Delay-Optimal Power and Subcarrier Allocation for OFDMA Systems via Stochastic Approximation

Vincent K. N. Lau, Senior Member, IEEE, Ying Cui, Student Member, IEEE
Department of ECE, The Hong Kong University of Science and Technology

Abstract—In this paper, we consider delay-optimal power and subcarrier allocation design for OFDMA systems with $N_F$ subcarriers, $K$ mobiles and one base station. There are $K$ queues at the base station for the downlink traffic to the $K$ mobiles with heterogeneous packet arrivals and delay requirements. We shall model the problem as a $K$-dimensional infinite horizon average reward Markov Decision Problem (MDP) where the control actions are assumed to be a function of the instantaneous Channel State Information (CSI) as well as the joint Queue State Information (QSI). This problem is challenging because it corresponds to a stochastic Network Utility Maximization (NUM) problem where general solution is still unknown. We propose an online stochastic value iteration solution using stochastic approximation. The proposed power control algorithm, which is a function of both the CSI and the QSI, takes the form of multi-level water-filling. We prove that under two mild conditions in Theorem 1 (One is the stepsize condition. The other is the condition on accessibility of the Markov Chain, which can be easily satisfied in most of the cases we are interested.), the proposed solution converges to the optimal solution almost surely (with probability 1) and the proposed framework offers a possible solution to the general stochastic NUM problem. By exploiting the birth-death structure of the queue dynamics, we obtain a reduced complexity decomposed solution with linear $O(KN_F)$ complexity and $O(K)$ memory requirement.

I. INTRODUCTION

There are plenty of literature on cross-layer optimization of power and subband allocation in OFDMA systems [1], [2]. Yet, all these works focused on optimizing the physical layer performance and the power/subband allocation solutions derived are all functions of the channel state information (CSI) only. On the other hand, real life applications are delay-sensitive and it is critical to consider the delay performance in addition to the conventional physical layer performance in OFDMA cross-layer design. A combined framework mining the optimal policy using the MDP approach is the huge state space involved. For instance, the state space is exponentially large in the number of users. Hence, brute force solution by value iteration or policy iteration is not applicable due to huge complexity and memory requirement involved.

4 channel states, the system state space contains $2^4 = 16$ states.

• The Curse of Dimensionality A more general approach is to model the problem as a Markov Decision Problem (MDP) [13], [14]. However, a primary difficulty in determining the optimal policy using the MDP approach is the huge state space involved. For instance, the state space is exponentially large in the number of users. Hence, brute force solution by value iteration or policy iteration is not applicable due to huge complexity and memory requirement involved.

• Issues of Convergence in Stochastic NUM Problem In conventional iterative solutions for deterministic NUM problems, the updates in the iterative algorithms (such as subgradient search) are performed within the coherence time of the CSI (the CSI remains quasi-static during the iteration updates). In stochastic NUM, however, convergence is not easily satisfied in most of the cases we are interested.)

1For a system with 4 users, 6 subcarriers, a buffer size of 50 per user and 4 channel states, the system state space contains $4^{4 \times 6} \times (50 + 1)^5$ states, which is already unmanageable.

2This poses a serious limitation on the practicality of the distributive iterative solutions because the convergence and the optimality of the iterative solutions are not guaranteed if the CSI changes significantly during the update.
the updates are performed over the ergodic realizations of the system states, and hence, the convergence proof is challenging (such as GPD in [11] and RT-SPD in [12]). When we consider the delay-optimal problem, the problem is stochastic and the control actions are defined over the ergodic realizations of the system states (CSI,QSI). Therefore, the convergence proof is also quite challenging.

- **Heterogeneous Users** There are some works that obtained a simple delay-optimal policy for multiple access channels by using majorization theory and exploiting symmetry between users [6], [7]. However, such simplifications cannot be extended easily to our case in which users have heterogeneous arrivals and delay requirements.

In this paper, we shall address the above issues by proposing a low-complexity solution to the delay-optimal OFDMA system. To address the open issue concerning the huge complexity involved in solving the \( K \)-dimensional MDP, we utilize the stochastic approximation (SA) techniques [16], [17] to derive a low complexity online stochastic value iteration algorithm. We shall show that under some mild conditions, the proposed online stochastic value iteration algorithm converges to the optimal solution almost surely (with probability 1). By exploiting the birth-death structure of the queue dynamics, we obtain a reduced complexity decomposed solution with linear \( O(KN_F) \) complexity and \( O(K) \) memory requirement.

II. System Models

In this section, we shall elaborate the system model, the OFDMA physical layer model as well as the underlying queueing model. Fig. 1 illustrates the top level system model. There are \( K \) user queues at the base station which buffer packets for the \( K \) mobile users in the OFDMA system. These \( K \) application streams may have different source arrival rates and delay requirements and this corresponds to a heterogeneous user situation. The base station has a cross-layer scheduler which takes the CSI and joint QSI as the inputs and produces a power allocation and subcarrier allocation action as outputs.

![Fig. 1. OFDMA physical layer and queueing model.](image)

A. OFDMA Physical Layer Model

We consider an OFDMA system with \( N_F \) subcarriers over a frequency selective fading channel with \( L \) independent multipaths. As a result, the received signal at the \( k \)-th mobile and \( n \)-th subcarrier is given by \( Y_{k,n}^{t} = H_{k,n}^{t}X_{k,n}^{t} + Z_{k,n} \), where \( X_{k,n}^{t} \) is the transmitted symbol and \( H_{k,n}, Z_{k,n} \sim CN(0,1) \) are fading coefficient (CSI) and noise respectively. Let \( s_{k,n} \in \{0,1\} \) denote the subcarrier allocation index for the \( k \)-th user at the \( n \)-th subcarrier. For simplicity, we assume powerful channel coding (such as LDPC) at the transmitter. Furthermore, the mobile receiver has perfect knowledge of CSI. Hence, the maximum achievable data rate of the \( k \)-th user is given by the mutual information between the channel inputs \( \{X_{k,n} : s_{k,n} = 1\} \) and the channel outputs \( \{Y_{k,n} : s_{k,n} = 1\} \), which is given by:

\[
R_k = \sum_{n=1}^{N_F} s_{k,n} I(X_{k,n}; Y_{k,n}|H_{k,n}) \\
= \sum_{n=1}^{N_F} s_{k,n} \log (1 + p_{k,n}|H_{k,n}|^2) 
\]

(1)

B. Queue Model, System Dynamics and Control Policy

In this paper, the time dimension is partitioned into scheduling slots indexed by \( t \). Let \( \tau \) (second/slot) be the slot duration.

**Assumption 1**: The joint CSI \( H(t) = \{H_{k,n}(t)\forall n,k\} \) remains quasi-static within a scheduling slot and i.i.d. between scheduling slots.

There are \( K \) queues at the base station transmitter for the \( K \) mobiles respectively. Data arrives in packets according to \( K \) random arrival processes and each packet is stored in one of the \( K \) queues according to its destination. Let \( \mathbf{A}(t) = (A_1(t), \ldots, A_K(t)) \) and \( \mathbf{N}(t) = (N_1(t), \ldots, N_K(t)) \) be the random new packet arrivals at the end of the \( t \)-th scheduling slot and the packet sizes of the packet in the front of the queues for the \( K \) users at the beginning of the \( t \)-th scheduling slot, respectively.

**Assumption 2**: The arrival process \( A_k(t) \) and random packet size \( N_k(t) \) are i.i.d. over scheduling slots.

Let \( \mathbf{Q}(t) = (Q_1(t), \ldots, Q_K(t)) \) be the joint QSI of the \( K \)-user OFDMA system, where \( Q_k(t) \) denotes the unfinished work (number of packets) in the \( k \)-th queue at the beginning of the \( t \)-th slot. Let \( \mathbf{R}(t) = (R_1(t), \ldots, R_K(t)) \) be the scheduled data rates (bit/second) of the \( K \) users where \( R_k(t) \) is given by \( \mathbf{I} \). At the beginning of the \( t \)-th scheduling slot, the cross-layer scheduler observes the CSI \( H(t) \) and the joint QSI \( \mathbf{Q}(t) \) and calculate the scheduled data rate \( R(t) \). We assume the scheduler at the transmitter is causal in the sense that new packet arrivals \( \mathbf{A}(t) \) at the \( t \)-th slot appears after the scheduler’s action. Hence, the queue dynamics is given by the following equation:

\[
Q_k(t+1) = \min\{(Q_k(t) - R_k(t)\tau/N_k(t))^+ + A_k(t), N_Q\} 
\]

(2)

The quasi-static assumption is realistic for pedestrian mobility users where the channel coherence time is around 50 ms but typical frame duration is less than 5ms in next generation wireless systems such as WiMAX. On the other hand, we assume the CSI is i.i.d. between slots (as in many other literature) in order to capture first order insights. Similar solution framework can be applied to deal with FSMC fading.
where $x^+ = \max\{x, 0\}$, $N_Q$ denotes the maximum buffer size (number of packets). Thus, the cardinality of the joint QSI is $I = (N_Q + 1)^K$ which grows exponentially with $K$.

For notation convenience, we denote $\chi(t) = (H(t), Q(t))$ to be the global system state at the $t$-th slot. Given an observed system state realization $\chi$, the transmitter may adjust the transmit power and subcarrier allocation according to a stationary power control and subcarrier allocation policy defined below.

**Definition 1: (Stationary Power Control and Subcarrier Allocation Policy)** A stationary transmit power and subcarrier allocation policy $\Omega = (\Omega_p, \Omega_s)$ is a mapping from the system state $\chi$ to the power and subcarrier allocation actions. A policy $\Omega$ is called feasible if the associated actions satisfy the average total transmit power constraint and the subcarrier assignment constraint. Specifically, $\Omega_p(\chi) = p$ and $\Omega_s(\chi) = s$, where $p = \{p_{k,n}\}$ are the power allocation actions satisfying

$$\sum_{k=1}^{K} \sum_{n=1}^{N_F} \mathbb{E}[p_{k,n}] \leq P_0, \quad p_{k,n} \geq 0$$

and $s = \{s_{k,n} \in \{0, 1\}\}$ are the subcarrier allocation actions satisfying

$$\sum_{k=1}^{K} s_{k,n} = 1 \quad \forall n \in \{1, N_F\}$$

Note that the $K$ queues are coupled together via the control policy $\Omega$ and the constraints in (3) and (4).

The average delay is related to the control actions (power and subcarrier allocation) via the packet service rates $R(t)$ and the queue dynamics in (1). The delay-optimal scheduler design can be formulated as the following optimization problem:

**Problem 1 (Delay-Optimal Policy):** For some constants $\beta = (\beta_1, \cdots, \beta_K)$ ($\beta_k > 0 \forall k$), we seek to find a stationary policy $\Omega$ that minimizes

$$J^\Omega_\beta = \sum_{k=1}^{K} \beta_k T_k(\Omega) + \gamma T_{tx}(\Omega)$$

$$= \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[ g(\chi(t), \Omega(\chi(t))) \right]$$

(7)

where $g(\chi, \{p, s\})$ is the per-stage reward given by:

$$g(\chi, \{p, s\}) = \sum_{k} \beta_k Q_k + \gamma \sum_{k,n} p_{k,n}$$

(8)

The positive weighting factors $\beta$ in (7) indicate the relative importance of buffer delay among the $K$ data streams and for each given $\beta$, the solution to (7) corresponds to a point on the Pareto optimal delay tradeoff boundary of a multi-objective optimization problem. The constant $\gamma > 0$ is the Lagrange multiplier for the average transmit power constraint in (6).

### III. MARKOV DECISION PROBLEM FORMULATION

In this section, we shall formulate the delay minimization problem in (7) as an infinite horizon Markov Decision Problem (MDP) and discuss the optimality condition.

#### A. MDP Formulation

A stationary control policy $\Omega$ induces a joint distribution for the random process $\{\chi(t)\}$. Since the system queue level $Q(t)$ evolves according to the system dynamics described in (2) and the arrival, departure and the CSI processes are memoryless, the random process $\chi(t)$ is a Markov chain and hence, the optimization problem in (7) can be modeled as a MDP with transition probability given by:

$$\Pr[\chi(t+1) | \chi(t), \Omega(\chi(t))] = \Pr[H(t+1) | \chi(t), \Omega(\chi(t))] \Pr[Q(t+1) | \chi(t), \Omega(\chi(t))]$$

(9)

As a result, the optimizing policy for the MDP in (7) can be obtained by solving the Bellman equation (13) (Page 308) recursively w.r.t. $(\theta, \{V(\cdot)\})$ as below:

$$\theta + V(\chi) = \min_{u(\chi)} \left[ g(\chi, u(\chi)) + \sum_{\chi'} \Pr[\chi' | \chi, u(\chi)] V(\chi') \right]$$

(10)

where $u(\chi) = \Omega(\chi) = \{p, s\}$ are the power control and subcarrier allocation actions taken in state $\chi$ and $g(\chi, u(\chi))$ given by (8) is the per-stage reward when the current state is $\chi$ and action $u(\chi)$ is taken. If there is a $\theta = J_\beta^\Omega$ satisfying (10), then $\theta = J_\beta^\Omega = \inf_{\Omega} J_\beta^\Omega$ is the optimal average reward per stage and the optimizing policy is given

- Given current queue state $Q(t)$, the departure $R(t)$ (which is determined by the control action $\Omega(\chi(t))$) and arrival $A(t)$, the future queue level $Q(t+1)$ is independent of the previous system states.
- Although the QSI $Q(t+1)$ and CSI $H(t+1)$ are correlated via the control action $\Omega(\chi(t))$, due to the causality of the control action, $H(t+1)$ is independent of $\chi(t)$ and $Q(t+1)$.
by $\Omega^\ast(\chi) = u^\ast(\chi)$ where $u^\ast(\chi)$ are the optimizing actions of $\text{II}$ at state $\chi$. Furthermore, since the induced Markov chain $\{\chi(i)\}$ is irreducible for any stationary policy considered, the solution to $\text{II}$ is unique up to one additive constant.

B. Reduced State Bellman Equation

Instead of working on the global state space $\chi = (H, Q)$, we shall derive a reduced-state Bellman equation from $\text{II}$ using partitioning of the control policy $\Omega$, which is based on partial system state $Q$ only. Specifically, we partition a unichain policy $\Omega$ into a collection of actions based on the QSI as follows:

Definition 2 (Conditional Actions): Given a control policy $\Omega$, we define $\Omega_i(Q) = \{(p, s) = \Omega(\chi) : \chi = (Q, H) \forall H\}$ as the collection of actions under a given QSI $Q$ for all possible CSI $H$. The policy $\Omega$ is therefore equal to the union of all the conditional actions. i.e. $\Omega = \bigcup_i \Omega_i(Q)$.

The following lemma summarizes the main result.

Lemma 1 (Equivalent Reduced-State Bellman Equation): The control policy obtained by solving the original Bellman equation in $\text{II}$ is equivalent to the control policy obtained by solving the following reduced state Bellman equation.

\[
\theta + \tilde{V}(Q^i) = \min_{u(Q^i)} \left[ \tilde{g}(Q^i, u(Q^i)) + \sum_{Q^j} \tilde{f}(Q^j|Q^i, u(Q^i)) \tilde{V}(Q^j) \right]
\]  

(11)

where $\tilde{V}(Q) = \mathbb{E}[V(\chi)|Q]$ is the conditional potential function, $u(Q) = \Omega(Q)$ is the collection of actions under a given QSI $Q$, $\tilde{g}(Q, u(Q)) = \mathbb{E}[g(\chi, u(\chi)|Q)]$ is the conditional per-stage reward, $\tilde{f}(Q^j|Q^i, u(Q^i)) = \mathbb{E}\left[ \Pr[Q^j|\chi^i, u(\chi^i)|Q^i] \right]$ is the conditional average transition kernel.

Proof: Please refer to Appendix A for the proof. □

A solution to (11) is still very complex due to the huge dimensionality of the state space ($I$ is exponential in $K$) and brute force value iteration or policy iteration [19] has exponential memory size requirement. As a result, it is desirable to obtain an online and low-complexity solution for the problem.

IV. GENERAL SOLUTION TO THE DELAY OPTIMAL PROBLEM

In this section, we shall derive a low complexity (but optimal) solution by proposing an online value iteration to solve the reduced state Bellman equation in (11). We shall also establish technical conditions for almost-sure convergence of the online value iteration.

A. Online value iteration via Stochastic Approximation

We shall propose an online sample-path-based iterative learning algorithm to estimate the performance potential and the control policy. Define a vector mapping $T : R^I \rightarrow R^I$ with the $i$-th component mapping ($1 \leq i \leq I$) as

\[
T_i(\tilde{V}) = \min_{u(Q^i)} \left\{ \tilde{g}(Q^i, u(Q^i)) + \sum_{Q^j} \tilde{f}(Q^j|Q^i, u(Q^i)) \tilde{V}(Q^j) \right\}
\]

(12)

Since the potential is unique up to a constant, we could set $T_i(\tilde{V}) - \tilde{V}(Q^i) = J_\beta$ for some reference state $Q^i$ ($1 \leq i \leq I$). Let $t = \{0, 1, 2, \ldots\}$ be the slot index and $\{Q(0), \ldots, Q(t), \ldots\}$ be the sample-path, i.e. the corresponding realizations of the system states. To perform the online value iteration, we divide the sample-path into regenerative periods. A regenerative period is defined as the minimum interval that each $Q$ state is visited at least once. Let $l_k(i)$ and $\tilde{V}_k$ be the times that $Q^i$ is visited and the estimated performance potential in the $k$-th regenerative period respectively. Let $n_0 = 0, n_{k+1} = \min\{t + 1 : t > n_k, \min_i l_k(i) = 1\}$ for $k \geq 0$. Then the sample path in the $k$-th regenerative period is $\{Q(n_k), \ldots, Q(n_{k+1} - 1)\}$. At the beginning of the $k$-th regenerative period ($n_k \leq t \leq n_{k+1} - 1$), initialize the following dummy variables as $S_{\tilde{g}_k}(i) = 0, S_{\tilde{V}_k}(i) = 0$ and $l_k(i) = 0$. Within the $k$-th regenerative period, we adopt policy $\Omega_k$. After observing the queue state $Q(t + 1)$ at the end of the $t$-th slot, update the following metric of the visited queue state $Q(t)$ according to

\[
\begin{align*}
S_{\tilde{g}_k}(i) &= S_{\tilde{g}_k}(i) + g(\chi(t), \Omega_k(\chi(t))) \\
S_{\tilde{V}_k}(i) &= S_{\tilde{V}_k}(i) + \tilde{V}_k(Q(t + 1)) & \text{if } Q^i = Q(t) \\
l_k(i) &= l_k(i) + 1 \\
\end{align*}
\]

At the end of the $k$-th regenerative period, using stochastic approximation algorithm [17], we update the estimated potential for the $(k + 1)$-th regenerative period, which is

\[
\tilde{V}_{k+1}(Q^i) = \tilde{V}_k(Q^i) + \epsilon_k Y_k(Q^i), \quad 1 \leq i \leq I
\]

(14)

where

\[
Y_k(Q^i) = \frac{S_{\tilde{g}_k}(i)}{l_k(i)} - \left( \frac{S_{\tilde{g}_k}(I)}{l_k(I)} - \tilde{V}_k(Q^I) \right) + \frac{S_{\tilde{V}_k}(i)}{l_k(i)} - \tilde{V}_k(Q^i)
\]

and $\epsilon_k$ is the step size of the stochastic approximation algorithm and $Q^I$ is the reference state. Accordingly, we update the policy for the $(k + 1)$-th regenerative period, which is given by

\[
\Omega_{k+1} = \arg \min T(\tilde{V}_{k+1})
\]

(15)

Therefore, we could construct an online value iteration algorithm as below.

Algorithm 1: Online Value Iteration

- Step 1 (Initialization): Set $t = 0$, and start the system at an initial state $Q(0)$. Set $k = 0$, initialize the potential $\tilde{V}_0$ and policy $\Omega_0 = \arg \min T(\tilde{V}_0)$ in the 0-th regenerative period.
- Step 2 (Online Potential Estimation): At the beginning of the $k$-th regenerative period, set $S_{\tilde{g}_k}(i) = 0, S_{\tilde{V}_k}(i) = 0$ and $l_k(i) = 0 \forall i$. Run the system with policy $\Omega_k$ to $n_{k+1} - 1$ and accumulate the information of the visited $Q$ from slot to slot according to (13). At $n_{k+1} - 1$, update the estimated potential $\tilde{V}_{k+1}$ for the $(k + 1)$-th regenerative period according to (14).
- Step 3 (Online Policy Improvement): Update the policy $\Omega_{k+1}$ for the $(k + 1)$-th regenerative period according to (15).
- Step 4 (Termination): If $||\tilde{V}_{k+1} - \tilde{V}_k|| < \delta_v$, stop; otherwise, set $k := k + 1$ and go to Step 2.

Fig. 2 illustrates the online value iteration algorithm with an example.

\[\text{Without loss of generality, we set the state that all } K \text{ buffers are empty as } Q^I \text{ and initialize } \tilde{V}_0(Q^I) = 0.\]
Ω is the total number of iterations within a CSI coherence time. The stepsize conditions satisfy the following conditions:

\[ \sum_{k} \epsilon_k = \infty, \quad \epsilon_k \geq 0, \quad \epsilon_k \to 0, \quad \sum_{k} \epsilon_k^2 < \infty \]  

Assume that for each set of control policies \( \Omega_0, \ldots, \Omega_m \), there exist a \( \delta_m = \mathcal{O}(\epsilon_m) > 0 \) and some positive integer \( m \) such that

\[ |A_m \cdots A_1|_{11} \geq \delta_m, \quad |B_m \cdots B_1|_{11} \geq \delta_m, \quad 1 \leq i \leq I \]  

where \([\cdot]_{11}\) denotes the element of the ith row and I-th column of the corresponding matrix, and \( e = [1, \ldots, 1]^T \) is the \( I \times 1 \) unit vector. For an arbitrary initial potential vector \( \hat{V}_0 \), we have \( \lim_{k \to \infty} \hat{V}_k = \hat{V}_\infty \) w.p.1, where the steady state potential \( \hat{V}_\infty \) satisfies:

\[ (T(\hat{V}_\infty) - \hat{V}_\infty(Q^I))e + \hat{V}_\infty = T(\hat{V}_\infty) \]  

Furthermore, the optimal reward of the delay optimal problem is \( J_\beta^* = T(\hat{V}_\infty) - \hat{V}_\infty(Q^I) \) and the optimal stationary policy is \( \Omega^* = \arg \min T(V) \).

**Proof:** Please refer to Appendix B.

**Remark 2:** Note that \( A_k \) and \( B_k \) are related to an equivalent transition matrix of the underlying Markov chain. It simply means that state \( Q^I \) is accessible from all the \( Q \) states after some finite number of transition steps. This is a very mild condition and will be satisfied in most of the cases we are interested.

**Remark 3:** Note that (18) is equivalent to the reduced state Bellman equation in (11). As a result, the converged solution of (18) corresponds to the solution of (11).
of two or more packets from the same queue or different queues and simultaneous arrival as well as departure in a slot are \( O((\lambda_\tau \tau)^2), \ O((\mu_\tau (\chi(t)) \tau)^2) \) and \( O((\lambda_\tau \tau \mu_\tau (\chi(t)) \tau) \) respectively, which are asymptotically negligible. Therefore, the vector queue dynamics are given by

\[
\tilde{f}(Q_1, \ldots, Q_k, \ldots, Q_{k+1}) = \lambda_\tau \tau, \ \forall k
\]

\[
\tilde{f}(Q_1^+, \ldots, Q_k^+, \ldots, Q_{k+1}^-) = \tilde{f}(Q_1^-, \ldots, Q_k^-, \ldots, Q_{k+1}^+) = \mu_\tau (\chi(t)) \tau, \ \forall k
\]

\[
\tilde{f}(Q_1^+, \ldots, Q_k, \ldots, Q_{k+1}^-) \approx 1 - \sum_k (\lambda_\tau \tau + \tilde{\pi}_k(Q_k^+)) \tau
\]

where \( x_{\tau\lambda N_Q} = \min(x, N_Q), Q_k^+ = (Q_1^+, \ldots, Q_k^+), \) and \( \tilde{\pi}_k(Q) = \mathbb{E}[\mu_\tau(\chi)|Q]. \) In what follows, we shall discuss the optimal solution and asymptotically optimal solution with only linear complexity and memory requirement of the OFDMA system with the conditional transition kernel given by \( Q. \)

A. Optimal Solution

Given \( Q, \) the optimization objective in the R.H.S. of Bellman equation in (11) becomes

\[
\mathbb{E} \left[ \gamma \sum_{k,n} p_{k,n}(X_i) - \sum_{k=1}^K \Delta_k V(Q_k^+) \frac{\tau}{N_k} \right.
\]

\[
\cdot \left( \sum_{n=1}^{N_F} s_{k,n} \log(1 + p_{k,n}|H_{k,n}|^2) \right) |Q|^2
\]

where \( \Delta_k V(Q_k^+) = \tilde{V}(Q_1^+, \ldots, Q_{k-1}^+, Q_k^+, \ldots, Q_{K}^+), \) Using standard optimization techniques, the closed-form solution during the policy improvement step \( (15) \) in the online-value iteration algorithm is summarized below.

**Lemma 2:** (Closed-Form Power Control and Subcarrier Allocation of Online Policy Improvement) Under the above setup, the optimizing power control and subcarrier allocation actions in the policy improvement step \( (15) \) for given CSI \( H \) and QSI \( Q \) are given by:

\[
p_{k,n}(H, Q) = s_{k,n}(H, Q) \left( \frac{\tilde{\pi}_k V(Q_k^+)}{\gamma} - \frac{1}{|H_{k,n}|^2} \right)^{+}
\]

\[
s_{k,n}(H, Q) = \begin{cases} 1, & \text{if } X_{k,n} = \max_{j} \{X_{j,n}\} > 0 \\ 0, & \text{otherwise} \end{cases}
\]

where \( \gamma \) satisfies \( (6) \) and \( X_{k,n} = \frac{\tilde{\pi}_k V(Q_k^+)}{\gamma} \log \left( 1 + |H_{k,n}|^2 \right) \left( \frac{\tilde{\pi}_k V(Q_k^+)}{\gamma} \right)^{+} - \frac{1}{|H_{k,n}|^2}. \)

**Proof:** Due to the combinatorial nature of the subcarrier allocation in the OFDMA system, i.e. \( s_{k,n}, \in \{0,1\} \) and \( (4), \) finding the optimal solution by brute force exhaustive search requires exponential complexity \( [20]. \) By applying continuous relaxation technique \( [20], [21], \) the original integer programming problem can be relaxed to a convex optimization problem and then the standard convex optimization techniques can be applied. It can be shown that the original problem and the relaxed problem share the same optimal solution in general. We omit the detailed proof here due to page limit.

**Remark 4:** (Structure of the Optimal Power Control and Subcarrier Allocation) The optimal power control and the subcarrier allocation solution in Lemma 2 are both functions of CSI and QSI where they depend on the QSI indirectly via the potential function \( \{V(Q)\}. \) For the power control solution, it has the form of multi-level water-filling where the power is allocated according to the CSI across subcarriers but the water-level is adaptive to the QSI. Similarly, the subcarrier is selected according to the metric \( X_{k,n} \) which depends on both the CSI and the QSI.

While the online value iteration is much simplified compared with the brute-force MDP solutions, it still suffers from exponentially large memory requirement and computational complexity for storing and computing the potential vector. In the next subsection, we shall exploit the birth-death dynamics and derive an asymptotically optimal solution with linear \( O(K N_F) \) computational complexity and \( O(K) \) memory requirement.

B. Asymptotically Optimal Solution

We first define a simplified subcarrier allocation policy below and summarize an important structural result of the Bellman equation in (11) under the simplified policy.

**Definition 3:** [CSI-Only Subcarrier Allocation Policy] A CSI-only subcarrier allocation policy is defined as \( \Omega_{s,H}(H) = \{s_{k,n}(H) \in \{0,1\}\} \sum_{k=1}^{K} s_{k,n} = 1 \ \forall n \} \).

**Theorem 2** (Additive Property of the Potential Function): Under the average power constraint in (3) and the CSI-only subcarrier allocation policy, the solution of the Bellman equation in (11) can be expressed into the form \( \tilde{V}(Q) = \sum_k \tilde{V}_k(Q_k) \) and \( \theta = \sum_k \theta_k, \) where \( \{\tilde{V}_k(Q_k), \theta_k\} \) is the solution of the \( k \)-th user’s reduced Bellman equation:

\[
\theta_k = \min_{u_k(Q_k)} \left\{ \tilde{\pi}_k(Q_k) + \lambda_\tau \tau \tilde{V}_k(Q_k + 1) \right\}
\]

where \( u_k(Q_k) = \{ (\tilde{p}(H, Q_k), \tilde{s}(H)) : (H, Q_k) \forall H \} \) is the set of the conditional control actions, \( \tilde{\pi}_k(Q_k) = \mathbb{E} \left[ \sum_{n=1}^{N_F} s_{k,n}(H, Q_k) \log(1 + \tilde{p}_{k,n}(H, Q_k)|H_{k,n}|^2)|Q_k|/N_k \right] \) is the conditional average service rate and \( \tilde{V}_k(Q_k) = \tilde{V}_k(Q_k) - \tilde{V}_k((Q_k - 1)^{+}) \) is the potential increment of the \( k \)-th queue.

**Proof:** Please refer to Appendix C for the proof.

As a result of Theorem 2 if we restrict the subcarrier allocation policy to the CSI-only subcarrier allocation policy, then the potential function \( \{\tilde{V}(Q)\} \) of the joint QSI for the original MDP problem can be decomposed into \( K \) individual potential functions \( \{\tilde{V}_k(Q_k)\} \) of the individual QSI for the \( K \) individual MDP’s. This could substantially simplify the potential estimation, the convergence speed as well as the memory requirement. In particular, to satisfy the condition of Theorem 2, we modify the optimal subcarrier allocation solution in Lemma 2 to a CSI-only policy which is given by

\[
\tilde{s}_{k,n}(H) = \begin{cases} 1, & \text{if } |H_{k,n}|^2 = \max_j \{ |H_{j,n}|^2 \} \\ 0, & \text{otherwise} \end{cases}
\]

While using \( (24) \) will result in some loss of optimality in strict sense, we shall show in the following Corollary that the solution is indeed asymptotically optimal.
Algorithm 2: Reduced Complexity Online Value Iteration

- Step 1 (Initialization): For each user $k = 1 : K$, initialize the potential $V_k^0$ and policy $Q_k^0 = \arg\min T(V_k^0)$ in the 0-th regenerative period. Set $t = 0$, and start the OFDMA at an initial QSI state $Q_k(0)$ for each user.
- Step 2 (Online Potential Estimation): For each user $k = 1 : K$ in its $l_k$-th regenerative period, set $S_{\hat{g}_k}(Q_k) = 0$, $S_{\hat{\psi}_k}(Q_k) = 0$ and $j_k^k(Q_k) = 0 \forall Q_k$. Perform power control and subcarrier allocation according to the policy $\hat{Q}_k^{l_k}$, and accumulate the information of the visited $Q_k$ for each user $k$ if $Q_k = Q_k(t)$ according to

$$
S_{\hat{g}_k}(Q_k) = S_{\hat{g}_k}(Q_k) + g_k(\chi(t), \Omega_k^t(\chi(t)))
$$

$$
S_{\hat{\psi}_k}(Q_k) = S_{\hat{\psi}_k}(Q_k) + V_k^t(\chi(t) + 1)
$$

$$
j_k^k(Q_k) = j_k^k(Q_k) + 1
$$

If the current slot corresponds to the end of any of the $K$ regenerative periods, then update the policy for its next regenerative period at the BS according to (24) and

$$
\hat{V}_k^{l_k+1}(Q_k) = \hat{V}_k^{l_k}(Q_k) + \epsilon_k \left( S_{\hat{g}_k}(Q_k) + S_{\hat{\psi}_k}(Q_k) - \left( \frac{S_{\hat{g}_k}(Q_k)}{l_k(Q_k)} + \frac{S_{\hat{\psi}_k}(Q_k)}{l_k(Q_k)} - \hat{V}_k^{l_k}(Q_k) - \hat{V}_k^{l_k}(Q_k) \right) \right)
$$

- Step 3 (Online Policy Improvement): For each user $k = 1 : K$, if the current slot corresponds to the end of any of the $K$ regenerative periods, then update the policy for its next regenerative period at the BS according to (25) and

$$
\hat{p}_{k,n}(H, Q_k) = \hat{p}_{k,n}(H) \left( \frac{\tau \hat{V}_k^{l_k+1}(Q_k)}{\gamma} - \frac{1}{|H_{k,n}|^2} \right) + \delta_k \forall k,
$$

- Step 4 (Termination): If $||\hat{V}_k^{l_k+1} - \hat{V}_k^{l_k}|| < \delta_k \forall k$, stop; otherwise, set $l_k := l_k + 1$ and go to Step 2.

Proof: Please refer to Appendix D for the proof.
of the Baseline 1 (M-LWDF), Baseline 2 (RT-SPD) and Baseline 3 (Round Robin) are 4%, 4%, 6% respectively.

![Figure 5](image)

Fig. 5. Average delay per queue versus the number of users. The buffer size $N_Q = 10$, the mean packet size $N_p = 39$ Kbyte/pck, the average arrival rate $\lambda_k = 20$ pck/s, the queue weight $\beta_k = 1$ at a transmit SNR = 10dB. The packet drop rate of the proposed scheme is 4%, while the packet drop rate of the Baseline 1 (M-LWDF), Baseline 2 (RT-SPD) and Baseline 3 (Round Robin) are 1%, 4%, 6% respectively.

![Figure 6](image)

Fig. 6. Illustration of convergence property. Potential function versus the number of iterations. The number of users $K = 8$, the buffer size $N_Q = 10$, the mean packet size $N_p = 39$ Kbyte/pck, the average arrival rate $\lambda_k = 20$ pck/s, the queue weight $\beta_k = 1$ at a transmit SNR = 10dB. The packet drop rate of the proposed scheme is 1%, while the packet drop rate of the Baseline 1 (M-LWDF), Baseline 2 (RT-SPD) and Baseline 3 (Round Robin) are 1%, 1%, 4% respectively.

even in 2 user-case. Fig. 4 depicts the average weighted delay versus SNR of 2 heterogeneous users with different queue weight. The average weighted delay of the reduced complexity solution is close to that of the optimal solution as well. Therefore, the proposed reduced complexity solution with linear $O(KN_F)$ memory requirement and computational complexity as well as near optimal performance is of great practical significance.

Fig. 5 illustrates the average delay per queue of the reduced complexity solution versus the number of users with equal queue weight at a transmit SNR = 10dB. It is obvious that the reduced complexity solution has great gain in delay over the three baselines in the whole user region.

Fig. 6 illustrates the convergence property of the proposed reduced complexity algorithm. We plot the average potential function of 8 users versus the number of iterations at a transmit SNR = 10dB. It can be seen that the reduced complexity algorithm converges quite fast. The average delay corresponding to the potential function at the 50-th iteration is 3.8 packets, which is much smaller than the other baselines.

VII. SUMMARY

In this paper, we propose a low-complexity solution to the delay-optimal power and subcarrier allocation design for OFDMA systems. We model the problem as a $K$-dimensional infinite horizon average reward MDP with the control actions based on CSI and joint QSI. We derive the equivalent reduced state Bellman equation and propose an online stochastic value iteration solution using stochastic approximation. We prove that under some mild condition, the proposed solution converges to the optimal solution almost surely (with probability 1). By exploiting the birth-death structure of the queue dynamics in the Poisson arrivals, we obtain a reduced complexity decomposed solution with linear $O(KN_F)$ complexity and $O(K)$ memory requirement.

APPENDIX

APPENDIX A: PROOF OF LEMMA 1

$$\theta + V(H, Q^i)$$

$$\forall H, 1 \leq i \leq I$$

$$= \min_{u(H, Q^i)} \left[ g((H, Q^i), u(H, Q^i)) + \sum_{H'} \Pr[H'|H, Q^i]V(H', Q^i) \right]$$

$$= \min_{u(H, Q^i)} \left[ g((H, Q^i), u(H, Q^i)) + \sum_{Q'} \Pr[Q'|H, Q^i]V(H', Q^i) \right]$$

$$\Rightarrow \theta + \bar{V}(Q^i)$$

$$\forall 1 \leq i \leq I$$

$$= \mathbb{E} \left[ \min_{u(H, Q^i)} \left[ g((H, Q^i), u(H, Q^i)) + \sum_{H'} \Pr[H'|H, Q^i]V(H', Q^i) \right] \right]$$

$$= \min_{u(H, Q^i)} \left[ g(Q^i, u(Q^i)) + \sum_{Q'} \Pr[Q'|Q^i]V(Q^i) \right]$$

$$= \min_{u(Q^i)} \left[ g(Q^i, u(Q^i)) + \sum_{Q'} \bar{f}(Q^i, Q') + \min_{u(Q^i)} \bar{f}(Q^i, u(Q^i)) \right]$$

where (a) is due to (9), (b) is due to the definition $\bar{V}(Q^i) \triangleq \mathbb{E}[V(Q^i) | Q^i]$, (c) is obtained by taking the conditional expectation (conditioned on $Q^i$) on both sides of (10) and (d) is due to the definition of “conditional actions”.

12 The mild conditions refer to the two conditions in Theorem 1. One is the stepsize condition. The other is the condition on accessibility of the Markov Chain, which can be easily satisfied in most of the cases we are interested.
and \( \tilde{g}(Q, u(Q)) = \mathbb{E}[g(x, u(x))|Q], \tilde{f}(Q^j|Q', u(Q')) = \mathbb{E}\left[\Pr[Q^j|x', u(x')]|Q\right] \).

### APPENDIX B: PROOF OF THEOREM 1

We shall first show the convergence of the martingale noise in (14). Let

\[
q_k(Q^i) = \mathbb{E}_k[Y_k(Q^i)] = \frac{Pr(T_i(V_k(Q^i)) - V_k(Q^i) - T_i(V_k(Q^i)) - V_k(Q^i))}{(27)}
\]

\(Y_k(Q^i)\) is the noise-corrupt observation of \(q_k(Q^i)\). The martingale difference noise is

\[
\delta M_k(Q^i) = Y_k(Q^i) - q_k(Q^i)
\]

\[
= \left(\frac{S_{\tilde{g}}(i)}{l_k(i)} - \tilde{g}(Q^i)\right) - \left(\frac{\tilde{S}_{\tilde{g}}(I)}{l(I)} - \tilde{g}(Q^i)\right)
\]

\[+ \left(\frac{S_{\tilde{g}}(i)}{l_k(i+1)} - \tilde{g}(Q^i)\right) - \left(\frac{\tilde{S}_{\tilde{g}}(I)}{l_k(i+1)} - \tilde{g}(Q^i)\right)
\]

\[+ \left(\frac{S_{\tilde{g}}(i)}{l_k(i+1)} - \tilde{g}(Q^i)\right) - \left(\frac{\tilde{S}_{\tilde{g}}(I)}{l_k(i+1)} - \tilde{g}(Q^i)\right)
\]

with property that \(\mathbb{E}_k[\delta M_k(Q^i)] = 0\) and \(\mathbb{E}[\delta M_k(Q^i)|\delta M_k(Q^j)] = 0, \forall k \neq k'\). For some \(j\), define \(M_k(Q^i) = \sum_{i=0}^j \epsilon_i \delta M_k(Q^i)\). Thus, from (14), we have

\[
\tilde{V}_{k+1}(Q^i) = \tilde{V}_k(Q^i) + \epsilon_k [q_k(Q^i) + \delta M_k(\tilde{V}_k(Q^i))]
\]

\[
= \tilde{V}_j(Q^i) + \sum_{i=0}^k \epsilon_i q_i(Q^i) + M_k(Q^i)
\]

(28)

Since \(\mathbb{E}_k[M_k(Q^i)] = M_{k-1}(Q^i), M_k(Q^i)\) is a martingale sequence. By martingale inequality, we have \(\mathbb{P}(\sup_{j\leq k} |M_j| \geq \lambda) \leq \mathbb{E}[|M_k(Q^i)|^2] / \lambda^2\). By the boundness assumption and the property of the martingale difference noise as well as the condition on the stepsize sequence, we have \(\mathbb{E}[|M_k(Q^i)|^2] = \mathbb{E}_k[\sum_{j=0}^k \epsilon_i \delta M_k(Q^i)] = \mathbb{E}_k[\sum_{j=0}^k \epsilon_i \delta M_k(Q^i)] = \lim_{k \to \infty} \mathbb{P}(\sup_{j\leq k} |M_j| \geq \lambda) = 0\). Thus, as \(\epsilon_k \to 0\), \(\tilde{V}_{k+1}(Q^i) = \tilde{V}_j(Q^i) + \sum_{i=0}^k \epsilon_i q_i(Q^i)\) with probability 1, the vector form of which is given by

\[
\tilde{V}_{k+1} = \tilde{V}_j + \sum_{i=0}^k \epsilon_i q_i(Q^i)
\]

(29)

Next, we shall show the convergence of (29) after the martingale noise are averaged out. In the following proof, we use \(i\) instead of \(Q^i\) for simplicity. Let \(\Omega_k(i)\) denote the optimal control action attaining the minimum in \(T_i(\tilde{V}_k)\). Let \(g\) and \(P\) denote the conditional per-stage reward vector and conditional average transition probability matrix under the optimal control policy \(\Omega_k\). Denote \(w_k = T_i(\tilde{V}_k) - \tilde{V}_i(I)\). We have

\[
q_k = g_{\Omega_k} + P_{\Omega_k} \tilde{V}_k - \tilde{V}_k - w_k\epsilon
\]

\[\leq g_{\Omega_k} + P_{\Omega_k} \tilde{V}_k - \tilde{V}_k - w_k\epsilon
\]

\[
q_{k-1} = g_{\Omega_{k-1}} + P_{\Omega_{k-1}} \tilde{V}_{k-1} - \tilde{V}_{k-1} - w_{k-1}\epsilon
\]

\[\leq g_{\Omega_k} + P_{\Omega_k} \tilde{V}_{k-1} - \tilde{V}_{k-1} - w_{k-1}\epsilon
\]

By iteration

\[
\leq A_{k-1} \cdots A_{k-m} q_{k-m} - (w_k - w_{k-m})\epsilon \leq q_k
\]

\[\leq B_{k-1} \cdots B_{k-m} q_{k-m} - (w_k - w_{k-m})\epsilon\]

From (27), we have \(q_k(I) = T_i(\tilde{V}_k(I)) - \tilde{V}_k(I) - (T_i(\tilde{V}_k(I)) - \tilde{V}_k(I)) = 0\) for all \(k\). By the assumption (17), we have

\[
(1 - \delta) \min_{i'} q_{k-m(i') - (w_k - w_{k-m})} \leq q_k(i)
\]

\[\leq (1 - \delta) \max_{i'} q_{k-m(i')} - (w_k - w_{k-m})\epsilon i
\]

\[
\leq \max_{i'} q_{k(i')} \leq \min_{i'} q_{k-m(i')} - (w_k - w_{k-m})\epsilon i
\]

\[\leq \min_{i'} q_{k-m(i')} - \min_{i'} q_{k-m(i')}
\]

\[\Rightarrow \max_{i'} q_{k(i')} = (1 - \delta) \max_{i'} q_{k-m(i')} - \min_{i'} q_{k-m(i')}
\]

(30)

where \(\phi_i > 0\). Since \(q_k(I) = 0\), we have \(\max_{i'} q_{k(i')} \geq 0\) and \(\min_{i'} q_{k(i')} \leq 0\). Thus, \(\forall k\), we have \(|q_k(i)| \leq \max_{i'} q_{k(i')} - \min_{i'} q_k(i') \leq \phi_k \lim_{l \to \infty} (1 - \delta_i + l m)\).

Therefore, as \(k \to \infty\), \(q_k \to 0\), i.e. \(\tilde{V}_k\) satisfies Bellman equation (13). By the Proposition 1 in Chapter 7 of [13], \(J_{i'} = \tilde{T}_i(\tilde{V}_k(I)) - \tilde{V}_k(I)\) is the optimal value and \(\tilde{V}_k\) is the potential vector, which is up to an constant vector. However, due to the property that \(q_k(I) = 0 \Rightarrow \tilde{V}_k(I) = \tilde{V}_k(I) = \tilde{V}_k(I)\forall k\), we have the convergence of the potential vector \(\tilde{V}_k = \lim_{k \to \infty} \tilde{V}_k\) and the optimal value \(J_{i'} = T_i(\tilde{V}_k(I)) - \tilde{V}_k(I)\) via the online value iteration via stochastic approximation algorithm. Accordingly, the optimal stationary policy is \(\Omega^* = \arg \min T(V^*_k)\).

### APPENDIX C: PROOF OF THEOREM 2

Solution of Bellman equation in (11) can be obtained by offline relative value iteration [19]. Without loss of generality, we set \(Q^i\) as the reference state. Hence, we have normalizing equation \(\tilde{V}_i(I) = 0 \forall i\). Assume \(\tilde{V}_i(Q^i) = \sum_{k=1}^K \tilde{V}_i(Q_k) \forall i\).

At the \((l-1)\)th iteration, updating policy according to (13) is equal to finding policy \(\Omega_{k'}\) which minimize the objective function (20). Under any given CSI-only subcarrier allocation policy, the optimal power actions for the \(l\)th iteration which
Bellman equation (23), the original MDP can be decoupled.

In addition, set $\theta^{k}(Q_k) = \{u_k^l(H, Q_k) | (H, Q_k) \in \mathcal{H}\}$ with $u_k^l(H, Q_k) = (\beta^l(H, Q_k), s^l(H), \Delta V^k_l(Q_k) = V^k_l(Q_k) - V^k_l((Q_k-1)^+) \}$ is the potential increment for the $k$-th queue.

At the $l$-th iteration, we determine the potential $\tilde{V}^k_l(Q)$ and $\theta^k_l$ by solving the normalizing equation $\tilde{V}^k_l(Q^l) = 0$ together with $I = (N_Q + 1)K$ Poisson equations

$$\theta^k_l + \tilde{V}^k_l(Q^l) = \sum_{k} \bar{g}_k(Q_k, u_k^l(Q_k)) + \tilde{V}^k_l(Q^l)$$

$$+ \sum_{k} \lambda_k \tau \Delta \tilde{V}^k_l(Q^l) + 1 - \sum_{k} \bar{p}_k^l(Q_k) \tau \Delta \tilde{V}^k_l(Q_k)$$

$$\Rightarrow \theta^k_l = \sum_{k} \theta^k_{k,l}, \forall 1 \leq i \leq I$$

where

$$\theta^k_{k,l} = \begin{cases} 0 & 0 \leq Q_k \leq N_Q \\ \bar{g}_k(Q_k, u_k^l(Q_k)) + \lambda_k \tau \Delta \bar{V}^k_l(Q^l) + 1 - \bar{p}_k^l(Q_k) \tau \Delta \bar{V}^k_l(Q_k) & \end{cases}$$

$$\bar{g}_k(Q_k, u_k^l(Q_k)) = \beta^l_k Q_k + \gamma \sum_{k} \bar{p}_k^l(Q_k) + \bar{H}_k^l(Q_k)$$

There are $I$ joint $Q = (Q_1, \cdots, Q_K)$ states, but there are only $N_Q + 1$ states for $Q_k \forall k$. Hence, in the original $N_Q + 1$ Poisson equations (30), there are only $N_Q + 1$ independent Poisson equations in (31) for the $k$-th $(1 \leq k \leq K)$ queue. In addition, set $\bar{V}^k_l(0) = 0 \forall k$ as the individual normalizing equation, which also satisfies $\bar{V}^k_l(Q^l) = \sum_{k} \bar{V}^k_l(0) = 0$.

Hence, in the $l$-th iteration, we can obtain $\{\bar{V}^k_l(Q_k), \theta^k_{k,l}\}$ by solving the $k$-th user’s reduced state Poisson equation in (31) together with its individual normalizing equation. Accordingly, $\{\bar{V}^k_l(Q), \theta^k_l\}$ is the solution of the original $(N_Q + 1)K$ Poisson equations (30), where $\bar{V}^k_l(Q)$ is the solution of the original $(N_Q + 1)K$ Bellman equations in (11), where $\bar{V}^k_l(Q) = \sum_{k} \bar{V}^k_l(Q_k)$ and $\theta = \sum_{k} \theta_k$, which are the potential for the joint $Q$ and the optimal average reward respectively.

APPENDIX D: PROOF OF COROLLARY 1

Since $\{\bar{V}^k_l(Q_k), \theta_k\}$ is the solution of the $k$-th user’s reduced Bellman equation (23), the original MDP can be decoupled into $K$ individual MDP’s with transition kernel is given by $f_k(Q_k + 1|Q_k, u_k(Q_k)) = \lambda_k \tau, f_k(Q_k - 1|Q_k, u_k(Q_k)) = \bar{p}_k(Q_k) \tau$ and $f_k(Q_k, u_k(Q_k)) = 1 - \lambda_k \tau - \bar{p}_k(Q_k) \tau$. Therefore, the online value iteration can be applied to $K$ individual MDP’s respectively. Under the same condition of Theorem (1), almost sure convergence of the online value iteration algorithm in Corollary 1 is guaranteed.

Next, we have to show that the converged solution is indeed asymptotically optimal. Denote $k^*_n \triangleq \arg \max_k |H_{k,n}|^2$. For large $K$, $|H_{k^*_n}|^2$ grows with $\log(K)$ by extreme value theory. Because the traffic loading remains unchanged as we scale up $K$, $|H_{k^*_n}|^2$ grows like $\log(\log(K))$. As $K \to \infty, \Pr[k^*_n = \arg \max_k H_{k,n} | x_k = 1] = 1$. Thus the subband allocation result of optimal subband allocation in (22) and the best CSI subband allocation in (24) will be the same for large $K$. Thus, (23) and (24) are asymptotically optimal. Therefore, with (24), following the proof of Theorem 2, we can prove $\sum_k \bar{V}^k(Q_k) \to V^*(Q)$, $\sum_k \theta_k \to \theta^*$ as $K \to \infty$, where $V^*(Q)$ and $\theta^*$ are the potential and optimal average reward under the global optimal power and subcarrier allocation given by Lemma 2.

REFERENCES

[1] K. Seong, M. Mohseni, and J. Cioffi, “Optimal resource allocation for OFDMA downlink systems,” in IEEE Int. Symp. Inform. Theory (ISIT), July 2006, pp. 1394 – 1398.

[2] C. Y. Wong, R.S. Cheng, K.B. Lataief, and R.D. Murch, “Multiuser OFDM with adaptive subcarrier, bit, and power allocation,” IEEE J. Select. Areas Commun., vol. 17, no. 10, pp. 1747 – 1758, Oct. 1999.

[3] D. Wu and R. Negi, “Effective Capacity: A Wireless Link Model for Support of Quality of Service,” IEEE Trans. Wireless Commun., vol. 2, pp. 630–643, July 2003.

[4] D. Hui and V. Lau, “Cross-Layer Design for OFDMA Wireless Systems With Heterogeneous Delay Requirements,” IEEE Trans. Wireless Commun., vol. 6, pp. 2872–2880, Aug. 2007.

[5] J. Tang and X. Zhang, “Quality-of-Service Driven Power and Rate Adaptation over Wireless Links,” IEEE Trans. Wireless Commun., vol. 6, pp. 3059–3068, Aug. 2007.

[6] E. M. Yeh, “Multiaccess and Fading in Communication Networks,” Ph.D. dissertation, MIT, September 2001.

[7] E. M. Yeh and A. Cohen, “Throughput and delay optimal resource allocation in multiaccess fading channels,” in Proc. ISIT, June-July 2003, p. 245.

[8] L. Georgiadis, M. J. Neely, and L. Tassiulas, “Resource allocation and cross-layer control in wireless networks,” Foundations and Trends in Networking, vol. 1, no. 1, pp. 1–144, 2006.

[9] M. J. Neely, “Order optimal delay for opportunistic scheduling in multi-user wireless uplinks and downlinks,” IEEE/ACM Trans. Networking, vol. 16, no. 5, pp. 1188–1199, 2008.

[10] M. Luo and A. Ephremides, “Stability of n interacting queues in random-access systems,” IEEE Trans. Inform. Theory, vol. 45, no. 5, pp. 1579 – 1587, July 1999.

[11] A. Stolyar, “Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm,” Queueing Systems, vol. 50, no. 4, pp. 401–457, 2005.

[12] X. Wang, G. B. Giannakis, and A. G. Marques, “A unified approach to qos-guaranteed scheduling for channel-adaptive wireless networks,” Proceedings of the IEEE, vol. 95, no. 12, pp. 2410–2431, 2007.

[13] D. P. Bertsekas, “Dynamic programming - deterministic and stochastic models,” Prentice Hall, New Jersey, NJ, USA, 1987.

[14] X. R. Cao, Stochastic Learning and Optimization: A Sensitivity-Based Approach, 1st ed. New York: Springer, 2007.

[15] D. P. Palomar and M. Chiang, “A tutorial on decomposition methods for network utility maximization,” IEEE J. Select. Areas Commun., vol. 24, no. 8, pp. 1439 – 1451, Aug. 2006.

[16] L. Ljung, G. Pfug, and H. Walk, Stochastic approximation and optimization of random systems, 1st ed. Berlin: Birkhasuer Verlag Basel, 1992.

[17] N. J. Kushner and G. G. Yin, Stochastic approximation and optimization of random systems, 2nd ed. New York: Springer, 2003.

[18] L. Kleinrock. Queueing Systems. Volume 1: Theory, 1st ed. New York: Wiley Interscience, 1975, ch. 2.

13It can be easily verify that all states in the birth-death chain are accessible.
Vincent K. N. Lau obtained B.Eng (Distinction 1st Hons) from the University of Hong Kong in 1992 and Ph.D. from Cambridge University in 1997. He was with PCCW as system engineer from 1992-1995 and Bell Labs - Lucent Technologies as member of technical staff from 1997-2003. He then joined the Department of ECE, HKUST as Associate Professor. His current research interests include the robust and delay-sensitive cross-layer scheduling, cooperative and cognitive communications as well as stochastic approximation and Markov Decision Process.

Ying Cui received B.Eng degree (first class honor) in Electronic and Information Engineering, Xian Jiaotong University, Xi’an, China in 2007. She is currently a Ph.D candidate in the Department of Electronic and Computer Engineering, the Hong Kong University of Science and Technology (HKUST). Her current research interests include cooperative and cognitive communications, delay-sensitive cross-layer scheduling as well as stochastic approximation and Markov Decision Process.