Research Article

On the Delay-Energy Tradeoff in Multiuser Fading Channels

Prasanna Chaporkar, Kimmo Kansanen, and Ralf R. Müller

Department of Electronics and Telecommunications, Norwegian University of Science and Technology,
7491 Trondheim, Norway

Correspondence should be addressed to Kimmo Kansanen, kimmo.kansanen@iet.ntnu.no

We consider the delay-energy tradeoff on a fading channel with multiuser diversity. For fixed arbitrary rates of the users, the total transmitted energy is minimized subject to a delay constraint. To achieve this goal we propose a scheme which schedules a subset of all users simultaneously. The scheduled users are allocated power to guarantee successful separation at the detector by successive decoding. In this way, we can benefit from both multiuser diversity and the near-far situation via scheduling and simultaneous transmission, respectively. We analytically show that when the number of users goes to infinity the energy required to guarantee the required user rates can be made as small as required at the cost of a higher delay "delay-energy tradeoff". We explicitly compute the delay under the proposed scheduling policy and discuss how delay differentiation can be achieved. We extend the results to multiband multiaccess channel. Finally, all the results can be generalized in a straightforward fashion to broadcast channel due to the Gaussian multiaccess-broadcast channel duality.

1. Introduction

A comprehensive treatment of multiaccess fading channels can be found in [1, 2]. In these papers, Tse and Hanly have characterized the so-called *throughput capacity* and *delay-limited capacity* of the multiaccess block fading channel with Gaussian noise assuming that perfect channel state information (CSI) is causally available at the transmitters and the receiver. The throughput capacity region quantifies the achievable rate region with average power constraint for ergodic fading. For the delay limited capacity, each user must be given the required rate irrespective of its fading state. The aim is to obtain a coding and power allocation scheme to minimize the energy while guaranteeing the rate in every slot. Here, *slot* refers to the time duration required to transmit a block of symbols over which the fading state remains unaltered. Thus, the slot duration is smaller than the channel coherence time.

The notion of throughput capacity leads to schemes that take advantage of the different channel qualities of the users *"multiuser diversity"*. Specifically, it has been shown that in the special case of single antenna transmission and reception on a frequency nonselective channel, sum rate in the system is maximized by letting only one user with the best channel quality to transmit. Such schemes that take current channel states into account while making scheduling decisions are referred to as *"Opportunistic Scheduling"* and may result in unfair rate allocation if the fading statistics are not symmetric. In wireless systems, the fading statistics are typically not symmetric because of many reasons that include the so-called *"near-far"* effect. To alleviate this limitation, several opportunistic scheduling schemes with fairness constraints have been designed [3, 4]. Different fairness objectives will in general result in different throughput performance of the system, along with changing other performance parameters of a system [5]. Among them, Proportional Fair Scheduling (PFS) is among the most well known and has many desirable properties including provable fairness guarantees and suitability for online implementation, that is, without prior knowledge of channel statistics [6]. In spite of these desirable features, PFS suffers from two limitations. First, PFS does not provide the required rate to the users,

but the rate allocation is done to maximize certain utility function and the allocated rates depend on the channel statistics of the users. Second, it does not guarantee delay as the users are scheduled in random slots, depending on their fading states and the resource allocation in the previous slots.

As discussed before, the notion of throughput capacity is relevant for delay-tolerant data applications. On the contrary, the notion of delay-limited capacity is relevant for the applications that have the strictest delay requirement, namely, the required rate should be given to each user in each slot irrespective of the channel conditions. The delay for such schemes is always one slot. Note that these schemes cannot make use of channel or multiuser diversity over time, but instead can benefit from near-far gain [7] by simultaneous transmission of users. It has been shown that simultaneous transmission (superposition coding in downlink) and successive decoding minimize energy for achieving the required rates [2]. In this signaling scheme, all the users transmit simultaneously with a coordinated power allocation, and the receiver decodes in the decreasing order of the users' channel states treating the undecoded signals as noise. One of the attractive features of this coding strategy is that given the ordered channel state sequence, the power allocation for the optimal signaling can be obtained using a greedy procedure. This has significant impact on practical implementation.

Note that opportunistic scheduling exploits channel and multiuser diversity to guarantee rates that maximize certain utility and/or guarantee fairness, while simultaneous transmission and successive decoding exploit superpositioning to guarantee the required rate and the strict delay of one slot in energy efficient fashion. Many applications cannot afford to sustain indefinite delay variability in attaining the fair rate (as in PFS) without suffering significant performance penalty [8], but they also do not need the strict delay of one slot (as in delay-limited schemes); that is, they have *limited* delay tolerance. With limited delay tolerance in multiaccess channel, it is possible to exploit both multiuser diversity and the superpositioning gain. *Our aim is to understand how delay tolerance of the application can be exploited so as to improve the energy efficiency (delay-energy tradeoff).*

In literature, the delay-energy tradeoff is typically studied assuming no multipath fading (AWGN channel) [9, 10] or assuming noncausal CSI at transmitters and the receiver [11–14]. In an AWGN channel, the problem of minimizing energy with strict deadline requirements has been addressed using filter theory [9] and network calculus [10]. In [11], the authors have studied the delay-energy tradeoff for a single user in fading channel with a strict deadline (say τ slots), while in [12–14] the multiaccess fading channel is considered. In all these works, optimal offline (noncausal CSI available) algorithms that iteratively solve the underlying optimization problem have been developed, and heuristic algorithms for noncausal CSI have been proposed. Because of the equivalence between the multiplexing in time and frequency, the algorithms with noncausal CSI are more relevant when τ orthogonal frequency bands are available and the delay requirement is one slot, that is, in the case of a delay limited multiband multiaccess system. In their seminal work, Cheng and Verdú have obtained the capacity region for the vector Gaussian multiaccess channel [15]. Here, authors have considered the scalar Gaussian channel with ISI, which reduces to the case of independent parallel memoryless Gaussian channels through the Karhunen-Loéve expansion. No multipath fading was assumed. Fundamental results for multiband fading channels were reported in [1, 2]. Perfect CSI at the transmitter and the receiver is assumed in these works. The capacity achieving power allocation is the multiuser waterfilling scheme. The exact characterization of this scheme is considered difficult, and only iterative algorithms and their convergence properties are known [16]. The problem of providing the desired throughput to each of the users has been addressed in multiband multiaccess fading channel with white Gaussian noise [17–19]. The approach taken in these works is one to obtain an approximate multiuser waterfilling solution efficiently.

In practice, CSI is only available causally. In spite of this obvious limitation, the case with noncausal CSI has been studied in literature because of the following two reasons. First, the optimum schemes with noncausal CSI (offline schemes) provide a benchmark for all the other schemes that guarantee the required delay. Also, the structural properties of the optimal offline schemes facilitate valuable insights for designing near-optimal online schemes (schemes with causal CSI). Second, in the fixed delay case, analytically it is difficult, if not impossible, to design optimal online schemes barring the trivial cases in which future fading states can be accurately predicted. This can be seen as follows. Since the required delay τ is finite, any fading realization is possible with positive probability. Thus, for any given online scheme, one can play the role of an adversary and orchestrate future fading states so as to make the scheme suboptimal compared to an optimal offline scheme. In view of these reasons, instead of a fixed delay constraint, an average delay constraint is considered while designing optimal online schemes [20, 21]. From the quality of service point of view, guaranteeing the fixed delay is more desirable than guaranteeing the average delay for the real-time applications. But, if the delay variance is small, then the higher system layers can employ mechanisms like playback buffer to cope up with the delay variability. In [20, 21], a single user fading channel is considered, and the minimization of average delay for the given average power constraint has been addressed using the framework of Markov Decision Processes (MDP). Again, only the structural properties of the optimal online scheme have been derived except in certain special cases. The exact online scheme has to be obtained by solving Bellman's equations, which is computationally expensive. The results in [20, 21] for a single user case are extended to the multiaccess channel by Neely [22]. Here, the author has considered the mean system delay, that is, the average delay over all the users. We note that guaranteeing the mean system delay does not guarantee the required delay performance to individual users. In fact, such schemes tend to favor users with a better channel by providing smaller delays to these at the expense of the users with a worse channel and still maintaining the desired mean system delay. Our aim is to quantify the delay-energy tradeoff while guaranteeing the average delay of each

of the users. In queueing theory literature, frameworks for designing scheduling schemes that minimize a certain utility (energy in our case) while guaranteeing bounded mean delay can be found in [23–26]. These frameworks guarantee the bounded delay but do not guarantee the desired delay to each of the users as we do.

From the previous discussion, it should be clear that the closed form expressions for the minimum energy required for guaranteeing the desired delay are not available for both online and offline schemes. Thus, the energy-delay tradeoff has to be quantified using numerical evaluations for specific scenarios of interest. Our aim is to obtain closed form expressions to quantify the delay-energy tradeoff for an average delay constraint in multiband multiaccess fading channels with causal CSI at the transmitters and the receiver. We consider a multiband system as it can model many practical systems including OFDM systems, channels with frequency selective fading, and ISI channels. The analysis targets specifically systems, where a large number of users must be supported with rate and delay guarantees. A motivating real world example could be a sensor network where sensing delays must be limited, data rates may be pre-allocated, and energy consumption should be minimal.

In our analysis, we allow for a general fading distribution and also consider random arrivals of data from the higher layers into the physical layer buffer to model real-time applications in which the symbols are generated in realtime. For analytical tractability, we use the following approach. We design a parametrized scheduling policy called *Opportunistic Superpositioning (OSP)* that exploits multiuser diversity by scheduling a set of users with high channel gains only, and among these users it uses simultaneous transmission and successive decoding to exploit superpositioning gain. One of the main challenges in designing such schemes is the quantification of performance. The quantification allows for the optimal and guaranteed control. We explicitly quantify the average per user delay in the general case, and the total energy requirement in the large system limit for the proposed policy. Thus, given the delay requirement, we can efficiently choose the appropriate parameter values so as to minimize the energy while guaranteeing the required delay. Using numerical computations, we show that allowing a little delay can yield significant energy savings. We also compare the performance of the proposed policy with PFS and the delay-limited schemes.

The paper is arranged as follows. In Section 2, we present our system model. In Section 3, we describe the OSP policy, and in Section 4 obtain analytical guarantees. In Section 5, we discuss extensions of OSP to multiband multiaccess channel and to provide delay differentiation. In Section 6, we compare the performance of OSP with PFS and delay-limited schemes using numerical computations. Finally, in Section 7, we conclude.

2. System Model

Consider the following system for the purpose of motivation. A sensor network is operating in an area, with a fusion center located in the center. Each sensor is stationary, but the propagation environment has enough inherent variation, that the channels between sensors and the fusion center can be considered slowly randomly varying over time. Each sensor must be polled within a reasonable time from making a measurement, which consists of a random number of bits. On the average, all sensors provide equally valuable data, and should receive an equal fraction of available system rate.

With the previous motivation in mind, we consider a multiaccess system with K users that are placed at random in a cell. Time is slotted. Each user i requires a certain fraction of the total rate provided in a system; that is, the required average rate of each user is $R = \Gamma/K$, where Γ then denotes the total average spectral efficiency of the system required to support the user rates. Alternatively, $R_i(t) = (\Gamma/K)\nu_i(t)$ denotes the arrivals for user i in slot t, where $\nu_i(t)$ is a random variable (r.v.). We assume that all the moments for $\nu_i(t)$ are finite and $\mathbb{E}[\nu_i(t)] = 1$ for every i. Note that random variables with finite support have all the moments finite. A sensor measurement usually has limited resolution due to measurement noise. In networks, the arrival rate is typically limited by the link capacities which may be large but finite. Moreover, we assume that the arrivals are independent and identically distributed (i.i.d.) across both slots and users. The arrivals are queued into an infinite buffer before served. Without loss of generality, let the system start in slot 1. Hence, users' buffers are empty at the beginning of slot 1.

We primarily discuss the uplink communication (multiaccess channel), but since the results consider total system energy expenditure, which can be interpreted as a total power constraint, the results can be generalized in a straightforward fashion to the downlink case (broadcast channel) using the Gaussian multiaccess-broadcast duality [27] in the hard fairness context [7].

Now, we describe the model for the multiaccess channel, described by the input (X) output (Y) relation

$$Y(t) = \sum_{i=1}^{K} \sqrt{d_i(t)} X(t) + N(t). \tag{1}$$

Each user i experiences the channel $d_i(t)$ in slot t. The channel $d_i(t)$ arises as the product of two effects assumed independent, namely, *path loss* (denoted by s_i) and *short-term fading* (denoted by $f_i(t)$), that is, $d_i(t) = s_i f_i(t)$. The path loss is a function of the distance between the transmitter-receiver pair. Typically, the distance between transmitter and receiver changes very slowly with respect to the signal bandwidth. Hence, we assume that the path loss is constant from slot to slot. On the contrary, $f_i(t)$ depends on the scattering environment around the user and changes in time depending on the channel Doppler bandwidth. We assume that $f_i(t)$ changes from slot to slot and is i.i.d. across both users and slots. This is referred to as the block fading model [28]. Let $E_i^R(t)$ ($E_i(t)$, resp.) denote the received (transmitted, resp.) energy from user i in slot t. Then, $E_i^R(t) = d_i(t)E_i(t)$. Note that the channel is given in terms of energy attenuation. Let $\vec{d}(t) = [d_1(t) \ \ldots \ d_K(t)]$. Note that the fading for users is *not* symmetric. The distribution of pathloss for a randomly placed user is denoted by $\Psi(\cdot)$, that is, $\mathbb{P}\{S \leq x\} = \Psi(x)$

where S is a generic r.v. indicating the pathloss of a randomly placed user. Note that $\Psi(\cdot)$ depends on the distribution of the users' placement and the path loss function. Also, let $\Phi(\cdot)$ denote the short-term fading distribution, that is, $\mathbb{P}\{f_i(t) \leq x\} = \Phi(x)$ for every user i and slot t. Let $N_0 = \mathrm{E}[|N(t)|^2]$ denote the noise power spectral density.

Definition 2.1 (scheduling policy). A *scheduling policy* Δ is an algorithm that in each slot t determines the rate vector $\overrightarrow{\rho}^{\Delta}(t) = [\rho_1^{\Delta}(t) \;\; \ldots \;\; \rho_K^{\Delta}(t)]$ and serves each user i with rate $\rho_i^{\Delta}(t)$.

We assume that the perfect CSI is available at the transmitters and the receiver in every slot t. Thus, a scheduling policy may adapt $\overrightarrow{\rho}^{\Delta}(t)$ to the channel state.

Now, we discuss the energy allocation for each user in order to achieve the desired rates $\overrightarrow{\rho}$. Fix an energy allocation given by a vector $\overrightarrow{E} = [E_1 \;\; \cdots \;\; E_K]$, where E_i denotes the energy of the ith user. The capacity region of the Gaussian multiaccess channel with the time invariant fading $\overrightarrow{d}$ for the energy allocation $\overrightarrow{E}$ is the set of rate vectors $\overrightarrow{\rho}$ that satisfy [29]

$$\sum_{i \in \mathcal{S}} \rho_i \leq \log\left(1 + \frac{\sum_{i \in \mathcal{S}} d_i E_i}{N_0}\right), \tag{2}$$

for every $\mathcal{S} \subseteq \{1, \ldots, K\}$. The rates are achieved with the standard random Gaussian codebook with variance E_i for user i. The signaling is as follows. All the users transmit simultaneously and the receiver decodes successively in the decreasing order of the channel states d_i. From this result, it is straightforward to see that a given rate vector $\overrightarrow{\rho}$ is feasible with energy allocation $\overrightarrow{E}$ in Gaussian multiaccess channel with fading $\overrightarrow{d}$ if and only if (iff) there exists a permutation $\pi_1, \ldots, \pi_K$ of $\{1, \ldots, K\}$ such that for every i

$$E_{\pi_i} \geq \frac{N_0}{d_{\pi_i}} \exp\left(\sum_{k < i} \rho_{\pi_k}\right) [\exp(\rho_{\pi_i}) - 1]. \tag{3}$$

Note that there may exist many energy allocations that can realize the rate vector $\overrightarrow{\rho}$. Since we aim to obtain energy efficient strategies, we seek an optimal energy allocation that minimizes the sum energy while achieving the desired rates. Such energy allocation can be explicitly obtained as follows [1]. Let $\overrightarrow{\pi}$ denote the permutation such that $d_{\pi_1} \leq d_{\pi_2} \leq \cdots \leq d_{\pi_K}$. Then, the optimal energy allocation is given by

$$E_{\pi_i} = \frac{N_0}{d_{\pi_i}} \exp\left(\sum_{k < i} \rho_{\pi_k}\right) [\exp(\rho_{\pi_i}) - 1], \tag{4}$$

for every $i \in \{1, \ldots, K\}$. Note that for the optimal signaling, the successive decoding order depends only on channel gains, but not on rates. For convenience, from now on, we consider users to be indexed according to the ordering. We assume that once a scheduling policy Δ determines the rate allocation, the corresponding power allocation is done as per (4).

We consider three performance measures, namely, stability, energy efficiency, and delay. Next, we formally define these.

Definition 2.2 (busy period). A *busy period* for user i is the set of consecutive slots in which its queue length is greater than zero.

Definition 2.3 (stability). Let $\mathcal{B}_{u,i}$ denote the length of the uth busy period of user i. The system is said to be *strongly stable* if for every user i,

$$\limsup_{U \to \infty} \frac{1}{U} \sum_{u=1}^{U} \mathcal{B}_{u,i} < \infty. \tag{5}$$

Strong stability, as defined here, guarantees that each user has a bounded average per-packet delay.

Definition 2.4 (energy efficiency). We define the energy efficiency in slot t as

$$\left(\frac{E_b(t)}{N_0}\right)_{\mathrm{sys}} \stackrel{\mathrm{def}}{=} \frac{\sum_{i=1}^{K} E_i(t)}{N_0 \Gamma}. \tag{6}$$

Then, the energy efficiency is defined as

$$\left(\frac{E_b}{N_0}\right)_{\mathrm{sys}} \stackrel{\mathrm{def}}{=} \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \left(\frac{E_b(t)}{N_0}\right)_{\mathrm{sys}}. \tag{7}$$

The energy efficiency measures the average transmitted energy the system uses per each transmitted bit.

Definition 2.5 (user delay). Delay for the uth arrival of user i (denoted by $D_{u,i}$) is the number of slots between its departure and arrival. The delay for user i is then defined as the limiting average of per-arrival delays

$$D_i \stackrel{\mathrm{def}}{=} \limsup_{U \to \infty} \frac{1}{U} \sum_{u=1}^{U} D_{u,i}. \tag{8}$$

We augment the notation to denote the dependence of various quantities on the scheduling policy Δ by using Δ as a superscript, for example, D_i^{Δ} will indicate the delay for user i under policy Δ.

3. Opportunistic Superpositioning (Δ^*)

The scheduling policy Δ^* is parametrized by a variable κ. To specify this dependence, we denote the OSP scheduling policy as $\Delta^*(\kappa)$. The scheduling decisions are taken as follows in every slot t.

 (i) Select all users i such that $f_i(t) > \kappa$.

 (ii) Allocate rate $\rho_i(t)$ to each of the chosen users i so that everything in its buffer is served, $\rho_i(t) = 0$ for others.

Note that $\Delta^*(\kappa)$ (OSP) selects users based on $f_i(t)$ and not on $d_i(t)$. This is for the reasons of fairness as $f_i(t)$ has the same distribution for every user, while the distribution

of $d_i(t)$ depends on the distance of user i from the base station and is biased towards stronger channels the closer the user is to the base station. Moreover, since the pathloss is constant for each user, it suffices to determine rate allocation in each slot depending only on short-term fading which is time varying, and thus, unlike pathloss, provides diversity which can be exploited.

We refer to κ as *opportunism threshold* as it dictates how opportunistic OSP is in exploiting the channel diversity. The opportunism threshold plays a key role in determining delay and energy consumption in the system. We intuitively explain how. Note that appropriate choice of κ allows for eliminating users that are in deep fade in a given slot. Typically, a few users with a very bad channel dominate the total energy consumption in the system for the delay-limited schemes. Thus, even a small value of the opportunism threshold should significantly improve energy efficiency of the system. Specifically, we can expect the energy consumption to decrease monotonically with an increase in κ. But, note that as κ increases, each user is scheduled less frequently. Thus, for satisfying the rate requirements, the rate given to the scheduled users increases monotonically with κ. And providing the higher rate requires higher energy. In fact, for fixed noise power spectral density, the required energy increases exponentially with an increase in the rate. Summarizing, the opportunism threshold in OSP allows us to save energy by eliminating the users in deep fade but requires higher energy to serve scheduled users. Thus, it is not clear if the opportunism threshold should improve the system performance. It, however, turns out that the energy consumption of OSP decreases monotonically as a function of κ. In other words, the energy savings caused by eliminating the worst users is more than the increase in the energy consumption to provide the higher rates to the scheduled users.

As discussed above, the users are scheduled less frequently by OSP when κ is large. So, clearly, the delay increases monotonically with an increase in κ. Thus, the opportunism threshold provides a way to achieve delay-energy tradeoff. Now, the main challenge is to quantify the delay and energy of OSP as a function of κ, so as to choose the optimal κ that guarantees the required delay while minimizing the energy.

4. Analytical Guarantees

In this section, we obtain analytical guarantees for $\Delta^*(\kappa)$. In Theorem 4.1, we show that $\Delta^*(\kappa)$ is strongly stable and has a finite average busy time. This results leads to Corollary 4.2, where we quantify the user delay under the proposed scheduler $\Delta^*(\kappa)$. In Theorem 4.5, we quantify the asymptotic energy efficiency of $\Delta^*(\kappa)$. In Theorem 4.6 we show that the asymptotic energy efficiency of $\Delta^*(\kappa)$ decreases monotonically with κ. In Section 4.1, we provide a demonstration of the results using important special cases and provide connections to recent results in the literature. Therein we provide Corollary 4.8, where we show that as delay grows without bound, the energy required to stabilize the system becomes equal to the minimum energy required

to provide the desired rates to each of the users in long term. Finally in Section 4.2, we discuss the implications of the analytical results.

First, let $\gamma \overset{\text{def}}{=} \mathbb{P}\{f_i > \kappa\}$.

Theorem 4.1 (strong stability). *For every κ such that $\gamma > 0$, $\Delta^*(\kappa)$ is strongly stable w.p.1.*

Proof. Fix any user i. Since $\{f_i(t)\}_{t \geq 1}$ is a sequence of i.i.d. random variables, the user is scheduled is each slot w.p. γ under $\Delta^*(\kappa)$. Moreover, each time the user is scheduled, $\Delta^*(\kappa)$ serves everything in its buffer. Also, since the arrivals are i.i.d. in slots, clearly, the busy periods are geometrically distributed i.i.d. random variables with mean $1/\gamma$. Thus, by the Strong Law of Large Numbers (SLLNs) for $\gamma > 0$

$$\limsup_{U \to \infty} \frac{1}{U} \sum_{u=1}^{U} \mathcal{B}_{u,i} = \lim_{U \to \infty} \frac{1}{U} \sum_{u=1}^{U} \mathcal{B}_{u,i} = \frac{1}{\gamma} \quad \text{w.p.1.} \quad (9)$$

Now, the result follows from Definition 2.3 as $\gamma > 0$. □

Corollary 4.2 (user delay). *The delay for any user i under $\Delta^*(\kappa)$ is*

$$D_i^{\Delta^*(\kappa)} = \frac{1}{\gamma} \quad w.p.1. \quad (10)$$

Proof. Because the arrival and transmission of each packet are mutually independent, and these are independent from the arrivals and transmissions of other packets, we can, without loss of generality, consider the delay of each packet separately. Conditioned on the arrival time, the waiting time follows the same geometric distribution as the busy time, and the SLLN proof of Theorem 4.1 applies to user delay directly. □

Note 1. Users' delays do not depend on the distribution of their arrival processes given that the mean is finite. Moreover, the delays for the users are not correlated. Thus, for a given κ, users may have different distributions for their respective arrival processes and yet receive the same delay as long as these processes are independent. Furthermore, user's delay guarantee is independent of the number of users in the system.

Let $\overline{F}_\kappa(\cdot)$ denote the fading distribution of user i who is randomly placed in a cell given that $f_i(t) > \kappa$, that is, $\overline{F}_\kappa(x) = \mathbb{P}\{s_i f_i(t) \leq x \mid f_i(t) > \kappa\}$, where s_i is an r.v. denoting the path loss of a randomly placed user i. Note that $\overline{F}_\kappa(\cdot)$ does not depend on time as the short-term fading is i.i.d. across the slots. Now, let $\mathcal{A}_K^\kappa(x, t)$ denote the set of users that are chosen for service under $\Delta^*(\kappa)$ in slot t such that their channel gains are less than or equal to x, that is, $\mathcal{A}_K^\kappa(x, t) = \{i \in \{1, \ldots, K\} : f_i(t) > \kappa \text{ and } d_i(t) \leq x\}$. Clearly, $\mathcal{A}_K^\kappa(\infty, t)$ denotes the set of all the users chosen by $\Delta^*(\kappa)$ in slot t. Let $|\mathcal{A}_K^\kappa(x, t)|$ denote the cardinality of $\mathcal{A}_K^\kappa(x, t)$.

Next, we obtain the energy efficiency of $\Delta^*(\kappa)$ in the asymptotic case when $K \to \infty$, where the system exhibits a self-averaging behavior, and the required energy to support the traffic demand converges to a deterministic limit.

Intuitively, this is due to the random processes of the asymptotically large system behaving deterministically according to their limiting probabilistic behavior, the distributions. Before we proceed, however, we first state two results that we use to prove the required. The first lemma shows how the sum rate of a set of users chosen by $\Delta^*(\kappa)$, having a channel state not larger than x, converges asymptotically to a limit. The second lemma shows how the rate of any arbitrary user selected by $\Delta^*(\kappa)$ asymptotically vanishes.

Lemma 4.3. *For every* t,

$$\lim_{K \to \infty} \sup_x \left| \sum_{k \in \mathcal{A}_K^\kappa(x,t)} \rho_k^{\Delta^*(\kappa)}(t) - \Gamma \overline{F}_\kappa(x) \right| = 0 \quad w.p.1. \quad (11)$$

Proof. The proof is presented in Appendix A. □

Lemma 4.4. *For every* t *and* x,

$$\lim_{K \to \infty} \sup_{i \in \mathcal{A}_K^\kappa(x,t)} \rho_i^{\Delta^*(\kappa)}(t) = 0 \quad w.p.1. \quad (12)$$

Proof. The proof is presented in Appendix B. □

Theorem 4.5 (energy efficiency). *Asymptotically, that is, as* $K \to \infty$, *the* $(E_b/N_0)_{sys}$ *under policy* $\Delta^*(\kappa)$ *is given by*

$$\left(\frac{E_b}{N_0} \right)_{sys}^{\Delta^*(\kappa)} = \mathbb{E}_{\overline{F}_\kappa} \left[\frac{e^{\Gamma \overline{F}_\kappa(x)}}{x} \right]$$

$$= \int_0^\infty \frac{1}{x} \exp\left(\Gamma \overline{F}_\kappa(x) \right) \mathbf{d}\overline{F}_\kappa(x) \quad w.p.1. \quad (13)$$

Proof. Fix arbitrary slot t. We show that under $\Delta^*(\kappa)$, $(E_b(t)/N_0)_{sys}^{\Delta^*(\kappa)}$ is the same in every slot t as $K \to \infty$. So, for convenience, we drop t in the notation.

Now, from (4) and Definition 2.4, it follows that

$$\left(\frac{E_b}{N_0} \right)_{sys}^{\Delta^*(\kappa)} = \sum_{i \in \mathcal{A}_K^\kappa(\infty)} \frac{1}{\Gamma d_i} e^{\sum_{k \in \mathcal{A}_K(d_i)} \rho_k^{\Delta^*(\kappa)}} \left(e^{\rho_i^{\Delta^*(\kappa)}} - 1 \right). \quad (14)$$

Since $\exp(\cdot)$ is a continuous function, Lemma 4.3 implies that there exists a sequence of nonnegative r.v.'s $\{\epsilon_K\}_{K \geq 1}$ independent of d_i such that $\lim_{K \to \infty} \epsilon_K = 0$ w.p.1, and

$$e^{-\epsilon_K} e^{\Gamma \overline{F}_\kappa(d_i)} \leq e^{\sum_{k \in \mathcal{A}(d_i)} \rho_k^{\Delta^*(\kappa)}} \leq e^{\epsilon_K} e^{\Gamma \overline{F}_\kappa(d_i)}. \quad (15)$$

Thus from (14) and (15),

$$\frac{e^{-\epsilon_K}}{\Gamma} \sum_{i \in \mathcal{A}_K^\kappa(\infty)} \frac{1}{d_i} e^{\Gamma \overline{F}_\kappa(d_i)} \left(e^{\rho_i^{\Delta^*(\kappa)}} - 1 \right) \leq \left(\frac{E_b}{N_0} \right)_{sys}^{\Delta^*(\kappa)}$$

$$\leq \frac{e^{\epsilon_K}}{\Gamma} \sum_{i \in \mathcal{A}_K^\kappa(\infty)} \frac{1}{d_i} e^{\Gamma \overline{F}_\kappa(d_i)} \left(e^{\rho_i^{\Delta^*(\kappa)}} - 1 \right). \quad (16)$$

Now, by Lemma 4.4, we conclude for a large enough K that the cost of providing user i with rate $\rho_i^{\Delta^*(\kappa)}$ approaches

$$\left(e^{\rho_i^{\Delta^*(\kappa)}} - 1 \right) \approx \rho_i^{\Delta^*(\kappa)} = \frac{\Gamma \overline{\nu}_i}{K}. \quad (17)$$

Note that the approximation becomes tighter for larger K. Thus,

$$\frac{e^{-\epsilon_K}}{\Gamma} \sum_{i \in \mathcal{A}_K^\kappa(\infty)} \frac{e^{\Gamma \overline{F}_\kappa(d_i)}}{d_i} \left(\frac{\Gamma \overline{\nu}_i}{K} \right) \leq \left(\frac{E_b}{N_0} \right)_{sys}^{\Delta^*(\kappa)}$$

$$\leq \frac{e^{\epsilon_K}}{\Gamma} \sum_{u=1}^\infty \sum_{i \in \mathcal{A}_K^\kappa(\infty)} \frac{e^{\Gamma \overline{F}_\kappa(d_i)}}{d_i} \left(\frac{\Gamma \overline{\nu}_i}{K} \right). \quad (18)$$

Let us consider the following term in (18):

$$\lim_{K \to \infty} e^{\epsilon_K} \sum_{i \in \mathcal{A}_K^\kappa(\infty)} \frac{e^{\Gamma \overline{F}_\kappa(d_i)}}{d_i} \frac{\overline{\nu}_i}{K}$$

$$= \lim_{K \to \infty} e^{\epsilon_K} \lim_{K \to \infty} \frac{|\mathcal{A}_K^\kappa(\infty)|}{K}$$

$$\times \lim_{|\mathcal{A}_K^\kappa(\infty)| \to \infty} \frac{1}{|\mathcal{A}_K^\kappa(\infty)|} \sum_{i \in \mathcal{A}_K^\kappa(\infty)} \frac{e^{\Gamma \overline{F}_\kappa(d_i)}}{d_i} \overline{\nu}_i$$

$$= \gamma \mathbb{E} \left[\frac{e^{\Gamma \overline{F}_\kappa(d_i)}}{d_i} \right] \mathbb{E}[\overline{\nu}_i] \quad w.p.1. \quad (19)$$

Similarly,

$$\lim_{K \to \infty} e^{-\epsilon_K} \sum_{i \in \mathcal{A}_K^\kappa(\infty)} \frac{e^{\Gamma \overline{F}_\kappa(d_i)}}{d_i} \frac{\overline{\nu}_i}{K} = \gamma \mathbb{E} \left[\frac{e^{\Gamma \overline{F}_\kappa(d_i)}}{d_i} \right] \mathbb{E}[\overline{\nu}_i] \quad w.p.1. \quad (20)$$

The first expectation in (19) and (20) is with respect to the distribution $\overline{F}_\kappa(\cdot)$, while the second is with respect to the distribution of the arrival process. The relations (19) and (20) hold because the channel gains d_i's of the chosen users can be viewed as i.i.d. variables drawn from the distribution $\overline{F}_\kappa(\cdot)$. This can be seen as follows. Since f_k's are i.i.d. irrespective of the distance between the user and receiver, the scheduling decision can be viewed as scheduling each user w.p. γ independently. Since the users are placed at random, s_k is a deterministic function of distance, and f_k and s_k are independent, we conclude that d_k's for the chosen users are i.i.d. and each d_k is distributed as $\overline{F}_\kappa(\cdot)$. Thus, $\{e^{\Gamma \overline{F}_\kappa(d_i)}/d_i\}_{i=1,\dots,|\mathcal{A}_K^\kappa(\infty)|}$ is an i.i.d. sequence. Thus, from (19) and (20), we have

$$\lim_{K \to \infty} e^{\epsilon_K} \sum_{i \in \mathcal{A}_K^\kappa(\infty)} \frac{e^{\Gamma \overline{F}_\kappa(d_i)}}{d_i} \frac{\overline{\nu}_i}{K}$$

$$= \lim_{K \to \infty} e^{-\epsilon_K} \sum_{i \in \mathcal{A}_K^\kappa(\infty)} \frac{e^{\Gamma \overline{F}_\kappa(d_i)}}{d_i} \frac{\overline{\nu}_i}{K} \quad (21)$$

$$= \int_0^\infty \frac{1}{x} \exp\left(\Gamma \overline{F}_\kappa(x) \right) \mathbf{d}\overline{F}_\kappa(x) \quad w.p.1.$$

Thus the required follows from Definition 2.4. □

Note 2. The energy efficiency does not depend on the distribution of the users' arrival processes but depends only on the mean as long as the processes are independent and

have all the moments finite. Next we will show that the energy efficiency behaves monotonically with respect to the opportunism threshold and provide a result for the rate of energy decrease with the threshold.

Theorem 4.6 (monotonicity of energy efficiency). *Asymptotically, that is, as $K \to \infty$, for every $\kappa' < \kappa$,*

$$\left(\frac{E_b}{N_0}\right)^{\Delta^*(\kappa)}_{sys} \leq \left(\frac{E_b}{N_0}\right)^{\Delta^*(\kappa')}_{sys} \quad w.p.1. \tag{22}$$

Moreover, the decrease in energy efficiency is $O(1/\kappa)$. Specifically,

$$\left(\frac{E_b}{N_0}\right)^{\Delta^*(\kappa)}_{sys} \leq \frac{1}{\kappa}\mathbb{E}_\Psi\left[\frac{e^{\Gamma\Psi(S)}}{S}\right] \leq \frac{e^\Gamma \mathbb{E}_\Psi[1/S]}{\kappa} \quad w.p.1. \tag{23}$$

First, we provide the intuition behind the result and then provide the proof.

Intuition. Note the following property of the $\log(\cdot)$ function. Let $\{a_i\}_{i\geq 1}$ denote any sequence of nonnegative real numbers and let Z be any constant. Then,

$$\sum_{i=1}^\infty \log\left(1 + \frac{a_i}{Z + \sum_{u<i} a_u}\right) = \log\left(1 + \frac{\sum_{i=1}^\infty a_i}{Z}\right). \tag{24}$$

Thus, under simultaneous transmission and successive decoding, we obtain

$$\sum_{i\in\mathcal{A}_\kappa(\infty)} \rho_i(t) = \log\left(1 + \frac{\sum_{i\in\mathcal{A}_\kappa(\infty)} E_i^R(t)}{N_0}\right), \tag{25}$$

where $\rho_i(t)$ denotes the rate requirement of user i in slot t and $\mathcal{A}_\kappa(\infty)$ denotes the number of scheduled users scheduled under $\Delta^*(\kappa)$.

Thus, from (25) we conclude that if the sum rate to be provided is the same, then the required sum received energy at the receiver remains the same and is independent of the individual rates. Now, note that the sum rate to be provided in every slot under $\Delta^*(\kappa)$ is asymptotically, that is, as $K \to \infty$, equal to Γ for every κ. This can be seen as follows:

$$\begin{aligned}
&\lim_{K\to\infty} \sum_{i\in\mathcal{A}_\kappa(\infty)} \rho_i(t) \\
&= \lim_{K\to\infty} \sum_{i\in\mathcal{A}_\kappa(\infty)} \frac{\Gamma}{K}\overline{\nu}_i(t) \\
&= \Gamma \lim_{K\to\infty} \frac{|\mathcal{A}_\kappa(\infty)|}{K} \lim_{|\mathcal{A}_\kappa(\infty)|\to\infty} \frac{1}{|\mathcal{A}_\kappa(\infty)|} \sum_{i\in\mathcal{A}_\kappa(\infty)} \overline{\nu}_i(t) \\
&= \Gamma \quad w.p.1.
\end{aligned} \tag{26}$$

Alternatively, one can apply Lemma 4.3 with $x \to \infty$. The converse of this result is naturally, that in the case of finite K the required rate served by the system is a random variable. From the above it follows that asymptotically the required

sum received energy under $\Delta^*(\kappa)$ is the same in every slot independent of κ. Now, the required sum transmit energy depends on the channel states. Since κ increases, only the users with a larger channel gains are selected, we intuitively expect the sum transmit energy required to achieve the given sum received energy to decrease monotonically. Now, we prove Theorem 4.6. First, we state a lemma that we use to prove the theorem.

Lemma 4.7. *Let $\mathcal{A}$ denote an ordered countable set of users, where the ordering is as per fading, that is, $d_i \leq d_j$ whenever $i < j$. Let $\overrightarrow{\rho}$ and $\overrightarrow{\rho}'$ denote two different rate requirement vectors satisfying*

$$\sum_{i=1}^k \rho_i \geq \sum_{i=1}^k \rho_i' \quad \text{for every } k. \tag{27}$$

Moreover, let $\overrightarrow{E}$ and $\overrightarrow{E}'$ denote the energy allocation as per (4) to realize $\overrightarrow{\rho}$ and $\overrightarrow{\rho}'$, respectively. Then, $\sum_{i\in\mathcal{A}} E_i \geq \sum_{i\in\mathcal{A}} E_i'$.

Proof. The proof is presented in Appendix C. □

Now, we prove Theorem 4.6.

Proof. Let us consider two identical copies of a sample path; that is, arrivals and fading for each of the users are the same in every slot. On the first sample path, the users are served as per $\Delta^*(\kappa)$, and on the second they are served as per $\Delta^*(\kappa')$, where $\kappa < \kappa'$. Note that $\mathcal{A}_K^{\kappa'}(x,t) \subseteq \mathcal{A}_K^\kappa(x,t)$ for every K, x and t, as $f_i(t) > \kappa'$ implies $f_i(t) > \kappa$.

Now, fix t and K and define $\mathcal{A} = \mathcal{A}_K^\kappa(\infty,t)$. Moreover, let $\overrightarrow{\rho}(t) = \{\rho_i^{\Delta^*(\kappa)}(t) : i \in \mathcal{A}\}$ and $\overrightarrow{\rho}'(t) = \{\rho_i^{\Delta^*(\kappa')}(t)\mathbf{1}_{\{i\in\mathcal{A}_K^{\kappa'}(\infty,t)\}} : i \in \mathcal{A}\}$. Now, we show that for every $i \in \mathcal{A}$, it holds that $\sum_{k<i} \rho_k \geq \sum_{k<i} \rho_k'$. By Lemma 4.3, $\sum_{k<i} \rho_k \to \Gamma\overline{F}_\kappa(d_i)$ and $\sum_{k<i} \rho_k' \to \Gamma\overline{F}_{\kappa'}(d_i)$ w.p.1 as $K \to \infty$. Moreover, for a randomly placed user i, $\{d_i \leq x \mid f_i > \kappa\} \supseteq \{d_i \leq x \mid f_i > \kappa'\}$ for every x whenever $\kappa \leq \kappa'$. Thus, by the monotonicity of the probability measure, $\overline{F}_{\kappa'}(x) \leq \overline{F}_\kappa(x)$ for every x. Thus, as $K \to \infty$, $\sum_{k<i} \rho_k \geq \sum_{k<i} \rho_{k'}$ on any nontrivial sample path. Thus, monotonicity property follows from Lemma 4.7 as t is arbitrary.

Now, we show the first inequality in (23). The inequality follows by observing that the fading of every chosen user i satisfies $d_i \geq \kappa s_i$. We consider energy allocation with these worse fading states (κs_i) for every chosen user. Now, observe that with this modified fading process, the randomness remains only in the pathloss s_i. Hence, the required follows using exactly the same arguments as that in Theorem 4.5. We, however, would like to mention that the inequality does not follow from the expressions stated in the statement of Theorem 4.5 as there the users are ordered as per $\overrightarrow{d}$ which may be different than that as per the pathloss considered here. Now, the second inequality in (23) follows by observing that $\Psi(x) \leq 1$ for every x in the first inequality in (23). □

4.1. Example Channels. In Theorems 4.1 to 4.6, we have not assumed anything about the distribution of f_i. Now, let us

consider an important special case where f_i has infinite support. This assumption holds for many distributions used to model multipath effects, for example, Rayleigh, Rician, and Nakagami fadings. Note that as $\gamma \to 0$ (equivalently, as delay goes to ∞), opportunism threshold $\kappa \to \infty$. Thus, from Theorem 4.6,

$$\lim_{\text{Delay} \to \infty} \left(\frac{E_b}{N_0} \right)_{\text{sys}}^{\Delta^*(\kappa)} = 0 \quad \text{w.p.1.} \tag{28}$$

The relation (28) holds for any $\Phi(\cdot)$ with unbounded support. Though the minimal value is 0 in all the cases, the delay-energy tradeoff, that is, how quickly energy approaches the minimum when the delay is increased, depends strongly on the distribution. It can be easily seen that if the tail of the distribution is heavy, then the delay increases at a smaller rate as κ increases significantly with the increase in the delay. On the contrary, if the tail is lighter, the energy decreases at a slower rate with the increase in the delay. We explain this by considering a special class of distributions parametrized by α. Fix $\alpha > 1$ and define $\Phi(x) = 0$ for $x < 1$ and $\Phi(x) = 1 - x^{(1-\alpha)}$ otherwise. Note that α determines how fast the tail of the distribution diminishes, for example, as α increases the tail diminishes faster. For this distribution, clearly, delay $D = 1/\gamma = \kappa^{(\alpha-1)}$. Thus, when the energy is within $O(1/\kappa)$ of the minimum, which is zero as $\Phi(\cdot)$ has unbounded support, the delay is $O(\kappa^{(\alpha-1)})$. If we pick $\alpha \in (1, 3/2)$, then the delay is $\omega(\sqrt{\kappa})$, that is, the delay increases at rate strictly smaller than $\sqrt{\kappa}$. Contrast this with the delay-energy tradeoff obtained in [22]. In [22], Neely considers a model in which the fast fading is modelled as a Markov Chain (MC) with finite state space. In these settings, Neely shows that the delay is at least of the order of $\sqrt{V}$ when the energy is in the order $1/V$ neighborhood of minimum energy required for the stability. Note that the result holds for any transition probability matrix, and hence the steady-state distribution of the MC is used to model the fast fading. But, we have shown that the result of [22] does not hold when the fast fading distribution has unbounded support.

In the following result, we show that $\Delta^*(\kappa)$ becomes energy optimal as the delay goes to infinity. From the above discussion it should be clear that $\Delta^*(\kappa)$ becomes energy optimal as the delay goes to infinity when $\Phi(\cdot)$ has unbounded support. Hence, we only focus on the case where $\Phi(\cdot)$ is supported on a compact set. Let B denote the supremum of the support.

Corollary 4.8 (energy optimality). *Let the short-term fading distribution $\Phi(\cdot)$ be supported on the interval $[0, B]$. As the number of users goes to infinity, for every policy Δ,*

$$\left(\frac{E_b}{N_0} \right)_{\text{sys}}^{\Delta} \geq \frac{1}{B} \mathbb{E}_\Psi \left[\frac{e^{\Gamma \Psi(S)}}{S} \right] \quad \text{w.p.1.} \tag{29}$$

Moreover,

$$\lim_{\text{Delay} \to \infty} \left(\frac{E_b}{N_0} \right)_{\text{sys}}^{\Delta^*(\kappa)} \leq \frac{1}{B} \mathbb{E}_\Psi \left[\frac{e^{\Gamma \Psi(S)}}{S} \right] \quad \text{w.p.1.} \tag{30}$$

Thus, $\Delta^*(\kappa)$ becomes asymptotically energy optimal as delay goes to ∞.

Proof. The relation (29) follows by observing that for user i, $d_i(t) \leq Bs_i$ for every t. Thus, the minimum sum energy required to support rates Γ/K for each user when fading is Bs_i in every slot provides a lower bound on the sum energy required to support the same rates with fading process $\vec{d}(t)$. Now, as shown in [1], optimal energy allocation in the multiaccess channel is given by (4). Now, the relation (29) follows using the arguments similar to those in Theorem 4.5, where fading process is now given by $d_i(t) = Bs_i$ for every i. Now, the relation (30) follows by taking the limit $\kappa \to B$ in (23), implying delay $\to \infty$, thus yielding (30). $\qquad \square$

4.2. Discussion on Analytical Results. Corollary 4.2 states that if the delay of $D \geq 1$ has to be provided, then the policy $\Delta^*(\kappa)$ can achieve it with any κ satisfying $\mathbb{P}\{f_i > \kappa\} \geq 1/D$. Now, Theorem 4.6 states that choosing the largest κ such that $\mathbb{P}\{f_i > \kappa\} = 1/D$ minimizes the system energy while providing the required delay when the number of users is large. Moreover, Theorem 4.5 quantifies the energy efficiency of the system for the given value κ. Note that Corollary 4.2 and Theorems 4.5 and 4.6 together quantify the delay-energy tradeoff in the multiaccess channel. Finally, Corollary 4.8 proves the energy optimality of $\Delta^*(\kappa)$ as the delay goes to ∞. Moreover, Theorem 4.6 shows that the decrease in energy efficiency with κ is at least linear. Thus, given any value of energy E greater than the minimum energy required to guarantee the rates to each of the users, there exists κ such that $\Delta^*(\kappa)$ provides the desired rate to each user while maintaining the required sum energy below E.

Caution has to be exercised while interpreting Corollary 4.8 as the limits are taken in two parameters, namely, the number of users $K \to \infty$ and the user delay $\to \infty$ (equivalently, $\kappa \to B$ or $\gamma \to 0$). Thus, the resulting limiting value depends on the relative rates at which these two parameters approach their respective limiting values. In Corollary 4.8, we first let $K \to \infty$ and subsequently, let $\gamma \to 0$. In other words, K and γ approach their respective limiting values, while $K\gamma$ is ∞. This can be clearly seen in (30) as we still see the superpositioning gain apparent in the term $e^{\Gamma \Psi(S)}$ as the optimal scheduling can take advantage of the variations in the pathloss values of different users. But, if we let $K \to \infty$ and $\gamma \to 0$, while ensuring $K\gamma \to 1$, then exactly one user will be scheduled under $\Delta^*(\kappa)$ for sufficiently large K. As a result, the superpositioning gain disappears. Thus, a question arises whether the energy optimality of $\Delta^*(\kappa)$ depends upon the relative rate at which K and γ approach their respective limits. Before answering this question, let us look at the lower bound (29). Note that this bound is tight if we split each of the K users in K different users (logical group) with the same pathloss, but i.i.d. short-term fading, and let these K users in a single logical group collectively desire the rate Γ/K, then let $K \to \infty$. Note that as $K \to \infty$, in each of the logical groups there exists a user with short-term fading B and scheduling such users from each of the logical groups at the rate Γ/K simultaneously

minimizes the sum energy required to achieve the desired rates. Thus, in this setting, the actual number of users, accounting for each user as K different users, goes to ∞ at rate K^2 and not at K. This shows that the tight lower bound also depends on the rate at which $K \rightarrow \infty$. Thus, for a fair comparison, the users in a logical group should scale at the rate γK. With these scaling laws the optimality of $\Delta^*(\kappa)$ should hold. Indeed, it can be shown that in a special case $K\gamma = 1$, while $K \rightarrow \infty$ and $\gamma \rightarrow 0$, the lower and upper bounds both equal $(e^\Gamma - 1/\Gamma B)\mathbb{E}[1/S]$ w.p.1. Note that this value corresponds to scheduling a single user with the best short-term fading value, which equals B as $K \rightarrow \infty$, in every slot.

5. Generalizations

Now, we discuss two important generalizations. First, we consider the system with multiple nonoverlapping bands. The required rate can be split over these bands. In Section 5.1, we discuss how the results in Section 4 can be generalized to this case. Second, we consider a case when the users need different delays. This is the case, when various types of applications are supported on a multiaccess channel, or when the multiaccess channel serves as an intermediate hop on the multiple hops traveled by the application in the network. In Section 5.2, we discuss how OSP can support this.

5.1. Multiband Multiaccess Channel. We consider multiaccess channel with M bands. We assume that the fading on these bands is statistically independent. Let f_i^m denote the short-term fading for user i on mth subband. Now, it is not immediately clear how the required rate should be split on the various bands in order to minimize the sum energy. But, fortunately, it has been shown that to minimize the sum energy required to realize a given rate vector on the multiband multiaccess channel, the total rate for a user should be supported on its best channel [7]. Let $f_i^*(t) \stackrel{\text{def}}{=} \max\{f_i^1(t)\ldots,f_i^M\}$. Thus, $\Delta^*(\kappa)$ has to be defined in terms $f_i^*(t)$ instead of $f_i(t)$; that is, $\Delta^*(\kappa)$ selects all the users with $f_i^*(t) > \kappa$ and provides the required rate on the best channel for every user. Now, Theorem 4.1 and Corollary 4.2 hold with $\gamma \stackrel{\text{def}}{=} \mathbb{P}\{f_i^* > \kappa\}$. In Theorem 4.5, the energy efficiency becomes

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}}^{\Delta^*(\kappa)} = \int_0^\infty \frac{1}{x}\exp\left(\frac{\Gamma}{M}\overline{F}_\kappa^*(x)\right)\mathbf{d}\overline{F}_\kappa^*(x) \quad \text{w.p.1,} \quad (31)$$

where $\overline{F}_\kappa^*(\cdot)$ denote the fading distribution of user i who is placed uniformly at random in a cell given that $f_i^*(t) > \kappa$. The additional factor of $1/M$ appears because only $1/M$ fraction of scheduled user transmit on a given band. Finally, Theorem 4.6 and Corollary 4.8 hold with f_i replaced by f_i^*.

5.2. Delay Differentiation. The users are divided into L classes based on their delay requirements. Let $\xi_1,\ldots,\xi_L$ represent the fraction of users that want delays $D_1,\ldots,D_L$, respectively. Let κ_l be the largest real number that satisfies
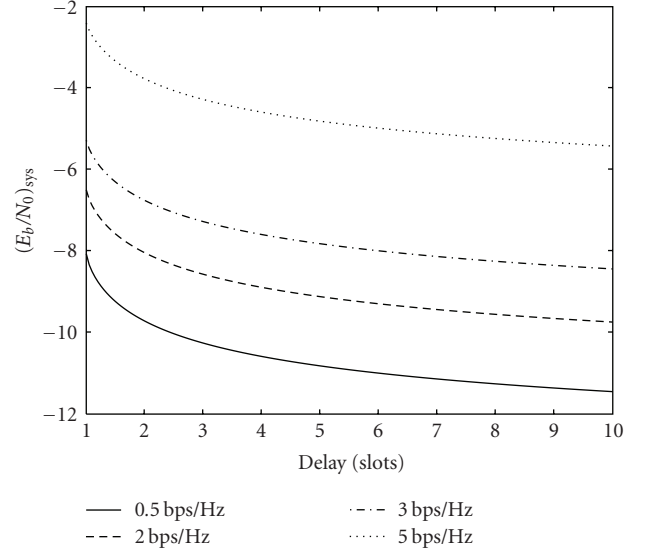

FIGURE 1: Total transmitted energy $(E_b/N_0)_{\text{sys}}$ as a function of delay.

$\mathbb{P}\{f_i > \kappa_l\} = 1/D_l$ for every $l \leq L$. Now, OSP can use κ_l instead of κ for users of class l. Clearly, Theorem 4.1 and Corollary 4.2 hold with $\gamma_l \stackrel{\text{def}}{=} \mathbb{P}\{f_i > \kappa_l\}$ for every l. Moreover, energy efficiency of each class l can be computed along the similar lines as the proof of Theorem 4.5. Now, the energy efficiency for the system is the weighted sum (with respect to ξ_l's) of the energy efficiency of each class. Finally, Theorem 4.6 and Corollary 4.8 can be shown to hold for each class individually. Since, the system energy efficiency is the convex combination of the energy efficiencies of the classes, Theorem 4.6 and Corollary 4.8 follow for the whole system.

6. Numerical Examples

We consider an example system where users are placed uniformly at random in a cell except for a forbidden region with radius $\delta = 0.01$ around the access point. The path loss exponent is two ($\alpha = 2$). All users experience short-term fading with exponential energy distribution with mean one on each of the ten ($M = 10$) independently fading bands. The explicit mathematical formulations for the channel models can be found in Appendix D. The path loss model is normalized to unity at cell edge, so that the results should be normalized with a corresponding factor. This, however, has no effect on the relative numerical results, nor on the delay-energy tradeoff we report.

Figure 1 demonstrates the delay energy tradeoff exhibited by OSP. For an increase in average delay from one to three slots, an energy saving of over 3 dB is gained. Thus, even the small delay tolerance of the application can be exploited to obtain significant improvement in the energy efficiency of the system. In terms of the sensor network application provided for motivation in Section 2, this would mean the halving of the transmitted energy by buffering of a few of the most recent samples. Moreover, the energy required
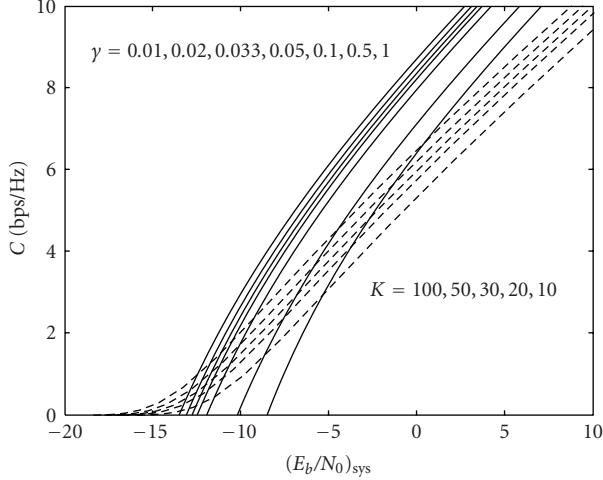
FIGURE 2: Comparison between PFS (dashed) and OSP (solid line).

FIGURE 3: Empirical convergence of extremes with a finite number of users.

to support the given rate decreases monotonically as delay increases, which verifies Theorem 4.6. Note that the gain exhibits very similar behavior across various spectral efficiencies. This implies that one can use a uniform threshold over a wide range of system configurations having different spectral efficiences, while maintaining a close to optimal efficiency.

Figure 2 provides a comparison between OSP and PFS. The mathematical expressions for computing the energy efficiency (E_b/N_0) of PFS can be found in [7, Theorem 1] and are given in Appendix E for completeness. In the delay-limited case with the strict delay constraint of a single slot (i.e., $\kappa = 0$), OSP can outperform PFS only at high spectral efficiencies. However, as delay tolerance of the application increases, OSP can outperform PFS over a wider range of spectral efficiencies. Also, the improvement in the energy efficiency of OSP over that of PFS increases monotonically with an increase in the delay tolerance. Note that the improvement happens while guaranteeing a required rate for each user, which is not the case for PFS. A notable feature of OSP is that changing the opportunism threshold results in an approximately horizontal shift of the performance curve, which again indicates that the energy-delay tradeoff behaves in a similar manner for a wide range of system spectral efficiencies.

An empirical verification for the convergence of the system energy can be found in Figure 3, where the smallest and largest empirically found energy efficiencies of an ensemble of 1000 simulated systems are reported. Each system has a different random pathloss vector. The system employs hard fairness, that is, it does not allow for delay, and data is assumed to arrive at users' transmit buffers in each slot. The extreme energies converge towards the asymptotic at $K = \infty$ as the user population grows.

Figure 3 indicates that the system energy efficiency deviates from asymptotic behavior at high spectral efficiency and with a small number of users. This behavior is due to the following. With a finite number of users, the transmitted rate is no longer deterministic as in the asymptotic c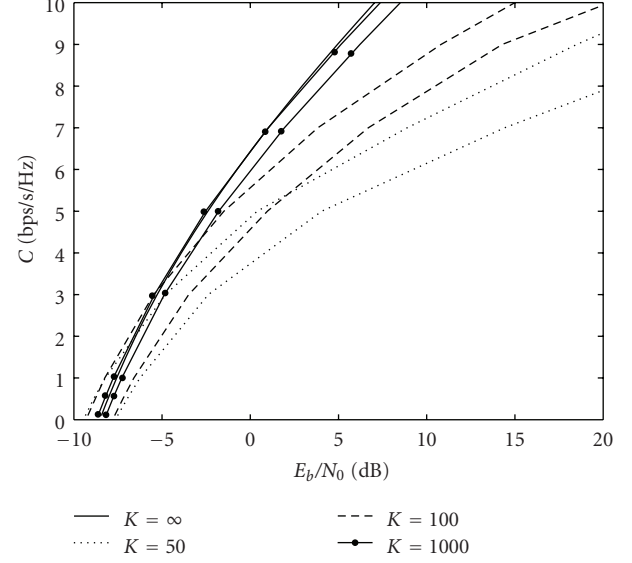ase. Instead, the number of simultaneously scheduled users and their buffer lengths vary from slot to slot. This results in energy loss due to the convexity of the exponential rate-energy function in (4). Since the rate-energy function has the spectral efficiency as a multiplicative factor in the exponential, the loss is greater at high spectral efficiency.

7. Conclusions

We showed that by opportunistically choosing a suitable fraction of users with the best channels in each slot, we can improve the energy efficiency of the system while providing the required delay to each user. Since the policy empties the scheduled users' queues, it has good stability properties. We showed that the expected user delay is inversely proportional to the scheduling fraction. Delay can then be adjusted simply by choosing an appropriate opportunism threshold, while delay differentiation can be achieved by applying different thresholds for different delay classes. Moreover, if the application does not need any delay guarantees, then OSP can achieve any required energy efficiency $(E_b/N_0 > 0)$ while maintaining system stability. The scheme performs well compared to PFS, while providing rate guarantees.

Appendices

A. Proof of Lemma 4.3

Proof. First, we show that for every x and t,

$$\lim_{K \to \infty} \sum_{k \in \mathcal{A}_K^\kappa(x,t)} \rho_k^{\Delta^*(\kappa)}(t) = \Gamma \overline{F}_\kappa(x) \quad \text{w.p.1.} \tag{A.1}$$

If x lies below the support of d_k the result trivially holds as both sides of (A.1) are equal to zero. Hence, we consider $x > \kappa \min\{s_k : k \in \mathcal{A}_K^\kappa(x,t)\}$ in the following.

For a chosen user k, the rate $\rho_k^{\Delta^*(\kappa)}$ is equal to its total buffer occupancy in slot t. Under $\Delta^*(\kappa)$, buffer occupancy in slot t is equal to the number arrivals since the last time k was scheduled. Let $\bar{\nu}_k \stackrel{\text{def}}{=} \sum_{u=\tau(t)+1}^t \nu_k(u)$, where $\tau(t) = \max\{u < t : f_k(u) > \kappa\}$. Note that $\nu_k(u)$ and the fading are i.i.d. across both the slots and users, and they are also mutually independent. So, clearly, $\bar{\nu}_k$ are i.i.d. across the chosen users and $\mathbb{E}[(\bar{\nu}_k)] = (1/\gamma)\mathbb{E}[(\nu_k)] = 1/\gamma$. Moreover, for a chosen user k,

$$\rho_k^{\Delta^*(\kappa)}(t) = \frac{\Gamma}{K}\bar{\nu}_k. \qquad (A.2)$$

We then have

$$\lim_{K \to \infty} \sum_{k \in \mathcal{A}_K^\kappa(x,t)} \rho_k^{\Delta^*(\kappa)}(t)$$

$$= \lim_{K \to \infty} \sum_{k \in \mathcal{A}_K^\kappa(x,t)} \frac{\Gamma}{K}\bar{\nu}_k \quad \text{(from (A.2))}$$

$$= \Gamma \lim_{K \to \infty} \frac{|\mathcal{A}_K^\kappa(\infty,t)|}{K} \lim_{K \to \infty} \frac{|\mathcal{A}_K^\kappa(x,t)|}{|\mathcal{A}_K^\kappa(\infty,t)|} \lim_{K \to \infty} \frac{\sum_{k \in \mathcal{A}_K^\kappa(x,t)} \bar{\nu}_k}{|\mathcal{A}_K^\kappa(x,t)|}$$

$$= \Gamma \overline{F}_\kappa(x) \quad \text{w.p.1.}$$
$$(A.3)$$

Note that $|\mathcal{A}_K^\kappa(x,t)| \to \infty$ as $K \to \infty$ for every t and $x > \kappa$ and then the first limit converges to γ, the second to $\overline{F}_\kappa(x)$ and the last one to $1/\gamma$. The final relation (A.3) follows as fading and arrivals are independent.

Now, we claim that the channel gains d_k's for the chosen users can be viewed as i.i.d. variables. Note that if $\Delta^*(\kappa)$ had scheduled users based on d_k rather than on f_k, then d_k for the chosen users would not be i.i.d. as the users that are nearer to the receiver are likely to be favored. Since f_k's are i.i.d. irrespective of the distance between the user and receiver, the scheduling decision can be viewed as scheduling each user independently w.p. γ. Since the users are placed at random, s_k is a deterministic function of distance, and f_k and s_k are independent, we conclude that d_k's for the chosen users are i.i.d. and each d_k is distributed as $\overline{F}_\kappa(\cdot)$. Finally, (11) follows from (A.3) using the Glivenko-Cantelli Theorem [30]. $\qquad \square$

B. Proof of Lemma 4.4

Proof. For simplicity, we prove the required when ν_i has finite support, say $\nu_{\max}$. Let $\tau_0(i)$ denote last time before t user i was scheduled, that is, $\tau_0(i) = \max_{u<t}\{f_i(u) > \kappa\}$. Moreover, let $T_i = t - \tau_0(i)$ for each user i. Note that $\{T_i\}_{i \in \{1,\dots,K\}}$ is i.i.d. sequence and $\mathbb{P}\{T_i > u\} = (1 - \gamma)^u$ for every i.

Fix any user $i \in \mathcal{A}_K^\kappa(x,t)$, and observe that

$$\rho_i^{\Delta^*(\kappa)}(t) = \frac{\Gamma}{K}\sum_{\tau_0(i)+1}^t \nu_i(t) \le \frac{\Gamma\nu_{\max}T_i}{K}. \qquad (B.1)$$

Next, fix $\epsilon > 0$, and consider

$$\mathbb{P}\left\{\sup_{i \in \mathcal{A}_K^\kappa(x,t)} \rho_i^{\Delta^*(\kappa)}(t) > \epsilon\right\}$$

$$\le \sum_{i \in \mathcal{A}_K^\kappa(x,t)} \mathbb{P}\left\{\rho_i^{\Delta^*(\kappa)}(t) > \epsilon\right\} \quad \text{(by union bound)}$$

$$\le \sum_{i \in \mathcal{A}_K^\kappa(x,t)} \mathbb{P}\left\{\frac{\Gamma\nu_{\max}T_i}{K} > \epsilon\right\} \quad \text{(by (B.1))}$$

$$\le \sum_{i=1}^K \mathbb{P}\left\{\frac{\Gamma\nu_{\max}T_i}{K} > \epsilon\right\}$$

$$\text{(as } |\mathcal{A}_K^\kappa(x,t)| \le K \; \forall x \text{ and } t)$$

$$= K\mathbb{P}\left\{T_i > \frac{\epsilon K}{\nu_{\max}\Gamma}\right\} \quad \text{(as } T_i's \text{ are i.i.d.)}$$

$$= K(1 - \gamma)^{\epsilon K/B\Gamma}. \qquad (B.2)$$

Since $\gamma > 0$, it follows that

$$\sum_{K=1}^\infty \mathbb{P}\left\{\sup_{i \in \mathcal{A}_K^\kappa(x,t)} \rho_i^{\Delta^*(\kappa)}(t) > \epsilon\right\} \le \sum_{K=1}^\infty K(1 - \gamma)^{\epsilon K/B\Gamma} < \infty. \qquad (B.3)$$

Now, the result follows by Borel-Cantelli Theorem [31] as ϵ was arbitrary. $\qquad \square$

C. Proof of Lemma 4.7

Proof. The proof is by construction. We construct a sequence of rate vectors $\{\vec{\rho}(u)\}_{u \ge 0}$ such that $\vec{\rho}(0) = \vec{\rho}$ and $\lim_{u \to \infty} \vec{\rho}(u) = \vec{\rho}\,'$. Let $\vec{E}(u)$ denote the energy allocation to realize $\vec{\rho}(u)$. Then, we show that $\sum_{i \in \mathcal{A}} E_i(u)$ is a nonincreasing function of u, and thus proving the required. The recursive procedure to construct $\vec{\rho}(u)$ is as follows.

Initialize: $\vec{\rho}(0) = \vec{\rho}$.

STEP u:

C1. $\rho_i(u) = \rho_i(u - 1)$ for every $i \notin \{u, u+1\}$,

C2. $\rho_u(u) = \rho_u'$,

C3. $\rho_{u+1}(u) = \rho_{u+1}(u - 1) + \rho_u(u - 1) - \rho_u'$.

Note that $\vec{\rho}(u)$ satisfies $\rho_i(u) = \rho_i'$ for every $i \le u$, and $\rho_i(u) = \rho_i$ for every $i > u + 1$. Thus, clearly, $\lim_{u \to \infty} \vec{\rho}(u) = \vec{\rho}'$. Now, using induction, we prove that for every u

$$\sum_{i=1}^{u+1} \rho_i(u) = \sum_{i=1}^{u+1} \rho_i. \tag{C.1}$$

Note that (C.1) holds for $u = 0$ as $\rho_1(0) = \rho_1$ by initialization. Now, let (C.1) hold for every $u = 0, \ldots, n-1$. We show (C.1) for $u = n$.

Since (C.1) holds for $u = n-1$, we know that

$$\sum_{i=1}^{n} \rho_i = \sum_{i=1}^{n} \rho_i(n-1)$$

$$= \sum_{i=1}^{n-1} \rho_i(n) + \rho_n(n-1) \quad \text{(by step C1.)}$$

$$= \sum_{i=1}^{n+1} \rho_i(n) - \rho_{n+1}(n-1) \quad \text{(from step C3.)}$$

$$= \sum_{i=1}^{n+1} \rho_i(n) - \rho_{n+1} \quad \text{(as } \rho_i(u) = \rho_i \text{ for every } i > u + 1). \tag{C.2}$$

This proves (C.1) for every u.

Now, we show that $\sum_{i \in \mathcal{A}} E_i(u)$ is nonincreasing function of u. From (4) and (C.1), it is clear that $E_i(u) = E_i(u-1)$ for every $i \notin \{u, u+1\}$. Thus it suffices to consider

$$[E_u(u) + E_{u+1}(u)] - [E_{u-1}(u) - E_{u+1}(u-1)]$$

$$= \frac{N_0}{d_u} \left[e^{\sum_i^u \rho_i(u)} - e^{\sum_i^u \rho_i(u-1)} \right] - \frac{N_0}{d_{u+1}} \left[e^{\sum_i^u \rho_i(u)} - e^{\sum_i^u \rho_i(u-1)} \right]$$

$$= N_0 \left[\frac{1}{d_u} - \frac{1}{d_{u+1}} \right] \left[e^{\sum_i^u \rho_i(u)} - e^{\sum_i^u \rho_i(u-1)} \right]$$

$$= N_0 \left[\frac{1}{d_u} - \frac{1}{d_{u+1}} \right] \left[e^{\sum_i^u \rho_i'} - e^{\sum_i^u \rho_i} \right]$$

$$\text{(by steps C1 and C2, and (C.1))}$$

$$\le 0. \tag{C.3}$$

The last inequality follows as $d_u \le d_{u+1}$ and $\sum_i^u \rho_i' \le \sum_i^u \rho_i$ by suppositions in the lemma.

Now, to prove the required, we need to show that $\lim_{u \to \infty} \sum_{i \in \mathcal{A}} E_i(u) = \sum_{i \in \mathcal{A}} E_i'$. First, note that if $\sum_{i \in \mathcal{A}} E_i = \infty$, then the result immediately follows. So, we consider a nontrivial case, $\sum_{i \in \mathcal{A}} E_i < \infty$. We know that for every u, $0 \le \sum_{i \in \mathcal{A}} E_i(u) \le \sum_{i \in \mathcal{A}} E_i$. Moreover, the sequence $\{\sum_{i \in \mathcal{A}} E_i(u)\}_{u \ge 0}$ is monotone. Thus, $\lim_{u \to \infty} \sum_{i \in \mathcal{A}} E_i(u)$

exists. Moreover, by dominated convergence theorem, we can exchange the limit and summation. Thus,

$$\lim_{u \to \infty} \sum_{i \in \mathcal{A}} E_i(u)$$

$$= \sum_{i \in \mathcal{A}} \lim_{u \to \infty} E_i(u)$$

$$= \sum_{i \in \mathcal{A}} \lim_{u \to \infty} \left[\frac{N_0}{d_i} e^{\sum_{k=1}^{i-1} \rho_k(u)} \left(e^{\rho_i(u)} - 1 \right) \right] \quad \text{(from (4))}$$

$$= \sum_{i \in \mathcal{A}} \left[\frac{N_0}{d_i} e^{\sum_{k=1}^{i-1} \rho_k'} \left(e^{\rho_i'} - 1 \right) \right]$$

$$\left(\text{as } \vec{\rho}(u) \to \vec{\rho}' \text{ and } \exp(\cdot) \text{ is continuous} \right)$$

$$= \sum_{i \in \mathcal{A}} E_i'. \tag{C.4}$$

This proves the required. □

D. Channel Statistics

Channel state is assumed to remain constant during one transmission slot and assume a new value for each slot independently for all slots and users. We, thus, avoid the dependence on users and time in the following.

Users are placed uniformly on a circular cell. The channel state of each user is the product of two independent ergodic random processes, path loss and short-term fading. Path loss is polynomially dependent on the distance of the user from the access point and is assumed to remain constant for a users across transmission slots. The distance dependency is parametrized by the path loss exponent α, usually ranging within the interval $[2, 4]$. To avoid a singularity for users next to the access point, a forbidden circular region of radius δ is created around the access point. This model results in the following cumulative distribution of the path loss:

$$P_s(x) = \begin{cases} 0, & x \le 1, \\ 1 - \dfrac{x^{-2/\alpha} - \delta^2}{1 - \delta^2}, & 1 \le x \le \delta^{-\alpha}, \\ 1, & x \ge \delta^{-\alpha}. \end{cases} \tag{D.1}$$

Note that the model is normalized to provide unit path loss at cell border. Thus, the results reported here must be rescaled in terms of $(E_b/N_0)_{\text{sys}}$ by a factor of $D_0^{-\alpha}$, where D_0 denotes the actual radius of the cell. Naturally, since all results behave accordingly, all comparisons remain valid.

The short-term fading process on each band m is modeled by a zero-mean Gaussian circularly symmetric random variable (i.e., Rayleigh fading), with an exponentially distributed squared envelope. It assumes a new value for each user in each slot. The multiple bands are assumed independently fading. In each slot, those users i whose maximum channel gain (over all M channels) $\max\{f_i^m(t) \ldots f_i^M(t)\} > \kappa$

are scheduled by OSP on each band, and the cumulative short-term fading distribution is given by

$$P_f(x) = 1 - \frac{1 - (1 - e^{-x})^M}{\gamma}, \quad x \in [\kappa, \infty), \qquad (D.2)$$

where $\gamma = \mathbb{P}(f_i > x) = 1 - (1 - e^{-\kappa})^M$. With some algebra, the cumulative distribution for the (product) channel can be expressed as

$$\overline{F}_\kappa^*(x)$$

$$= \begin{cases} \dfrac{1 - \kappa/x}{1 - \delta^2} - \dfrac{1}{\gamma x(1-\delta^2)} \displaystyle\sum_{i=1}^{M} \dfrac{1}{i} \left[(1 - \exp(-x))^i - (1 - \exp(-\kappa)^i) \right] \\ \hfill x < \kappa\delta^{-2}, \\[2mm] 1 - \dfrac{1}{\gamma x(1-\delta^2)} \displaystyle\sum_{i=1}^{M} \dfrac{1}{i} \left[(1 - \exp(-x))^i - (1 - \exp(-x\delta^2))^i \right] \\ \hfill x \geq \kappa\delta^{-2}. \end{cases}$$
$$(D.3)$$

E. Proportional Fair Scheduling

We follow the approach of [7] in the evaluation. The average spectral efficiency of the system with PFS is given implicitly by [7]

$$C = \int_0^\infty \log_2(1 + x\,\mathrm{SNR}) dF_{s\max\{f\},K}(x),$$
$$\left(\frac{E_b}{N_0}\right)_{\mathrm{sys}}^{\mathrm{PFS}} = \frac{\mathrm{SNR}}{C}, \qquad (E.1)$$

where $F_{s\max\{f\},K}(x)$ denotes the distribution of the product of the random path loss and the maximum of K users' short-term fading coefficients, $s\max\{f_k, \ldots, f_K\}$, and is given by

$$F_{s\max\{f\},K}(x) = 1 - \frac{1}{x(1-\delta^2)}$$
$$\times \sum_{i=1}^{K} \frac{1}{i} \left[(1 - \exp(-x))^i - (1 - \exp(-x\delta^2))^i \right], \qquad (E.2)$$

which is identical to (D.3) when $\gamma = 1$, $\kappa = 0$, and $M = K$.

Acknowledgments

References

[1] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels-part I: polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, 1998.

[2] S. V. Hanly and D. N. C. Tse, "Multiaccess fading channels-part II: delay-limited capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2816–2831, 1998.

[3] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: a bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Communications Magazine*, vol. 38, no. 7, pp. 70–77, 2000.

[4] D. N. C. Tse, "Multiuser diversity through proportional fair scheduling," in *Proceedings of the Communication Theory Workshop (CTW '01)*, May 2001.

[5] E. A. Jorswieck, A. Sezgin, and X. Zhang, "Throughput versus fairness: channel-aware scheduling in multiple antenna downlink," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 271540, 13 pages, 2009.

[6] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, 2002.

[7] G. Caire, R. R. Müller, and R. Knopp, "Hard fairness versus proportional fairness in wireless communications: the single-cell case," *IEEE Transactions on Information Theory*, vol. 53, no. 4, pp. 1366–1385, 2007.

[8] T. E. Klein, K. K. Leung, and H. Zheng, "Improved TCP performance in wireless IP networks through enhanced opportunistic scheduling algorithms," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '04)*, vol. 5, pp. 2744–2748, Dallas, Tex, USA, November 2004.

[9] M. A. Khojastepour and A. Sabharwal, "Delay-constrained scheduling: power efficiency, filter design, and bounds," in *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '04)*, vol. 3, pp. 1938–1949, Hong Kong, March 2004.

[10] M. A. Zafer and E. Modiano, "A calculus approach to minimum energy transmission policies with quality of service guarantees," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05)*, vol. 1, pp. 548–559, Miami, Fla, USA, March 2005.

[11] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1468–1489, 1999.

[12] E. Uysal-Biyikoglu and A. El Gamal, "On adaptive transmission for energy efficiency in wireless data networks," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3081–3094, 2004.

[13] W. Chen, M. J. Neely, and U. Mitra, "Energy efficient scheduling with individual packet delay constraints: offline and online results," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 1136–1144, Anchorage, Alaska, USA, May 2007.

[14] W. Chen, U. Mitra, and M. J. Neely, "Energy-efficient scheduling with individual delay constraints over a fading channel," in *Proceedings of the 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt '07)*, pp. 1–10, April 2007.

[15] R. Cheng and S. Verdú, "Gaussian multi-access channel with ISI: capacity region and multi-user waterfilling," *IEEE Transactions on Information Theory*, vol. 39, pp. 773–785, 1993.

[16] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, 2004.

[17] J. Oh, S.-J. Kim, and J. M. Cioffi, "Optimum power allocation and control for OFDM in multiple access channels," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '04)*, vol. 60, pp. 774–778, Los Angeles, Calif, USA, 2004.

[18] T. Michel and G. Wunder, "Solution to the sum power minimization problem under given rate requirements for the OFDM multipleaccess channel," in *Proceedings of the Annual Allerton Conference on Communications, Control and Computing*, September 2004.

[19] D. D. Yu and J. M. Cioffi, "Iterative water-filling for optimal resource allocation in OFDM multiple-access and broadcast channels," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '06)*, pp. 1–5, November-December 2006.

[20] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135–1149, 2002.

[21] M. Goyal, A. Kumar, and V. Sharma, "Power constrained and delay optimal policies for scheduling transmission over a fading channel," in *Proceedings of the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '03)*, vol. 1, pp. 311–320, San Francisco, Calif, USA, March-April 2003.

[22] M. J. Neely, "Optimal energy and delay tradeoffs for multi-user wireless downlinks," in *Proceedings of the 25th IEEE International Conference on Computer Communications (INFOCOM'06)*, pp. 1–13, Barcelona, Spain, April 2006.

[23] P. Chaporkar and S. Sarkar, "Stable scheduling policies for maximizing throughput in generalized constrained queueing systems," in *Proceedings of the 25th IEEE International Conference on Computer Communications (INFOCOM '06)*, pp. 1–13, Barcelona, Spain, April 2006.

[24] M. Neely, "Energy optimal control for time varying wireless networks," in *Proceedings of the Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'05)*, Miami, Fla, USA, March 2005.

[25] A. L. Stolyar, "Maximizing queueing network utility subject to stability: greedy primal-dual algorithm," *Queueing Systems*, vol. 50, no. 4, pp. 401–457, 2005.

[26] P. Chaporkar, K. Kansanen, and R. R. Müller, "Power optimal scheduling for guaranteed throughput in multi-access fading channels," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '07)*, pp. 2961–2965, Nice, France, June 2007.

[27] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 768–783, 2004.

[28] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2006.

[30] G. Shorak and J. Wellner, *Empirical Processes with Applications to Statistics*, John Wiley & Sons, New York, NY, USA, 1986.

[31] W. Feller, *An Introduction to Probability Theory and Its Application*, John Wiley & Sons, New York, NY, USA, 1961.