



HAL
open science

Learning to be green: Robust energy efficiency maximization in dynamic MIMO-OFDM systems

Panayotis Mertikopoulos, Elena Veronica Belmega

► **To cite this version:**

Panayotis Mertikopoulos, Elena Veronica Belmega. Learning to be green: Robust energy efficiency maximization in dynamic MIMO-OFDM systems. *IEEE Journal on Selected Areas in Communications*, 2016, 34 (4), pp.743 - 757. 10.1109/jsac.2016.2544600 . hal-01382276

HAL Id: hal-01382276

<https://hal.science/hal-01382276v1>

Submitted on 4 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning to Be Green: Robust Energy Efficiency Maximization in Dynamic MIMO-OFDM Systems

Panayotis Mertikopoulos, *Member, IEEE*, and E. Veronica Belmega, *Member, IEEE*

Abstract

In this paper, we examine the maximization of energy efficiency (EE) in next-generation multi-user MIMO-OFDM networks that evolve dynamically over time – e.g. due to user mobility, fluctuations in the wireless medium, modulations in the users' load, etc. Contrary to the static/stationary regime, the system may evolve in an arbitrary manner so, targeting a fixed optimum state (either static or in the mean) becomes obsolete; instead, users must adjust to changes in the system “on the fly”, without being able to predict the state of the system in advance. To tackle these issues, we propose a simple and distributed online optimization policy that leads to *no regret*, i.e. it allows users to match (and typically outperform) even the best fixed transmit policy in hindsight, irrespective of how the system varies with time. Moreover, to account for the scarcity of perfect channel state information (CSI) in massive MIMO systems, we also study the algorithm's robustness in the presence of measurement errors and observation noise. Importantly, the proposed policy retains its no-regret properties under very mild assumptions on the error statistics and, on average, it enjoys the same performance guarantees as in the noiseless, deterministic case. Our analysis is supplemented by extensive numerical simulations which show that, in realistic network environments, users track their individually optimum transmit profile even under rapidly changing channel conditions, achieving gains of up to 600% in energy efficiency over uniform power allocation policies.

Index Terms

Energy efficiency; imperfect CSI; MIMO; OFDM; 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 systems under enormous strain and is one of the driving forces behind the transition to fifth generation (5G) mobile networks [1]. In this way, the ICT industry is faced with a formidable mission: data rates must be increased significantly in order to meet the soaring demand for wireless broadband, but this task must be accomplished under an extremely tight energy budget. Thus, to achieve the seamless integration of a diverse set of mobile users, applications and services, current design requirements for 5G systems target a dramatic decrease in energy-per-bit consumption of the order of 1,000× or more [2, 3].

P. Mertikopoulos is with the French National Center for Scientific Research (CNRS) and the Laboratoire d'Informatique de Grenoble (LIG), F-38000, Grenoble, France. E. V. Belmega is with ETIS / ENSEA - UCP - CNRS, Cergy-Pontoise, France.

This research was supported in part by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM# (contract no. 318306), by the French National Research Agency projects NETLEARN (ANR-13-INFR-004) and GAGA (ANR-13-JS01-0004-01) and by ENSEA, Cergy-Pontoise, France. Part of this was presented at VTC2015-Spring.

A contending technology to achieve these design targets is the emerging massive MIMO (multiple-input and multiple-output) paradigm. Coupled with the use of multiple carrier frequencies via orthogonal frequency division multiplexing (OFDM), massive MIMO “goes large” by employing inexpensive service antennas to focus energy into ever smaller regions of space [4–6]. As a result, very large MIMO arrays can greatly enhance the reliability of wireless connections and increase throughput and energy efficiency (EE) by a factor of $10\times$ to $100\times$ without requiring the deployment of expensive new air interfaces [1, 6]. However, due to the massive complexity and variability of such systems, a crucial challenge that arises is that wireless users must also be capable of adapting to a dynamic spectrum landscape “on the fly”, usually with minimal coordination and limited information at the device end.

An added challenge in the above considerations is that wireless users often do not have access to perfect channel state information (CSI) and co-channel interference (CCI) measurements, especially at the transmitter end – for instance, due to pilot contamination in massive MIMO systems [6]. In particular, if the system operates in the presence of uncertainty (imperfect CSI, observation noise, etc.), optimization techniques that rely on a greedy, “one-off” calculation of optimal transmit characteristics (such as water-filling) are no longer suitable because stochastic fluctuations could lead the system to a suboptimal state [7, 8]. On that account, our main objective in this paper will be to provide an adaptive transmit policy for energy efficiency maximization in dynamic MIMO–OFDM networks that are subject to uncertainty, measurement errors and/or other unpredictable changes in the wireless medium.

In the general context of MIMO–OFDM systems, the vast majority of works on energy efficiency maximization and energy-efficient power allocation have focused on two limit cases [9]. In the static regime [10–15], the attributes of the wireless system under study (channel gains, user load, etc.) are assumed effectively static and the system’s analysis revolves around techniques from the theory of non-cooperative games and optimization (continuous or discrete). At the other end of the spectrum, in the ergodic regime [12, 16], the wireless medium is assumed to evolve over a very fast time scale, typically following a sequence of independent and identically distributed (i.i.d.) random variables; consequently, the figure of merit in problems of this type is the stochastic average of the users’ energy efficiency function. All these works study the trade-off between the Shannon achievable rate and power consumption either for a single user (via fractional programming) or multiple ones (using the theory of non-cooperative games). Finally, in the static channel regime, [17–22] consider a throughput model that depends on the connection’s bit error rate (BER) and use tools from game theory to characterize the system’s stable (equilibrium) states.

In this paper, we focus squarely on dynamic MIMO–OFDM systems that evolve *arbitrarily* over time (e.g. due to channel variability, fading, mobility, etc.), and we make no statistical hypotheses regarding the dynamics that govern the network’s evolution (such as stationarity or ergodicity). As opposed to the stationary/ergodic regime discussed above, static solution concepts such as Nash/correlated equilibria are no longer relevant because there is no underlying target state to attain (either static or in the mean); as such, no conclusions can be drawn from the existing literature on energy-efficient power allocation. Instead, users have to optimize their transmit characteristics on the fly, based only on locally available information of the past state of the system, and hoping to track (or at least emulate the performance of) the *a posteriori* optimum transmit policy.

The most widely used optimization criterion in this setting is that of *regret minimization*, a seminal notion which

was first introduced by Hannan [23] and which has since given rise to a vigorous literature at the interface of machine learning, optimization, statistics, and game theory – for a comprehensive survey, see e.g. [24–26]. More precisely, in the language of game theory, a user’s (cumulative) regret over a given time horizon is simply the difference between his average payoff (over the time horizon in question) and the payoff that he would have obtained if he had employed the best possible fixed action in hindsight. Accordingly, in our case, regret minimization corresponds to dynamic transmit policies that are asymptotically optimal in hindsight, irrespective of how the users’ effective wireless medium evolves over time.

A regret-based approach was recently employed by the authors of [27] who studied the problem of power control in infrastructureless wireless networks and proposed an algorithm that minimizes the users’ (internal) regret to attain the system’s equilibrium. In a similar vein, [28] studied the transient phase of the Foschini–Miljanic (FM) power control algorithm in static environments and used the notion of *swap regret* [29] to propose alternative convergent power control schemes; even more recently, [30] showed that the FM dynamics lead to no regret, so they retain their optimality properties in dynamic environments. Finally, [31] and [32] used online optimization techniques and a methodology based on matrix exponential learning [7, 8, 33, 34] to derive a no-regret adaptive transmit policy for power control and throughput maximization in cognitive radio networks respectively. However, the proposed policies drive wireless users to transmit at either full or minimum power (subject to their rate requirements), so they cannot be applied to minimize energy-per-bit consumption in dynamic MIMO–OFDM systems.

Summary of results and paper outline

In this paper, we formulate the maximization of energy efficiency in dynamic MIMO–OFDM systems as an online semidefinite optimization problem and, drawing on Zinkevich’s seminal online gradient ascent (OGA) methodology [35], we propose an adaptive transmit policy which is asymptotically optimal in hindsight – i.e. that leads to no regret. In particular, we show that the proposed algorithm guarantees an $\mathcal{O}(T^{-1/2})$ regret bound after T update epochs (transmission frames), and this bound tightens to $\mathcal{O}(\log T/T)$ if the users’ channel gains always remain above a given level. Furthermore, to address the lack of perfect measurements and channel state information at the transmitter (CSIT), we show that the proposed algorithm retains its optimality properties under very mild statistical hypotheses that are satisfied by the vast majority of error distributions. Specifically, as long as *a*) there is no *systematic* error in the measurement process; and *b*) the probability of observing very large errors (z) is not higher than $\mathcal{O}(1/z^2)$, the proposed policy leads to no regret and enjoys a mean bound of the same order as in the deterministic case.

The performance of the proposed transmit policy is validated by means of extensive numerical simulations modeling a cellular orthogonal frequency-division multiple access (OFDMA) network with multiple base stations and mobile MIMO users with realistic wireless propagation, fading and mobility features. Our results show that the proposed policy represents a scalable and flexible method that allows users to attain very high energy efficiency levels, with gains of up to 600% over uniform/fixed power allocation policies and with surprisingly modest feedback requirements.

Our work here greatly extends our recent conference paper [36] where we derived a continuous-time exponential learning method for energy efficiency maximization in dynamic single-input and single-output (SISO) systems. Com-

pared to [36], the current paper provides a bona fide learning algorithm for multiple-antenna systems, with discrete-time updates and performance guarantees, and with proven robustness in the presence of uncertainty and observation noise.

The rest of our paper is structured as follows: in Section II, we present our wireless system model and we formulate the problem of dynamic energy efficiency maximization as an online semidefinite program. In Section III, we derive our online learning policy, and we establish its no-regret properties and performance guarantees under both perfect and imperfect CSI. Finally, our theoretical analysis is supplemented by extensive numerical simulations in Section IV where we illustrate the gains of the proposed policy under realistic channel gain and mobility conditions.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a wireless network consisting of several point-to-point connections $u \in \mathcal{U} = \{1, \dots, U\}$ (the system's *users*) that are established over a set of orthogonal subcarriers $k \in \mathcal{K} \equiv \{1, \dots, K\}$. Each connection $u \in \mathcal{U}$ represents a pair of communicating wireless multi-antenna devices with M_u antennas at the transmitter and N_u antennas at the receiver. Thus, focusing on the uplink case, if $\mathbf{x}_k^u \in \mathbb{C}^{M_u}$ and $\mathbf{y}_k^u \in \mathbb{C}^{N_u}$ denote the signals transmitted and received over connection u on subcarrier k , we obtain the familiar baseband signal model:

$$\mathbf{y}_k^u = \mathbf{H}_k^{uu} \mathbf{x}_k^u + \sum_{u' \neq u} \mathbf{H}_k^{u'u} \mathbf{x}_k^{u'} + \mathbf{z}_k^u, \quad (1)$$

where $\mathbf{H}_k^{u'u} \in \mathbb{C}^{N_u \times M_{u'}}$ denotes the transfer matrix between the u' -th transmitter and the u -th receiver over subcarrier k while \mathbf{z}_k^u is the ambient noise over the channel (including thermal and atmospheric effects, and modeled as a circularly symmetric Gaussian complex vector). In this way, the multi-user interference-plus-noise (MUI) at the intended receiver of the u -th connection will be:

$$\mathbf{w}_k^u = \sum_{u' \neq u} \mathbf{H}_k^{u'u} \mathbf{x}_k^{u'} + \mathbf{z}_k^u, \quad (2)$$

so (1) may be written more simply as:

$$\mathbf{y}_k^u = \mathbf{H}_k^{uu} \mathbf{x}_k^u + \mathbf{w}_k^u. \quad (3)$$

In the rest of this paper, we will focus on a specific connection $u \in \mathcal{U}$ and we will treat the MUI vector \mathbf{w}_k as an aggregate noise variable whose covariance depends on the wireless medium and the transmit characteristics of all other users. As such, if we drop the user index u for notational convenience, the signal model (3) attains the more compact form:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k. \quad (4)$$

Hence, assuming Gaussian input and single user decoding (SUD) at the receiver, the Shannon rate at the focal connection will be given by the well-known expression [37]:¹

$$R(\mathbf{Q}) = \sum_{k \in \mathcal{K}} \left[\log \det \left(\mathbf{W}_k + \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^\dagger \right) - \log \det \mathbf{W}_k \right], \quad (5)$$

¹For the sake of simplicity, constant multiplicative factors such as the bandwidth of the connection have been dropped in (5); these factors are reinstated in the numerical analysis of Section IV.

where:

- 1) $\mathbf{Q}_k = \mathbb{E}[\mathbf{x}_k \mathbf{x}_k^\dagger] \in \mathbb{C}^{M \times M}$ is the user's input signal covariance matrix over subcarrier k .²
- 2) $\mathbf{Q} = \text{diag}(\mathbf{Q}_1, \dots, \mathbf{Q}_K)$ is the power profile of the focal user over all subcarriers.
- 3) $\mathbf{W}_k = \mathbb{E}[\mathbf{w}_k \mathbf{w}_k^\dagger] \in \mathbb{C}^{N \times N}$ is the MUI covariance matrix of the co-channel interference plus noise affecting the focal connection (obviously, \mathbf{W}_k depends on all other users in the network).

Remark 1. The Gaussian input and noise assumptions are fairly standard in the literature: in particular, Gaussian noise is known to be the worst additive noise distribution with respect to the Shannon achievable rate [38] while Gaussian input is optimal against a Gaussian environment [37]. Finally, regarding the decoding technique, SUD has the advantage of being simple, distributed, and scalable as it does not require any coordination or signaling among the interfering users.

In view of the above, if we let

$$\tilde{\mathbf{H}}_k = \mathbf{W}_k^{-1/2} \mathbf{H}_k \quad (6)$$

denote the *effective channel matrix* of the focal user over subcarrier k , the user's Shannon rate (7) can be written more concisely as:

$$R(\mathbf{Q}) = \sum_{k \in \mathcal{K}} \log \det (\mathbf{I} + \tilde{\mathbf{H}}_k \mathbf{Q}_k \tilde{\mathbf{H}}_k^\dagger) = \log \det (\mathbf{I} + \tilde{\mathbf{H}} \mathbf{Q} \tilde{\mathbf{H}}^\dagger), \quad (7)$$

where $\tilde{\mathbf{H}} = \text{diag}(\tilde{\mathbf{H}}_1, \dots, \tilde{\mathbf{H}}_K)$ is the block-diagonal sum of the user's effective channel matrices over all subcarriers. Thus, following [11–14], the user's energy efficiency function is defined as his Shannon rate per unit of consumed power, i.e.

$$\text{EE}(\mathbf{Q}) = \frac{R(\mathbf{Q})}{P_c + \text{tr}(\mathbf{Q})} = \frac{\log \det (\mathbf{I} + \tilde{\mathbf{H}} \mathbf{Q} \tilde{\mathbf{H}}^\dagger)}{P_c + \text{tr}(\mathbf{Q})}, \quad (8)$$

where $\text{tr}(\mathbf{Q}) = \sum_k \text{tr}(\mathbf{Q}_k)$ is the user's total transmit power while P_c denotes the total power dissipated in all other circuit components of the transmitting device (mixer, frequency synthesizer, digital-to-analog converter, etc.). This efficiency function (which, formally, has units of bits/Joule) has been widely studied in the literature [12, 19, 39] and it captures the fundamental trade-off between higher spectral efficiency and increased battery life. Consequently, in the context of power-limited, energy-aware users, we obtain the maximization problem:

$$\begin{aligned} & \text{maximize} && \text{EE}(\mathbf{Q}), \\ & \text{subject to} && \mathbf{Q} \in \mathcal{Q}, \end{aligned} \quad (9)$$

where

$$\mathcal{Q} = \{\text{diag}(\mathbf{Q}_1, \dots, \mathbf{Q}_K) : \mathbf{Q}_k \succeq 0, \sum_k \text{tr}(\mathbf{Q}_k) \leq P_{\max}\}, \quad (10)$$

and P_{\max} denotes the user's maximum transmit power.

Of course, the user's energy efficiency function depends not only on the transmitter's signal covariance profile \mathbf{Q} , but also on the transmit characteristics of all other users via the effective channel matrices $\tilde{\mathbf{H}}_k$: in particular, $\tilde{\mathbf{H}}$ collects

²In the above, expectations are taken over the users' codebooks (assumed Gaussian).

all sources of noise and interference that cannot be controlled by the focal transmit/receive pair, so the user's energy efficiency objective may vary itself over time in an unpredictable way. On that account, since we wish to focus on dynamic networks that evolve in an arbitrary fashion, we will not be making any specific postulates regarding the behavior of other users in the network and/or the evolution of the user's actual channel matrix \mathbf{H} . The only generic assumptions that we will make regarding the effective channel matrices $\tilde{\mathbf{H}}$ are:

- (A1) $\tilde{\mathbf{H}}$ remains bounded over the entire transmission horizon (e.g. due to the minimum distance between transmitter and receiver, RF circuit losses, antenna directivity, etc.).
- (A2) The variability of $\tilde{\mathbf{H}}$ within each transmission frame is sufficiently slow so that the standard caveats of information theory remain valid.

Consequently, if $\tilde{\mathbf{H}}(t)$ is the user's effective channel matrix at time t , we obtain the following *online* energy efficiency problem:

$$\begin{aligned} & \text{maximize} && \text{EE}(\mathbf{Q}; t), \\ & \text{subject to} && \mathbf{Q} \in \mathcal{Q}, \end{aligned} \tag{OEE}$$

where, in obvious notation:

$$\text{EE}(\mathbf{Q}; t) = \frac{\log \det (\mathbf{I} + \tilde{\mathbf{H}}(t)\mathbf{Q}\tilde{\mathbf{H}}^\dagger(t))}{P_c + \text{tr}(\mathbf{Q})} \tag{11}$$

denotes the user's energy efficiency function at time t . Thus, given that the user cannot predict the state of the system ahead of time, we will focus on the following sequence of events:

- 1) At each update period $n = 1, 2, \dots$, the user selects a transmit power profile $\mathbf{Q}(n) \in \mathcal{Q}$.
- 2) The user's energy efficiency over the current period is determined by the effective channel matrix $\tilde{\mathbf{H}}(n)$ at the time of transmission.
- 3) At the end of the period, the user selects a new signal covariance profile $\mathbf{Q}(n+1)$ seeking to maximize his *a priori unknown* objective function $\text{EE}(\mathbf{Q}; n+1)$, and the process repeats.

Of course, the key challenge in this dynamic framework is that the user does not know ahead of time the effective channel matrix $\tilde{\mathbf{H}}(n+1)$ that determines his energy efficiency function at stage $n+1$, so he must try to adapt to the changing network conditions "on the fly". To be sure, if the user had perfect foresight and knowledge of the evolution of $\tilde{\mathbf{H}}(n)$ in advance, the (fixed) power profile that maximizes the user's average energy efficiency over a given transmission horizon T would be the solution to the (offline) maximization problem:

$$\max_{\mathbf{Q} \in \mathcal{Q}} \frac{1}{T} \sum_{n=1}^T \text{EE}(\mathbf{Q}; n). \tag{12}$$

Obviously however, this "oracle" solution cannot be computed without precognitive abilities, so we will focus on adaptive transmit policies $\mathbf{Q}(n)$ that approach the maximal value of (12) asymptotically, *irrespective of the system's evolution over time*.

To make this analysis precise, we define the user's (cumulative) *regret* at time T as the cumulative difference between the user's achieved EE and the solution of the maximization problem (12), i.e. we let:

$$\text{Reg}(T) = \max_{\mathbf{Q} \in \mathcal{Q}} \sum_{n=1}^T [\text{EE}(\mathbf{Q}; n) - \text{EE}(\mathbf{Q}(n); n)]. \tag{13}$$

We then say that a dynamic transmit policy $\mathbf{Q}(n)$ leads to *no regret* if

$$\limsup_{T \rightarrow \infty} T^{-1} \text{Reg}(T) \leq 0, \quad \text{or, equivalently:} \quad \text{Reg}(T) = o(T), \quad (14)$$

independently of the evolution of the user’s energy efficiency function. In this way, a no-regret policy $\mathbf{Q}(n)$ is *asymptotically optimal in hindsight* in that it provides an asymptotic solution to the average energy efficiency maximization problem (12), without requiring any oracle-like capabilities from the user.

The seminal notion of regret was first introduced in a game-theoretic setting by Hannan [23] and it has since given rise to a vast corpus of research at the interface of optimization, statistics and machine learning – for a recent survey, see e.g. [24, 25]. In particular, if the user’s energy efficiency function does not vary with time (i.e. if the user’s effective channels are static), standard arguments from the theory of online optimization [24] can be used to show that no-regret policies converge to the set $\mathcal{Q}^* = \arg \max_{\mathbf{Q}} \text{EE}(\mathbf{Q})$ of maximally energy-efficient power profiles that solve the (static) problem (9). Likewise, if the user could somehow predict an instantaneous optimum policy $\mathbf{Q}^*(n) \in \arg \max_{\mathbf{Q}} \text{EE}(\mathbf{Q}; n)$ ahead of every stage $n = 1, 2, \dots, T$, we would have $\text{Reg}(T) \leq 0$ for all T ; by this token, the no-regret requirement (14) is a crucial indicator that $\mathbf{Q}(n)$ tracks the optimal solution of (OEE) as it evolves over time. The quality of this tracking can be quantified by more sophisticated regret notions such as adaptive [40] or shifting [41] regret. We focus here on the simpler case of external regret minimization due to space limitations; however, in Section IV, we explore this issue via extensive numerical simulations.

Remark. We should also note here that the no-regret property (14) is a “worst-case” guarantee that carries no assumptions on the evolution of the user’s environment over time: the user’s channels could evolve randomly (following some stationary, ergodic process, as in the case of fast-fading), adversarially (e.g. if the user is subject to jamming), or not at all (in the static regime). As such, in the special case where the wireless medium is affected only by the behavior of other users in the network, a natural question that arises is whether the use of a no-regret policy by all users leads to an equilibrium of the underlying game.³ We address this issue in more detail in Section IV.

III. ONLINE LEARNING

A first idea to achieve no regret in the online energy efficiency maximization problem (OEE) would be to calculate at each stage the power profile that maximizes energy efficiency based on the latest available information at the previous stage. However, as can be seen by a standard online optimization argument, this policy may lead to *positive* regret: for instance, when the user’s channel alternates every other period between two values – say \mathbf{H}_a and \mathbf{H}_b with corresponding optimal power profiles \mathbf{Q}_a^* and \mathbf{Q}_b^* – best-responding to the last observed system state performs strictly worse than the *fixed* policy $(\mathbf{Q}_a^* + \mathbf{Q}_b^*)/2$ [24]. With this in mind, we propose in this section an *adaptive* power allocation policy that utilizes *all* past information in a recursive way based on Zinkevich’s seminal OGA method [35].

For simplicity, we first consider the case where the transmitter has access to perfect CSI and MUI measurements and we derive an anytime $\mathcal{O}(T^{1/2})$ bound for the user’s regret; we then show that this bound can be tightened to

³For instance, it is well-known that internal regret minimization implies convergence to the set of correlated equilibria [42].

$\mathcal{O}(\log T)$ if the user's effective channels always remain above a certain threshold and the algorithm's step-size is chosen accordingly. The robustness of these guarantees in the presence of noise and uncertainty is then discussed in Section III-C.

A. Energy efficiency maximization as an online concave problem

The first difficulty in designing a no-regret policy for the online fractional program (OEE) is that the user's energy efficiency function is not concave. This is perhaps most easily seen in the SISO case where the user's energy efficiency objective becomes:

$$\text{EE}(\mathbf{p}) = \frac{\sum_k \log(1 + \tilde{g}_k p_k)}{P_c + \sum_k p_k}, \quad (15)$$

where $\mathbf{p} = (p_1, \dots, p_K)$ denotes the user's power allocation vector and \tilde{g}_k is the effective channel gain of channel k . Clearly, the fractional objective (15) is not concave with respect to any p_k ; however, $\text{EE}(\mathbf{p})$ can be recast as a concave function by employing the so-called Charnes–Cooper transformation [43] for turning fractional programs into concave ones.⁴ Specifically, if we set

$$x_0 = (P_c + \sum_k p_k)^{-1}, \quad \mathbf{x} = x_0 \cdot \mathbf{p}, \quad (16)$$

we readily obtain $\text{EE}(\mathbf{p}) = x_0 \sum_k \log(1 + \tilde{g}_k x_k / x_0)$, and this last function is concave because the summands $x_0 \log(1 + \tilde{g}_k x_k / x_0)$ are jointly concave in x_0 and x_k . We may then get rid of the parameter x_0 by noticing that $x_0 = \frac{1}{P_c} (1 - \sum_k x_k)$ which restricts the energy efficiency in (x_0, x_k) to an affine set on which it remains concave. Thus, by rewriting \mathbf{x} as $\mathbf{x} = \mathbf{p} / (P_c + \sum_k p_k)$, solving for \mathbf{p} and substituting in $\text{EE}(\mathbf{p})$ to obtain a concave reformulation of (15).

In the general MIMO framework, this procedure amounts to the change of variables:

$$\mathbf{X} = \frac{P_c + P_{\max}}{P_{\max}} \frac{\mathbf{Q}}{P_c + \text{tr}(\mathbf{Q})}, \quad (17)$$

where we have introduced the normalization constant $(P_c + P_{\max})/P_{\max}$ in order to have $\text{tr}(\mathbf{X}) \leq 1$ for all $\mathbf{Q} \in \mathcal{Q}$ (with equality if and only if $\text{tr}(\mathbf{Q}) = P_{\max}$). Solving for \mathbf{Q} then yields

$$\mathbf{Q} = \frac{P_c P_{\max}}{P_c + P_{\max}(1 - \text{tr}(\mathbf{X}))} \mathbf{X}, \quad (18)$$

so, after substituting in (8), we obtain the maximization objective

$$u(\mathbf{X}) = \text{EE}(\mathbf{Q}) = \frac{P_c + P_{\max}(1 - \text{tr}(\mathbf{X}))}{P_c(P_c + P_{\max})} \log \det \left(\mathbf{I} + \frac{P_c P_{\max} \tilde{\mathbf{H}} \mathbf{X} \tilde{\mathbf{H}}^\dagger}{P_c + P_{\max}(1 - \text{tr}(\mathbf{X}))} \right), \quad (19)$$

while the corresponding feasible region of (9) attains the simple form:

$$\mathcal{X} = \{\text{diag}(\mathbf{X}_1, \dots, \mathbf{X}_k) : \mathbf{X}_k \succeq 0 \text{ and } \sum_k \text{tr}(\mathbf{X}_k) \leq 1\}. \quad (20)$$

Given that $R(\mathbf{Q})$ is concave in \mathbf{Q} , the function $F(\mathbf{X}, x) = \frac{P_{\max}}{P_c + P_{\max}} x \cdot R(\mathbf{X}/x)$ will be jointly concave in \mathbf{X} and x [44], so $u(\mathbf{X})$ will also be concave in \mathbf{X} as the restriction of $F(\mathbf{X}, x)$ to the convex set $P_c P_{\max} x = P_c + P_{\max}(1 - \text{tr}(\mathbf{X}))$. In this

⁴See also [12] for a similar use of the Charnes–Cooper transformation in the context of energy efficiency maximization.

way, (OEE) boils down to the online concave maximization problem:

$$\begin{aligned} & \text{maximize} && u(\mathbf{X}; n), \\ & \text{subject to} && \mathbf{X} \in \mathcal{X}, \end{aligned} \quad (21)$$

where, as before, the dependence on $n = 1, 2, \dots$, reflects the evolution of the user's effective channel matrices over time. Thus, in view of all this, we will first derive a no-regret transmit policy $\mathbf{X}(n)$ for the online concave problem (21) and we will then use the inverse transformation (18) to obtain a no-regret policy for (OEE).

B. Learning with perfect CSI

Building on Zinkevich's online gradient ascent method [35], the core idea of our approach will be to track the gradient matrix $\mathbf{V} = \nabla u$ of the user's (time-varying) utility function and then project back to the problem's feasible region when the user's power constraints are violated. To that end, some straightforward matrix calculus yields:

$$\mathbf{V} = \nabla u = \frac{P_{\max}}{P_c + P_{\max}} \left[\mathbf{A} + \frac{\text{tr}(\mathbf{A}\mathbf{Q}) - R(\mathbf{Q})}{P_c} \cdot \mathbf{I} \right], \quad (22)$$

where \mathbf{Q} is calculated in terms of \mathbf{X} via (18) and

$$\mathbf{A} \equiv \nabla R(\mathbf{Q}) = \tilde{\mathbf{H}}^\dagger [\mathbf{I} + \tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger]^{-1} \tilde{\mathbf{H}}. \quad (23)$$

The above expression shows that \mathbf{V} can be calculated at the transmitter as a function of the connection's effective channel matrix $\tilde{\mathbf{H}}$ (which, in turn, can be estimated at the receiver end and then fed back to the transmitter via a dedicated backbone channel or as part of a TDD downlink subframe). Moreover, since \mathbf{V} is a bounded function of $\tilde{\mathbf{H}}$ and the channel matrices $\tilde{\mathbf{H}}(n)$ are assumed bounded, the induced sequence of gradient matrices $\mathbf{V}(n) \equiv \nabla u(\mathbf{X}(n); n)$ will also be bounded. We will therefore assume that there exists a constant $V_0 > 0$ such that

$$\|\mathbf{V}(n)\| \leq V_0 \quad \text{for all } n = 1, 2, \dots, \quad (24)$$

where $\|\mathbf{V}\| = \text{tr}(\mathbf{V}^\dagger \mathbf{V})^{1/2}$ denotes the Frobenius (matrix) norm of \mathbf{V} .

In view of the above, and assuming for the moment perfect knowledge of $\mathbf{V}(n)$ at the transmitter, we will consider the matrix-valued online gradient ascent scheme:

$$\mathbf{X}(n+1) = \mathbf{\Pi}(\mathbf{X}(n) + \gamma_n \mathbf{V}(n)), \quad (\text{OGA})$$

where $\gamma_n > 0$ is a nonincreasing step-size sequence and $\mathbf{\Pi}$ denotes the matrix projection map:

$$\mathbf{\Pi}(\mathbf{Y}) = \arg \min_{\mathbf{X} \in \mathcal{X}} \|\mathbf{X} - \mathbf{Y}\|^2. \quad (25)$$

As we show in Appendix A, the matrix projection $\mathbf{\Pi}(\mathbf{Y})$ can be calculated by the simple expression:

$$\mathbf{\Pi}(\mathbf{Y}) = \mathbf{U} \cdot \text{diag}(\boldsymbol{\pi}(\mathbf{y})) \cdot \mathbf{U}^\dagger, \quad (26)$$

where the tuple (\mathbf{y}, \mathbf{U}) diagonalizes \mathbf{Y} (i.e. $\mathbf{Y} = \mathbf{U} \cdot \text{diag}(\mathbf{y}) \cdot \mathbf{U}^\dagger$) and

$$\pi_i(\mathbf{y}) = \begin{cases} 0 & \text{if } y_i < 0, \\ y_i & \text{if } y_i \geq 0 \text{ and } \sum_j [y_j]_+ < 1, \\ [y_i - \lambda]_+ & \text{if } y_i \geq 0 \text{ and } \sum_j [y_j]_+ \geq 1, \end{cases} \quad (27)$$

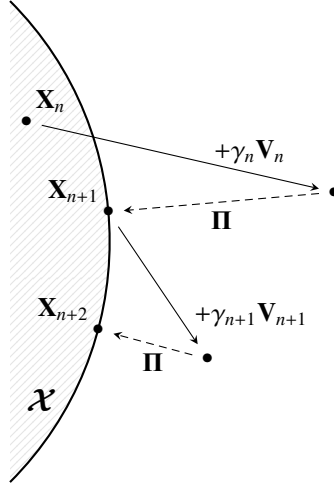


Fig. 1. Schematic representation of the recursive learning scheme (OGA).

with $\lambda > 0$ chosen so that $\sum_{i: y_i \geq 0} [y_i - \lambda]_+ = 1$.⁵

The iterative process (OGA) will be the main focus of our paper, so we proceed with some remarks (see also Fig. 1 for a schematic representation and Alg. 1 for a pseudocode version):

a) Implementation properties: From a practical point of view, (OGA) has the following desirable properties:

(P1) *Distributedness:* users require the same information as in distributed water-filling [45, 46].

(P2) *Statelessness:* users do not need to know the state of the system (e.g. the number of users in the network or its topology).

(P3) *Reinforcement:* users tend to become more energy-efficient based on their past observations.

(P4) *Asynchronicity:* the algorithm does not require a global update timer or any further signaling/coordination between users.

b) Computational complexity: From a computational standpoint (which is crucial in massive MIMO systems), each iteration of Algorithm 1 requires a number of elementary binary operations which is polynomial (with a low degree) on the number of transmit/receive antennas and the number of subcarriers. Specifically, letting $S = \max\{M, N\}$ and recalling that $\tilde{\mathbf{H}}$ and \mathbf{Q} consist of K diagonal blocks, the required matrix multiplication and inversion steps for \mathbf{A} and \mathbf{V} carry a complexity of $\mathcal{O}(KS^\omega)$, with the complexity exponent ω being as low as 2.373 if fast Coppersmith–Winograd multiplication methods are employed [47]. As for the projection step $\mathbf{X} = \mathbf{\Pi}(\mathbf{Y})$, Eqs. (26) and (27) show that it can also be carried out in $\mathcal{O}(KS^\omega)$ operations: the diagonalization in (26) involves $\mathcal{O}(KS^\omega)$ steps while (27) only requires $\mathcal{O}(KM)$ operations for calculating the projection to the simplex [48].

With all this in mind, our main result for (OGA) is as follows:

⁵Recall here that $\mathbf{Y}(n)$ is Hermitian (because $\mathbf{V}(n)$ is Hermitian for all n), so its eigenvalues are real. Just as in water-filling methods [45], the Lagrange multiplier $\lambda > 0$ can then be calculated by sorting \mathbf{y} and performing a line search for λ .

Algorithm 1 Online gradient ascent (OGA) for dynamic energy efficiency maximization.

Parameter: variable step-size sequence $\gamma_n > 0$.

Initialize: $n \leftarrow 0$; $\mathbf{X} \leftarrow \mathbf{0}$.

Repeat

$n \leftarrow n + 1$;

 { Pre-transmission phase: set signal covariance matrix }

$\mathbf{Q} \leftarrow P_c P_{\max} / (P_c + P_{\max}(1 - \text{tr } \mathbf{X})) \cdot \mathbf{X}$;

transmit;

 { Post-transmission phase: receive feedback and update }

 get $\tilde{\mathbf{H}}$;

$\mathbf{A} \leftarrow \tilde{\mathbf{H}}^\dagger [\mathbf{I} + \tilde{\mathbf{H}} \mathbf{Q} \tilde{\mathbf{H}}^\dagger]^{-1} \tilde{\mathbf{H}}$;

$\mathbf{V} \leftarrow P_{\max} / (P_c + P_{\max}) (\mathbf{A} + [\text{tr}(\mathbf{A} \mathbf{Q}) - R(\mathbf{Q})] / P_c \cdot \mathbf{I})$;

$\mathbf{X} \leftarrow \Pi(\mathbf{X} + \gamma_n \mathbf{V})$;

until transmission ends.

Theorem 1. Assume that (OGA) is run with a variable step-size γ_n such that $\gamma_n \rightarrow 0$ and $n\gamma_n \rightarrow \infty$. Then, the induced transmit policy $\mathbf{Q}(n)$ leads to no regret in the online energy efficiency maximization problem (OEE); specifically, (OGA) enjoys the cumulative regret bound:

$$\text{Reg}(T) \leq \frac{1}{\gamma_T} + \frac{1}{2} V_0^2 \sum_{n=1}^T \gamma_n, \quad (28)$$

or, using a step-size sequence of the form $\gamma_n = \gamma n^{-1/2}$:

$$\text{Reg}(T) \leq \frac{1 + \gamma^2 V_0^2}{\gamma} \sqrt{T}. \quad (29)$$

Proof: See Appendix B. ■

The anytime regret bound (28) will be our core performance guarantee for Algorithm 1, so some remarks are in order:

a) *Fine-tuning γ_n :* Theorem 1 shows that taking $\gamma_n \propto n^{-\alpha}$ for some $\alpha \in (0, 1)$ leads to a regret guarantee that is $\mathcal{O}(T^\omega)$ with $\omega = \max\{\alpha, 1 - \alpha\}$;⁶ as such, (29) captures the optimal asymptotic behavior of the bound (28) for step-size sequences of the form $\gamma_n = \gamma/n^\alpha$. In fact, if V_0 can be estimated by the transmitter beforehand, the step-size parameter γ can be fine-tuned further in order to minimize the coefficient of $T^{1/2}$ in (29). Doing just that gives $\gamma = 1/V_0$ and provides the optimized bound:

$$\text{Reg}(T) \leq 2V_0 \sqrt{T}. \quad (30)$$

Since V_0 is a bound on the Frobenius norm of the block-diagonal gradient matrices $\mathbf{V}(n)$, the guarantee (30) becomes $\mathcal{O}(KM^2)$ so the algorithm's overall regret will be at most linear in the number of subcarriers and quadratic in the

⁶To see this, simply note that $\sum_{n=1}^T n^{-\alpha} = \mathcal{O}(T^{1-\alpha})$ for large T and $\alpha \in (0, 1)$.

number of antennas. This guarantee is key for massive MIMO systems (where the number of transmit/receive antennas can grow to be quite large) because it provides a worst-case estimate for the system’s equilibration time. That being said, (30) only becomes tight in adversarial environments (e.g. in the presence of jamming); in typical scenarios, the user’s regret usually decays much faster and the system attains a stable, no-regret state within a few iterations, even for large numbers of antennas per user – cf. the detailed discussion in Sec. IV.

b) The static case: If the user’s effective channels remain static over time and $\mathbf{Q}(n)$ is a no-regret policy, a straightforward concavity argument can be used to show that $\max_n \text{EE}(\mathbf{Q}(n))$ converges to the solution of the (static) EE maximization problem (9) [24, 49]. In this way, (OGA) can also be seen as a provably convergent low-cost algorithm for solving (9); furthermore, as we show in what follows, this convergence result continues to hold even in the presence of imperfect CSIT and measurement errors.

c) Initialization: The agnostic initialization $\mathbf{X}(0) = 0$ of Algorithm 1 means that the focal transmitter remains effectively silent during the first transmission frame (recall that $\mathbf{Q} \propto \mathbf{X}$). As such, the first iteration of (OGA) can be seen as a “handshake” that allows the transmitter to estimate his effective wireless medium before starting the bona fide transmission of data frames. If the transmitter begins with a given belief regarding his effective channel conditions, the algorithm can be initialized more aggressively in a manner consistent with the user’s initial expectations (setting for instance $\mathbf{X} = (KM)^{-1}\mathbf{I}$ for uniform power allocation across subcarriers and antennas). In so doing, the regret bound (28) can be tightened further but this only makes a significant difference if the transmission horizon T is very short (in the order of a few frames).

d) Logarithmic regret under fair channel conditions: As stated, Theorem 1 provides a worst-case guarantee which holds without any further caveats on the evolution of the channels from one stage to the next (other than basic information-theoretic hypotheses that allow the receiver to decode the transmitter’s signal). As such, another important question that arises is whether we can achieve stronger performance guarantees under the additional hypothesis that channel conditions do not become too bad.

To quantify this, note first that the Shannon rate function $R(\mathbf{Q}) = \log \det(\mathbf{I} + \widetilde{\mathbf{H}}\mathbf{Q}\widetilde{\mathbf{H}}^\dagger)$ is strongly concave in \mathbf{Q} with a strong concavity constant that is an increasing function of the singular values of $\widetilde{\mathbf{H}}$.⁷ Accordingly, since the user’s energy efficiency function $\text{EE}(\mathbf{Q})$ can be expressed as a perspective transformation $R(\mathbf{Q}) \mapsto xR(\mathbf{X}/x)$, the same will also hold for the strong concavity constant of $u(\mathbf{X})$ over \mathcal{X} [33, 44]. On that account, if we assume that:

$$\text{Hess}(u(\mathbf{X}; n)) \leq -a\mathbf{I} \quad \text{for some } a > 0 \text{ and for all } n = 1, 2, \dots, \mathbf{X} \in \mathcal{X}, \quad (31)$$

we obtain the following stronger result:

Proposition 1. *Assume that (OGA) is run with the step-size sequence $\gamma_n = \gamma/n$ for some $\gamma \geq a^{-1}$ with a as in (31). Then, the induced transmit policy $\mathbf{Q}(n)$ enjoys the logarithmic regret bound:*

$$\text{Reg}(T) \leq \frac{1}{2}\gamma V_0^2(1 + \log T). \quad (32)$$

⁷Recall here that a function f is strongly concave with constant $c > 0$ if $\text{Hess}(f) \preceq -c\mathbf{I}$.

Proof: See Appendix B. ■

Proposition 1 provides us with an important rule of thumb for choosing the step-size sequence of Algorithm 1. On the one hand, if the user expects that his effective channel can become arbitrarily bad (e.g. due to network congestion or deep fading events), the OGA algorithm should be run with a $n^{-1/2}$ step-size sequence that allows higher adaptability to strongly varying channel conditions. Otherwise, if the user expects reasonable channel quality over his transmission horizon, the “softer” step-size choice $\gamma_n \propto n^{-1}$ minimizes the danger of overcompensating for transmit directions that appear suboptimal and allows the user to converge to a no-regret state faster.

C. Learning under uncertainty

A key assumption in our analysis so far is that the transmitter has access to perfect CSI and MUI measurements with which to calculate the gradient matrices $\mathbf{V}(n)$ at each stage; in practice however, factors such as pilot contamination, sparse feedback and imperfect channel sampling could have a deleterious effect on the algorithm’s performance. As such, our goal in this section will be to analyze (OGA) in the presence of uncertainty and measurement errors.

To formalize this, we will assume that, at each stage $n = 1, 2, \dots$, the transmitter observes a noisy estimate $\hat{\mathbf{V}}(n)$ of $\mathbf{V}(n)$ satisfying the statistical hypotheses:

(H1) *Unbiasedness:*

$$\mathbb{E}[\hat{\mathbf{V}}(n) | \mathbf{Q}(n-1)] = 0. \quad (\text{H1})$$

(H2) *Tame error tails:*

$$\mathbb{P}(\|\hat{\mathbf{V}}(n) - \mathbf{V}(n)\| \geq z) \leq B/z^\beta \quad \text{for some } B > 0 \text{ and for some } \beta > 2. \quad (\text{H2})$$

Clearly, both hypotheses are quite mild from a practical point of view. First, the unbiasedness hypothesis (H1) simply amounts to asking that there is no *systematic* error in the user’s measurements. Likewise, Hypothesis (H2) is a bare-bones assumption on the probability of observing very high errors and is satisfied by the vast majority of statistical error distributions (including uniform, Gaussian, log-normal, and all Lévy-type error processes); in particular, we will *not* be assuming that the error process $\mathbf{Z}(n) = \hat{\mathbf{V}}(n) - \mathbf{V}(n)$ is i.i.d., state-independent, or even a.s. bounded.

Remarkably, under these minimal hypotheses, we have:

Theorem 2. *Assume that (OGA) is run with noisy measurements $\hat{\mathbf{V}}(n)$ satisfying Hypotheses (H1) and (H2), and with a variable step-size sequence of the form $\gamma_n = \gamma/n^\alpha$ for some $\alpha \in (2/\beta, 1)$. Then, the induced transmit policy $\mathbf{Q}(n)$ leads to no regret (a.s.) and enjoys the mean regret bound:*

$$\mathbb{E}[\text{Reg}(T)] \leq \frac{1}{\gamma_T} + \frac{\hat{V}_0^2}{2} \sum_{n=1}^T \gamma_n, \quad (33)$$

where $\hat{V}_0^2 = \sup_n \mathbb{E}[\|\hat{\mathbf{V}}(n)\|^2]$.

Theorem 2 (proven in Appendix C) will be our main result in the context of dynamic energy efficiency maximization under imperfect CSI, so a few remarks are in order:

a) *Step-sizes vs. large error probabilities:* The requirement $\alpha\beta > 2$ of Theorem 2 indicates a trade-off between the probability of observing very large errors and achieving low regret (33). Specifically, if the error distribution of $\mathbf{Z}(n) = \hat{\mathbf{V}}(n) - \mathbf{V}(n)$ has very heavy tails (i.e. (H2) does not hold for $\beta \gg 2$), Algorithm 1 must be bootstrapped with a conservative step-size sequence $\gamma_n \propto 1/n^\alpha$ for some $\alpha \approx 1$; in so doing however, the first term of (33) becomes almost linear, so the user might experience relatively high regret on average (due to the high probability of observing very large errors). On the other hand, if the tails of $\hat{\mathbf{V}}(n)$ are lighter (for instance, the standard case of normally distributed errors exhibits exponentially thin tails, so (H2) holds for all β), Algorithm 1 can be employed with a more adaptive step-size sequence that guarantees a lower regret bound.

In particular, if (H2) holds for some $\beta > 4$, (OGA) can be used with a step-size sequence of the form $\gamma_n = \gamma n^{-1/2}$ which achieves the optimal behavior of (33), viz.

$$\mathbb{E}[\text{Reg}(T)] \leq \frac{1 + \gamma^2 \hat{V}_0^2}{\gamma} \sqrt{T}. \quad (34)$$

Thus, if the mean square bound $\hat{V}_0^2 = \sup_n \mathbb{E}[\|\hat{\mathbf{V}}(n)\|^2]$ can be estimated ahead of time,⁸ the step-size sequence γ_n can be optimized further. More precisely, working as in the deterministic case, the coefficient of $T^{1/2}$ in (34) is minimized when $\gamma = 1/\hat{V}_0$, so we obtain the optimized bound:

$$\mathbb{E}[\text{Reg}(T)] \leq 2\hat{V}_0 \sqrt{T}. \quad (35)$$

b) *The estimation process:* The no-regret properties of (OGA) under uncertainty rely on the availability of statistically unbiased measurements $\hat{\mathbf{V}}$ of \mathbf{V} . In turn, given that users have perfect knowledge of their individual transmit covariance matrices, this requirement boils down to constructing an unbiased estimator of the matrix $\mathbf{A} = \tilde{\mathbf{H}}^\dagger (\mathbf{I} + \tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger)^{-1} \tilde{\mathbf{H}}$. In our Gaussian context, this can be accomplished via the statistical sampling process of [7, 8] which provides an unbiased estimator of \mathbf{A} with exponentially decaying error tails (i.e. (H2) holds for all $\beta > 2$). However, due to space limitations we will not address this question in more detail here.

c) *Fair channel conditions and noise:* As before, the regret guarantee (33) can be tightened significantly if the user's effective channel conditions satisfy (31). In that case, running (OGA) with step-sizes $\gamma_n \propto 1/n$, we obtain the following stochastic analogue of Proposition 1:

Proposition 2. *With notation as in Theorem 2, assume that (OGA) is run with noisy measurements and a variable step-size sequence $\gamma_n = \gamma/n$ for some $\gamma \geq a^{-1}$ with a defined as in (31). Then, the induced transmit policy $\mathbf{Q}(n)$ leads to no regret (a.s.) and enjoys the mean guarantee:*

$$\mathbb{E}[\text{Reg}(T)] \leq \frac{1}{2} \gamma \hat{V}_0^2 (1 + \log T). \quad (36)$$

Proof: See Appendix C. ■

As in the perfect CSI case, Proposition 2 provides a rule of thumb for achieving lower regret faster when the user's (effective) wireless medium is not too bad: as long as (31) holds for some $a > 0$, the user can achieve logarithmic regret, even with very noisy measurements.

⁸Note here that \hat{V}_0 is guaranteed to be finite on account of Hypothesis (H2).

TABLE I
WIRELESS NETWORK SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
Number of cells	19 (hexagonal)	Cell radius	1 km
User density	500 users/km ²	Time frame duration	5 ms
Propagation model	COST Hata	BS/MS antenna height	32 m / 1.5 m
Central frequency	2.5 GHz	Total bandwidth	11.2 MHz
OFDM subcarriers	1024	Subcarrier spacing	11 kHz
Spectral noise density (20 °C)	-174 dBm/Hz	User speed	[3, 130] km/h
Maximum transmit power	$P_{\max} = 33$ dBm	Non-radiative power	$P_c = 20$ dBm
Transmit antennas per device	$M = 4$	Receive antennas per link	$N = 8$

IV. NUMERICAL RESULTS

In this section, we assess the practical performance aspects of the OGA algorithm via numerical simulations. For presentational clarity, we only present here a representative subset of these results but our conclusions apply to a wide range of wireless network parameters and specifications.

Our setup is as follows: we consider a cellular OFDMA wireless network occupying a 10 MHz band divided into 1024 subcarriers around a central frequency of $f_c = 2.5$ GHz. Wireless signal propagation is modeled following the well-known COST Hata model [50, 51] and the spectral noise density is taken to be -174 dBm/Hz at 20°C (for a detailed overview of simulation parameters, see Table I). Network coverage is provided by 19 hexagonal cells (each with a radius of 1 km) that form a honeycomb pattern spanning an urban area with wireless user density $\rho = 500$ users/km². To minimize complexity, OFDM subcarriers are allocated to wireless users in each cell following a simple randomized access scheme that assigns different users to disjoint subcarrier sets [52]; as such, the main sources of CCI are connections in neighboring cells that utilize the same subcarriers.

To model this, we focus on a set of $U = 15$ transmitting users that are located in different cells (following a Poisson point process sampling) and share $K = 8$ common subcarriers. Each wireless transmitter is further assumed to have $M = 4$ transmit antennas, a maximum transmit power of $P_{\max} = 40$ dBm and circuit (non-radiative) power consumption of $P_c = 20$ dBm; at the receiver end, we consider $N = 8$ receive antennas per connection and a receiver noise figure of 7 dB. Finally, communication occurs over a time-division duplexing (TDD) scheme with frame duration $T_f = 5$ ms: transmission takes place during the uplink (UL) subframe while the receivers process the received signal and provide feedback during the downlink (DL) subframe; upon reception of the feedback, the users update their power profiles following Alg. 1 and the process repeats.

A. Static channels

For benchmarking purposes, our first simulation scenario addresses the case of stationary users with static channel conditions (so the variability of a user's effective channel matrix is only due to the modulation of the interfering users' transmit characteristics). Each user is assumed to run (OGA) with a variable step-size of the form $\gamma_n \propto 1/\sqrt{n}$ and an agnostic initialization with initial transmit power $P_0 = P_{\max}/2 = 26$ dBm spread evenly across antennas and

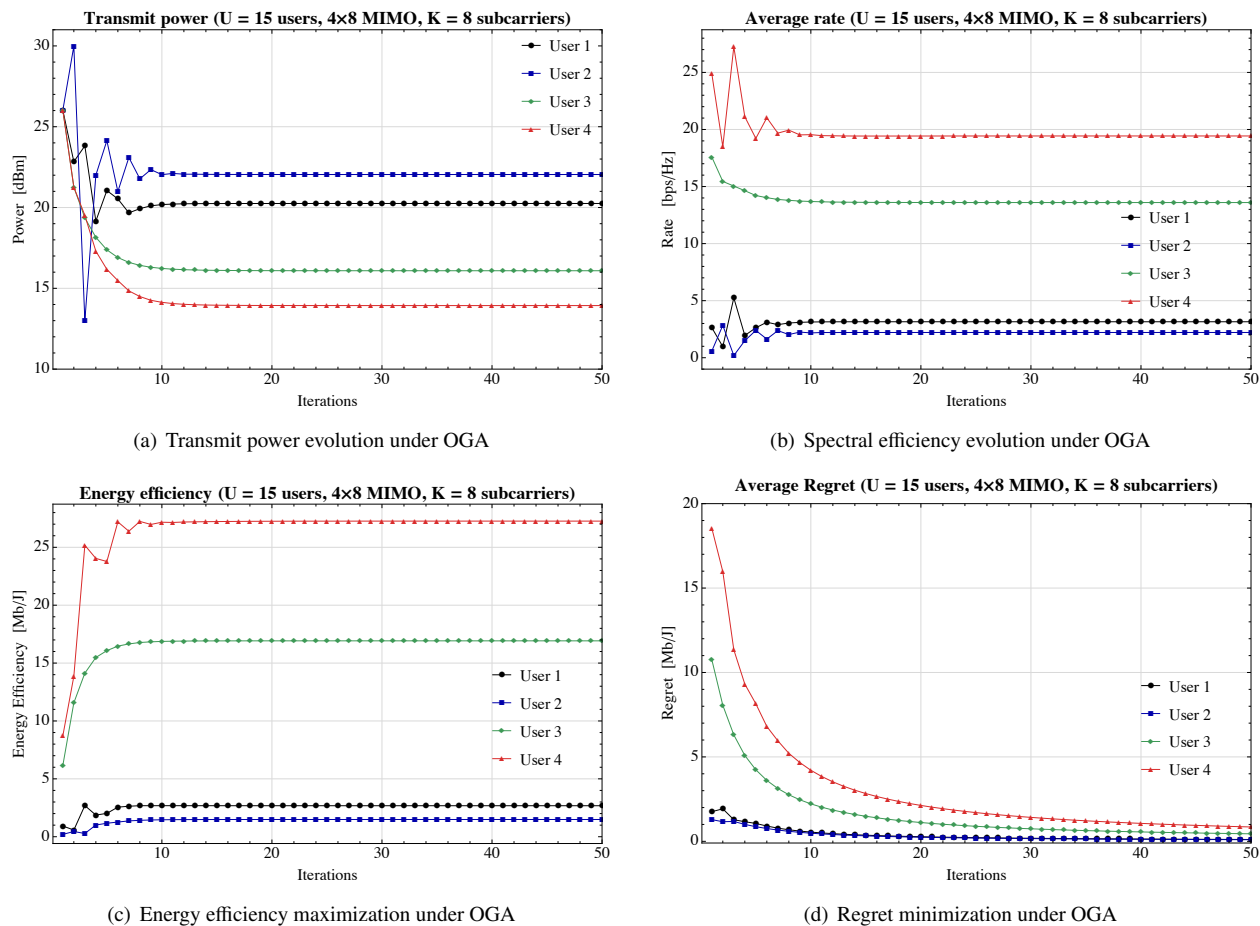


Fig. 2. Performance of the Algorithm 1 under static channel conditions. The system converges within a few iterations to an equilibrium state (Fig. 2(c)) where users experience no regret (Fig. 2(d)).

subcarriers. Our simulation results are presented in Fig. 2 where, to minimize graphical clutter, we only plot the relevant information for 4 users with diverse channel characteristics.

First, in Fig. 2(a), we plot the users' transmit power under (OGA). As can be seen, even though users change their power by several dBm, the algorithm quickly equilibrates after an initial transient phase. Similarly, in Fig. 2(b), we plot the users' transmit rate over all subcarriers (normalized by the bandwidth and thus measured in bps/Hz). We see here that users who reduce power by more than 10 dBm (Users 3 and 4) experience a commensurate drop in spectral efficiency (of the order of a few bps/Hz); on the other hand, users that decrease power only by a little achieve higher rates because the OGA algorithm leads to a more efficient allocation of power over subcarriers and antennas. Nonetheless, in all cases we observe a dramatic increase in energy efficiency over the users' initial (uniform) power allocation policy, ranging from $\approx 200\%$ to more than 600% .⁹

In fact, as we see in Fig. 2(c), after some slight oscillations during the first few iterations (the algorithm's transient

⁹Contrary to Fig. 2(b), we do not normalize the users' energy efficiency by the bandwidth, so it is measured in Mb/J.

phase), the system rapidly equilibrates and reaches a state where users have no incentive to change their individual power profiles (a *Nash equilibrium*). This equilibration is consistent with the no-regret properties of the OGA algorithm: as predicted by Theorem 1 and shown in Fig. 2(d), the users' regret quickly decays to zero even though the algorithm's agnostic – and, in hindsight, suboptimal – initialization leads to high regret in the first few iterations.

B. Time-varying channels and mobility

To account for dynamic network conditions, we also consider in Fig. 3 the case of mobile users whose channels vary with time due to Rayleigh fading, path loss fluctuations, etc. For simulation purposes, we used the extended typical urban (ETU) model for the users' environment and the extended pedestrian A (EPA) and extended vehicular A (EVA) models to simulate pedestrian (3–5 km/h) and vehicular (30–130 km/h) movement respectively [53]; for reference, the focal users' channel gains ($\text{tr}(\mathbf{H}\mathbf{H}^\dagger)$) have been plotted in Fig. 3(a). Despite the channels' variability, Fig. 3(b) shows that the users attain a no-regret state in a few iterations, even under rapidly changing channel conditions (cf. the case of Users 2 and 4 with an average speed of 30 km/h and 130 km/h respectively). For completeness, we also plot in Figs. 3(c) and 3(d) the achieved energy efficiency for a pedestrian and a vehicular user, and we compare it to its instantaneous maximum value, the users' initial (uniform) power allocation policy, and the "oracle" solution which corresponds to the best fixed transmit profile in hindsight (i.e. the solution of the offline maximization problem (12) which posits that users can predict the system's evolution in advance). Remarkably, even under rapidly changing channel conditions, the users' achieved energy efficiency tracks its (evolving) maximum value remarkably well and consistently outperforms even the oracle solution (a fact which is consistent with the negative regret observed in Fig. 3(b)).

An intuitive explanation for the adaptability of OGA is provided by Figs. 3(e) and 3(f) where we plot the transmit power of the optimum policy, the OGA scheme, and the oracle solution for the same users as in Figs. 3(c) and 3(d). Even though the optimum covariance matrix $\mathbf{Q}^*(n)$ may change significantly from one frame to the next, $\text{tr}(\mathbf{Q}^*(n))$ remains roughly constant (within a few dBm) over the entire transmission horizon. The OGA algorithm then learns this power level in a few iterations and stays close to it throughout the transmission horizon; as a result, the users' achieved energy efficiency remains itself very close to its maximum value for all time.

C. Robustness to observation noise and scalability for large antenna numbers

Finally, to assess the robustness of the OGA algorithm in the presence of observation noise and measurement errors, the simulation cycle above was repeated in Fig. 4 for the case where users only have access to noisy gradient observations as in Section III-C. Also, to study the algorithm's scalability in the massive MIMO regime (large number of antennas), we increased the number of transmit antennas to $M = 8$ and the number of receive antennas to $N = 128$; otherwise, for comparison purposes, we used the same network simulation parameters as in Fig. 2.

The intensity of the measurement noise was quantified via the relative error level of the estimator $\hat{\mathbf{V}}$, i.e. its standard deviation over its mean (so a relative error level of $\eta\%$ means that the observed matrix $\hat{\mathbf{V}}$ lies within $\eta\%$ of its true value). We then plotted the users' achieved energy efficiency under the OGA algorithm for noise levels $\eta = 20\%$

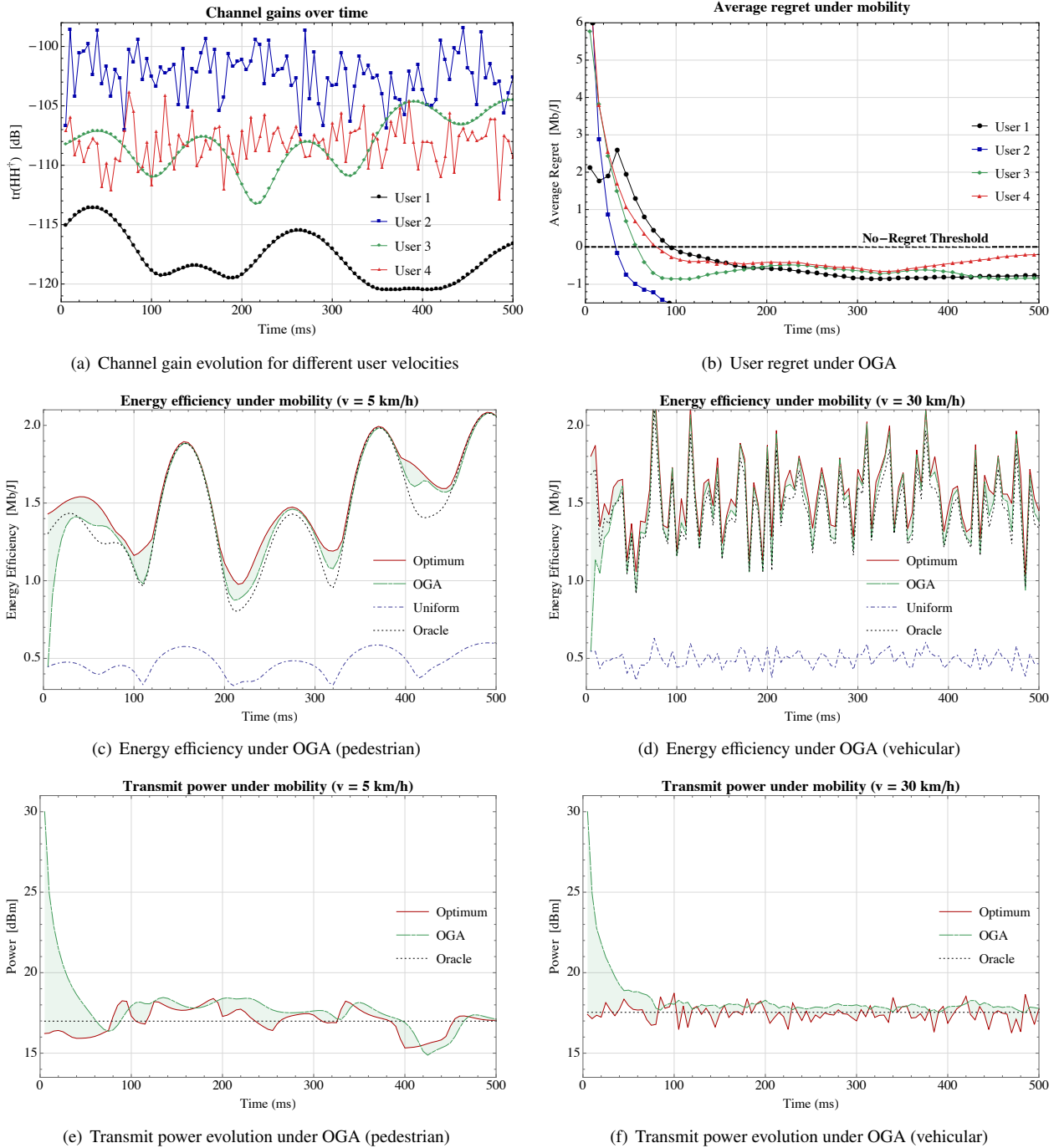


Fig. 3. Performance of the OGA algorithm in a dynamic setting with mobile users moving at $v = \{3, 30, 5, 130\}$ km/h. The users' achieved energy efficiency tracks its (evolving) maximum value remarkably well, even under rapidly changing channel conditions.

(moderate-to-high uncertainty) and $\eta = 100\%$ (very high uncertainty). As can be seen in Fig. 4, the system's rate of equilibration is adversely affected by the intensity of the noise; however, the system still equilibrates within a few tens of iterations and the users exhibit a drastic increase in energy efficiency (of the order of 150% and higher), even in the

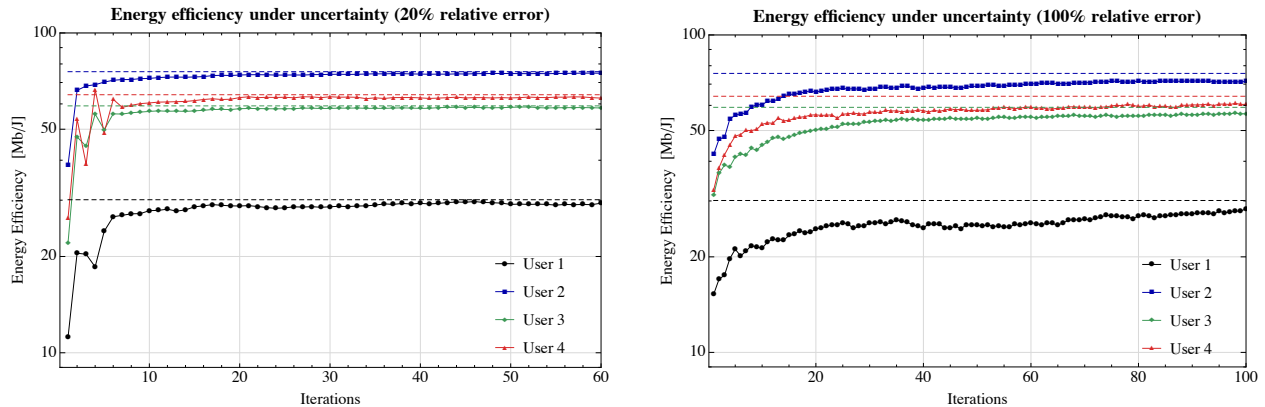


Fig. 4. Performance of Algorithm 1 with imperfect measurements and observation errors. Remarkably, even under very high uncertainty, the system converges within a few tens of iterations to a stable state (dashed lines) where users experience no regret.

presence of very high measurement noise.

V. CONCLUSIONS AND PERSPECTIVES

In the context of multi-carrier, massive MIMO systems where numerous interfering mobile users co-exist, the temporal variability of the system cannot be ignored when targeting highly energy-efficient communications. To tackle these issues, we introduced an online semidefinite optimization framework for the study of energy efficiency maximization in dynamically varying networks, and we proposed an adaptive transmit policy that allows users to attain a “no-regret” state. Importantly, the proposed policy is distributed, asynchronous, computationally simple, and it only requires minimal, strictly causal and (potentially) noisy channel state information at the transmitter. Specifically, under very mild assumptions for the statistics of the error process, we showed that the users’ average regret after T epochs decays as $\mathcal{O}(T^{-1/2})$, a bound which is further improved to $\mathcal{O}(\log T/T)$ if the users do not experience arbitrarily bad channel conditions. As a result, users are able to track their most efficient transmit power profile with modest feedback requirements, even under rapidly changing channel conditions (corresponding to highly mobile users): specifically, our simulations show that users could gain up to 600% in energy efficiency over fixed/uniform power allocation policies in realistic network environments.

An important theoretical question which arises is whether the system converges to an equilibrium state if all users employ a no-regret policy (our numerical simulations show that this indeed the case over a wide region of system parameters). Additionally, different throughput-per-power models accounting for the probability of outage can also be considered and would require a modification of the proposed transmit policy. We intend to explore these directions in future work.

APPENDIX

TECHNICAL PROOFS

Throughout this appendix (and unless explicitly mentioned otherwise), all matrices are assumed to be Hermitian and of dimension $D = KM$. Additionally, the stage variable n will be written as a subscript instead of as a parenthetical

argument – i.e. we will write \mathbf{X}_n and $u_n(\mathbf{X})$ instead of $\mathbf{X}(n)$ and $u(\mathbf{X}; n)$ respectively. We do so in order to reduce the notational clutter caused by an overflow of parentheses; since we will not require a subcarrier index, there is no fear of ambiguity.

A. Matrix projections

We first prove that the projection map $\mathbf{\Pi}(\mathbf{Y})$ is given by the explicit formula (26). To that end, simply note that $\mathbf{\Pi}(\mathbf{Y})$ can be expressed equivalently as the solution to the maximization problem

$$\begin{aligned} & \text{maximize} && \text{tr}(\mathbf{Y}\mathbf{X}) - \frac{1}{2} \|\mathbf{X}\|^2, \\ & \text{subject to} && \mathbf{X} \in \mathcal{X}. \end{aligned} \quad (37)$$

However, if $\mathbf{Y} = \mathbf{U}\mathbf{\Delta}\mathbf{U}^\dagger$ is a diagonalization of \mathbf{Y} , the objective of (37) can be written as:

$$\text{tr}(\mathbf{Y}\mathbf{X}) - \frac{1}{2} \|\mathbf{X}\|^2 = \text{tr}(\mathbf{\Delta}\mathbf{U}^\dagger\mathbf{X}\mathbf{U}) - \frac{1}{2} \text{tr}(\mathbf{U}^\dagger\mathbf{X}\mathbf{U}\mathbf{U}^\dagger\mathbf{X}\mathbf{U}). \quad (38)$$

Thus, given that $\mathbf{X} \in \mathcal{X}$ if and only if $\mathbf{U}\mathbf{X}\mathbf{U}^\dagger \in \mathcal{X}$, we readily get $\mathbf{\Pi}(\mathbf{Y}) = \mathbf{U}\mathbf{\Pi}(\mathbf{\Delta})\mathbf{U}^\dagger$, so it suffices to prove (37) for diagonal \mathbf{Y} .

We first show that $\mathbf{\Pi}(\mathbf{Y})$ is itself diagonal if $\mathbf{Y} = \text{diag}(\mathbf{y})$ for some $\mathbf{y} \in \mathbb{R}^D$. Indeed, we have:

$$\text{tr}(\mathbf{Y}\mathbf{X}) - \frac{1}{2} \|\mathbf{X}\|^2 = \sum_i y_i X_{ii}^2 - \frac{1}{2} \sum_{i,j} |X_{ij}|^2 \leq \sum_i \left(y_i X_{ii}^2 - \frac{1}{2} X_{ii}^2 \right), \quad (39)$$

with equality if and only if \mathbf{X} is diagonal. As a result, if \mathbf{X} is a solution of (37), the diagonal matrix \mathbf{X}' which coincides with \mathbf{X} on the diagonal and has zero entries otherwise will also be a solution of (37); since (37) admits a unique solution, we conclude that $\mathbf{\Pi}(\mathbf{Y})$ must also be diagonal, as claimed. We are thus left to solve the maximization problem

$$\begin{aligned} & \text{maximize} && \sum_j y_j x_j - \frac{1}{2} \sum_j x_j^2, \\ & \text{subject to} && x_j \geq 0, \quad \sum_j x_j \leq 1. \end{aligned} \quad (40)$$

Writing $\lambda_j \geq 0$ and $\lambda \geq 0$ for the Lagrange multipliers of the constraints $x_j \geq 0$ and $\sum_j x_j \leq 1$ respectively, the first-order Karush–Kuhn–Tucker (KKT) conditions for (40) become:

$$y_j = x_j + \lambda - \lambda_j, \quad (41a)$$

$$\lambda_j x_j = 0, \quad \lambda (1 - \sum_j x_j) = 0. \quad (41b)$$

Thus, to obtain the first branch of (27), simply note that if $y_j \leq 0$ but $x_j > 0$, we will also have $\lambda_j = 0$, so (41a) gives $y_j = x_j + \lambda > 0$, a contradiction. Likewise, if $\sum_{j:y_j \geq 0} y_j \leq 1$, setting $\lambda_j = \lambda = 0$ and $x_j = y_j$ for all j such that $y_j \geq 0$ is obviously a solution of (41), so we obtain the second branch of (27). Finally, to obtain the third branch of (27), note first that $\sum_{j:y_j \geq 0} x_j = 1$ if $\sum_{j:y_j \geq 0} y_j \geq 1$; otherwise, we would have $\lambda = 0$ and (41a) would give $y_j = x_j - \lambda_j \leq x_j$ whenever $y_j \geq 0$, implying in turn that $\sum_{j:y_j \geq 0} y_j \leq \sum_{j:y_j \geq 0} x_j < 1$, a contradiction. Accordingly, we are left to project the vector with components $y_i^+ = [y_i]_+$ to the unit simplex $\Delta = \{\mathbf{x} \in \mathbb{R}^D : x_j \geq 0 \text{ and } \sum_j x_j = 1\}$; this projection simply gives $x_i = [y_i^+ - \lambda]_+$ with $\lambda \geq 0$ such that $\sum_i [y_i^+ - \lambda]_+ = 1$ [48], so (26) follows.

B. No regret with perfect CSI

The key step in bounding the user's regret is the inequality:

$$\mathbb{E}E(\mathbf{Q}^*) - \mathbb{E}E(\mathbf{Q}) = u(\mathbf{X}^*) - u(\mathbf{X}) \leq \text{tr}[\mathbf{V} \cdot (\mathbf{X}^* - \mathbf{X})], \quad (42)$$

which is a simple consequence of the fact that $u(\mathbf{X})$ is concave in \mathbf{X} . Our proof follows the methodology of [35] where OGA methods were used in a vector (as opposed to matrix) setting with the special step-size sequence $\gamma_n \propto n^{-1/2}$ (as opposed to general γ_n). To be precise, we will establish the no-regret properties of (OGA) by showing that $\sum_{n=1}^T \text{tr}[\mathbf{V}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)] = o(T)$ for all $\mathbf{X}^* \in \mathcal{X}$ and for every matrix sequence \mathbf{V}_n .

Proof of Theorem 1: Letting $D_n = \frac{1}{2} \|\mathbf{X}^* - \mathbf{X}_n\|^2$, we get:

$$D_{n+1} = \frac{1}{2} \|\mathbf{X}^* - \mathbf{X}_{n+1}\|^2 = \frac{1}{2} \|\mathbf{X}^* - \mathbf{\Pi}(\mathbf{X}_n + \gamma_n \mathbf{V}_n)\|^2 \leq \frac{1}{2} \|\mathbf{X}^* - \mathbf{X}_n - \gamma_n \mathbf{V}_n\|^2, \quad (43)$$

on account of the definition of $\mathbf{\Pi}(\mathbf{Y})$ as the closest point to \mathbf{Y} on \mathcal{X} . In this way, (43) yields:

$$D_{n+1} \leq D_n - \gamma_n \text{tr}[\mathbf{V}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)] + \frac{1}{2} \gamma_n^2 \|\mathbf{V}_n\|^2, \quad (44)$$

and hence, after rearranging and summing over n , we obtain:

$$\begin{aligned} \sum_{n=1}^T \text{tr}[\mathbf{V}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)] &\leq \sum_{n=1}^T \gamma_n^{-1} (D_n - D_{n+1}) + \frac{1}{2} \sum_{n=1}^T \gamma_n \|\mathbf{V}_n\|^2 \\ &\leq \gamma_1^{-1} D_1 + \sum_{n=2}^T (\gamma_n^{-1} - \gamma_{n-1}^{-1}) D_n + \frac{1}{2} \sum_{n=1}^T \gamma_n \|\mathbf{V}_n\|^2 \\ &\leq \gamma_1^{-1} \Omega + \sum_{n=2}^T (\gamma_n^{-1} - \gamma_{n-1}^{-1}) \Omega + \frac{V_0^2}{2} \sum_{n=1}^T \gamma_n = \frac{1}{\gamma_T} + \frac{V_0^2}{2} \sum_{n=1}^T \gamma_n, \end{aligned} \quad (45)$$

where $\Omega \equiv \frac{1}{2} \max_{\mathbf{X}, \mathbf{X}' \in \mathcal{X}} \|\mathbf{X} - \mathbf{X}'\|^2 = 1$. The fact that \mathbf{X}_n leads to no regret then follows by noting that $1/(T\gamma_T) \rightarrow 0$ (by assumption) and that $T^{-1} \sum_{n=1}^T \gamma_n \rightarrow 0$ (since $\gamma_n \rightarrow 0$). ■

Proof of Proposition 1: We note first that the a -strong concavity assumption (31) for u gives:

$$u_n(\mathbf{X}^*) - u_n(\mathbf{X}_n) \leq \text{tr}[\mathbf{V}_n(\mathbf{X}^* - \mathbf{X}_n)] - \frac{1}{2\gamma} \|\mathbf{X}^* - \mathbf{X}_n\|^2, \quad (46)$$

where we have used the fact that $a \geq \gamma^{-1}$. Thus, by summing over n and using (45), we obtain:

$$\begin{aligned} \text{Reg}(T) &\leq \frac{1}{2} \sum_{n=2}^T (\gamma_n^{-1} - \gamma_{n-1}^{-1} - \gamma^{-1}) \|\mathbf{X}^* - \mathbf{X}_n\|^2 + \frac{1}{2} \sum_{n=1}^T \gamma_n \|\mathbf{V}_n\|^2 \\ &\leq \frac{1}{2} \gamma V_0^2 \sum_{n=1}^T n^{-1} \leq \frac{1}{2} \gamma V_0^2 (1 + \log T), \end{aligned} \quad (47)$$

where, in the second line, we used the fact that $\gamma_n^{-1} - \gamma_{n-1}^{-1} = n\gamma^{-1} - (n-1)\gamma^{-1} = \gamma^{-1}$. ■

C. The case of imperfect CSI

Proof of Theorem 2: As before, we begin with the basic inequality:

$$\text{Reg}(T) = \max_{\mathbf{X}^* \in \mathcal{X}} \sum_{n=1}^T [u_n(\mathbf{X}^*) - u_n(\mathbf{X}_n)] \leq \max_{\mathbf{X}^* \in \mathcal{X}} \sum_{n=1}^T \text{tr}[\mathbf{V}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)], \quad (48)$$

with \mathbf{X}_n defined via the (stochastic) recursion:

$$\mathbf{X}_{n+1} = \Pi(\mathbf{X}_n + \gamma_n \hat{\mathbf{V}}_n). \quad (49)$$

Thus, for the first part of the theorem, we need to show that:

$$\sum_{n=1}^T \text{tr}[\hat{\mathbf{V}}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)] + \sum_{n=1}^T \text{tr}[\mathbf{Z}_n \cdot (\mathbf{X}_n - \mathbf{X}^*)] = o(T) \quad (\text{a.s.}), \quad (50)$$

where $\mathbf{Z}_n = \hat{\mathbf{V}}_n - \mathbf{V}_n$. The first term of (50) then becomes:

$$\begin{aligned} \sum_{n=1}^T \text{tr}[\hat{\mathbf{V}}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)] &\leq \gamma_1^{-1} \Omega + \sum_{n=2}^T (\gamma_n^{-1} - \gamma_{n-1}^{-1}) D_n + \frac{1}{2} \sum_{n=1}^T \gamma_n \|\hat{\mathbf{V}}_n\|^2 \\ &\leq \gamma_1^{-1} \Omega + \sum_{n=2}^T (\gamma_n^{-1} - \gamma_{n-1}^{-1}) \Omega + \frac{1}{2} \sum_{n=1}^T \gamma_n (\|\mathbf{V}_n\|^2 + 2 \text{tr}(\mathbf{V}_n \mathbf{Z}_n) + \|\mathbf{Z}_n\|^2) \\ &= \frac{1}{\gamma_T} + \frac{V_0^2}{2} \sum_{n=1}^T \gamma_n + \mathcal{O}\left(\sum_{n=1}^T \gamma_n \|\mathbf{Z}_n\|^2\right), \end{aligned} \quad (51)$$

where, as before, $\Omega \equiv \frac{1}{2} \max_{\mathbf{X}, \mathbf{X}' \in \mathcal{X}} \|\mathbf{X} - \mathbf{X}'\|^2 = 1$. We now claim that $\lim_{T \rightarrow \infty} T^{-1} \sum_{n=1}^T \gamma_n \|\mathbf{Z}_n\|^2 \rightarrow 0$ (a.s.). Indeed, let $z_n = \|\mathbf{Z}_n\|$ and choose $\varepsilon > 0$ such that $4\varepsilon \leq \alpha\beta - 2$ (recall that $\alpha\beta > 2$); Hypothesis (H2) then implies that $\mathbb{P}(z_n \geq n^{\alpha/2 - \varepsilon/\beta}) \leq B/n^{\alpha\beta/2 - \varepsilon} \leq B/n^{1+\varepsilon}$ for all n , so we obtain:

$$\sum_{n=1}^{\infty} \mathbb{P}(z_n \geq n^{\alpha/2 - \varepsilon/\beta}) = \sum_{n=1}^{\infty} \mathcal{O}(1/n^{1+\varepsilon}) < \infty, \quad (52)$$

and hence, by the Borel-Cantelli lemma, we conclude that $\mathbb{P}(z_n \geq n^{\alpha/2 - \varepsilon/\beta} \text{ for infinitely many } n) = 0$. In turn, this implies that $z_n^2 = \mathcal{O}(n^{\alpha - 2\varepsilon/\beta})$ almost surely, so we get:

$$\sum_{n=1}^T \gamma_n \|\mathbf{Z}_n\|^2 = \mathcal{O}\left(\sum_{n=1}^T n^{-\alpha} n^{\alpha - 2\varepsilon/\beta}\right) = \mathcal{O}\left(\sum_{n=1}^T 1/n^{2\varepsilon/\beta}\right) = o(T) \quad (\text{a.s.}). \quad (53)$$

For the second term of (50), let $\xi_n = \text{tr}[\mathbf{Z}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)]$. Then, given that \mathbf{X}_n is a deterministic function of \mathbf{X}_{n-1} and $\hat{\mathbf{V}}_{n-1}$, we will also have $\mathbb{E}[\xi_n | \mathbf{X}_{n-1}] = 0$, i.e. ξ_n is a sequence of martingale differences. By the strong law of large numbers for martingale differences [54, Theorem 2.18], it then follows that $\lim_{T \rightarrow \infty} T^{-1} \sum_{n=1}^T \xi_n = 0$ (a.s.). As a result, combining this with (53), we get

$$\sum_{n=1}^T \text{tr}[\hat{\mathbf{V}}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)] = o(T) \quad (\text{a.s.}), \quad (54)$$

i.e. (49) leads to no regret, as claimed.

Finally, for the mean regret bound (33), taking the expectation of the first line of (51) yields:

$$\mathbb{E}[\text{Reg}(T)] \leq \gamma_1^{-1} \Omega + \sum_{n=2}^T (\gamma_n^{-1} - \gamma_{n-1}^{-1}) D_n + \frac{1}{2} \sum_{n=1}^T \gamma_n \mathbb{E}[\|\hat{\mathbf{V}}_n\|^2] \leq \frac{1}{\gamma_T} + \frac{\hat{V}_0^2}{2} \sum_{n=1}^T \gamma_n, \quad (55)$$

where we used the fact that $\text{Reg}(T) \leq \mathbb{E}[\mathbf{V}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)] = \mathbb{E}[\hat{\mathbf{V}}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)]$ for the LHS, and the assumption that $\mathbb{E}[\|\hat{\mathbf{V}}_n\|^2] \leq \hat{V}_0^2$ for the RHS. ■

Proof of Proposition 2: By reasoning as in the proof of Proposition 1, we readily obtain:

$$\sum_{n=1}^T \text{tr}[\hat{\mathbf{V}}_n \cdot (\mathbf{X}^* - \mathbf{X}_n)] \leq \frac{1}{2} \sum_{n=1}^T \gamma_n \|\hat{\mathbf{V}}_n\|^2, \quad (56)$$

so (36) follows by taking expectations on both sides as in the proof of Theorem 2. That \mathbf{X}_n leads to no regret then follows by noting that (53) holds even for $\alpha = 1$, so the RHS of (56) is $o(T)$. ■

REFERENCES

- [1] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.
- [2] Qualcomm, "The 1000x data challenge," White paper, 2013.
- [3] Huawei Technologies, "5G: A technology vision," White paper, 2013.
- [4] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE Trans. Wireless Commun.*, vol. 31, no. 2, pp. 160–171, February 2013.
- [5] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, pp. 40–60, January 2013.
- [6] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, February 2014.
- [7] P. Couchene, B. Gaujal, and P. Mertikopoulos, "Distributed optimization in multi-user MIMO systems with imperfect and delayed information," in *ISIT '14: Proceedings of the 2014 IEEE International Symposium on Information Theory*, 2014.
- [8] P. Mertikopoulos and A. L. Moustakas, "In an uncertain world: Distributed optimization in MIMO systems with imperfect information," <http://arxiv.org/abs/1502.01653>, 2015.
- [9] D. Feng, C. Jiang, G. Lim, L. J. Cimini Jr., G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 167–178, 2013.
- [10] 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.
- [11] 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.
- [12] 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, August 2012.
- [13] X. Ge, X. Huang, Y. Wang, M. Chen, Q. Li, T. Han, and C.-X. Wang, "Energy efficiency optimization for MIMO-OFDM mobile multimedia communication systems with qos constraints," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2127–2138, June 2014.
- [14] G. Bacci, E. V. Belmega, P. Mertikopoulos, and L. Sanguinetti, "Energy-aware competitive power allocation for heterogeneous networks under QoS constraints," *IEEE Trans. Wireless Commun.*, to appear.
- [15] Z. Wang, I. Stupia, and L. Vandendorpe, "Energy efficient precoder design for MIMO-OFDM with rate-dependent circuit power," 2015, <http://arxiv.org/abs/1503.05070>.
- [16] E. V. Belmega and S. Lasaulce, "Energy-efficient precoding for multiple-antenna terminals," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 329–340, 2011.
- [17] Y. J. A. Zhang and K. B. Letaief, "An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 107–116, 2005.
- [18] 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, November 2005.
- [19] F. Meshkati, M. Chiang, H. V. Poor, and S. C. Schwartz, "A game-theoretic approach to energy-efficient power control in multicarrier CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 1115–1129, 2006.
- [20] E. V. Belmega, S. Lasaulce, and M. Debbah, "A survey on energy-efficient communications," in *PIMRC'10: 21st annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2010, pp. 289–294.
- [21] E. Eraslan, C.-Y. Wang, and B. Daneshrad, "Practical energy-aware link adaptation for MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 246–258, 2014.
- [22] X. Huang, X. Ge, Y. Wang, F. Y. Li, and J. Zhang, "Energy-efficient binary power control with bit error rate constraint in MIMO-OFDM wireless communication systems," in *VTC2012-Fall: Proceedings of the 2012 IEEE Vehicular Technology Conference*, 2012, pp. 1–5.

- [23] 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.
- [24] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.
- [25] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge University Press, 2006.
- [26] E. Hazan, "A survey: The convex optimization approach to regret minimization," in *Optimization for Machine Learning*, S. Sra, S. Nowozin, and S. J. Wright, Eds. MIT Press, 2012, pp. 287–304.
- [27] S. Maghsudi and S. Stanczak, "Joint channel selection and power control in infrastructureless wireless networks: A multi-player multi-armed bandit framework," *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, 2014.
- [28] J. Dams, M. Hoefler, and T. Kesselheim, "Convergence time of power-control dynamics," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2231–2237, December 2012.
- [29] A. Blum and Y. Mansour, "From external to internal regret," *Journal of Machine Learning Research*, vol. 8, pp. 1307–1324, December 2007.
- [30] I. Stiakogiannakis, P. Mertikopoulos, and C. Touati, "No regrets: Distributed power control under time-varying channels and QoS requirements," in *Allerton '14: Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing*, 2014.
- [31] P. Mertikopoulos and E. V. Belmega, "Transmit without regrets: online optimization in MIMO-OFDM cognitive radio systems," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 1987–1999, November 2014.
- [32] I. Stiakogiannakis, P. Mertikopoulos, and C. Touati, "Adaptive power allocation and control in time-varying multi-carrier MIMO networks," 2015, <http://arxiv.org/abs/1503.02155>.
- [33] S. M. Kakade, S. Shalev-Shwartz, and A. Tewari, "Regularization techniques for learning with matrices," *The Journal of Machine Learning Research*, vol. 13, pp. 1865–1890, 2012.
- [34] K. Tsuda, G. Rätsch, and M. K. Warmuth, "Matrix exponentiated gradient updates for on-line Bregman projection," *Journal of Machine Learning Research*, vol. 6, pp. 995–1018, 2005.
- [35] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *ICML '03: Proceedings of the 20th International Conference on Machine Learning*, 2003.
- [36] E. V. Belmega and P. Mertikopoulos, "Energy-efficient power allocation in dynamic multi-carrier systems," in *VTC Spring 2015: Proceedings of the 2015 IEEE Vehicular Technology Conference*, Glasgow, Scotland, May 2015.
- [37] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications and Related Technologies*, vol. 10, no. 6, pp. 585–596, 1999.
- [38] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *Information Theory, IEEE Transactions on*, vol. 47, no. 7, pp. 3072–3081, 2001.
- [39] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, August 2004.
- [40] J. Hofbauer and W. H. Sandholm, "Stable games and their dynamics," *Journal of Economic Theory*, vol. 144, pp. 1710–1725, 2009.
- [41] N. Cesa-Bianchi, P. Gaillard, G. Lugosi, and G. Stoltz, "Mirror descent meets fixed share (and feels no regret)," in *Advances in Neural Information Processing Systems*, 989–997, Ed., vol. 25, 2012.
- [42] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, September 2000.
- [43] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Research Logistics Quarterly*, vol. 9, pp. 181–196, 1962.
- [44] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [45] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, 2004.
- [46] G. Scutari, D. P. Palomar, and S. Barbarossa, "The MIMO iterative waterfilling algorithm," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1917–1935, May 2009.
- [47] A. M. Davie and A. J. Stothers, "Improved bound for complexity of matrix multiplication," *Proceedings of the Royal Society of Edinburgh, Section A: Mathematics*, vol. 143, no. 2, pp. 351–369, 4 2013.
- [48] N. Maculan and G. G. de Paula Jr., "A linear-time median-finding algorithm for projecting a vector on the simplex of \mathbb{R}^n ," *Operations Research*

Letters, vol. 8, pp. 219–222, 1989.

- [49] J. Kwon and P. Mertikopoulos, “A continuous-time approach to online optimization,” 2014, <http://arxiv.org/abs/1401.6956>.
- [50] M. Hata, “Empirical formula for propagation loss in land mobile radio services,” *IEEE Trans. Veh. Technol.*, vol. 29, no. 3, pp. 317–325, August 1980.
- [51] COST Action 231, “Digital mobile radio towards future generation systems,” European Commission, final report, 1999.
- [52] I. Stiakogiannakis, D. Zorbouti, G. Tsoulos, and D. Kaklamani, “Subcarrier allocation algorithms for multicellular OFDMA networks without channel state information,” in *ISWPC 2008: Proceedings of the 3rd International Symposium on Wireless Pervasive Computing*, 2008.
- [53] 3GPP, “User equipment (UE) radio transmission and reception,” White paper, Jun. 2014.
- [54] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*, ser. Probability and Mathematical Statistics. New York: Academic Press, 1980.