Energy-Efficient Power and Subcarrier Allocation for OFDMA Systems with Value Function Approximation Approach

Kai Bi, Qinghai Yang, Fenglin Fu
State Key Laboratory on ISN, School of Telecomm.
Engineering, Xidian University, No.2 Taibainan-lu,
Xi'an, 710071, Shaanxi, China.
Email:qhyang@xidian.edu.cn

Kyung Sup Kwak
Graduate School of IT and Telecomm.,
Inha University, #253 Yonghyun-dong,
Nam-gu, Incheon, 402-751, Korea.
Email: kskwak@inha.ac.kr

Abstract—In this paper, we conceive an energy-efficient power and subcarrier allocation scheme for downlink OFDMA systems under average delay constraints. The services associated with multi-users have different packet arrivals and delay requirements. The problem of dynamic power and subcarrier allocation is formulated as a constrained Markov decision process (CMDP) with control actions based on the joint states of channel state information (CSI) and the queue state information (QSI). An online learning algorithm is developed for solving the CMDP problem with the aid of stochastic approximation and the value function approximation approach.

Index Terms—Energy-efficient, resource allocation, constrained Markov decision process, stochastic approximation, value function approximation.

I. INTRODUCTION

With the rapid development of wireless systems in terms of the number of subscribers and applications, two key design challenges lie in the power-efficient transmission of portable handsets, as well as the delay-sensitive multimedia transmission applications such as video conference, remote teaching/training, etc. Both of the two challenges require highly efficient scheduling and resource allocation schemes.

Many efforts have been made on the cross-layer optimization of power and subcarrier allocation for orthogonal frequency division multiple access (OFDMA) systems. The traditional view of considering the energy-efficient and delay-sensitive application was to treat the wireless system as a layered network [1]. In the single user scenario, the pioneer work of [2] investigated the energy efficient delay-constraint cross-layer scheduling scheme under the channel variation and buffer dynamics. Following the model of Berry-Gallager, the energy efficient transmission was investigated under the average delay constraint in the constrained Markov decision process (CMDP) framework [3-5]. Literature [5] analyzed the structural properties of optimal transmission policies over a randomly varying channel. A stable online algorithm was developed in [6], which is interpreted as "one component at a time", via introducing a middle state called post-decision state. The work of [7] extended the model in [6] to a single-cell
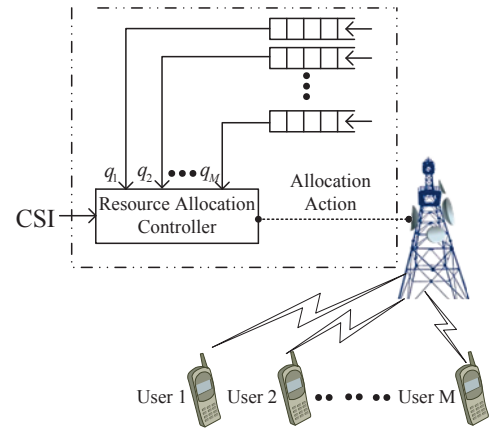

Fig. 1.   System model

multiuser wireless uplink system and provided the convergence analysis and queue stability proof.

In this paper, we investigate an energy-efficient resource allocation with value function approximation under average delay constraint. It proves that the post-decision value function for the problem of the energy efficient scheduling under average delay constraint is convex on the backlog at the BS. And the post-decision value function is approximated using the piecewise linear function to attain quick convergence.

II. SYSTEM MODEL AND STATE DEFINITION

A. System Model

We consider an OFDMA downlink system with M users, N_F subcarriers, and one base station (BS), as depicted in Fig. 1. BS manages M queues, which are associated with M users, respectively. In a non-homogenous user scenario, different application streams may have different arrival rates, and delay requirements. The subcarrier and power allocation is performed with respect to the effects of both channel state information (CSI) and queue state information (QSI).

We assume the channel is time-varying but keeps constraint over the interested time interval. We use a finite-state Markov

model to describe the fading channel. The corresponding one-step transition probability is computed by the method in [9].

We assume that, the packets from higher layers arrive at the end of each slot and their arrivals of the packets are independently identically distributed (i.i.d). The packet arrival process and the channel state evolution process are assumed to be stationary and independent. The packets of different users are stored in the corresponding queues at BS for transmission, respectively. We denote q_i^t and U_i^t as the queue length in the buffer of user i and the number of transmitted packets at time slot t. If the number of data arrivals is A_i^{t+1}, the queue length at the next time slot can be expressed as:

$$q_i^{t+1} = q_i^t - U_i^t + A_i^{t+1}. \tag{1}$$

Supposing the buffer size is N_Q in a practical system, the queue length at time $t+1$ is $q_i^{t+1} = \min(q_i^t - U_i^t + A_i^{t+1}, N_Q)$. If N_Q is large enough, the buffer overflow can be neglected. Next, let denote $\bar{q}$ and $\bar{D}$ as the average queue length and average delay, respectively. According to Little's Theorem, it holds $\bar{q} = \bar{a}\bar{D}$ where $\bar{a}$ is a parameter. That is, the delay constraint of each user can be mapped to that of the queue length.

B. State Definition

The cardinality of the set of QSI is $L = (N_Q + 1)^M$, which grows exponentially with M. Denote the system state as: $s^t = (s_1^t, s_2^t, \cdots, s_L^t)$ with $s_i^t = (q_i^t, h_i^t)$, where h_i^t represents the channel gain of user i at time slot t. The joint allocation scheduling policy of M queues is stated as $\pi^t = (\pi_{1,1}^t, \pi_{1,2}^t, \cdots, \pi_{1,N_F}^t; \cdots; \pi_{i,1}^t, \pi_{i,2}^t, \cdots, \pi_{i,N_F}^t; \cdots; \pi_{M,1}^t, \pi_{M,2}^t, \cdots, \pi_{M,N_F}^t)$. $\pi_{i,n}^t$ is an indicator that indicates whether queue i gets the n-th subcarrier at time slot t. $\pi_{i,n}^t = 1$ means queue i gets the n-th subcarrier at current time slot, and $\pi_{i,n}^t = 0$ represents waiting for transmission in the next slot. The queue length in the next time slot is affected by the current queue length and the number of packets to be transmitted. We assume the channel state acts as an ergodic Markov chain and inevitably transits into another state at the next time slot. Thus, it can be formulated as a (CMDP) problem.

We denote the state transition from time t to $t+1$ by $s^t \rightarrow s^{t+1}$, represented by:

$$s_i^t = (q_i^t, h_i^t) \xrightarrow{y_i^t} s_{i'}^{t+1} = (q_{i'}^{t+1}, h_{i'}^{t+1}).$$

Assuming that the queue length and the channel fading are independent of each other, the probability of state transition is expressed by:

$$\begin{aligned} p(s_{i'}^{t+1}|s_i^t, y_i^t) &= p(q_{i'}^{t+1}|q_i^t, y_i^t)p(h_{i'}^{t+1}|h_i^t) \\ &= p(A_i^{t+1})p(h_{i'}^{t+1}|h_i^t). \end{aligned} \tag{2}$$

III. Problem Formulation

Given the defined the system state $s_i^t = (q_i^t, h_i^t)$, the transmitter may adjust the transmit power and subcarrier allocation according to a stationary policy $\pi^t = (\pi_1^t, \pi_2^t, \cdots, \pi_{N_F}^t)$. A transmission policy π^t is a mapping from a system state s^t to

a power control and subcarrier allocation actions. A policy π^t is called feasible if the allocation actions satisfy the average total transmit power constraint and the subcarrier assignment constraint. Power allocation should satisfy:

$$\sum_{i=1}^{M} \sum_{n=1}^{N_F} E[P_{i,n}] \leq P_0, \quad P_{i,n} \geq 0, \tag{3}$$

where P_0 denotes the maximum power that allowed to transmit packets at BS.

The subcarrier allocation should satisfy the constraint:

$$\sum_{i=1}^{M} \pi_{i,n} = 1, \quad \forall n \in \{1, N_F\}. \tag{4}$$

We write the infinite horizon discounted power expenditure that queue i consumed:

$$P_i = \mathcal{E}\left[\sum_{t=1}^{\infty} \alpha^t P(U_i^t, h_i^t)\right]. \tag{5}$$

Similarly, the average delay of queue i is expressed as:

$$Q_i = \mathcal{E}\left[\sum_{t=1}^{\infty} \alpha^t q_i^t\right], \tag{6}$$

where α is the discounted factor in the range of $[0, 1)$.

Accordingly, the problem may be formulated as:

$$\begin{aligned} &\min \ P_i \\ &\text{s. t.} \quad Q_i \leq \zeta_i, \quad i = 1, \cdots, M. \end{aligned} \tag{7}$$

where ζ_i denotes the maximum delay that the i-th user could tolerate.

IV. Allocation Policy

The resource allocation controller computes which queue will be allocated one or more subcarriers for transmission based on CSI at each time slot. The allocation decision in a time slot affects the buffer occupancy of all the queues in future slots, in other words, the average delay of everybody. At each time slot, the controller calculates the transmission rate based on QSI and power expenditure for each queue online based on multiple traffic's delay constraint. The rate for each queue is denoted as y_i^t. Combining this rate with CSI at current slot, we can compute the corresponding power expenditure P_i^t. In this way, the controller analyzes all the informations and finally makes the allocation decision. In the next time slot, the controller remake the allocation decision based on the decision made at the last time slot as well as the new CSI and QSI.

V. Online Learning with Value Function Approximation

A. Lagrangian Approach

We define the utility function as the long-term average power expenditure under average delay constraints. The controller computes the required transmission rate for an individual queue online based not only on the current queue length and channel state but also on the future likely states. While

introducing Lagrange factor, the constrained problem can be converted to a non-constrained problem.

It is equivalent to maximize:

$$L(\lambda_i^t, q_i^t, h_i^t, y_i^t) = \mathcal{E}\Big[\sum_{t=1}^{\infty} \alpha^t \times$$
$$\big[P(h_i^t, y_i^t) - \lambda_i^t(Q(q_i^t, y_i^t) - \zeta_i)\big]\Big], \tag{8}$$

where y_i^t is the required transmission rate for queue i. For a certain λ_i^t, this problem is a non-constrained MDP problem. Depriving the delay deadline $\zeta_i, i = 1, \cdots, N$, according to [10], the non-constrained Markov decision process problem can be solved by:

$$\mu_\zeta^*(s_0) = \min_{\lambda_i^t \geq 0} \max_{y_i^t \in \Phi} N_{y_i^t, \lambda_i^t}(s_0) + \lambda_i^t \zeta_i, \tag{9}$$

where the so-called normal state value function is expressed as:

$$N_{y_i^t, \lambda_i^t}(s_0) = \mathcal{E}\Big[\sum_{t=1}^{\infty} \alpha^t \times$$
$$\big[P(h_i^t, y_i^t) - \lambda_i^t(Q(q_i^t, y_i^t))|s_0\big]\Big]. \tag{10}$$

Further based on dynamic optimization theory, it is equivalent to solve:

$$N_{i*, \lambda_i^t}(s) = \min_{y_i^t \in \Phi_i} \Big[P\big(y_i^t(s), h_i^t\big) - \lambda_i^t Q\big(q_i^t, y_i^t(s)\big)$$
$$+\alpha \sum_{s'} p\big(s'|s, y_i^t(s)\big) N_{i*, \lambda_i^t}(s')\Big], \forall s. \tag{11}$$

B. Structural Analysis of Post-Decision State Value Function

The classic approaches to solve CMDP are based value iteration, and policy iteration, etc, in which the knowledge of a prior probability is necessary to be known. However, this requires the probability distributions of the arrivals and channel states, which are generally unknown in practical communication system. Meanwhile, when determining the optimal solution, the optimal Lagrange factor is necessary to be found.

From [6], after a post-decision state is introduced, the conditional expectation and the maximization operator can be swapped. Besides, the prior probability information is separated from the normal state value function. Thus, we can represent the normal state value function through post-decision state function explicitly. The following analysis shows that the post-decision state function is coupled with the normal state value function.

The definition of the post-decision state is given by:

$$s_i^t = (q_i^t, h_i^t) = (q_i^t - y_i^t, h_i^t). \tag{12}$$

The queue length denotes the buffer occupancy after the data transmitted and before the new data comes. The post-decision channel state inherits the same as the current one.

Remark 1: Due to the competition for the common channel, not all the requested transmission y_i^t is guaranteed, so the actual queue length variation follows: $q_i^t = q_i^t - \sum_{n=1}^{N_F} \pi_{i,n}^t y_i^t$;

The post-decision state value function $V_{*,\lambda}(q, h)$ represents the optimal long-term utility starting from the post decision state,

$$V_{*,\lambda}(q, h) = \sum_A \sum_{h' \in H} p(A)p(h'|h)N_{*,\lambda}\big((q + A), h'\big). \tag{13}$$

The corresponding optimization problem can be expressed as:

$$N_{*,\lambda}(q, h) = \min_{0 \leq y \leq q} \big[P(h, y) - \lambda Q(q, y)$$
$$+\alpha V_{*,\lambda}(q - y, h)\big]. \tag{14}$$

Based on the relative value iterative algorithm, we can update the components of post-decision state function in a "one component at a time" way as [6]. However, this approach requires relatively large storage space. This prompts us to analyze the structure properties of the optimal required transmission rates. Meanwhile, we suppose the post-decision state value function preserves some properties as the function in [8].

The following process shows the properties of the post-decision state value function. From (13), we note that to compute the post-decision state value function, it is necessary to know the underlying dynamics, i.e. the distribution of the traffic arrivals and the channel transition probability; that is, to perform the expectation on all possible states over the normal-state value function (c.f. Eq.(10)). For simplicity, the expectation can be replaced via the averaging over time. The update of the post-decision state value function is given by:

$$V_{t,\lambda}(\tilde{q}, h^{t-1}) = (1 - \beta^t)V_{t-1,\lambda}(\tilde{q}, h^{t-1})$$
$$+ \beta^t J_{t,\lambda}((\tilde{q} + A), h^t), \forall \tilde{q} \tag{15}$$

$$V_{t,\lambda}(\tilde{q}, \tilde{h}) = V_{t-1,\lambda}(\tilde{q}, \tilde{h}), \forall \tilde{q}, \tilde{h} \neq h^{t-1}.$$

Similar to the analysis in [8], some results on the post-decision state value function are developed given the objective and constraints in the paper. Here, we shall approximate the value function using piecewise linear function [11].

Proposition: The maximum transmission rate for each queue $y_i(q, h), i = 1, \cdots, N$ is non-decreasing with q and the post-decision state value function $V_{*,\lambda}(\tilde{q}, \tilde{h})$ is convex of q.

Proof: First, we construct the power function and the delay function as $P(h, y) = (\sigma^2/|h|^2)(2^y - 1)$ and $Q(q, y) = q - y$, respectively. The optimization function is given by:

$$N_{*,\lambda}(q, h) = \max_{0 \leq y \leq q} \Big[-P(h, y) - \lambda Q(q, y)$$
$$+\alpha V_{*,\lambda}(q - y, h)\Big]. \tag{16}$$

Subsequently, we demonstrate $-P(h, y) - \lambda Q(q, y)$ satisfies the definition of supermodularity.

Supermodularity (definition): For any $y' \geq y, \ x' \geq x$, if the following expression holds:

$$f(x', y') - f(x', y) \geq f(x, y') - f(x, y). \tag{17}$$

Then, the function $f : X \times Y \to R$ is a super modularity on the pair of (x, y).

Let $f(q, y) = -P(h, y) - \lambda Q(q, y)$, for any $q \geq q', y \geq y'$. Then, it yields

$$f(q', y') - f(q', y)$$

$$= -P(h, y') - \lambda Q(q', y') - [-P(h, y) - Q(q', y)] \quad (*1)$$

$$= -\frac{\delta^2}{|h|^2}(2^{y'} - 1) + \frac{\delta^2}{|h|^2}(2^y - 1) - \lambda(q' - y') + \lambda(q' - y),$$

$$f(q, y') - f(q, y)$$

$$= -P(h, y') - \lambda Q(q, y') - [-P(h, y) - Q(q, y)] \quad (*2)$$

$$= -\frac{\delta^2}{|h|^2}(2^{y'} - 1) + \frac{\delta^2}{|h|^2}(2^y - 1) - \lambda(q - y') + \lambda(q - y).$$

Obviously, $(*1) - (*2) = 0$. Thus, we verified that $-P(h, y) - \lambda Q(q, y)$ is supermodularity in pair of (q, y).

Next, we shall show that $V_{m-1,\lambda}(q - y, h)$ is supermodularity in pair of (q, y) and is also a convex function on q.

For any $V_{0,\lambda}(q, y)$, the value iteration ensures that the post-decision state function converges to the $V_{*,\lambda}(q, y)$[8]. Supposing that $V_{0,\lambda}(q, y) = q$ and $V_{m-1,\lambda}(q, h)$ is monotonic and convex on q, we have

$$V_{m-1,\lambda}(q_1, h) + V_{m-1}^{\lambda}(q_2, h)$$

$$\leq V_{m-1,\lambda}((1 - \eta)q_1 + \eta q_2, h) + V_{m-1,\lambda}(\eta q_1 + (1 - \eta)q_2, h),$$

$$\forall \eta \in [0, 1].$$

considering $q' \geq q, y' \geq y$ and $q \geq y'$.

Let $q_1 = q' - y$, $q_2 = q - y'$ and $\eta = \dfrac{y' - y}{q' - q + y' - y}$. The following two expressions hold:

$$(1 - \eta)q_1 + \eta q_2 = q' - y'.$$

$$\eta q_1 + (1 - \eta)q_2 = q - y.$$

Thus, we obtain

$$V_{m-1,\lambda}(q' - y, h) + V_{m-1,\lambda}(q - y', h)$$

$$\leq V_{m-1,\lambda}(q - y, h) + V_{m-1,\lambda}(q' - y', h).$$

Now, we observe that $V_{m-1,\lambda}(q - y, h)$ is supermodularity in the pair of (q, y).

To demonstrate that $V_{m,\lambda}(q, h)$ is convex on q, we first verify $N_{m,\lambda}(q, h) = \max_{0 \leq y \leq q} T_{m,\lambda}(q, h, y)$ is convex on q. For any q_1 and q_2, we assume that the optimal scheduling policies are $y_1 = y^*(q_1, h)$ and $y_2 = y^*(q_2, h)$. For all $\eta \in [0, 1]$, we

have

$$N_{m,\lambda}(q_1, h) + N_{m,\lambda}(q_2, h)$$

$$= -P(h, y_1) - \lambda Q(q_1, y_1) + \alpha V_{m-1}^{\lambda}(q_1 - y_1, h) - P(h, y_2)$$

$$\quad - \lambda Q(q_2, y_2) + \alpha V_{m-1,\lambda}(q_2 - y_2, h)$$

$$\leq -P(h, \eta y_1 + (1 - \eta)y_2) - \lambda Q(\eta q_1 + (1 - \eta)q_2, \eta y_1 + (1 - \eta)y_2)$$

$$\quad + \alpha V_{m-1,\lambda}(\eta(q_1 - y_1) + (1 - \eta)(q_2 - y_2), h)$$

$$= T_{m,\lambda}(\eta q_1 + (1 - \eta)q_2, h, \eta y_1 + (1 - \eta)y_2)$$

$$\leq N_{m,\lambda}(\eta q_1 + (1 - \eta)q_2, h).$$

On the hypothesis that $Q(q, y)$ is jointly convex in the pair of (q, y) (obviously, the hypothesis holds), we know $N_{m,\lambda}(q, h)$ is convex on q. Meanwhile, $V_{m,\lambda}(q, h) = \sum_A \sum_{h' \in H} p(A)p(h'|h) J_{m,\lambda}(q, h)$. Thus, $V_{m,\lambda}(q, h)$ is convex on q. ∎

C. Online Learning Algorithm with Value Function Approximation

Based on the above analysis, we know $V_{m,\lambda}(q, h)$ is convex in the buffer backlog. The normal state value function $N_{*,\lambda}(q, h)$ can be expressed by $V_{m,\lambda}(q, h)$ explicitly. We note that if we derive the expression $V_{m,\lambda}(q, h)$, the maximum required transmission rate can be easily guaranteed via Eq. (16). However, the closed-form expression of the post-decision state value function is still unknown. Literature [11] introduces a method to approximate the univariate convex function using only function value evaluations. Thus, we can approximate the post-decision state value function using piecewise linear function. Setting a threshold error $\delta_{\max}$, we can adaptively select the points $(q_1, q_2, \cdots, q_{M_{\delta_{max}}})$ to evaluate the value of the post-decision state value function which is convex in the backlog q.

Now, we shall develop the algorithm with value function approximation approach.

---

1) Initialization: initialize the potential post-decision state value matrix for all channel states and queue lengths, the slot counter $t = 0$, post-decision state $\widetilde{s}_0 = (q_0, h_0)$, Lagrange factor $\lambda = 0$ in the 0-th regenerative period.
2) At the beginning of the k-th regenerative period, determine the new channel state q_t and incoming data A_{t-1} at current time slot and compute the normal state $s_t = (\min(q_{t-1} + A_{t-1}, N_Q), h_t)$.
3) Run the system with policy π^t, judge the current post-decision channel state and approximate the corresponding post-decision state value for all queue lengths at least once using a piecewise linear function.
4) At the beginning of the next regeneration, update the

post-decision state value function:

$$V_\lambda^t(q, h^{t-1}) = (1 - \beta^t)V_\lambda^{t-1}(q, h^{t-1})$$

$$+ \beta^t N_\lambda^t(min(q, +a^{t-1}, B), h^t), \forall q.$$

$$V_\lambda^t(q, h'') = V_\lambda^{t-1}(q, h''), \forall q, h'' \neq h^{t-1}$$

and $N_{t,\lambda}(q, h^t) = \min_{0 \leq y \leq q} \Big[P(h^t, y) - \lambda Q(q, y) + \alpha V_\lambda^{t-1}(q - y, h^t) \Big].$

Compute the optimal required transmission rate based on the current normal state $s_t = (min(q_{t-1} + A_{t-1}, N_Q), h_t)$ through

$$N_\lambda^t(q, h)$$
$$= \min_{0 \leq y \leq q} \Big[P(h, y) - \lambda Q(q, y) + \alpha V_\lambda^{t-1}(q - y, h) \Big].$$

5) Suppose BS transmits data at the expected transmission rate y_*^t.
   Update the post-decision state

   $$\widetilde{s}^t = (min(q^{t-1} + a^{t-1}, B) - y_*^t, h^t).$$

   Update the Lagrange factor

   $$\lambda_{n+1,i} = \max[\lambda_{n,i} + \frac{1}{n}(Q_{n,i} - \zeta_i), 0].$$

6) If $\|V_\lambda^{k+1} - V_\lambda^k\| < \delta_v$, stop; Otherwise, set k=k+1 and go to step 2 for a new regeneration.

Remark 2: When implementing this algorithm, we update the post-decision state function simultaneously for all states while remaining the channel state $\hat{h}^{t-1}$ unchanged, instead of updating the post-decision state value function only at current state $\left(\hat{q}^{t-1}, \hat{h}^{t-1}\right)$. This is known as "batch update", which greatly accelerates the convergence rates for the post-decision state value function.

VI. SIMULATION RESULTS

In this section, simulations are provided to illustrate our proposed schemes. The power expenditure function for transmitting the amount of y (in packets) traffic at the channel state h is given by $P(h, y) = \frac{\sigma^2}{|h|^2}\left(2^{\frac{yl}{W}} - 1\right)$, where σ^2 is the variance of the white Gaussian noise, l denotes the packet length and W is the bandwidth. We assume $\frac{\bar{h}^2}{\sigma^2} = 0.14$ with $\bar{h}$ being the average channel gain. In the simulation, we assume there are 2 users, 4 subcarriers with total bandwidth 31.25KHZ, $N_Q = 4$ and $\lambda_i = 80$ packet/s.

Fig. 2 depicts the trend of the norm of the post-decision state value function's different between two adjacent generations. As expected, it has a downward trend and an excellent allocation policy will be found if δ_v sets appropriated. Fig. 3 illustrates the average number of subcarriers varying with regenerations periods of the two users. Fig. 4 and Fig. 5 demonstrate the power expenditure and average delay cost varying with
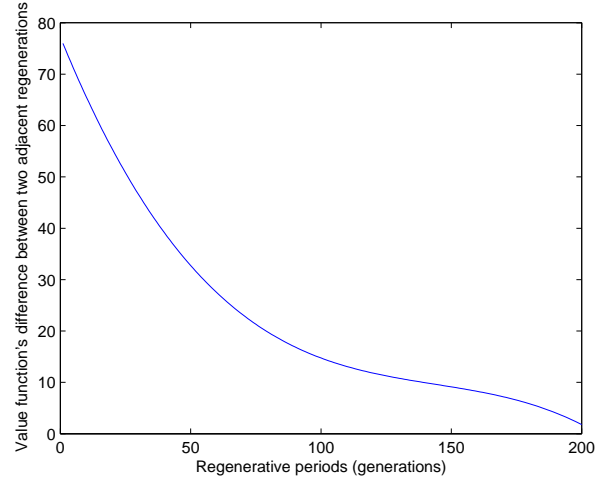

Fig. 2. Graphics trends for $\|V_\lambda^{k+1} - V_\lambda^k\|$ with 200 regenerations by value evaluations.
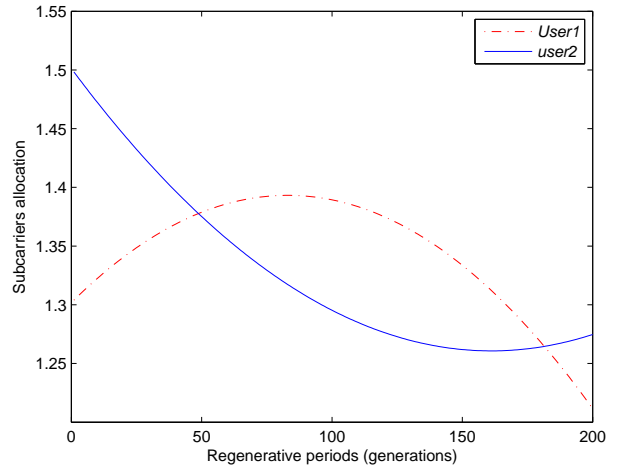

Fig. 3. Average subcarriers allocation versus regenerations periods of the two users.

regenerations periods. In Fig. 4, the average power allocation has the similar trend with the norm of the post-decision state value function's different that showed in Fig. 2. In Fig. 5, as the power allocation decreases, the average queue delay becomes higher. As the iteration number increases, we observe that the average power expenditure converges to a nearly fixed level. Fig. 6 illustrates the average power allocation versus delay with the regenerative periods. It verifies that the algorithm presents a significant energy reduction within average delay constraints. In practical applications, it is unnecessary to wait the algorithm to reach the convergence. Once the delay cost and the power expenditure fall in an acceptable region, it is reasonable to terminate the algorithm. It's not hard to find a sensational resource allocation scheme with low power requirement and non-homogenous delay constraint. Simulation results demonstrate that the algorithm converges to the optimal
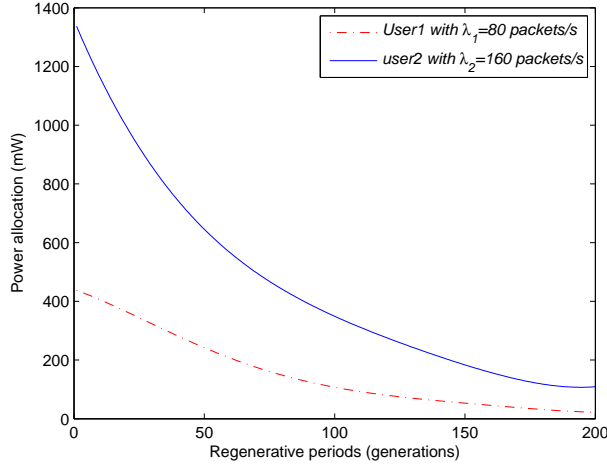
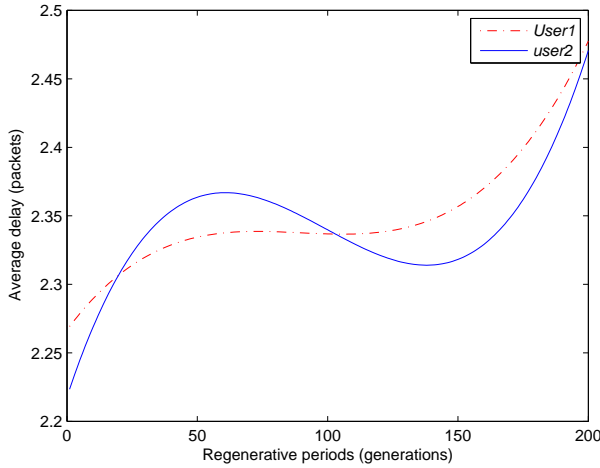Fig. 4. Average power expenditure versus regenerations periods of the two users.

Fig. 6. Average power allocation versus delay.

Fig. 5. Average delay cost versus regenerations periods of the two users.

solution after the acceptable iterations almost surely.

VII. CONCLUSIONS

In this paper, we conceive an energy-efficient power and subcarrier allocation scheme for downlink OFDMA systems under average delay constraints. The problem of dynamic power and subcarrier allocation is formulated as a constrained Markov decision process (CMDP) with the control actions based on CSI and QSI. We prove that the post-decision value function for the problem of the energy efficient scheduling under average delay constraint is convex on the backlog. We derive the equivalent state-reduced Bellman equation and propose an online stochastic value iteration solution to accelerate the algorithm convergence speed.

REFERENCES

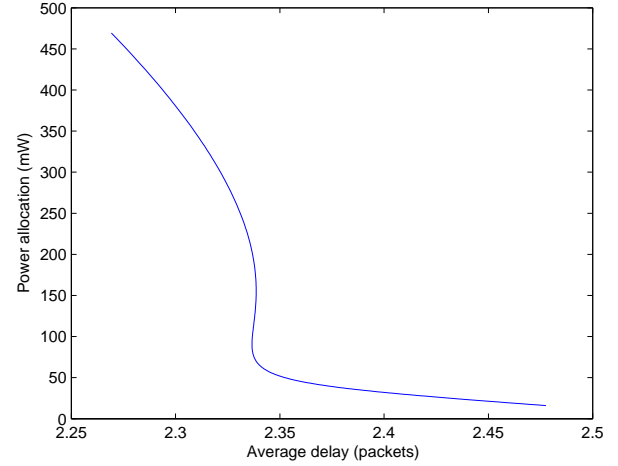[1] D. P. Bertsekas and R. Gallager, "Data Networks", *Prentice Hall*, 1987.

[2] R. A. Berry and R. G. Gallager, "Communication over Fading Channels with Delay Constraints," *IEEE Transactions on Information Theory*, Vol.48, No.5, pp.135-1149,2002.

[3] H. Wang and N. Mandayam, "A Simple Packet Transmission Scheme for Wireless Data over Fading Channels," *IEEE Transcations on Communications*, Vol.52, No.7, pp.1055-1059, 2004.

[4] M. Goyal, A. Kumar, and V. Sharma, "Power Constrained and Delay Optimal Policies for Scheduling Transmissions over a Fading Channel," *Proceedings of IEEE INFOCOM*, Vol.1, pp.311-320, March.2003.

[5] M. Agarwal, A. Karandikar, and V. S. Borkar, "Structural Properties of Optimal Transmission Policies over a Randomly Varying channel," Accepted for publication, *IEEE Transactions on Automatic Control*, 2007.

[6] N. Salodkar, A. Karandikar, and V. S. Borkar, "An On-Line Learning Algorithm for Energy Efficient Delay Constrained Scheduling over a Fading Channel," *IEEE Journal on Selected Areas in Com*, Vol.26, No.4, May.2008.

[7] N. Salodkar, A. Karandikar, and V. S. Borkar, "A Stable Online Algorithm for Energy-Efficient Multiuser Scheduling," *IEEE Transactions on mobile computing*, Vol.9, No.10, Oct.2010.

[8] Fangwen Fu and Mihaela van der Schaar," Structure-Aware Stochastic Control for Transmission Scheduling".

[9] Q. Zhang and S. A. Kassam, "Finite-State Markov Model for Rayleigh Fading Channels," *IEEE Transactions on Communications*, Vol.47, No.11, Nov.1999.

[10] E. Altman, "Constrained Markov decision processes: stochastic modeling," *London: Chapman and Hall*, CRC, 1999.

[11] A. Siem, D. Hertog and A. Hoffmann, "A Method for Approximating Univariate Convex Functions Using Only Function Value Evaluations,"Online, available at SSRN:http://ssrn.com/abstract=1012289.

[12] V. K. N. Lau and Y. Cui, "Delay-Optimal Power and Subcarrier Allocation for OFDMA Systems via Stochastic Approximation," *IEEE Transactions on Wireless Communications*, Vol. 9, No.1, Jan.2010.