Energy-Efficient Power Allocation in Dynamic Multi-Carrier Systems

E. Veronica Belmega
ETIS/ENSEA – UCP – CNRS
Cergy-Pontoise, France
E-mail: belmega@ensea.fr

Panayotis Mertikopoulos
CNRS (French National Center for Scientific Research)
and Univ. Grenoble Alpes, LIG, F-38000, Grenoble, France
E-mail: panayotis.mertikopoulos@imag.fr

***Abstract*—We propose an online power allocation algorithm for optimizing energy efficiency (throughput per unit of transmit power) in multi-user, multi-carrier systems that evolve dynamically over time (e.g. due to changes in the wireless environment or the users' load). Contrary to the static/ergodic regime, a fixed optimal power allocation profile (either static or in the mean) does not exist, so we draw on exponential learning techniques to design an algorithm that is able to adapt to system changes "on the fly". Specifically, the proposed transmit policy leads to *no regret*, i.e., it is asymptotically optimal in hindsight, irrespective of the system's evolution over time. Importantly, despite the algorithm's simplicity and distributed nature, users are able to track their individually optimum transmit profiles as they vary with time, even under rapidly changing network conditions.**

***Index Terms*—energy efficiency, multi-carrier systems, no regret, online optimization.**

I. Introduction

The wildfire spread of Internet-enabled mobile devices and the exponential growth of bandwidth-hungry applications is putting existing wireless networks under enormous strain and is steering the transition to fifth generation (5G) mobile networks. In this context, one of the most pressing challenges faced by the wireless industry is the development of cost-efficient and environment-friendly communication schemes that ensure fiber-like data rates under a tight energy budget. Undoubtedly, meeting this formidable challenge calls for significant advances in wireless technology and hardware system design, but the inherent limitations in upgrading an ageing wireless infrastructure also heighten the need for distributed resource allocation policies that are provably energy-efficient – i.e. that maximize the users' achieved rate per unit of transmit power.

Prior work on energy efficiency [1–4] has focused on the static regime where there are no changes to the wireless network over time. In this setting, optimizing the users' energy efficiency amounts to solving a static optimization problem (or game) whose control variables depend on the system model under study. However, given the dynamic spectrum landscape of current and emerging wireless systems, flexible next-generation multi-user networks must be capable of "on-the-fly" adaptation to a time-varying environment – often with very limited (and potentially obsolete) information at the device end. Accordingly, we focus here on dynamic multi-carrier networks that evolve in an arbitrary, unpredictable fashion (e.g. due to fading, user mobility or fluctuations in the users' load), and we employ techniques and ideas from online optimization to quantify how well users adapt to changes in the wireless medium.

In this dynamic framework, static solution concepts (such as Nash/Debreu equilibria etc.) are no longer relevant, so we focus on the criterion of *regret minimization*, a seminal concept which was first introduced in game theory by Hannan [5], and which has since given rise to a vigorous literature at the interface of machine learning, optimization, statistics and game theory – for a recent survey, see [6, 7]. In game-theoretic parlance, the notion of regret compares a user's cumulative payoff over a given time horizon to the cumulative payoff that the user would have obtained by employing the *a posteriori* best possible action over the time horizon in question. As such, in the context of energy efficiency, regret minimization corresponds to dynamic transmit policies that are asymptotically optimal in hindsight, irrespective of how the user's environment evolves over time.

In this paper, we focus on multi-user orthogonal frequency-division multiple access (OFDMA) systems that evolve arbitrarily over time (for instance, due to fading, intermittent user connectivity, etc.), and we seek to provide an adaptive power allocation scheme that allows users to optimize their energy efficiency ratio "on the fly", based only on locally available channel state information (CSI). To that end, drawing on fractional programming techniques [4] and the method of exponential learning [8], we derive an online power allocation policy that is:

a) *Distributed:* each user updates his individual power variables based on local CSI.
b) *Asynchronous:* there is no need for a global update timer.
c) *Stateless:* users do not need to know the global state of the network or its topology.

The proposed online power allocation algorithm leads to no regret independently of the system's evolution; moreover, our numerical results show that users track the system's most energy-efficient state under realistic fading conditions and rapidly changing channels.

II. System Model and Problem Formulation

Consider a wireless communication system consisting of several interfering wireless connections $u \in \{1, \ldots, U\}$, each one associated with a transmit-receive pair of users sharing a common set of orthogonal subcarriers $k \in \mathcal{K} = \{1, \ldots, K\}$. Then, if $x_k^u \in \mathbb{C}$ and $y_k^u \in \mathbb{C}$ denote respectively the signals transmitted and received over connection u on subcarrier k, we obtain the familiar signal model:

$$y_k^u = h_k^{uu} x_k^u + \sum\nolimits_{v\neq u} h_k^{vu} x_k^v + z_k^u, \tag{1}$$

where $z_k^u \in \mathbb{C}$ denotes the ambient noise over subcarrier k and $h_k^{vu} \in \mathbb{C}$ denotes the channel transfer coefficient between the transmitter of the v-th connection and the receiver of the u-th connection. In view of the above, given that the received signal y_k^u is affected by the ambient noise and interference due to all other connections on subcarrier k, we will write

$$w_k^u = \sum\nolimits_{v\neq u} h_k^{vu} x_k^v + z_k^u \tag{2}$$

for the multi-user interference-plus-noise (MUI) at the receiver of connection u. Thus, if we focus on a specific connection u and drop the index u for clarity, the signal model (1) can be written in more concise form as:

$$y_k = h_k x_k + w_k. \tag{3}$$

In this context, assuming Gaussian input and noise and single user decoding (SUD) at the receiver (i.e. the MUI from all other users is treated as additive noise), the transmission rate of the focal connection will be:

$$R(\mathbf{p}) = \sum\nolimits_{k\in\mathcal{K}} \log\left(1 + g_k(t) p_k\right), \tag{4}$$

where $p_k = \mathbb{E}[|x_k|^2]$ denotes the focal user's transmit power on subcarrier k, $g_k = |h_k|^2 / \mathbb{E}[|w_k|^2]$ denotes the focal user's effective channel gain over subcarrier k and the expectations are taken over the users' (Gaussian) codebooks. Obviously, under dynamic network conditions, the effective channel gain coefficient g_k will vary with time, both due to changes in the wireless medium (fading, mobility, etc.), but also due to the behavior of the other users in the network. Accordingly, if $g_k(t)$ denotes the effective channel gain of the focal user over subcarrier k at time t, his energy efficiency function will be given by the expression:

$$u(\mathbf{p}; t) = \frac{R(\mathbf{p})}{p_c + \sum_k p_k} = \frac{\sum_k \log(1 + p_k g_k(t))}{p_c + \sum_k p_k}, \tag{5}$$

where $p_c > 0$ represents the circuit power consumption of the user's wireless device.

The above considerations lead to the following online optimization problem for maximizing energy efficiency in dynamic multi-carrier networks:

$$\begin{aligned} &\text{maximize} && u(\mathbf{p}; t), \\ &\text{subject to} && \mathbf{p} \in \mathcal{P}. \end{aligned} \tag{6}$$

where

$$\mathcal{P} \equiv \left\{\mathbf{p} : p_k \geq 0 \text{ and } \sum\nolimits_k p_k \leq P_{\max}\right\}, \tag{7}$$

denotes the prolbem's state space (i.e. the space of admissible radiated power profiles) and $P_{\max}$ is the user's maximum transmit power. Specifically, given that the user has no control over his effective channels $g_k(t)$, the sequence of events that we envision is as follows:

1) At each discrete time instance $t = 1, 2, \ldots$, the user selects a transmit power profile $\mathbf{p}(t) \in \mathcal{P}$.
2) The user's energy efficiency $u(\mathbf{p}(t); t)$ is determined by the state of the network and the behavior of all other users via the effective channel gain coefficients $g_k(t)$.
3) The user selects a new transmit power profile $\mathbf{p}(t+1) \in \mathcal{P}$ at stage $t+1$ seeking to maximize the (*a priori unknown*) objective $u(\mathbf{p}; t+1)$, and the process repeats.

In this dynamic framework, the key challenge lies in the user's inability to predict the effective channels that he will face so as to choose an optimal power profile. As a result, static solution concepts (such as Nash, Debreu, or correlated equilibria) are no longer relevant because, in general, there is no fixed optimum system state to target – either static or in the mean. On that account, given a dynamic transmit policy $\mathbf{p}(t) \in \mathcal{P}$, we will compare the user's cumulative "payoff" $\sum_{t=1}^T u(\mathbf{p}(t); t)$ up to time T to what the user would have obtained if he had chosen best possible transmit profile in hindsight – i.e. had he been able to predict in advance the channel conditions for the entire communication horizon. Following [5, 7], we thus define the user's *regret* as:

$$\mathrm{Reg}(T) = \max_{\mathbf{p}'\in\mathcal{P}} \sum\nolimits_{t=1}^T \left[u(\mathbf{p}'; t) - u(\mathbf{p}(t); t)\right]. \tag{8}$$

and, in what follows, our aim will be to derive a policy that leads to *no regret*, i.e.

$$\limsup\nolimits_{T\to\infty} \mathrm{Reg}(T)/T \leq 0, \tag{9}$$

irrespective of the system's evolution over time.

Intuitively, positive regret implies that the focal user would have gained more in terms of energy efficiency by employing a transmit power profile $\mathbf{p}^*$ which is optimal in hindsight – i.e. that maximizes the user's aggregate utility $\sum_{t=1}^T u(\mathbf{p}(t); t)$ over the time horizon in question. Of course, such a profile is a purely theoretical contraption because it requires perfect forecasting of global state information (for the entire transmission horizon), which is impossible to obtain in an evolving, decentralized environment. Instead, by being "asymptotically optimal in hindsight", an adaptive no-regret policy provides the best possible practical approximation to this theoretical target [6, 7].

III. Energy Efficiency via Exponential Learning

To derive a no-regret policy for the online problem (6), we begin by exploiting the concavity of the Shannon throughput function and the so called Charnes–Cooper transform for recasting the fractional objective of (6) as a concave one [9] – for a similar approach in the static channel regime, see [4]. More precisely, we introduce the auxiliary variables:

$$\lambda = p_c/(p_c + \sum\nolimits_k p_k) \quad \text{and} \quad \mathbf{x} = \lambda\mathbf{p}/p_c, \tag{10}$$

so the problem's state space becomes

$$\mathcal{X} \equiv \left\{(\lambda, \mathbf{x}) : \lambda \geq \lambda_c,\ x_k \geq 0, \text{ and } \lambda + \sum\nolimits_k x_k = 1\right\} \tag{11}$$

where $\lambda_c = p_c/(p_c + P_{\max})$. The focal user's energy efficiency function may then be rewritten as:

$$\tilde{u}(\lambda, \mathbf{x}; t) = (\lambda/p_c) \cdot \sum\nolimits_k \log\left[1 + p_c g_k(t) x_k/\lambda\right], \tag{12}$$

an expression which is jointly concave in λ and $\mathbf{x}$ owing to the concavity of $R(\mathbf{p})$ and the properties of the perspective map [4].

Our dynamic transmit policy is based on the following two steps for updating $\lambda(t)$ and $\mathbf{x}(t)$: *a*) we track the gradient of the concave function $\tilde{u}$ in a dual, unconstrained space; and *b*) we map the resulting iterates back to the problem's state space $\mathcal{X}$ via an exponential regularization step. Specifically, we have:

Step 1. *Track the gradient:*

$$\gamma \leftarrow \gamma + \frac{\partial}{\partial \lambda}\tilde{u}(\lambda, \mathbf{x}; t), \tag{13a}$$

$$y_k \leftarrow y_k + \frac{\partial}{\partial x_k}\tilde{u}(\lambda, \mathbf{x}; t), \tag{13b}$$

where, after a quick calculation:

$$\frac{\partial}{\partial \lambda}\tilde{u}(\lambda, \mathbf{x}; t) = \frac{R(\mathbf{p})}{p_c} - \frac{1}{\lambda}\sum\nolimits_k \frac{g_k(t) x_k}{\lambda + p_c g_k(t) x_k} \tag{14a}$$

$$\frac{\partial}{\partial x_k}\tilde{u}(\lambda, \mathbf{x}; t) = \frac{\lambda g_k(t)}{\lambda + p_c g_k(t) x_k} \tag{14b}$$

Step 2. *Return to the problem's feasible space:*

$$\lambda \leftarrow \lambda_c + (1 - \lambda_c) \cdot e^{\eta\gamma} / (e^{\eta\gamma} + \sum\nolimits_k e^{\eta y_k}), \tag{15a}$$

$$x_k \leftarrow (1 - \lambda_c) \cdot e^{\eta y_k} / (e^{\eta\gamma} + \sum\nolimits_k e^{\eta y_k}), \tag{15b}$$

where $\eta \equiv \eta(t) > 0$ is a learning sensitivity parameter that can be tuned freely by the user.

With this two-step update process at hand, the user can set his transmit power profile to $\mathbf{p}(t) = p_c \mathbf{x}(t)/\lambda(t) \in \mathcal{P}$ (for an algorithmic implementation, see Alg. 1). Moreover, as can be seen by (14), the gradient tracking step (13) can be calculated based only on locally available information at the transmitter, so the proposed online power allocation algorithm has the following desirable properties:

(P1) It is *distributed*: each transmitter updates his own power profile based on local CSI.

(P2) It is *asynchronous*: updates do not require user coordination or a global update timer.

(P3) It is *stateless*: users do not need to know the global state (or the topology) of the network.

In this context, our main theoretical result is as follows:

Theorem 1. *The online power allocation policy induced by Algorithm 1 with $\eta(t) = \eta_0/\sqrt{t}$ leads to no regret; specifically:*

$$\mathrm{Reg}(T)/T = \mathcal{O}\left(T^{-1/2}\log(1 + K)\right). \tag{16}$$

This result relies on two important components of the problem (6): *a*) the concavity of the transformed energy efficiency function (12) (which provides an intuitive explanation of why tracking the gradient (13) is a reasonable policy); and *b*) the properties of the exponential mapping (15) which maps the gradient tracking process back to the problem's (primal) state space $\mathcal{X}$. The proof of this result proceeds in two steps. First, assuming that the updates in Algorithm 1 are performed in continuous time, we show that the proposed policy leads to vanishing regret. Then, the passage from continuous to discrete updates is proved by employing the methods introduced recently in [10]. More details can be found in the Appendix.

Algorithm 1 Exponential learning for energy efficiency.

Parameter: $\eta_0 > 0$.

Initialize: $t \leftarrow 1$; $\gamma \leftarrow 0$; $\mathbf{y} \leftarrow 0$

Repeat

  $t \leftarrow t + 1$;

  { *Pre-transmission phase: set transmit powers* }

  foreach subcarrier $k \in \mathcal{K}$ **do**

    calculate auxiliary variables λ and x_k using (15);

    set transmit power $p_k \leftarrow p_c x_k / \lambda$;

  transmit;

  { *Post-transmission phase: receive feedback* }

  foreach subcarrier $k \in \mathcal{K}$ **do**

    measure $g_k(t)$;

    update score variables γ and y_k using (13);

until transmission ends.

IV. Numerical Simulations

To validate the theoretical analysis of Section III, we conducted extensive numerical simulations over a wide range of design parameters and specifications; in what follows, we present a representative sample of these results.

Our simulations focus on a typical cellular OFDMA network occupying a 10 MHz band divided into 1024 subcarriers around a central frequency $f_c = 2$ GHz. Subcarriers in each cell are allocated to each user in a random fashion, and we focus on $U = 7$ users that are located at different, neighboring cells, and that have been allocated the same set of $K = 8$ subcarriers. We focus on the uplink, so receivers (base-stations) are assumed stationary while transmitters are assumed mobile. Communication occurs over a time-division duplexing (TDD) scheme with frame duration $T_f = 5$ ms and the users' channels were modeled following the well-known Jakes model for Rayleigh fading [11] and the extended pedestrian A (EPA) and extended vehicular A (EVA) models for pedestrian ($v = 5$ km/h) and high-speed vehicular movement ($v = 130$ km/h) [12, 13].

In this dynamic setting, the main challenge for the wireless users is to track the transmit power profile that optimizes their energy efficiency ratio over time. Thus, to evaluate the performance of the proposed algorithm in tracking the instantaneous optimum power profile, we plot in Fig. 1 the users' instantaneous energy efficiency under Algorithm 1 against its (evolving) optimum; for comparison purposes, we also plot the energy efficiency achieved by uniform beamforming. Remarkably, even under rapidly varying channels (with a coherence time of the order of a few ms), the wireless users track their optimum transmit profiles very closely (the tracking time is often within a single transmit frame), and the proposed policy consistently outperforms uniform beamforming by a factor of 100% (and often reaching gains up to 500% or more).

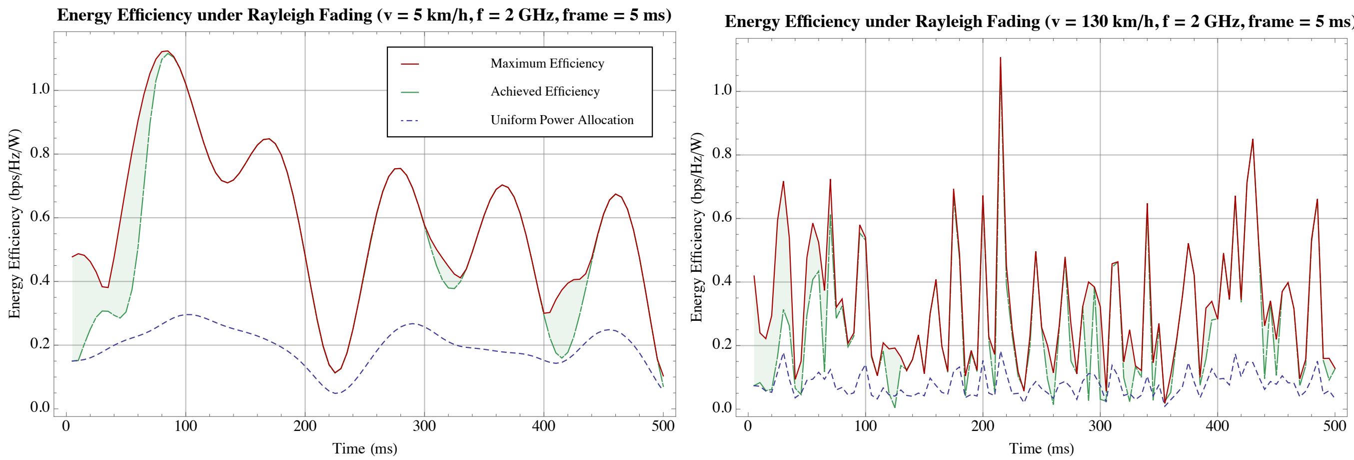

Fig. 1: Achieved energy efficiency (long-dashed green line) under Algorithm 1. We simulated a multi-carrier system with $K = 8$ subcarriers per user operating at a base frequency $f = 2\,\text{GHz}$ (user power characteristics: $P_{\max} = 40\,\text{dBm}$, $p_c = 20\,\text{dBm}$); for comparison purposes, we considered both pedestrian mobility ($v = 5\,\text{km/h}$, Fig. 1(a)) and high-speed vehicular mobility ($v = 130\,\text{km/h}$, Fig. 1(b)). Despite the channel variability, users are able to track the most energy-efficient transmit power profile as it evolves over time (solid red line), gaining a significant advantage over e.g. a uniform beamforming policy (dashed blue line).

V. Conclusions

No regret learning algorithms and online optimization seem very promising and powerful tools to design energy efficient communications in dynamic multi-user networks, in which classical optimization and game theoretic tools are no longer applicable. Indeed, if by the time the users converge to the temporary optimum or Nash equilibrium resource allocation policy, the communication environment has changed (in the network topology, channel conditions, etc.), these classical algorithms become obsolete. For such dynamic networks, we propose an exponential learning policy that leads to no regret, i.e., that is on average as energy efficient as the best power allocation policy in an ideal case in which complete network and channel information is perfectly known in advance (and for the entire communication horizon). Moreover, the dynamic policy we propose is simple, distributed, asynchronous and requires only local network and channel state information. Our numerical results illustrate that users are able to remarkably track their instantaneous energy efficient allocation policies. The properties of the proposed algorithm offers a huge potential in a wide palette of dynamic and decentralized systems: small cell networks, cognitive opportunistic communications, machine-to-machine communications etc.

Appendix
Proof of Theorem 1

We start by proving that the continuous-time analogue of Algorithm 1 leads to no regret.

Theorem 2. *The online power allocation policy* $\mathbf{p}(t) = p_c\mathbf{x}(t)/\lambda(t)$ *induced by the recursion* (13)–(15) *with* $\eta(t) = \eta_0/\sqrt{t}$ *leads to no regret; specifically:*

$$\frac{\text{Reg}(t)}{t} \le \frac{1}{\eta_0\sqrt{t}}\frac{P_{\max}}{P_{\max}+p_c}\left(\log(1+K) + \log\frac{P_{\max}+p_c}{P_{\max}}\right).$$

For the continuous-time dynamics, the regret is defined as:

$$\text{Reg}(t) = \int_0^t \left[u(\mathbf{p}^*; s) - u(\mathbf{p}(s); s)\right]\, ds,$$

where $\mathbf{p}^*$ denotes the optimal fixed policy in hindsight:

$$\mathbf{p}^* = \arg\max_{\mathbf{q}\in\mathcal{P}} \int_0^t u(\mathbf{q}; s)\, ds.$$

First, using the variables introduced in (10) and given the joint concavity of the utility function (12), we can upper-bound the regret as follows:

$$\begin{aligned}
\text{Reg}(t) \le & \int_0^t \partial_\lambda \tilde{u}(\lambda, \mathbf{x}; s)(\lambda^* - \lambda(s))\, ds \\
& + \int_0^t \nabla_{\mathbf{x}} \tilde{u}^T(\lambda, \mathbf{x}; s)(\mathbf{x}^* - \mathbf{x}(s))\, ds \\
& = \int_0^t \dot{\gamma}(s)(\lambda^* - \lambda(s)) + \dot{\mathbf{y}}(s)^T(\mathbf{x}^* - \mathbf{x}(s))\, ds \\
& = \lambda^*\gamma(s) + \mathbf{x}^{*T}\mathbf{y}(s) - \int_0^t \dot{\gamma}(s)\lambda(s) + \dot{\mathbf{y}}(s)^T\mathbf{x}(s)\, ds \qquad (17)
\end{aligned}$$

where the first equality follows from the updating rules (13). By using the chain rule, the above integral term writes as:

$$\begin{aligned}
A \quad \triangleq & \quad \int_0^t \dot{\gamma}(s)\lambda(s) + \dot{\mathbf{y}}(s)^T\mathbf{x}(s)\, ds \\
= & \quad \int_0^t \frac{1}{\eta}\frac{d(\gamma\eta)}{ds}\lambda + \frac{1}{\eta}\frac{d(\eta\mathbf{y})^T}{ds}\mathbf{x}\, ds - \\
& \quad \int_0^t \frac{\dot{\eta}}{\eta}(\gamma\lambda + \mathbf{y}^T\mathbf{x})\, ds
\end{aligned}$$

We define the following function $g : \mathbb{R}^{K+1} \to \mathbb{R}$:

$$g(\eta\gamma, \eta\mathbf{y}) = \lambda_c\eta\gamma + (1-\lambda_c)\log\left(e^{\eta\gamma} + \sum_{i=1}^{K} e^{\eta y_i}\right). \qquad (18)$$

Using the time derivative of this function and the mapping onto the feasible space (15), we can rewrite the expression A as follows:

$$\begin{aligned} A &= \int_0^t \frac{1}{\eta}\frac{dg}{ds}(\eta\gamma, \eta\mathbf{y})\, ds - \int_0^t \frac{\dot\eta}{\eta}(\gamma\lambda + \mathbf{y}^T\mathbf{x})\, ds \\ &= \frac{1}{\eta}g(\eta\gamma,\eta\mathbf{y}) - \frac{1}{\eta(0)}g(\eta(0)\gamma(0), \eta(0)\mathbf{y}(0)) + \\ &\quad \int_0^t \frac{\dot\eta}{\eta^2}\left(g(\eta\gamma,\eta\mathbf{y}) - \eta\gamma\lambda - \eta\mathbf{y}^T\mathbf{x}\right) ds. \end{aligned}$$

By replacing back the term A in (17), the bound on the regret becomes:

$$\begin{aligned} \mathrm{Reg}(t) &\le \lambda^*\gamma(s) + \mathbf{x}^{*T}\mathbf{y}(s) - \frac{1}{\eta}g(\eta\gamma,\eta\mathbf{y}) + \\ &\quad \frac{1}{\eta(0)}g(\eta(0)\gamma(0), \eta(0)\mathbf{y}(0)) - \\ &\quad \int_0^t \frac{\dot\eta}{\eta^2}\left(g(\eta\gamma,\eta\mathbf{y}) - \eta\gamma\lambda - \eta\mathbf{y}^T\mathbf{x}\right) ds \end{aligned} \tag{19}$$

Notice that $\eta(t) \ge 0$ and $\dot\eta(t) \le 0$ which implies $-\dot\eta/\eta \ge 0$.

The last term in (19) can be bounded using the following proposition.

Proposition 1. *The maximum value of the function $h : \mathbb{R}^n \to \mathbb{R}$ defined by*

$$h(x) = \log\left(\sum_{i=1}^n e^{x_i}\right) - \frac{\sum_{i=1}^n x_i e^{x_i}}{\sum_{j=1}^n e^{x_j}} \tag{20}$$

equals $h^* = \log(n)$.

The proof of this result relies on the variable change: $v_j = e^{x_j}/\sum_{i=1}^n e^{x_i},\ \forall j$ and on the fact that maximum entropy is achieved by a uniform probability distribution (which gives the optimal point $v_j^* = 1/n,\ \forall j$).

Replacing λ and $\mathbf{x}$ in (19) by the updates (15) and using Proposition 1, we further obtain:

$$\begin{aligned} \mathrm{Reg}(t) &\le \lambda^*\gamma(s) + \mathbf{x}^{*T}\mathbf{y}(s) - \frac{1}{\eta}g(\eta\lambda,\eta\mathbf{y}) + \\ &\quad \frac{1}{\eta(0)}g(\eta(0)\lambda(0), \eta(0)\mathbf{y}(0)) - \\ &\quad (1-\lambda_c)\int_0^t \frac{\dot\eta}{\eta^2}\log(1+K)\, ds. \end{aligned}$$

Assuming that $\lambda(0) = 0$, $\mathbf{y}(0) = \mathbf{0}$ and $\eta(0) < +\infty$, we obtain:

$$\begin{aligned} \mathrm{Reg}(t) &\le \frac{1-\lambda_c}{\eta}\log(1+K) + \frac{1}{\eta}\Big[(\lambda^*-\lambda_c)\eta\gamma + \eta\mathbf{x}^{*T}\mathbf{y} - \\ &\quad (1-\lambda_c)\log\left(e^{\eta\gamma} + \sum_{i=1}^K e^{\eta y_i}\right)\Big]. \end{aligned}$$

The first term depends only on the parameters of the system. The second term can be upper-bounded using the following variable change:

$$\begin{aligned} \delta &= \lambda_c + (1-\lambda_c)\frac{e^{\eta\gamma}}{e^{\eta\gamma} + \sum_i e^{\eta y_i}} \\ u_k &= (1-\lambda_c)\frac{e^{\eta y_k}}{e^{\eta\gamma} + \sum_i e^{\eta y_i}}, \end{aligned}$$

and is written as

$$\begin{aligned} B &\triangleq (\lambda^*-\lambda_c)\eta\gamma + \eta\mathbf{x}^{*T}\mathbf{y} - (1-\lambda_c)\log\left(e^{\eta\gamma} + \sum_{i=1}^K e^{\eta y_i}\right) \\ &= \log\left(\frac{e^{\eta\gamma} + \sum_{i=1}^K e^{\eta y_i}}{1-\lambda_c}\right)\Big[\lambda^* - \lambda_c + \sum_j x_j^* - (1-\lambda_c)\Big] + \\ &\quad (\lambda^*-\lambda_c)\log(\delta-\lambda_c) \\ &\quad + \sum_j x_j^*\log(u_j) - (1-\lambda_c)\log(1-\lambda_c) \\ &\le -(1-\lambda_c)\log(1-\lambda_c). \end{aligned} \tag{21}$$

The last inequality follows from the fact that $\lambda^* + \sum_j x_j^* = 1$, $\log(\delta-\lambda_c) \le 0$ and $\log(u_j) \le 0$. Theorem 2 follows simply by replacing $\lambda_c = p_c/(p_c + P_{\max})$.

The remaining step to prove Theorem 1 is the passage from continuous to discrete updates which will be detailed in an extended version of this paper and follows via the framework developed in [10].

References

[1] F. Meshkati, H. V. Poor, S. C. Schwartz, and N. B. Mandayam, "An energy-efficient approach to power control and receiver design in wireless data networks," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1885–1894, 2005.

[2] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, 2010.

[3] G. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Distributed interference-aware energy-efficient power optimization," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1323–1333, 2011.

[4] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946–2957, 2012.

[5] J. Hannan, "Approximation to Bayes risk in repeated play," in *Contributions to the Theory of Games, Volume III*, ser. Annals of Mathematics Studies, M. Dresher, A. W. Tucker, and P. Wolfe, Eds. Princeton, NJ: Princeton University Press, 1957, vol. 39, pp. 97–139.

[6] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[7] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[8] P. Mertikopoulos and E. V. Belmega, "Transmit without regrets: Online optimization in mimo-ofdm cognitive radio systems," 2014.

[9] S. Schaible, "Parameter-free convex equivalent and dual programs of fractional programming problems," *Zeitschrift für Operations Research*, vol. 18, pp. 187–196, 1974.

[10] J. Kwon and P. Mertikopoulos, "A continuous-time approach to online optimization," *ArXiv, http://arxiv.org/abs/1401.6956*, pp. 1–24, 2014.

[11] G. Calcev, D. Chizhik, B. Göransson, S. Howard, H. Huang, A. Kogiantis, A. F. Molisch, A. L. Moustakas, D. Reed, and H. Xu, "A wideband spatial channel model for system-wide simulations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, p. 389, March 2007.

[12] "User equipment (UE) radio transmission and reception," 3GPP, Technical Specification 36.101 V12.4.0, Jun. 2014. [Online]. Available: http://www.3gpp.org

[13] "Base station (BS) radio transmission and reception," 3GPP, Technical Specification 36.104 V12.4.0, Jun. 2014. [Online]. Available: http://www.3gpp.org