Forty-Eighth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 29 - October 1, 2010

Adaptive Distributed Algorithms for Optimal Random Access Channels

Yichuan Hu and Alejandro Ribeiro
Department of Electrical and Systems Engineering, University of Pennsylvania
200 South 33rd Street, Philadelphia, PA 19104
Email: {yichuan, aribeiro}@seas.upenn.edu

***Abstract*—We develop adaptive scheduling and power control algorithms for random access in a multiple access channel where terminals acquire instantaneous channel state information but do not know the probability distribution of the channel. In each time slot, terminals measure the channel to the common access point. Based on the observed channel value, they determine whether to transmit or not and, if they decide to do so, adjust their transmitted power. We remark that there is no coordination between terminals and that adaptation is to the local channel value only. It is shown that the proposed algorithm almost surely maximizes a proportional fair utility while adhering to instantaneous and average power constraints. Important properties of the algorithm are adaptivity, low computational complexity and the ability to handle non-convex rate functions. Numerical results on a randomly generated network with heterogeneous users corroborate theoretical results.** [1]

I. Introduction

This paper considers wireless random access channels in which terminals contend for access to a common access point (AP) as is the case in wireless local area networks and cellular systems. To exploit favorable channel conditions terminals adapt their transmitted power and access decisions to the state of the random fading channels linking them to the AP. The challenges in developing this adaptive scheme are that terminals have access to their own channel state information (CSI) only, and that the probability distribution function (pdf) of the fading channel is unknown. The goal of this paper is to develop a distributed learning algorithm to determine optimal transmitted power and channel access decisions relying on local CSI only.

The idea of adapting medium access and power control to CSI was first explored in [1] where terminals are allowed to transmit when channels exceed a predefined threshold. These threshold-based rules were later proved optimal in [2] and further extended to networks with different packet reception models, e.g., [3]–[9]. To compute optimal thresholds, however, terminals are assumed to know the pdf of their fading channels. This is a restrictive assumption because the channel fading distribution is usually unknown and can only be estimated based on channel observations. Overcoming this limitation motivates the development of adaptive algorithms to learn optimal operating points based on local CSI [10], [11]. The work in [10] proposes a heuristic adaptive algorithm for threshold-based schedulers in which the thresholds are tuned based on local channel realizations in a time window. The work in [11] develops an online learning algorithm for transmission probability and power control under rate constraints using game-theoretic approaches. However, neither [10] nor [11] guarantees global optimality.

The contribution of this paper is to develop an optimal distributed adaptive algorithm for scheduling and power control. At each time slot terminals observe their channel states, based on which they decide whether to transmit or not and select a power for their communication attempt. As time progresses, power budgets are satisfied almost surely, while the network almost surely maximizes a weighted proportional fair utility.

The rest of the paper is organized as follows. System model and problem formulation are presented in Section II. Section III develops optimal distributed adaptive algorithms whose almost sure feasibility and optimality are proved. Numerical results are provided in Section IV and concluding remarks in Section V.

II. System Model and Problem Formulation

Consider a multiple access channel with n terminals contending to communicate with a common AP. Time is divided in slots identified by an index t. We assume a backlogged system, i.e., all terminals always have packets to transmit in each time slot. The time-varying nonnegative channel $h_i(t) \in \mathbf{R}^+$ between terminal i and the AP at time t is modeled as block fading. Channel gains $h_i(t)$ and $h_j(t)$ of different terminals $i \neq j$ are assumed independent, while channels $h_i(t_1)$ and $h_i(t_2)$ of terminal i at different time slots $t_1 \neq t_2$ are assumed independent and identically distributed. Channels $h_i(t)$ take continuous values with no channel realization having strictly positive measure. This latter assumption holds true for models used in practice, e.g., Rayleigh, Rician and Nakagami. We assume each terminal i has access to its channel gain $h_i(t)$ at each time slot t. There are various ways for terminals to obtain channel conditions. For example, by having the AP send a beacon signal at the beginning of each time slot, terminals can estimate their channel gains.

Based on its channel state $h_i(t)$, terminal i decides whether to transmit or not in time slot t by determining the value of a scheduling function $q_i(t) := Q_i(h_i(t)) : \mathbf{R}^+ \to \{0,1\}$. Node i transmits in time slot t if $q_i(t) = 1$ and remains silent if $q_i(t) = 0$. Besides channel access decisions, terminals also adapt transmission power to their channel gains through a power control function $p_i(t) := P_i(h_i(t)) : \mathbf{R}^+ \to [0, p_i^{\text{inst}}]$, where $p_i^{\text{inst}} \in \mathbf{R}^+$ represents instantaneous power constraints. If node i transmits in time slot t, $p_i(t)$ and $h_i(t)$ jointly determine the transmission rate through a function $C_i(h_i(t)p_i(t)) : \mathbf{R}^+ \to \mathbf{R}^+$. The exact form of $C_i(h_i(t)p_i(t))$ depends on how the signal is modulated and coded at the physical layer. To keep the analysis general we do not restrict $C_i(h_i(t)p_i(t))$ to take any specific form. It

[1]This work was supported by ARO P-57920-NS and NSF CAREER CCF-0952867.

is only assumed that $C_i(h_i(t)p_i(t))$ is a nonnegative increasing function of the product of $h_i(t)$ and $p_i(t)$ that takes finite values for finite arguments. This assumption is lax enough to allow for discontinuous rate functions that arise in, e.g., systems that use adaptive modulation and coding (AMC); see Section IV

Since terminals contend for channel access, transmission of terminal i in a time slot t is successful if and only if $q_i(t) = 1$ and $q_j(t) = 0$ for all $j \neq i$. As as consequence, the instantaneous transmission rate for terminal i in time slot t is

$$r_i(t) = C_i\left(h_i(t)p_i(t)\right) q_i(t) \prod_{j=1,j\neq i}^{n} \left[1 - q_j(t)\right]. \qquad (1)$$

Assuming ergodic operation, quality of service is determined by the long term behavior of $r_i(t)$, implying that system performance is determined by the ergodic limits $r_i := \lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t} r_i(u)$ that we can write as [cf. (1)]

$$r_i = \lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t}\left[C_i\left(h_i(u)p_i(u)\right) q_i(u) \prod_{j=1,j\neq i}^{n}\left[1 - q_j(u)\right]\right]. \quad (2)$$

Assuming ergodicity of schedules $q_i(t) = Q_i(h_i(t))$ and power allocations $p_i(t) = P_i(h_i(t))$, the limit r_i can be written as a expected value over channel realizations,

$$r_i = \mathbb{E}_{\mathbf{h}}\left[C_i(h_iP_i(h_i))Q_i(h_i) \prod_{j=1,j\neq i}^{n}\left[1 - Q_j(h_j)\right]\right], \qquad (3)$$

where we have defined the vector $\mathbf{h} = [h_1, \cdots, h_n]^T$ grouping all channels h_i. Since terminals are required to make channel access and power control decisions independently of each other, $Q_i(h_i)$ and $P_i(h_i)$ are independent of $Q_j(h_j)$ and $P_j(h_j)$ for all $i \neq j$. This allows us to rewrite r_i as

$$r_i = \mathbb{E}_{h_i}\left[C_i(h_iP_i(h_i))Q_i(h_i)\right] \prod_{j=1,j\neq i}^{n}\Big[1 - \mathbb{E}_{h_j}[Q_j(h_j)]\Big]. \quad (4)$$

In order to allow terminals to know if their transmissions are successful or not, we assume a $(1, 0)$ feedback is available in each time slot, where 1 and 0 represent successful transmission and collision, respectively. If a terminal transmits a packet but detects a collision, then it can schedule a retransmission in the next time slot. Since we assume a backlogged system, feedback only tells terminals if they should retransmit previous packets or not, it does not enforce terminals to make channel access and power allocation decisions based on feedback information. Therefore, the assumption that different users operate independently of each other still holds.

In addition to instantaneous power constraints $p_i(t) \leq p_i^{\text{inst}}$, terminals adhere to average power constraints $p_i^{\text{avg}} \in \mathbf{R}^+$. This average power constraint restricts the long term average of transmitted power p_i defined by

$$p_i := \lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t} q_i(u)p_i(u) = \mathbb{E}_{h_i}[Q_i(h_i)P_i(h_i)]. \qquad (5)$$

With rates r_i given as in (4), the objective is to maximize a weighted proportional fair (WPF) utility defined as

$$U(\mathbf{r}) = \sum_{i=1}^{n} w_i \log(r_i), \qquad (6)$$

where $\mathbf{r} = [r_1, \cdots, r_n]^T$ is the vector of rates and $w_i \in \mathbf{R}^+$ is a weight coefficient for terminal i. Setting $w_i = w_j$ for all $i \neq j$ in a homogenous system with all channels having the same pdf, the WPF utility is equivalent to maximizing the sum of throughputs. In a heterogeneous network where channel pdfs vary among users, maximizing $U(\mathbf{r})$ yields solutions that are fair since it prevents users from having very low transmission rates. Grouping the objective in (6) with the constraints in (4) and (5), optimal adaptive random access is formulated as the following optimization problem

$$\begin{aligned}
\mathsf{P} = \max\ & U(\mathbf{r}) \\
\text{s.t. } & r_i = \mathbb{E}_{h_i}\left[C_i(h_iP_i(h_i))Q_i(h_i)\right] \prod_{j=1,j\neq i}^{n}\Big[1 - \mathbb{E}_{h_j}[Q_j(h_j)]\Big], \\
& \mathbb{E}_{h_i}\left[Q_i(h_i)P_i(h_i)\right] \leq p_i^{\text{avg}}, \\
& Q_i(h_i) \in \mathcal{Q}, P_i(h_i) \in \mathcal{P}_i, \qquad (7)
\end{aligned}$$

where the constraints are required for all terminals i, $\mathcal{Q}$ denotes the set of functions $\mathbf{R}^+ \to \{0, 1\}$ and $\mathcal{P}_i$ represents the set of functions $\mathbf{R}^+ \to [0, p_i^{\text{inst}}]$.

III. Adaptive Algorithms For Decentralized Channel-Aware Random Access

To develop adaptive decentralized algorithms, begin by separating (7) in per terminal subproblems. To do so, substitute (4) into (6) and express the logarithm of a product as a sum of logarithms and rewrite the global utility as

$$U(\mathbf{r}) = \sum_{i=1}^{n} w_i\left[\log \mathbb{E}_{h_i}[C_i(h_iP_i(h_i))Q_i(h_i)] + \sum_{j=1,j\neq i}^{n} \log\Big[1 - \mathbb{E}_{h_j}[Q_j(h_j)]\Big]\right]. \quad (8)$$

Since each summand in (8) involves variables related to a particular node only, we can reorder summands in (8) by node index. Further defining $\tilde{w}_i := \sum_{j=1,j\neq i}^{n} w_i$, we rewrite (8) as

$$U(\mathbf{r}) = \sum_{i=1}^{n}\left[w_i \log\Big[\mathbb{E}_{h_i}[C_i(h_iP_i(h_i))Q_i(h_i)]\Big] + \tilde{w}_i \log\Big[1 - \mathbb{E}_{h_i}[Q_i(h_i)]\Big]\right] := \sum_{i=1}^{n} U_i, \quad (9)$$

where we have defined the local utilities U_i. Since U_i only involves variables that are related to terminal i, it can be regarded as a utility function for terminal i. To maximize $U(\mathbf{r})$ for the whole system it suffices to separately maximize U_i for each terminal i. Introducing auxiliary variables $x_i = \mathbb{E}_{h_i}[C_i(h_iP_i(h_i))Q_i(h_i)]$ and $y_i = \mathbb{E}_{h_i}[Q_i(h_i)]$, it follows that (7) is equivalent to the following per terminal subproblems

$$\begin{aligned}
\mathsf{P}_i \quad = \quad & \max\ w_i \log x_i + \tilde{w}_i \log(1 - y_i) \qquad (10) \\
& \text{s.t. } x_i \leq \mathbb{E}_{h_i}\left[C_i(h_iP_i(h_i))Q_i(h_i)\right], \\
& \qquad y_i \geq \mathbb{E}_{h_i}\left[Q_i(h_i)\right], \\
& \qquad \mathbb{E}_{h_i}\left[Q_i(h_i)P_i(h_i)\right] \leq p_i^{\text{avg}}, \\
& \qquad x_i \geq 0, 0 \leq y_i \leq 1, Q_i(h_i) \in \mathcal{Q}, P_i(h_i) \in \mathcal{P}_i,
\end{aligned}$$

where we wrote the equality constraints as inequalities which can be done without loss of optimality. Finding optimal solutions of (10) for all terminals i is equivalent to solving (7). Note, however, that as is the case with (7), solving (10) is difficult because: (i) The optimization space in (10) includes functions $Q_i(h_i)$ and $P_i(h_i)$ that are defined on $\mathbf{R}^+$, implying that the dimension of the problem is infinite. (ii) The rate function $C_i(h_i P_i(h_i))$ is in general non-concave with respect to $h_i P_i(h_i)$, and may be even discontinuous as in systems that use AMC. (iii) The constraints involve expected values over random variables h_i whose pdfs are unknown.

An important observation is that the number of constraints in (10) is finite. This implies that while there are infinite variables in the primal domain, there are a finite number of variables in the dual domain. This observation suggests that (10) is more tractable in the dual space. Introduce Lagrange multipliers $\boldsymbol{\lambda}_i = [\lambda_{i1}, \lambda_{i2}, \lambda_{i3}]^T$ associated with the first three inequality constraints in (10); define vectors $\mathbf{x}_i := [x_i, y_i]^T$ and $\mathbf{P}_i(h_i) := [Q_i(h_i), P_i(h_i)]^T$; and write the Lagragian of the optimization problem in (10) as

$$\begin{aligned} &\mathcal{L}_i(\mathbf{x}_i, \mathbf{P}_i(h_i), \boldsymbol{\lambda}_i) \qquad (11)\\ &= w_i \log x_i + \tilde{w}_i \log(1-y_i) \\ &\quad + \lambda_{i1} \left[\mathbb{E}_{h_i}\left[Q_i(h_i) C_i(h_i P_i(h_i))\right] - x_i\right] \\ &\quad + \lambda_{i2}\left[y_i - \mathbb{E}_{h_i}\left[Q_i(h_i)\right]\right] + \lambda_{i3}\left[p_i^{\text{avg}} - \mathbb{E}_{h_i}\left[Q_i(h_i)P_i(h_i)\right]\right] \end{aligned}$$

Reordering terms in (11) we can rewrite the Lagrangian as

$$\begin{aligned} &\mathcal{L}_i(\mathbf{x}_i, \mathbf{P}_i(h_i), \boldsymbol{\lambda}_i) \qquad (12)\\ &= \lambda_{i3} p_i^{\text{avg}} + [w_i \log x_i - \lambda_{i1} x_i] + [\tilde{w}_i \log(1-y_i) + \lambda_{i2} y_i] \\ &\quad + \mathbb{E}_{h_i}\left[Q_i(h_i)\left[\lambda_{i1} C_i(h_i P_i(h_i)) - \lambda_{i2} - \lambda_{i3} P_i(h_i)\right]\right]. \end{aligned}$$

whose separable structure is leveraged later on. The dual function is then defined as the maximum of the Lagrangian over the set of feasible $\mathbf{x}_i$ and $\mathbf{P}_i(h_i)$, i.e.,

$$\begin{aligned} g_i(\boldsymbol{\lambda}_i) &:= \max\ \mathcal{L}_i(\mathbf{x}_i, \mathbf{P}_i(h_i), \boldsymbol{\lambda}_i) \qquad (13)\\ &\text{s.t.}\quad x_i \geq 0, 0 \leq y_i \leq 1, Q_i(h_i) \in \mathcal{Q}, P_i(h_i) \in \mathcal{P}_i, \end{aligned}$$

and the dual problem as the minimization of $g_i(\boldsymbol{\lambda}_i)$, i.e.,

$$\mathsf{D}_i = \underset{\boldsymbol{\lambda}_i \geq 0}{\text{minimize}} \quad g_i(\boldsymbol{\lambda}_i). \qquad (14)$$

In general, the optimal dual value D_i provides an upper bound for the optimal primal value P_i, i.e., $\mathsf{D}_i \geq \mathsf{P}_i$. While the inequality is typically strict for non-convex problems, for the problem in (10) $\mathsf{P}_i = \mathsf{D}_i$ as long as the fading distribution has no realization with positive measure [12]. This lack of duality gap implies that the finite dimensional convex dual problem is equivalent to the infinite dimensional nonconvex primal problem. However, this does not necessarily mean that solving the dual problem is easy because evaluation of the dual function's value requires maximization of the Lagrangian. In particular, this maximization includes an expected value over the unknown channel distribution. Still, convexity of the dual function allows the use of descent algorithms in the dual domain because any local optimal solution is a global optimal solution $\boldsymbol{\lambda}_i^* = [\lambda_{i1}^*, \lambda_{i2}^*, \lambda_{i3}^*]^T$. This property is exploited next to develop a stochastic subgradient descent algorithm that solves (14) using observations of instantaneous channel realizations $h_i(t)$.

A. Adaptive Algorithms Using Stochastic Subgradient Descent

Instead of directly finding optimal $x_i, y_i, Q_i(h_i)$ and $P_i(h_i)$ for the primal problem (10), the proposed algorithm exploits the lack of duality gap to use a stochastic subgradient descent in the dual domain. Starting from given dual variables $\boldsymbol{\lambda}_i(t)$, the algorithm computes instantaneous primal variables $x_i(t)$, $y_i(t)$, $q_i(t)$ and $p_i(t)$ based on channel realization $h_i(t)$ in time slot t, and uses these values to update dual variables $\boldsymbol{\lambda}_i(t+1)$. Specifically, the algorithm starts finding primal variables that optimize the summands of the Lagrangian in (12) (the operator $[\cdot]^+$ denotes projection in the positive orthant)

$$x_i(t) = \underset{x_i \geq 0}{\operatorname{argmax}} \{w_i \log x_i - \lambda_{i1}(t) x_i\} = \frac{w_i}{\lambda_{i1}(t)}, \qquad (15)$$

$$y_i(t) = \underset{0 \leq y_i \leq 1}{\operatorname{argmax}} \{\tilde{w}_i \log(1-y_i) + \lambda_{i2}(t) y_i\} = \left[1 - \frac{\tilde{w}_i}{\lambda_{i2}(t)}\right]^+, \qquad (16)$$

$$\begin{aligned} &\{q_i(t), p_i(t)\} = \\ &\underset{q_i \in \{0,1\}, p_i \in [0, p_i^{\text{inst}}]}{\operatorname{argmax}} \{q_i\left[\lambda_{i1}(t) C_i(h_i(t) p_i) - \lambda_{i2}(t) - \lambda_{i3}(t) p_i\right]\}, \qquad (17) \end{aligned}$$

The maximization in (17) determines schedules and transmitted powers associated with current channel realization $h_i(t)$. Since q_i in (17) takes values on {0,1}, we can alternatively compute the maximizing arguments in (17) as

$$\begin{aligned} p_i(t) &= \underset{p_i \in [0, p_i^{\text{inst}}]}{\operatorname{argmax}} \{\lambda_{i1}(t) C_i(h_i(t) p_i) - \lambda_{i2}(t) - \lambda_{i3}(t) p_i\}, \\ q_i(t) &= H\big(\lambda_{i1}(t) C_i(h_i(t) p_i(t)) - \lambda_{i2}(t) - \lambda_{i3}(t) p_i(t)\big), \qquad (18) \end{aligned}$$

where $H(a)$ denotes Heaviside's step function with $H(a) = 1$ for $a > 0$ and $H(a) = 0$ otherwise.

Since $\tilde{w}_i$, i.e., the sum of w_j for $j \neq i$, is needed for terminal i to compute instantaneous primal variable $y_i(t)$ [cf. (16)], terminals need to exchange their weight coefficients w_i with each other before the algorithm begins. This can be done by letting terminals send w_i to AP and AP then broadcast to all terminals.

Based on $x_i(t)$, $y_i(t)$, $q_i(t)$ and $p_i(t)$, define the stochastic subgradient $\mathbf{s}_i(t) = [s_{i1}(t), s_{i2}(t), s_{i3}(t)]^T$ with components

$$s_{i1}(t) = q_i(t) C_i(h_i(t) p_i(t)) - x_i(t), \qquad (19)$$

$$s_{i2}(t) = y_i(t) - q_i(t), \qquad (20)$$

$$s_{i3}(t) = p_i^{\text{avg}} - q_i(t) p_i(t). \qquad (21)$$

The algorithm is completed with the introduction of a constant step size ϵ and a descent update in the dual domain along the stochastic subgradient $\mathbf{s}_i(t)$

$$\lambda_{il}(t+1) = [\lambda_{il}(t) - \epsilon s_{il}(t)]^+, \quad \text{for } l = 1, 2, 3. \qquad (22)$$

Notice that computation of variables in (15)-(22) does not require information exchanges between terminals. This guarantees $Q_i(h_i)$ and $P_i(h_i)$ to be independent of $Q_j(h_j)$ and $P_j(h_j)$ for all $i \neq j$ as required by problem definition.

The proposed adaptive scheduling and power control algorithm is summarized below:

Algorithm 1: Adaptive scheduling and power control at terminal i

Initialize Lagrangian multipliers $\boldsymbol{\lambda}_i(0)$;
for $t = 0, 1, 2, \cdots$ **do**
  Compute primal variables as per (15), (16), and (18):
    $x_i(t) = \dfrac{w_i}{\lambda_{i1}(t)}$;
    $y_i(t) = \left[1 - \dfrac{\tilde{w}_i}{\lambda_{i2}(t)}\right]^+$;
    $p_i(t) = \underset{p_i \in [0, p_i^{\text{inst}}]}{\text{argmax}} \{\lambda_{i1}(t) C_i(h_i(t)p_i) - \lambda_{i2}(t) - \lambda_{i3}(t)p_i\}$;
    $q_i(t) = H\big(\lambda_{i1}(t) C_i(h_i(t)p_i(t)) - \lambda_{i2}(t) - \lambda_{i3}(t)p_i(t)\big)$;
  if $q_i(t) = 1$ **then**
    Transmit with power $p_i(t)$;
  end
  Compute stochastic subgradients as per (19)-(21):
    $s_{i1}(t) = q_i(t) C_i(h_i(t)p_i(t)) - x_i(t)$;
    $s_{i2}(t) = y_i(t) - q_i(t)$;
    $s_{i3}(t) = p_i^{\text{avg}} - q_i(t)p_i(t)$;
  Update dual variables as per (22):
    $\lambda_{il}(t+1) = [\lambda_{il}(t) - \epsilon s_{il}(t)]^+, \quad$ for $l = 1, 2, 3$;
end

To analyze convergence of (15)-(22) let us start by showing that $\mathbf{s}(t)$ is indeed a stochastic subgradient of the dual function as stated in the following proposition.

Proposition 1: Given $\boldsymbol{\lambda}_i(t)$, the expected value of the stochastic subgradient $\mathbf{s}_i(t)$ is a subgradient of the dual function in (13), i.e., $\forall \boldsymbol{\lambda}_i \geq 0$,

$$\mathbb{E}_{h_i}\left[\mathbf{s}_i^T(t)|\boldsymbol{\lambda}_i(t)\right](\boldsymbol{\lambda}_i(t) - \boldsymbol{\lambda}_i) \geq g_i(\boldsymbol{\lambda}_i(t)) - g_i(\boldsymbol{\lambda}_i). \quad (23)$$

In particular,

$$\mathbb{E}_{h_i}\left[\mathbf{s}_i^T(t)|\boldsymbol{\lambda}_i(t)\right](\boldsymbol{\lambda}_i(t) - \boldsymbol{\lambda}_i^*) \geq g_i(\boldsymbol{\lambda}_i(t)) - \mathsf{D}_i \geq 0. \quad (24)$$

Proof: See [13]. ■

Proposition 1 states that the average of the stochastic subgradient $\mathbf{s}_i(t)$ is a subgradient of the dual function. We can then think of an alternative algorithm by replacing $\mathbb{E}_{h_i}\left[\mathbf{s}_i(t)\middle|\boldsymbol{\lambda}_i(t)\right]$ for $\mathbf{s}_i(t)$ in the dual iteration step (22), which would amount to a subgradient descent algorithm for the dual function. Since, $\mathbb{E}_{h_i}\left[\mathbf{s}_i(t)\middle|\boldsymbol{\lambda}_i(t)\right]$ points towards $\boldsymbol{\lambda}^*$ – the angle between $\mathbb{E}_{h_i}\left[\mathbf{s}_i(t)\middle|\boldsymbol{\lambda}_i(t)\right]$ and $\boldsymbol{\lambda}_i(t) - \boldsymbol{\lambda}_i^*$ is positive as indicated by (24) –, it is not difficult to prove that $\boldsymbol{\lambda}_i(t)$ eventually approaches $\boldsymbol{\lambda}_i^*$. However, since we assume the pdf of h_i is unknown, the subgradient $\mathbb{E}_{h_i}\left[\mathbf{s}_i(t)\middle|\boldsymbol{\lambda}_i(t)\right]$ can only be approximated using past channel realizations $h_i(1), \ldots, h_i(t)$. While this approach is possible, it is computationally costly.

The computation of the stochastic subgradient $\mathbf{s}_i(t)$, on the contrary, is simple because it only depends on the current channel state $h_i(t)$. Furthermore, since $\mathbf{s}_i(t)$ points towards the set of optimal dual variables $\boldsymbol{\lambda}_i^*$ on average [cf. (24)] it is reasonable to expect the stochastic subgradient descent iterations in (22) to also approach $\boldsymbol{\lambda}_i^*$ in some sense. This can be proved true and leveraged to prove almost sure convergence of primal iterates $x_i(t)$, $y_i(t)$, $p_i(t)$ and $q_i(t)$ to an optimal operating point in an ergodic sense [14]. Specifically, Theorem 1 of [14] assumes as hypotheses that the second moment of the norm of the stochastic subgradient $\mathbf{s}_i(t)$ is finite, i.e., $\mathbb{E}_{h_i}\left[\|\mathbf{s}_i(t)\|^2\middle|\boldsymbol{\lambda}_i(t)\right] \leq \hat{S}_i^2$, and that there exists a set of strictly feasible primal variables that satisfy the constraints in (10) with strict inequality. If these hypotheses are true, primal iterates of dual stochastic subgradient descent are almost surely feasible in an ergodic sense. For the particular case of the problem in (10), [14, Theorem 1] implies that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} q_i(u)p_i(u) \leq p_i^{\text{avg}} \quad \text{a.s.}, \quad (25)$$

$$\lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} x_i(u) \leq \lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} q_i(u) C_i(h_i(u)p_i(u)) \quad \text{a.s.}, \quad (26)$$

$$\lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} y_i(u) \geq \lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} q_i(u) \quad \text{a.s.} \quad (27)$$

It also follows from [14, Theorem 1] that $x_i(t)$ and $y_i(t)$ yield ergodic utilities that are almost surely within $\epsilon\hat{S}_i^2/2$ of optimal,

$$\mathsf{P}_i - \lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} \left[w_i \log x_i(u) + \tilde{w}_i \log(1 - y_i(u))\right] \leq \frac{\epsilon \hat{S}_i^2}{2} \quad \text{a.s.} \quad (28)$$

From (25) we can conclude that the ergodic limit of the power allocated by the proposed algorithm satisfies the average power constraint. However, (28) does not imply that the scheduling and power allocation variables $p_i(t)$ and $q_i(t)$ are optimal. The optimality claim in (28) is for the auxiliary variables $x_i(t)$ and $y_i(t)$ but the goal here is to claim optimality of the scheduling and power allocation variables $p_i(t)$ and $q_i(t)$. To prove optimality of the algorithm, we need to show that the ergodic transmission rate r_i of (2), achieved by allocations $q_i(t)$ and $p_i(t)$ is optimal in the sense of maximizing the throughput utility $U(\mathbf{r}) = \sum_{i=1}^{n} w_i \log(r_i)$. This mismatch can be addressed to prove the following theorem.

Theorem 1: Consider a random multiple access channel with n terminals using schedules $q_i(t)$ and power allocations $p_i(t)$ generated by the algorithm defined by (15)-(22) resulting in instantaneous transmission rates $r_i(t)$ as given by (1) and ergodic rates r_i as defined by (2). Define vector $\mathbf{r} := [r_1, \ldots, r_n]^T$, and let $U(\mathbf{r})$ be the weighted proportional fair utility in (6). Assume that the second moment of the norm of the stochastic subgradient $\mathbf{s}_i(t)$ with components as in (19)-(21) is finite, i.e., $\mathbb{E}_{h_i}\left[\|\mathbf{s}_i(t)\|^2\middle|\boldsymbol{\lambda}_i(t)\right] \leq \hat{S}_i^2$, and that there exists a set of strictly feasible primal variables that satisfy the constraints in (10) with strict inequality. Then, the average power constraint is almost surely satisfied

$$\lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} q_i(u)p_i(u) \leq p_i^{\text{avg}} \quad \text{a.s.}, \quad (29)$$

and the utility of the ergodic limit of the transmission rates almost surely converges to a value within $\epsilon/2 \sum_{i=1}^{n} \hat{S}_i^2$ of optimal,

$$\mathsf{P} - U(\mathbf{r}) := P - \sum_{i=1}^{n} w_i \log\left(\lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} r_i(u)\right) \leq \frac{\epsilon}{2} \sum_{i=1}^{n} \hat{S}_i^2. \quad (30)$$

Proof: The hypotheses of Theorem 1 are chosen to satisfy the hypotheses guaranteeing convergence of ergodic stochastic optimization algorithms [14, Theorem 1]. Thus, almost sure feasibility and almost sure near optimality of iterates $x_i(t)$, $y_i(t)$, $p_i(t)$

and $q_i(t)$ follows in the sense of (25)-(28). To establish almost sure satisfaction of average power constraints as per (29) just notice that this inequality coincides with the one in (25). To establish (30) start by rearranging terms in (28) to conclude that $\mathsf{P}_i - \epsilon\hat{S}_i^2/2 \leq \lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}[w_i\log x_i(u) + \tilde{w}_i\log(1-y_i(u))]$. Due to continuity and concavity of the logarithm function we can further bound $\mathsf{P}_i - \epsilon\hat{S}_i^2/2$ as

$$\mathsf{P}_i - \frac{\epsilon\hat{S}_i^2}{2} \leq w_i\log\left[\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}x_i(u)\right] + \tilde{w}_i\log\left[1-\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}y_i(u)\right]. \quad (31)$$

The limits in (31) are equal to the limits in the left hand sides of the inequalities in (26) and (27). Thus, using this almost sure ergodic feasibility results $\mathsf{P}_i - \epsilon\hat{S}_i^2/2$ is bounded as

$$\mathsf{P}_i - \frac{\epsilon\hat{S}_i^2}{2} \leq w_i\log\left[\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}q_i(u)C_i(h_i(u)p_i(u))\right] + \tilde{w}_i\log\left[1-\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}q_i(u)\right]. \quad (32)$$

Ergodicity, possibly restricted to an ergodic component, allows replacement of the ergodic limits in (33) by the corresponding expected values, leading to the bound

$$\mathsf{P}_i - \frac{\epsilon\hat{S}_i^2}{2} \leq w_i\log\mathbb{E}_{h_i}\left[Q_i(h_i)C_i(h_iP_i(h_i))\right] + \tilde{w}_i\log\mathbb{E}_{h_i}[1-Q_i(h_i)]. \quad (33)$$

Recall that $\mathsf{P} = \sum_{i=1}^{n}\mathsf{P}_i$ per definition, and consider the sum of the inequalities in (33) for all terminals i so as to write

$$\begin{aligned}\mathsf{P} - \sum_{i=1}^{n}\frac{\epsilon\hat{S}_i^2}{2} &\leq \sum_{i=1}^{n}w_i\log\mathbb{E}_{h_i}\left[Q_i(h_i)C_i(h_iP_i(h_i))\right] + \tilde{w}_i\log\mathbb{E}_{h_i}[1-Q_i(h_i)]\\ &\leq \sum_{i=1}^{n}w_i\log\Bigg[\mathbb{E}_{h_i}\left[Q_i(h_i)C_i(h_i(t)P_i(h_i))\right]\prod_{j=1,j\neq i}^{n}\mathbb{E}_{h_j}[1-Q_j(h_j)]\Bigg], \end{aligned} \quad (34)$$

where the second inequality follows by using the definition $\tilde{w}_i := \sum_{j=1,j\neq i}^{n}w_i$, reordering terms in the sum, and rewriting a sum of logarithms as the logarithm of a product.

The fundamental observation in this proof is that the scheduling function $Q_i(h_i)$ and the power allocation function $P_i(h_i)$ are independent of the corresponding $Q_j(h_j)$ and $P_j(h_j)$ of other terminals. This is not a coincidence, but the intended goal of reformulating (7) as (10). Using this independence, the product of expectations in (34) can be written as single expectation over the vector channel $\mathbf{h}$ to yield

$$\mathsf{P} - \sum_{i=1}^{n}\frac{\epsilon\hat{S}_i^2}{2} \leq \sum_{i=1}^{n}w_i\log\Bigg[\mathbb{E}_{\mathbf{h}}\Bigg(Q_i(h_i)C_i(h_iP_i(h_i))\prod_{j=1,j\neq i}^{n}(1-Q_j(h_j))\Bigg)\Bigg]. \quad (35)$$

To finalize the proof use ergodicity, possibly restricted to an ergodic component, to substitute the expectation in (35) by an ergodic limit to yield

$$\mathsf{P} - \sum_{i=1}^{n}\frac{\epsilon\hat{S}_i^2}{2} \leq \sum_{i=1}^{n}w_i\log\Bigg[\lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}q_i(u)C_i(h_i(u)p_i(u))\prod_{j=1,j\neq i}^{n}(1-q_j(u))\Bigg] := U(\mathbf{r}), \quad (36)$$

where we have used the definitions of the ergodic rate in (2) and of the utility in (6). The result in (30) follows after reordering terms in (35). ■

Theorem 1 states that the stochastic dual descent algorithm in (15)-(22) computes schedules $q_i(t)$ and power allocations $p_i(t)$ yielding rates $r_i(t)$ that are almost surely near optimal in an ergodic sense. It also states that $p_i(t)$ satisfies the average power constraint with probability 1. Notice that the stochastic dual descent algorithm in (15)-(22) does not compute the optimal scheduling and power control functions for each terminal. Rather, it draws schedules $q_i(t)$ and power allocations $p_i(t)$ that are close to the optimal functions. This is not a drawback because the latter property is sufficient for a practical implementation. Further note that the use of constant step sizes ϵ endows the algorithm with adaptability to time-varying channel distributions. This is important in practice because wireless channels are non-stationary due to user mobility and environmental dynamics. The gap between $U(\mathbf{r})$ and P can be made arbitrarily small by reducing ϵ.

IV. Numerical Results

To illustrate performance of the proposed algorithms, we conduct numerical experiments on a network with $n = 20$ terminals randomly placed in a square with side $L = 100$ m and a common AP located at the center of the square. Numerical experiments here utilize the realization of this random placement shown in Fig. 1. Communication between terminals and the AP is over a bandlimited Gaussian channel with bandwidth B and noise power spectral density N_0. We set $B = 1$ so that capacities are measured in bits per second per Hertz (b/s/Hz) and $N_0 = 10^{-10}$ W. Channel gains $h_i(t)$ are Rayleigh distributed with mean $\bar{h}_i$ and are independent across terminals and time. The average channel gain $\bar{h}_i := \mathbb{E}[h_i]$ follows an exponential pathloss law, $\bar{h}_i = \alpha d_i^{-\beta}$ with $\alpha = 10^{-6}\text{m}^{-1}$ and $\beta = 2$ constants and d_i denoting the distance in meters between terminal i and the AP. All weights in the proportional fair utility in (6) are set to $w_i = 1$. Throughout, the performance metric of interest is the average transmission rate $\bar{r}_i(t)$ of terminal i at time t defined as

$$\bar{r}_i(t) = \frac{1}{t}\sum_{u=1}^{t}r_i(u), \quad (37)$$

where $r_i(u)$ is normalized so that it represents bits/s/Hz. The system's throughput utility by time t is then defined in terms of $\bar{r}_i(t)$ as $\bar{U}(t) := \sum_{i=1}^{n}w_i\log(\bar{r}_i(t))$.

The algorithm in (15)-(22) is first tested in a network where nodes use capacity achieving codes and have instantaneous power constraints but do not have average power constraints; see Section

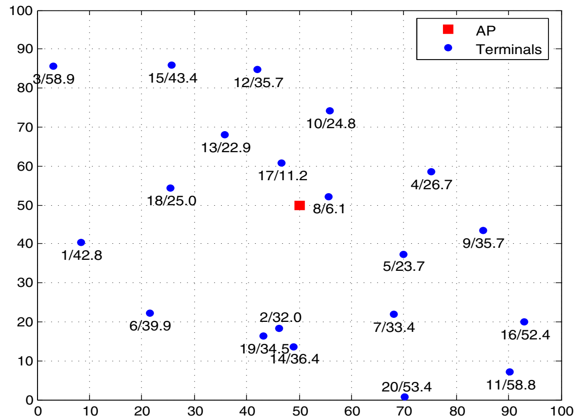


Fig. 1. An example multiple access channel with $n = 20$ nodes communicating with a common access point (AP). Nodes are randomly placed in a 100 m × 100 m square and the AP is located at the center of the square. Nodes' labels represent indexes and distances to the AP. Subsequent numerical experiments use this realization of the random placement.

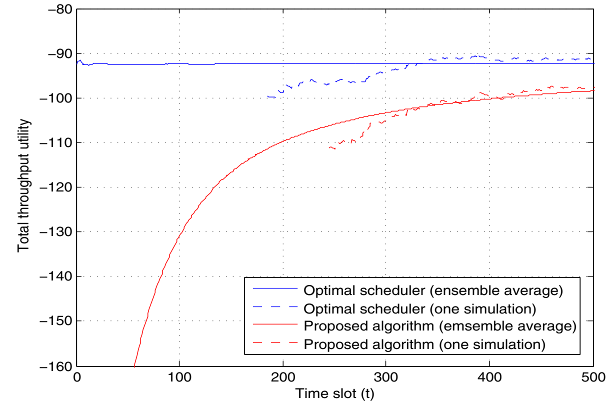


Fig. 2. Convergence of the proposed algorithm to near optimal utility with instantaneous power constrains but no average power constraints. Throughput utility of the proposed adaptive algorithm and of the optimal offline scheduler are shown as functions of time for one realization and for the ensemble average of realizations. In steady state the adaptive algorithm operates with minimal performance loss with respect to the optimal offline scheduler. A utility gap smaller than 10 is achieved in about 350 iterations. Power constraint $p_i^{\text{inst}} = 100$ mW, step size $\epsilon = 0.1$, capacity achieving codes.

IV-A. We then consider nodes that have average as well as instantaneous power constraints using AMC; see Section IV-B.

A. System with Instantaneous Power Constraint

Assume the use of capacity achieving codes so that the rate function for terminal i takes the form

$$C_i(h_i(t)p_i(t)) = B \log\left(1 + \frac{h_i(t)p_i(t)}{BN_0}\right). \tag{38}$$

Further assume that there is an instantaneous power constraint $p_i^{\text{inst}} = 100$ mW for each terminal, but that there is no average power constraint. Since the rate function is a nonnegative increasing function of power it is optimal for each terminal to transmit with its maximum allowed instantaneous power every time it decides to transmit. Therefore, the power control function is a constant $p_i(t) = p_i^{\text{inst}}$ and the system's performance depends solely on the terminals' scheduling functions $q_i(t)$. In this simplified setting, a closed form solution for $q_i(t)$ is known if the channel pdf is available [2]. Our interest in this simplified problem is that it allows a performance comparison between the schedules yielded by (15)-(22) and those of the optimal offline scheduler.

Convergence of (15)-(22) to a near optimal operating point is illustrated in Fig. 2 for step size $\epsilon = 0.1$. The ergodic utility $\bar{U}(t)$ is shown through 500 iterations and is compared with the utility of the optimal offline scheduler. When using (15)-(22) the total throughput utility converges to a value with negligible optimality gap with respect to the offline scheduler. Observe that convergence is fast as it takes less than 180 iterations to reach a utility with optimality gap smaller than 20 and 360 iterations to get an optimality gap smaller than 10. Figs. 3 and 4 respectively show average rates and transmission probabilities after 500 iterations for each terminal. Observe in Fig. 3 that all terminals achieve average rates that are very close to the optimal ones. Further observe that even though terminals experience different channel conditions, fair schedules are obtained as a consequence of the use of a logarithmic utility. Indeed, as seen in Fig. 4, average transmission probabilities are close for all

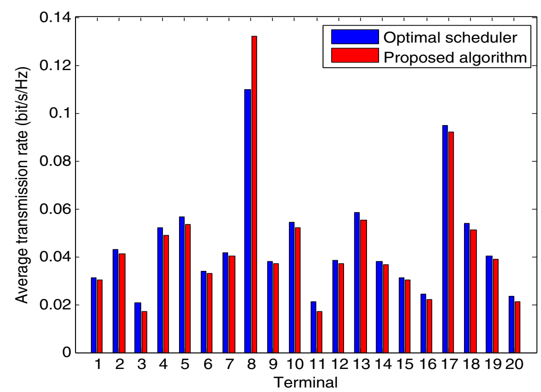


Fig. 3. Average transmission rates (bits/s/Hz) in 500 time slots, i.e., $\bar{r}_i(500)$ as defined in (37), for all terminals. The optimal offline scheduler and the proposed adaptive algorithm yield similar close to optimal average rates. The variation in achieved rates is commensurate with the variation in average signal to noise ratios (SNRs) due to different distances to the access point. For the network in Fig.1 and the pathloss and power parameters used here, average signal to noise ratios vary between 0.4 and 10. Instantaneous power constraint $p_i^{\text{inst}} = 100$ mW, step size $\epsilon = 0.1$, capacity achieving codes.

terminals. Note, however, that the achieved rates shown in Fig. 3 are different because terminals have different average channels.

To test how the optimality gap changes as the step size ϵ varies, we ran the algorithm (15)-(22) with different step sizes. Fig. 5 shows the optimality gap when the step size ϵ varies between 10^{-2} to 10^{-1}. The optimality gap indeed decreases as the step size ϵ is reduced. This corroborates the result of Theorem 1 that ensures a vanishing optimality gap as $\epsilon \to 0$. Using smaller step size, however, leads to slower convergence. This tradeoff between convergence speed and optimality gap determines the choice of ϵ for practical implementations.

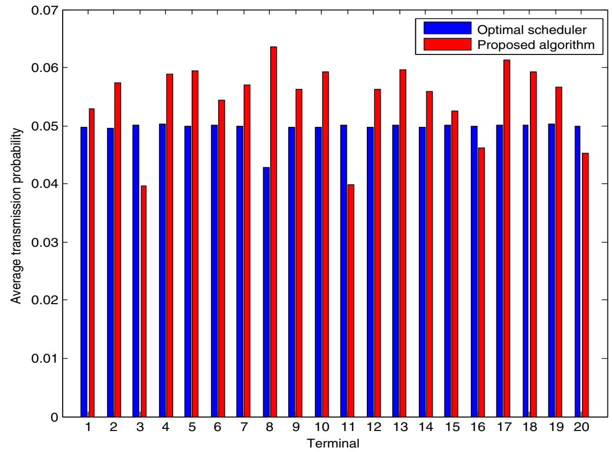


Fig. 4. Average transmission probabilities in 500 time slots for all terminals. Offline and adaptive optimal schedulers shown. Despite different channel conditions all terminals transmit with a similar probability close to $1/n = 0.05$. This is consistent with the use of a logarithmic, i.e., proportional fair, utility. Instantaneous power constraint $p_i^{\text{inst}} = 100$ mW, step size $\epsilon = 0.1$, capacity achieving codes.

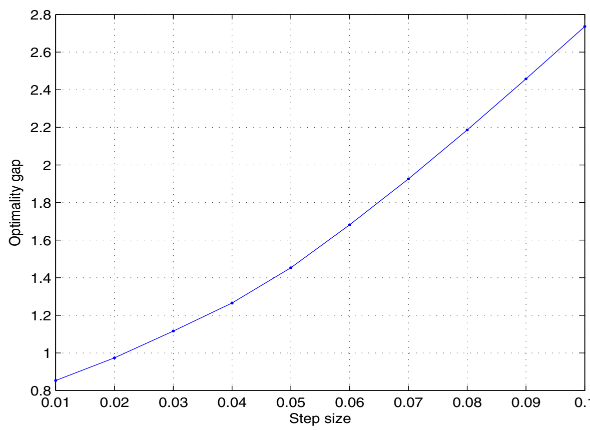


Fig. 5. Steady state optimality gap between proposed adaptive algorithm and optimal offline scheduler as a function of step size ϵ. Values of ϵ between 10^{-2} and 10^{-3} shown. As the step size decreases, the optimality gap decreases. The optimality gap can be made arbitrarily small by reducing ϵ. Instantaneous power constraint $p_i^{\text{inst}} = 100$ mW, capacity achieving codes.

B. System with Average Power Constraint

For the same network, consider now the case in which each terminal adheres to both, instantaneous and average power constraints. Assume AMC with M transmission modes is used in the physical layer. The mth mode affords communication rate τ_m and is used when the signal to noise ratio (SNR) $h_i(t)p_i(t)/BN_0$ is between η_m and η_{m+1}. The rate function is therefore

$$C_i(h_i(t)p_i(t)) = \sum_{m=1}^{M} \tau_m \mathbb{I}\left(\eta_m \leq \frac{h_i(t)p_i(t)}{BN_0} \leq \eta_{m+1}\right), \quad (39)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Each terminal has $M = 4$ AMC modes with respective rates $\tau_1 = 1$ bits/s/Hz, $\tau_2 = 2$ bits/s/Hz, $\tau_3 = 3$ bits/s/Hz, and $\tau_4 = 4$ bits/s/Hz. The transitions between AMC modes are at SNRs $\eta_1 = 1$, $\eta_2 = 4$, $\eta_3 = 8$, and

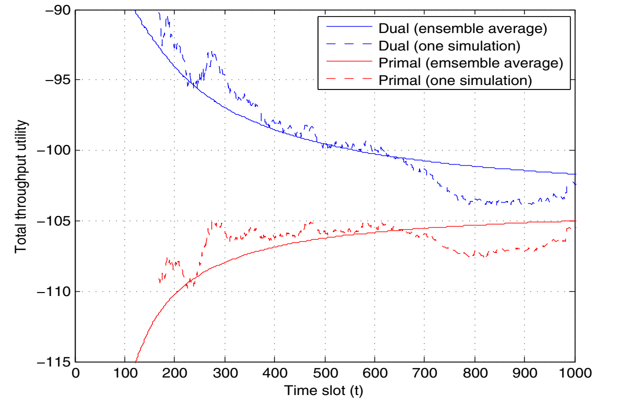


Fig. 6. One realization and ensemble average of primal and dual objectives. As time grows the duality gap decreases, eventually approaching a small positive constant and implying near optimality of the achieved rates.

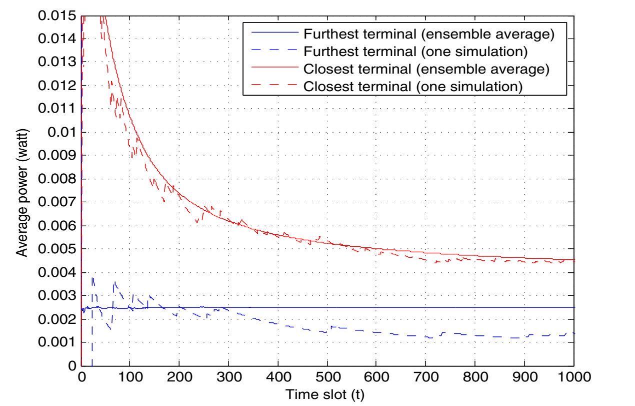
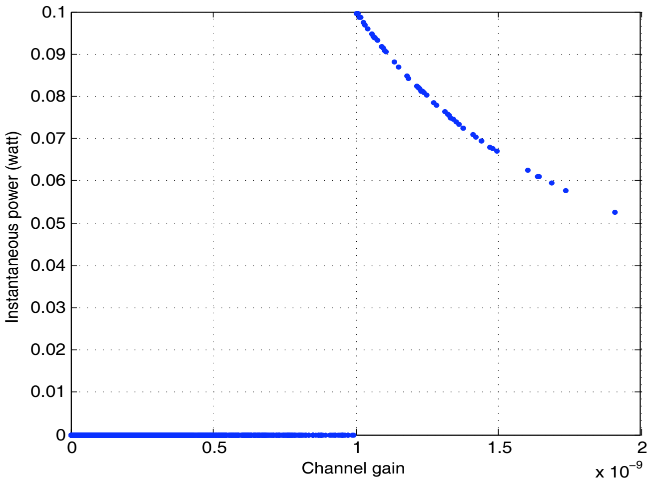


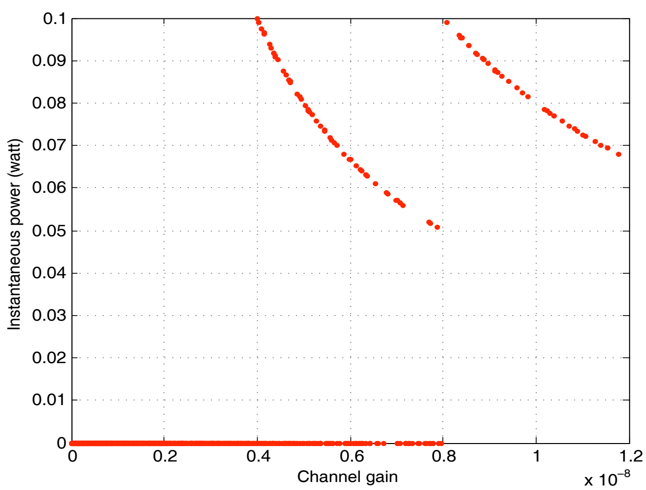
Fig. 7. Average power consumption for the closest and furthest terminals. Average power constraints $p_i^{\text{avg}} = 5$ mW are satisfied as time grows. Power $\bar{p}_f(t)$ consumed by the furthest terminal is smaller than the allowed budget p_f^{avg} due to unfavorable channel conditions. The closest terminal adheres to its power budget after approximately 600 iterations.

$\eta_4 = 16$. We set the instantaneous power constraint to $p_i^{\text{inst}} = 100$ mW and the average power constraint to $p_i^{\text{avg}} = 5$ mW for all terminals i.

To demonstrate optimality of the proposed algorithm, we compute the primal objective $\bar{U}(t)$, the dual value $D(t) = \sum_{i=1}^{n} g_i(\boldsymbol{\lambda}_i(t))$, and examine the duality gap between them. Fig. 6 shows $\bar{U}(t)$ and $D(t)$ for 10^3 time slots. As time grows, the duality gap decreases and eventually approaches a small positive constant, implying near optimality of the proposed algorithm. To test the satisfaction of the average power constraint, define the average power consumption of terminal i by time t as $\bar{p}_i(t) = \frac{1}{t}\sum_{u=1}^{t} p_i(u)$. Average power consumptions $\bar{p}_f(t)$ and $\bar{p}_c(t)$ for the terminals that are furthest and closest to the AP, are shown in Fig. 7. Observe that in both cases the average power constraints are satisfied as time increases. For the furthest terminal, $\bar{p}_f(t)$ is always smaller than p_f^{avg} since channel conditions are unfavorable, resulting in this terminal utilizing only mode 1 for communication to the AP. For the closest terminal, $\bar{p}_c(t)$ falls below p_c^{avg} after 600 iterations. This is as expected due to the almost sure feasibility result of Theorem 1. Fig. 8 (a) and (b) illustrate



(a) Furthest terminal



(b) Closest terminal

Fig. 8. Instantaneous power allocations $p_i(t)$ for the closest and furthest terminals plotted against the channel realization $h_i(t)$. Notice that the channel axes scales are different in (a) and (b). In both cases, no power is allocated when channel realizations are bad. The furthest terminal uses only the AMC mode with the lowest rate $\tau_1 = 1$ bits/s/Hz, while the closest terminal uses two modes with rates $\tau_2 = 2$ bits/s/Hz and $\tau_3 = 3$ bits/s/Hz. This happens because the terminal closer to the AP, has a better average channel.

the relationship between instantaneous power allocations $p_i(t)$ and instantaneous channel gains $h_i(t)$ for the furthest and closest terminal, respectively. Channel access is opportunistic since no power is allocated when channel realizations are below average. Further note that the furthest terminal only uses the AMC mode with the lowest rate $\tau_1 = 1$ bits/s/Hz while the closest terminal uses two modes with rates $\tau_2 = 2$ bits/s/Hz and $\tau_3 = 3$ bits/s/Hz. This happens because the terminal closer to the AP, has a better average channel.

V. Conclusions

We developed optimal adaptive scheduling and power control algorithms for random multiple access channels. Terminals are assumed to know their local channel state information but have no access to the probability distribution of the channel or the channel state of other terminals. In this setting, the proposed online algorithm determines schedules and transmitted powers that maximize a global proportional fair utility. The global utility maximization problem was decomposed in per-terminal utility maximization subproblems. Adaptive algorithms using stochastic subgradient descent in the dual domain were then used to solve these local optimizations. Almost sure convergence and almost sure near optimality of the proposed algorithm was established. Important properties of the algorithm are low computational complexity and the ability to handle non-convex rate functions. Numerical results for a randomly generated network using adaptive modulation and coding corroborated theoretical results.

References

[1] X. Qin and R. A. Berry, "Distributed approaches for exploiting multiuser diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 392–413, Feb. 2006.

[2] Y. Yu and G. B. Giannakis, "Opportunistic medium access for wireless networking adapted to decentralized csi," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1445–1455, Jun. 2006.

[3] S. Adireddy and L. Tong, "Expoiting decentralized channel state information for random access," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 537–561, Feb. 2005.

[4] K. Bai and J. Zhang, "Opportunistic multichannel aloha: Distributed multiaccess control scheme for ofdma wireless networks," *IEEE Trans. Veh. Technol.*, vol. 55, no. 3, pp. 848–855, May 2006.

[5] Y. Xue, T. Kaiser, and A. B. Gershman, "Channel-aware aloha-based ofdm subcarrier assignment in single-cell wireless communications," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 953–962, May 2007.

[6] X. Qin and R. A. Berry, "Distributed power allocation and scheduling for parallel channel wireless networks," *Journal Wireless Networks*, no. 5, pp. 601–613, Oct. 2008.

[7] M. H. Ngo, V. Krishnamurthy, and L. Tong, "Optimal channel-aware aloha protocol for random access in wlans with multipacket reception and decentralized channel state information," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2575–2588, Jun. 2008.

[8] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad hoc networks with random access: an optimal stopping approach," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 205–222, Jan. 2009.

[9] G. Miao, G. Y. Li, and A. Swami, "Decentralized optimization for multichannel random access," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3012–3023, Oct. 2009.

[10] Y. Al-Harthi and S. Borst, "Distributed adaptive algorithms for optimal opportunistic medium access," in *Proc. WiOpt'09*, Seoul, Korean, Jun. 2009.

[11] N. Salodkar and A. Karandikar, "Random access algorithm with power control and rate guarantees over a fading wireless channel," *IEEE Trans. Wireless Commun.*, 2009, under review. [Online]. Available: http://www.ee.iitb.ac.in/k̃arandi/pubs_dir/preprints/nitin_karandikar_ieeetwc.pdf

[12] A. Ribeiro and G. B. Giannakis, "Separation principles of wireless networking," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4488 – 4505, Sep. 2010.

[13] Y. Hu and A. Ribeiro, "Distributed adaptive algorithms for optimal random access channels," *IEEE Trans. Wireless Commun.*, Feburary 2010, submitted. [Online]. Available: http://www.seas.upenn.edu/ỹichuan/preprints/ara_twc10.pdf

[14] A. Ribeiro, "Ergodic stochastic optimization algorithms for wireless communication and networking," *IEEE Trans. Signal Process.*, Aug. 2009, submitted. [Online]. Available: http://www.seas.upenn.edu/ãribeiro/preprints/eso_main.pdf