s)$ with respect to $\boldsymbol{P}$ can be computed using the algorithm described in Section III-A. In this case the estimated function $\tilde{f}(\mathcal{K})$ is the weighted throughput (14) corresponding to the evaluated powers $\boldsymbol{P}^{(N_{\mathrm{PC}})}(\mathcal{K})$, for UO $\mathcal{K}$.

The resulting algorithm is shown in Table I, where N_{DSA} is the number of iterations for the DSA loop.

TABLE I
GENERAL (I)DSA ALGORITHM FOR UO AND PC.

Initialization
1) Choose at random an order $\mathcal{K} \in \mathcal{S}$. Set $\mathcal{K}_{\text{curr}}^{(0)} = \mathcal{K}$.
2) Set the probability vector $\pi^{(0)}(\mathcal{K}') = 0, \forall \mathcal{K}' \neq \mathcal{K}$. Set $\pi^{(0)}(\mathcal{K}) = 1$.
3) Set $n = 1$.
DSA loop
4) Select at random an order $\mathcal{K}_{\text{rand}} \in \mathcal{S}$.
5) Select the UO with highest probability $\mathcal{K}_{\text{prob}}^{(n)} = \arg\max_{\mathcal{K}} \pi^{(n-1)}(\mathcal{K})$.
PC loop
6) Iterate N_{PC} times the PC algorithm of Section III-A to compute the weighted throughput $T(\boldsymbol{P}^{(N_{\text{PC}})}(\mathcal{K}), \mathcal{K}, s)$ for UO's $\mathcal{K}_{\text{rand}}$, $\mathcal{K}_{\text{curr}}^{(n-1)}$ and $\mathcal{K}_{\text{prob}}^{(n)}$.
Comparison step
7) Set $\mathcal{K}_{\text{curr}}^{(n)} = \arg\max_{\mathcal{K} \in \{\mathcal{K}_{\text{rand}}, \mathcal{K}_{\text{curr}}^{(n-1)}, \mathcal{K}_{\text{prob}}^{(n)}\}} T(\boldsymbol{P}^{(N_{\text{PC}})}(\mathcal{K}), \mathcal{K}, s)$.
8) Update the probability vector $\pi^{(n)}$ according to (21).
9) **End of DSA iteration**
10) Increase n.
11) If $n < N_{\text{DSA}}$ go to step 6), else end.

B. DSA with integrated power control and user order

It is seen that the above algorithm may have a very slow convergence, due to the two iterative loops, the outer for UO and the inner for PC. Hence, we propose to integrate the two adaptive processes in a loop where at each iteration n we store both the current UO $\mathcal{K}_{\text{curr}}^{(n)}$ and the corresponding allocated power vector $\boldsymbol{P}^{(n)}$. The estimated function is $\tilde{f}(\mathcal{K}) = T(\boldsymbol{P}^{(n-1)}, \mathcal{K}, s)$.

In particular, once the current state $\mathcal{K}_{\text{curr}}^{(n)}$ has been selected, $\boldsymbol{P}^{(n)}$ is computed by just a one-step application of the PC iterative algorithm of Section III-A. We denote the DSA with integrated PC and UO as IDSA.

We observe that at each iteration, IDSA compares the weighted throughput for UOs $\mathcal{K}_{\text{rand}}$, $\mathcal{K}_{\text{curr}}^{(n-1)}$ and $\mathcal{K}_{\text{prob}}^{(n)}$, using the same allocated power $\boldsymbol{P}^{(n)}$. As a result, the selection procedure may turn out to be biased toward a suboptimal UO, and not leading to the optimal solution. Indeed, by simulations it is seen that in the cases of interest IDSA converges to the DSA solution. As a matter of fact, for an equal number of total iterations, $N_{\text{PC}} \cdot N_{\text{DSA}}$, IDSA explores more UOs than DSA, while slowly adapting the allocated power. Hence, in general, IDSA shows a faster convergence than DSA.

C. IDSA with matrix implementation (IDSA-MI)

Concerning the implementation complexity, we observe that set $\mathcal{S}$ on which DSA is applied has $K!$ states and for each state it is necessary to store and update its probability. This results in a factorial memory requirement. Hence, the conventional DSA approach translates the $\mathcal{NP}$ problem (11) into a problem requiring an exponential memory with respect to the number of active users.

In order to obtain a feasible approach we consider an implementation based on a factorization of the state probabilities. We consider the probability that user i precedes user j in the UO, i.e.,

$$\Psi_{i,j} = \mathrm{P}[p < q : k_p = i, k_q = j\,,\ \text{with } \mathcal{K} = [k_1, k_2, \ldots, k_K]]\,, \tag{24}$$

with $i, j = 1, 2, \ldots, K$, $i \neq j$. At each IDSA iteration we store an estimate of $\{\Psi_{i,j}\}$ into the $K \times K$ matrix $\mathbf{\Phi}^{(n)}$. The diagonal of $\mathbf{\Phi}^{(n)}$ is set to zero and its values are never used.

When UO $\mathcal{K}_{\text{curr}}^{(n)}$ is selected, $\mathbf{\Phi}^{(n)}$ is updated as follows. Let us consider the first user k_1 in the ordered vector $\mathcal{K}_{\text{curr}}^{(n)}$. This user precedes all the other users in $\mathcal{K}_{\text{curr}}^{(n)}$ and correspondingly entries $\mathbf{\Phi}_{k_1,j}^{(n)}$, for all $j \neq k_1$, are increased in the probability matrix. For the second user k_2, all entries $\mathbf{\Phi}_{k_2,j}^{(n)}$, for all $j \neq k_1, k_2$, are increased, and so on. In general, let us define the set of $[K(K-1)]/2$ couples

$$\mathcal{I}(\mathcal{K}_{\text{curr}}^{(n)}) = \{(k_i, k_j) : i < j,\ k_i, k_j \in \mathcal{K}_{\text{curr}}^{(n)}\}\,. \tag{25}$$

The updated probability matrix is

$$\mathbf{\Phi}_{i,j}^{(n)} = \begin{cases} (1-\mu^{(n)})\mathbf{\Phi}_{i,j}^{(n-1)} + \mu^{(n)}, & (i,j) \in \mathcal{I}(\mathcal{K}_{\text{curr}}^{(n)}) \\ (1-\mu^{(n)})\mathbf{\Phi}_{i,j}^{(n-1)}, & (i,j) \notin \mathcal{I}(\mathcal{K}_{\text{curr}}^{(n)})\,. \end{cases} \tag{26}$$

Now, we approximate the probability of UO $\mathcal{K}$ as the product of probabilities $\mathbf{\Phi}_{i,j}^{(n)}$ relative to order $\mathcal{K}$, i.e., the probabilities that user k_p precedes users $k_{p+1}, k_{p+2}, \ldots, k_K$, for $p = 1, 2, \ldots, K$ and we write

$$\pi^{(n)}(\mathcal{K}) \approx \prod_{p=1}^{K} \prod_{j=p+1}^{K} \mathbf{\Phi}_{k_p,k_j}^{(n)}\,. \tag{27}$$

In order to find the UO with the highest probability [step 5) of Table I] using matrix $\mathbf{\Phi}^{(n-1)}$, we resort to a greedy algorithm that, instead of maximizing over the probabilities (27), maximizes the probability of each pair user-position. As a first step, we determine the first user in $\mathcal{K}_{\text{prob}}^{(n)}$ as the one having the highest probability, i.e., for $p = 1$ we determine

$$k_1 = \arg\max_k \prod_{j \neq k} \mathbf{\Phi}_{k,j}^{(n-1)}\,. \tag{28}$$

Then, the row k_1 of $\mathbf{\Phi}^{(n-1)}$ is set to zero and the column k_1 of $\mathbf{\Phi}^{(n-1)}$ is set to one. Next, the second user in $\mathcal{K}_{\text{prob}}^{(n)}$ is selected by using (28) on the reduced matrix. In general, at iteration p, after the selection of user k_p by (28), the row k_p of $\mathbf{\Phi}_{k,j}^{(n-1)}$ is set to zero and the column k_p of $\mathbf{\Phi}_{k,j}^{(n-1)}$ is set to one. The new matrix is used for the choice of the next user and the procedure is iterated K times to determine all the users of $\mathcal{K}_{\text{prob}}^{(n)}$.

V. NUMERICAL RESULTS

We assess the performance of the proposed cross-layer RA technique in terms of both network and physical layer parameters. For the network parameters we consider the packet delay, the packet loss probability and the total network throughput. As physical layer parameters we consider the consumed power per user.

We compare the proposed optimization algorithm against two approaches: the *maximum-power weight-ordering* (MPWO) technique and the power ordering (PO) technique. In MPWO [3] all users with $w_{k_p}(s) > 0$ are at the maximum power, i.e., $P_{k_p} = P_{\max}$ for $p = 1, 2, \ldots, K$, and users are ordered with increasing weights, i.e., $w_{k_p}(s) < w_{k_{p+1}}(s)$. In this case the SNIR becomes

$$SNIR_{k_p} = \frac{P_{\max}|R_{k_p,k_p}|^2}{P_{\max}\sum_{q=p+1}^{K}|R_{k_q,k_q}|^2 + R_{k_p,k_p}N_0}\,. \tag{29}$$

TABLE II
PARAMETERS OF THE SIMULATION SCENARIO.

PARAMETER	VALUE
channel rms_{ds}	0.5 μs
cell radius	580 m
T_{S}	1 ms
T_c	50 ns
pathloss exponent	3.5
std. deviation shadowing	6 dB
N_S	32
$P_{\max}$	10 mW
Γ_{gap}	1
S_v	10
σ	4 kbps
a	3.9 s^{-1}
$L_{\min}$	815 bytes
$L_{\max}$	320000 bytes
T_{mes}	296 slots
T_{rs}	1000 slots
T_{pc}	300 slots
$D_{\max,\mathrm{RT}}$	2%
$D_{\max,\mathrm{NRT}}$	20%
$W_{\mathrm{RT}} = W_{\mathrm{NRT}}$	100ms

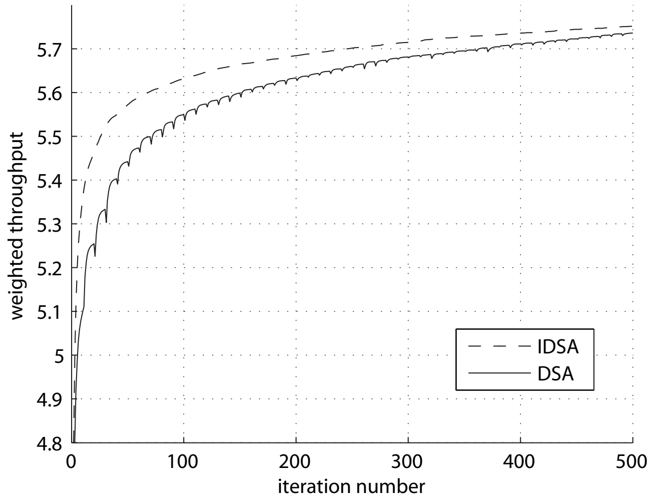

Fig. 1. Weighted throughput as a function of the iteration number per slot for DSA and IDSA algorithms. RT traffic with an average rate $\bar{m} = 250$ kbps and $K = 32$ active users.

It has been shown [3] that MPWO is optimum when the mutual interference among users is a constant, i.e., $R_{i,j} = R_{j,j}$ for all $i \neq j$.

In the power ordering (PO) approach users are ordered with decreasing channel gains, i.e., $R_{k_j,k_j} \geq R_{k_i,k_i}$, $j < i$, $i, j = 1, 2, \ldots, K$, and $P_{k_i} = P_{\max}$ for all users with $w_{k_i}(s) > 0$.

For a fair comparison, weights $w_{k_p}(s)$ are updated as described in Appendix I for both MPWO and PO.

Simulation scenario. The parameters of the simulation scenario are reported in Table II. We consider a UMTS-like environment where the channel includes pathloss, log-normal shadowing and an exponentially decaying power delay profile. The channel is block fading as each slot is characterized by independently generated channel. MTs are uniformly distributed within the cell and each terminal is assigned a different Hadamard code. Transmission from terminals is asynchronous with uniform random delays. We assume that the noise is such that a terminal transmitting with maximum power from the border of the cell achieves an average SNR of 10 dB.

As examples of RT and NRT traffics, we consider video and web sources, respectively, while the extension of the proposed models to other sources is straightforward. A description of the traffic generation is reported in Appendix II while its parameters are reported in Table II. The average rate of NRT traffic is 123 kbps.

For DSA and IDSA algorithms the coefficient for the update of the probability vector is $\mu^{(n)} = 1$ [22].

Comparison between DSA and IDSA. We first compare the DSA and IDSA algorithms for RT traffic with an average rate $\bar{m} = 250$ kbps. For each slot, for DSA we considered $N_{\mathrm{DSA}} = 50$ and $N_{\mathrm{PC}} = 10$ while for IDSA we considered $N_{\mathrm{DSA}} = 500$ and obviously $N_{\mathrm{PC}} = 1$. Fig. 1 shows the weighted throughput as a function of the iteration number for $K = 32$ active users. From the figure we see that IDSA has a much faster convergence than DSA, because it explores more UOs than DSA. Hence, forthcoming results are presented only for IDSA with $N_{\mathrm{DSA}} = 250$ iterations per slot. Indeed in practice, results shown in Fig. 1 represent an upper bound on the convergence time, associated with the worst case scenario of a block fading channel with independent realizations at each slot. In fact, when channels of adjacent slots are correlated, the iterative RA algorithm can be initialized with the solution obtained at the previous slot. In this case the number of iterations needed to achieve convergence is smaller. Note also that when the channel is varying during a slot, the capacity (13) should be replaced by the ergodic capacity.

Packet delivery probability and excess delay. A first QoS parameter is the percentage of packets delivered to the destination. We remember that RT packets have a deadline time after which they are discarded from the queue. Fig. 2 shows the probability of discarded packets for a RT transmission, as a function of the RT traffic average rate $\bar{m}$, while the rate of NRT traffic is as from Table II. We see that the proposed DSA-based RA has a significantly higher probability of delivering RT packets than both MPWO and PO. Concerning the NRT traffic, Fig. 3 shows the probability the NRT packets have an excessive delay, as a function of the offered RT traffic. We observe that IDSA is able to significantly reduce the probability of packets with excessive delay with respect to existing techniques.

Average delay. Another interesting QoS parameter is the average delay for both RT and NRT traffic. Figures 4 and 5 show the mean delay of RT and NRT traffic, respectively, as a function of the average rate of RT traffic $\bar{m}$. We observe that in general IDSA yields a much lower mean delay than both PO and MPWO.

Mean allocated power. In the considered multiuser scenario, increasing the transmit power yields potentially a higher throughput but also a higher interference level for other users. Fig. 6 compares the different RA schemes in terms of mean allocated power. The result is that IDSA has a reduced power consumption with respect to both PO and MPWO. Hence we conclude that the proposed technique is not only able to guarantee a better fairness among users but it also provides a lower power consumption for MTs, which are usually strongly limited by battery duration.

Network goodput. Fig. 7 shows the goodput of the net-

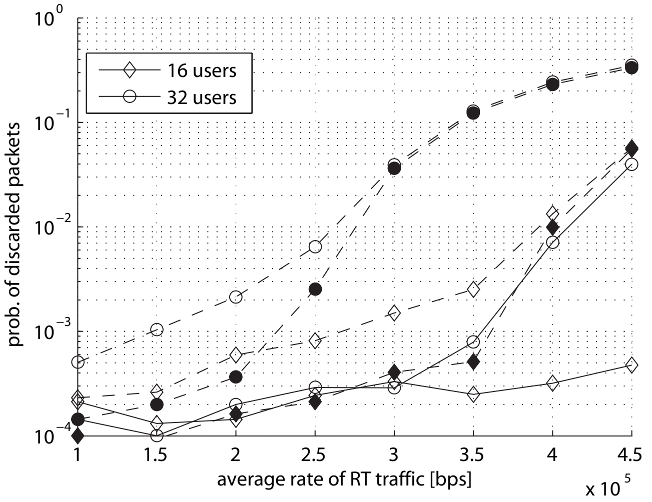

Fig. 2. Probability of discarded packet as a function of the RT traffic average rate $\bar{m}$. Solid lines: IDSA. Dashed lines: PO (empty markers), MPWO (filled markers).

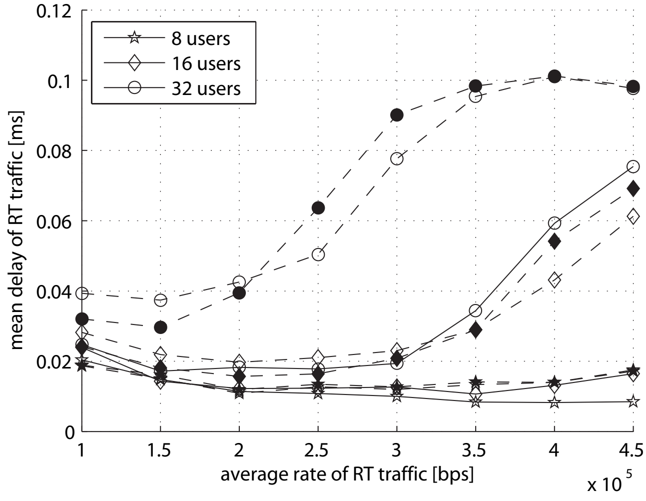

Fig. 4. Mean delay of RT traffic as a function of the RT traffic average rate $\bar{m}$. Solid lines: IDSA. Dashed lines: PO (empty markers), MPWO (filled markers).

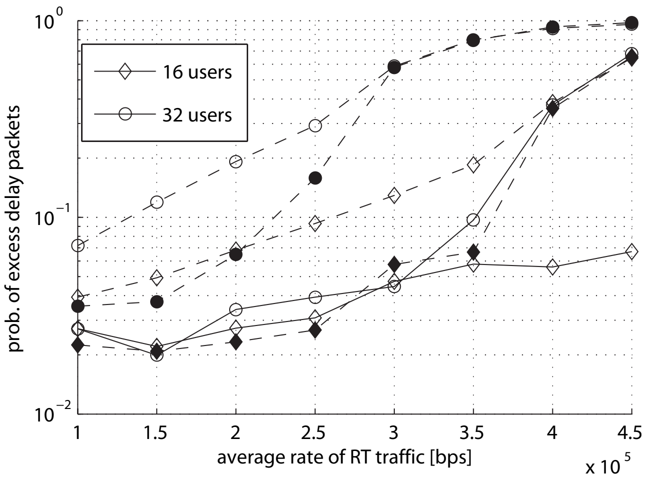

Fig. 3. Probability of excessive delay for NRT packets as a function of the RT traffic average rate $\bar{m}$. Solid lines: IDSA. Dashed lines: PO (empty markers), MPWO (filled markers).

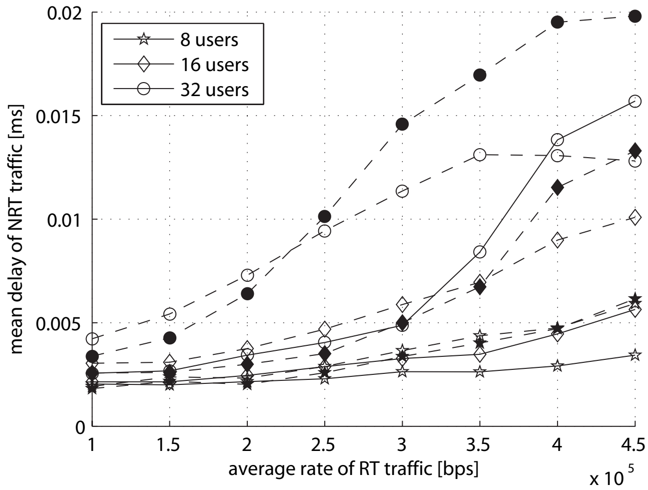

Fig. 5. Mean delay of NRT traffic as a function of the RT traffic average rate $\bar{m}$. Solid lines: IDSA. Dashed lines: PO (empty markers), MPWO (filled markers).

work, i.e., the number of bits per second that effectively reach the application. Goodput does not include the partial transmission of RT packets that are discarded because not completely transmitted before the deadline. Note that the target of the RA is not to increase goodput but to balance goodput, priority and fairness among users. Still, from Fig. 7 we observe that the IDSA, which provides the better fairness, also yields a similar or even better goodput than the other two techniques. Lastly, note that goodput shows a floor for a high traffic rate, due to the mutual interference among users, as to be expected in a CDMA uplink.

Computational complexity and memory requirements. Computational complexity and memory requirements of the proposed schemes are now provided. In order to assess the computational complexity, we compare both the number of complex multiplications and the number of comparisons, since UO requires a large number of comparisons. For the memory occupation, we provide the order of required memory cells with respect to the number of users. Table III summarizes the complexity for various UO and PC schemes. We observe that the complexity of (I)DSA is dominated by the square of the number of explored UOs (N_{DSA}^2) and the square of the number of users (K^2). As the memory occupation, for (I)DSA it grows proportionally to KN_{DSA}, while for IDSA-MI it grows as K^2, i.e., the size of the probability matrix. A comparison of (I)DSA(-MI) techniques with MPWO and PO shows that the proposed techniques have a much higher complexity and memory occupation than existing techniques. However, since K is in general not very large, complexity may not be a problem. On the other hand, the performance advantage of the proposed technique is significant.

VI. Conclusions

We proposed an iterative technique for optimizing PC and UO in SIC for differentiated traffic, suited to enable multimedia applications in multiuser scenarios. By means of a cross-layer optimization, based on a discrete stochastic approximation method, we were able to optimize each user

TABLE III
COMPUTATIONAL COMPLEXITY AND MEMORY OCCUPATION OF UO AND PC SCHEMES.

Technique	Complex Multiplications	Comparisons	$\mathcal{O}$(Memory)
DSA, IDSA	$N_{\rm DSA}N_{\rm PC}\left[3K+5\frac{K(K-1)}{2}\right]+$ $N_{\rm DSA}^2+N_{\rm DSA}K$	$N_{\rm DSA}N_{\rm PC}K+$ $3N_{\rm DSA}+N_{\rm DSA}^2$	$KN_{\rm DSA}$
IDSA-MI	$N_{\rm DSA}N_{\rm PC}\left[3K+5\frac{K(K-1)}{2}\right]+N_{\rm DSA}K^2+$ $N_{\rm DSA}\left[K^2+\frac{K(K-1)}{2}\right]+N_{\rm DSA}K$	$N_{\rm DSA}N_{\rm PC}K+$ $\left[\frac{K(K-1)}{2}+3\right]N_{\rm DSA}$	K^2
MPWO, PO	$K+\frac{K(K-1)}{2}$	$K\log_2(K)$	K

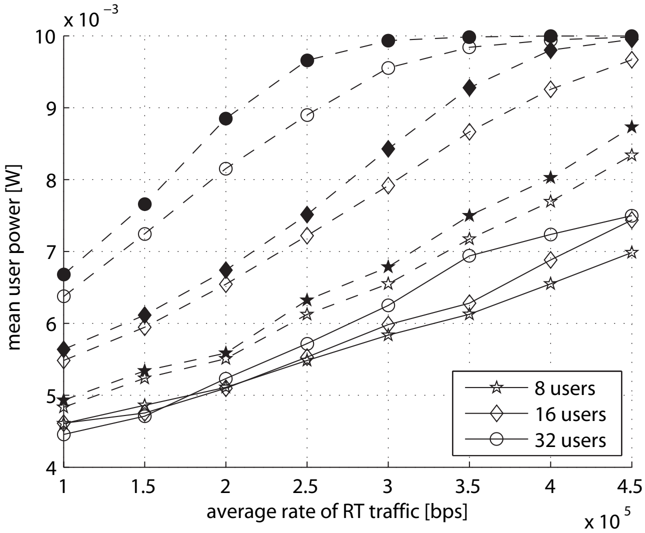

Fig. 6. Mean allocated power per user as a function of the RT traffic average rate $\bar{m}$. Solid lines: IDSA. Dashed lines: PO (empty markers), MPWO (filled markers).

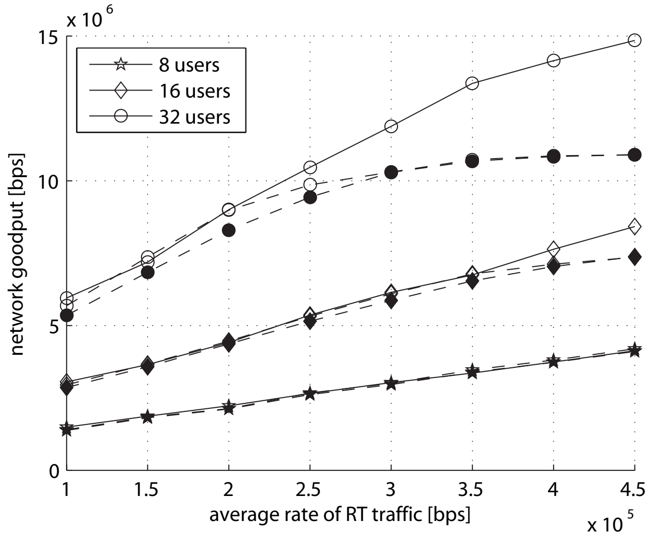

Fig. 7. Network goodput as a function of the RT traffic average rate $\bar{m}$. Solid lines: IDSA. Dashed lines: PO (empty markers), MPWO (filled markers).

power/rate and at the same time guarantee fairness and reduce the delay of packet transmissions. For multimedia transmissions over a UMTS-like scenario, the proposed technique allows to increase the goodput of RT traffic by about 50% and decrease the mean delay of NRT traffic by about 25%, with respect to existing techniques. This significant performance advantage comes at the price of a much higher computational complexity due to the large number of iterations per slot needed to achieve convergence, at least in a block fading channel with independent realizations.

APPENDIX I
CHOICE OF THE WEIGHTS

We choose the weights according to [32], which aims at satisfying the QoS requirements in terms of minimum delay while exploiting the channel conditions. We consider three indices, two for fairness and one for the channel quality. Concerning the channel condition at slot s, we consider the channel gain at the detection point normalized to the sum of all other user channel gains [33], i.e.,

$$I_{\rm ch}(k,s)=\frac{R_{k,k}}{\sum_{j=1}^{K}R_{j,j}}\,. \tag{30}$$

The normalized channel index $I_{\rm ch}(k,s)$ provides a first assessment of channel quality independently of the allocated powers, while using (9) as channel index would require to know in advance the user powers.

We account for fairness, based on the delay that each user's packet experiences at the MT, with two indices. The first index takes care of the packets in the queue, while the second index is referred to the average user performance. In particular, let us indicate with $D_{\max,{\rm RT}}(k)$ ($D_{\max,{\rm NRT}}(k)$) the deadline, in slots, of the oldest packet of user k in the RT (NRT) queue and with $\varsigma_{\rm RT}(k)$ ($\varsigma_{\rm NRT}(k)$) the corresponding slot number when the packet entered into the queue. For RT and NRT queues of user k at slot s we define a first fairness index [32]

$$\xi_{\rm RT}(k,s)=\frac{[s-\varsigma_{\rm RT}(k)]}{D_{\max,{\rm RT}}(k)}\,, \tag{31}$$

$$\xi_{\rm NRT}(k,s)=\min\left(\frac{[s-\varsigma_{\rm NRT}(k)]}{D_{\max,{\rm NRT}}(k)},1\right)\,. \tag{32}$$

Note that when $\xi_{\rm RT}(k,s)=1$, RT packets are dropped because the deadline is strictly enforced. For NRT packets instead, even when the deadline is reached packets are still kept in the queue for future transmission, but they are considered to have an *excessive delay*.

The second fairness index accounts for the history of user k, i.e., for its average performance up to slot s. Let us indicate with $L_{\rm RT}(k,s)$ ($L_{\rm NRT}(k,s)$) the percentage of discarded (delayed) RT (NRT) packets of user k up to slot s, averaged over a window $W_{\rm RT}$ ($W_{\rm NRT}$). Let $L_{\max,\rm RT}(k)$ ($L_{\max,\rm NRT}(k)$) be a parameter indicating the maximum tolerated discard (excess delay) rate for user k. The second fairness index is defined as

$$\eta_{\rm RT}(k,s) = \min\left(\frac{L_{\rm RT}(k,s)}{L_{\max,\rm RT}(k)}, 1\right), \tag{33}$$

$$\eta_{\rm NRT}(k,s) = \min\left(\frac{L_{\rm NRT}(k,s)}{L_{\max,\rm NRT}(k)}, 1\right) \tag{34}$$

for RT and NRT traffic queues, respectively.

From indices $\xi_{\rm RT}(k,s)$, $\xi_{\rm NRT}(k,s)$, $\eta_{\rm RT}(k,s)$ and $\eta_{\rm NRT}(k,s)$ we obtain a fairness priority index for each of the two queues, relative to RT and NRT traffic, respectively, as

$$\phi_{\rm RT}(k,s) = \begin{cases} \frac{\xi_{\rm RT}(k,s)+\eta_{\rm RT}(k,s)}{4} & \text{if } \eta_{\rm RT}(k,s) < 1 \\ 1 & \text{if } \eta_{\rm RT}(k,s) = 1, \end{cases} \tag{35}$$

$$\phi_{\rm NRT}(k,s) = \begin{cases} \frac{\xi_{\rm NRT}(k,s)+\eta_{\rm NRT}(k,s)}{4} & \text{if } \eta_{\rm NRT}(k,s) < 1 \\ 1 & \text{if } \eta_{\rm NRT}(k,s) = 1. \end{cases} \tag{36}$$

The fairness user index is then derived as

$$\eta(k,s) = \frac{\phi_{\rm NRT}(k,s) + \phi_{\rm RT}(k,s)}{\sum_{\ell=1}^{K} \phi_{\rm RT}(\ell,s) + \phi_{\rm NRT}(\ell,s)}. \tag{37}$$

Lastly, indices $I_{\rm ch}$ and η are combined to obtain the weight for user k as

$$w_k(s) = \begin{cases} \eta(k,s) + [1-\eta(k,s)]I_{\rm ch}(k,s) & \text{if } \eta(k,s) > 0 \\ 0 & \text{if } \eta(k,s) = 0, \end{cases} \tag{38}$$

and if $\eta(k,s) \leq 0$ the power of user k is set to zero.

Once PC and UO have been performed, for each user packets are transmitted by choosing the RT traffic first if $\phi_{\rm RT}(k,s) \geq \phi_{\rm NRT}(k,s)$, or the NRT traffic first if $\phi_{\rm RT}(k,s) < \phi_{\rm NRT}(k,s)$.

APPENDIX II
TRAFFIC MODEL

Each video source is obtained by aggregating the traffic of S_v independent minisources, each characterized by an ON-OFF Markov chain [27], [28]. When in the ON state, the minisource generates bits at a constant rate R_v, while no traffic is generated in the OFF state. The number of slots spent in the ON (OFF) state is geometrically distributed with mean $T_{\rm ON}$ ($T_{\rm OFF}$). We indicate with $\bar{m}$ and σ the mean and standard deviation of the bit-rate of a video source. Then the parameters of the video source are obtained as

$$T_{\rm ON} = \frac{1}{aT_{\rm S}}\left(1 + \frac{\bar{m}^2}{S_v\sigma^2}\right) \tag{39a}$$

$$T_{\rm OFF} = \frac{1}{aT_{\rm S}}\left(1 + \frac{S_v\sigma^2}{\bar{m}^2}\right) \tag{39b}$$

$$R_v = \frac{\bar{m}}{S_v} + \frac{\sigma^2}{\bar{m}}, \tag{39c}$$

where a characterizes the slope of the auto covariance of the bit-rate.

For each web source we consider a Markov chain with two states: packet call and reading [27]. When in the packet call state, the source generates a number of messages geometrically distributed with mean $T_{\rm pc}$. Each message has a duration in bytes ranging from $L_{\min}$ to $L_{\max}$ and the duration is generated as $L = \lfloor x \rfloor$, where x is random variable having as probability density function a truncated Pareto function

$$\begin{aligned} p(x) = & \frac{\zeta L_{\min}^{\zeta}}{x^{\zeta+1}}\left[1(x-L_{\min}) - 1(x-L_{\max})\right] \\ & + \left(\frac{L_{\min}}{L_{\max}}\right)^{\zeta}\delta(x-L_{\max}), \end{aligned} \tag{40}$$

with $1(x)$ the step function, $\delta(x)$ the Dirac delta function and $\zeta = 1.1$. The inter-arrival time between messages is geometrically distributed with mean $T_{\rm mes}$. The number of slots spent in the reading state is also geometrically distributed with mean $T_{\rm rs}$ and in this state no traffic is generated.

REFERENCES

[1] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushyana, and S. Viterbi, “CDMA/HDR: a bandwidth efficient high speed wireless data service for nomadic users,” *IEEE Commun. Mag.* vol. 38, no. 7, pp. 70–77, July 2000.

[2] K. Kumaran and L. Qian, “Uplink scheduling in CDMA packet-data systems,” in *Proc. 22nd IEEE Conf. on Computer and Commun. (INFOCOM)*, San Francisco, CA, Mar. 2003.

[3] K. Kumaran and L. Qian, “Scheduling on uplink of CDMA packet data network with successive interference cancellation,” in *Proc. Wireless Commun. Networking (WCNC) 2003*, vol. 3, pp. 1645–1650, Mar. 2003.

[4] F. Khan, “A time-orthogonal CDMA high-speed uplink data transmission scheme for 3G and beyond,” *IEEE Commun. Mag.*, pp. 88–94, Feb. 2005.

[5] L. Wang, A. H. Aghvami, and W. G. Chambers, “Design issues of uplink media access control (MAC) protocols for multimedia traffic over DS-CDMA systems,” *IEEE Trans. Multimedia*, vol. 7. no. 3, pp. 551-562, June 2005.

[6] H. Jiang, W. Zhuang, X. (S) Shen, and Q. Bi, “Quality of service provisioning and efficient resource utilization in CDMA cellular communications,” *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, pp. 4–15, Jan. 2006.

[7] “3GPP, Technical Specification Group Radio Access Network; Spreading and Modulation (TDD),” 3G TS 25.223 version 3.5.0., 1999.

[8] V. K. Garg, *IS-95 CDMA and CDMA 2000: Cellular/PCS Systems Implementation*. Englewood Cliffs, NJ: Prentice Hall, 1999.

[9] M. Kim, C. G. Kang, I.-C. Choi, and R. R. Rao, “Ordered packet length based groupwise transmission scheme for rate scheduling in burst switching DS/CDMA system,” *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1426–1437, July 2005.

[10] A. Bedekar, S. Borst, K. Ramanan, P. Whiting, and E. Yeh, “Downlink scheduling in CDMA data networks,” in *Proc. IEEE Globecom*, 1999.

[11] F. Berggren, S.-L. Kim, R. Jantti, and J. Zander, “Joint power control and intracell scheduling of DS-CDMA nonreal time data,” *IEEE J. Select. Areas Commun.*, vol. 19, no. 10, pp. 1860-1870, Oct. 2001.

[12] P. Patel and J. Holzman, “Analysis of a simple successive interference cancellation scheme in a DS/CDMA system,” *IEEE J. Select. Areas Commun.*, vol. 12, pp. 727–807, June 1994.

[13] J. Hou, J. E. Smee, H. Pfister, and S. Tomasin, “Implementing interference cancellation to increase the EV-DO RevA reverse link capacity,” *IEEE Commun. Mag.*, vol. 44, no. 2, pp. 58–64, Feb. 2006.

[14] J. C. Andrews and T. H. Meng, “Optimum power control for successive interference cancellation with imperfect channel estimation,” *IEEE Trans. Commun.*, vol. 2, pp. 375–383, Mar. 2003.

[15] A. Agrawal, J. Andrews, J. Cioffi, and T. Meng, “Iterative power control for imperfect successive interference cancellation,” *IEEE Trans. Wireless Commun.*, vol. 4, pp. 878–884, May 2005.

[16] R. M. Buehrer, “Equal BER performance in linear successive interference cancellation for CDMA systems,” *IEEE Trans. Commun.*, vol. 49, pp. 1250–1258, July 2001.

[17] D. Warrier and U. Madhow, “On the capacity of cellular CDMA with successive decoding and controlled power disparities,” in *Proc. Vehic. Tech. Conf. (VTC)*, vol. 3, Ottawa, Canada, pp. 1873–1877, May 1998.

[18] F. Berggren and S. Ben Slimane, "Successive interference calcellation in multi-rate DS-CDMA systems," in *Proc. IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun.*, Beijing, China, vol. 2, pp. 1752–1756, Sept. 2003.

[19] K. Puttegowda, G. Verma, S. Bali, and R. M. Buehrer, "On the effect of cancellation order in successive interference cancellation for CDMA systems," in *Proc. Vehic. Tech. Conf. (VTC)*, Orlando, FL, pp. 1035–1039, Oct. 2003.

[20] N. Benvenuto, G. Carnevale, and S. Tomasin, "Optimum power control and ordering in SIC receivers for uplink CDMA systems," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Seoul, Korea, May 2005.

[21] N. Benvenuto, G. Carnevale, and S. Tomasin, "Joint power control and receiver optimization of CDMA transceivers using successive interference cancelation," *IEEE Trans. Commun.*, vol. 55, no. 3, pp. 563-573, Mar. 2007.

[22] S. Andradottir, "A global search method for discrete stochastic optimization," *Siam. J. Optimization*, vol. 6, no. 2, pp. 513-530, May 1996.

[23] S. Andradottir, "A method for discrete stochastic optimization," *Manag. Sci.*, vol. 41, no. 12, pp. 1946-1961, May 1995.

[24] V. Krishmaurty, X. Wang, and G. Yin, "Spreading code optimization via discrete stochastic approximation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1927-1941, Sept. 2004.

[25] C. Li and X. Wang, "Adaptive opportunistic fair scheduling over multiuser spatial channels," *IEEE Trans. Commun.*, vol. 53, no. 10, pp. 1708–171, Oct. 2005.

[26] V. Krishnamurthy, C. R. N. Athaudage, and H. Dawei, "Adaptive OFDM synchronization algorithms based on discrete stochastic approximation," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1561–1574, Apr. 2005.

[27] A. Andreadis and G. Giambene, *Protocols for High-Efficiency Wireless Networks*. Boston, MA: Kluwer, 2002.

[28] C. Blondia and O. Casals, "Performance analysis of statistical multiplexing of VBR sources," in *Proc. Infocom '92*, pp. 828–838, Florence, Italy, 1992.

[29] M. Tan and Y. Bar-Ness, "Equal BER power control for uplink MC-CDMA with MMSE successive interference cancellation," *IEEE Commun. Lett.*, vol. 8, no. 6, pp. 348–350, June 2004.

[30] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling for maximum throughput in multihop radio networks," *IEEE Trans. Automatic Control*, vol. 37, no. 12, pp. 1936–1949, Dec. 1992.

[31] L. Tassiulas, "Scheduling and performance limits of networks with constantly changing topology," *IEEE Trans. Inform. Theory*, vol. 43, no. 3, pp. 1067–1073, May 1997.

[32] F. De Angelis, I. Habib, G. Giambene, and S. Giannetti, "Scheduling for differentiated traffic types in HSDPA cellular systems," in *Proc. Globecom 2005*, St. Louis, MO, Dec. 2005.

[33] A. Farrokh, F. Blomer, and V. Krishnamurthy, "A comparison of opportunistic scheduling algorithms for streaming media in high-speed downlink packet access (HSDPA)," in *Proc. MIPS 2004* Grenoble, France, Nov. 2004.

[34] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 884–894, June 1999.

[35] W. Rudin, *Principles of Mathematical Analysis.* New York: McGraw-Hill, 1976.

[36] L. A. Wolsey and G. L. Nemhauser, *Integer and Combinatorial Optimization*. New York: Wiley-Interscience, 1999.

[37] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: Freeman, 1979.

[38] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *Proc. Vehic. Tech. Conf. 2000 Spring*, Tokyo, Japan, vol. 3, pp. 1854–1858, May 2002.

Daniele Veronesi was born in Rovereto, Italy, on August 4, 1979. He received the Laurea degree in Telecomunication Engineering (2003) from the University of Padova, Italy. From January 2004 to December 2006, he has been a Ph.D. student at the University of Padova. During the academic year 2005, he has been six months at the University of Massachusetts, Amherst (USA) as exchanging Ph.D. student. After receiving the Ph.D. degree (2006) he joined STMicroelectronics as a consulting digital design engineer.

Stefano Tomasin (S'99-M'04) received the Laurea degree and the Ph.D. degree in Telecommunications Engineering from the University of Padova, Italy, in 1999 and 2002, respectively. In the Academic year 1999–2000 he was on leave at the IBM Research Laboratory, Zurich, Switzerland, doing research on signal processing for magnetic recording systems. In the Academic year 2001–2002 he was on leave at Philips Research, Eindhoven, the Netherlands, studying multicarrier transmission for mobile applications. He joined the University of Padova first as contractor researcher for a national research project (2002) and then as Assistant Professor (2005). In the second half of 2004 he was visiting faculty at Qualcomm, San Diego (CA) doing research on receiver design for mobile cellular systems. His current research interests include signal processing for wireless communications, access technologies for multiuser/multiantenna systems and cross–layer protocol design and evaluation.

Nevio Benvenuto (S'81-M'82-SM'88) received the Laurea degree from the University of Padova, Padova, Italy, and the Ph.D. degree from the University of Massachusetts, Amherst, in 1976 and 1983, respectively, both in electrical engineering. From 1983 to 1985 he was with AT&T Bell Laboratories, Holmdel, NJ, working on signal analysis problems. He spent the next three years alternating between the University of Padova, where he worked on communication systems research, and Bell Laboratories, as a Visiting Professor. From 1987 to 1990, he was a member of the faculty at the University of Ancona. He was a member of the faculty at the University of L'Aquila from 1994 to 1995. Currently, he is a Professor in the Electrical Engineering Department, University of Padova. His research interests include voice and data communications, digital radio, and signal processing.