Channel and Multiuser Diversities in Wireless Systems: Delay-Energy Tradeoff

Prasanna Chaporkar, Kimmo Kansanen, Ralf R. Müller
Institute of Electronics and Telecommunications
Norwegian University of Science and Technology

Abstract— **We consider a communication system with multi-access fading channel. Each user in the system requires certain rate guarantee. Our main contribution is to devise a scheduling scheme called "Opportunistic Super-position Coding"that satisfies the users' rate requirements. Using mean-field analysis, i.e., when the number of users go to infinity, we analytically show that the energy required to guarantee the required user rate 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 multi-access channel. Finally, all the results can be generalized in a straightforward fashion to broadcast channel because of the uplink-downlink duality.**

Index Terms— **Super-position coding, opportunistic scheduling, multi-access channel, rate guarantees, delay guarantees, stability.**

I. Introduction

Comprehensive treatment of multi-access fading channels can be found in [1], [2]. In these papers, Tse and Hanly have characterized so called *throughput capacity* and *delay-limited capacity* of the multi-access fading channel with Gaussian noise. 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 states. The aim here is to obtain a coding and power allocation scheme to minimize the energy while guaranteeing the rate in every slot.

The notion of throughput capacity leads to schemes that take advantage of the differential channel qualities (*"channel diversity"*) of the users. Specifically, it has been shown that the sum rate in the system is maximized by letting only one user with the best channel 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 so called *"near-far"* effect. To alleviate this limitation, several opportunistic scheduling schemes with fairness constraints have been designed [3], [4]. Among them, Proportional Fair Scheduling (PFS) has many desirable properties including provable fairness guarantees and suitability for on-line implementation, i.e., without prior knowledge of channel statistics [5]. 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 the random slots, depending on their fading states and the resource allocation in the previous slots.

As discussed above, 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 diversity over time. Instead these schemes use *"multiuser diversity"*, i.e., they exploit variation in channel quality of various users in the same slot. It has been shown that super-position coding and successive decoding minimizes energy for achieving the required rates [2]. In this signaling scheme, all the users transmit simultaneously, and the receiver decodes in the decreasing order of the received power treating the undecoded signal as the noise. One of the attractive feature 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 the opportunistic scheduling exploits channel diversity to guarantee rates that maximize certain utility or/and guarantee fairness, while super-position coding and successive decoding exploits multiuser diversity 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 [6], but they also do not need the strict delay of one slot (as in delay-limited schemes), i.e., they have *limited* delay tolerance. *Our aim is to design a scheduling policy that makes use of both channel and multiuser diversities so as to improve the energy efficiency by exploiting the limited delay tolerance of the applications*.

Caire *et al*. have analyzed PFS scheduling and delay-limited schemes for block fading multi-access channel [7]. When the number of users become large, the authors have obtained the closed form expressions for energy consumed per transmitted bit under these schemes. Interestingly, the paper shows that for small spectral efficiency PFS is more energy efficient, however, as the spectral efficiency increases the delay-limited schemes are more energy efficient. In other words, for small (large, resp.) spectral efficiency, utilizing channel (multiuser, resp.) diversity dominates energy efficiency. Here, we show that efficiently exploiting the limited delay tolerance of the applications enables us to design schemes that are energy efficient irrespective of the spectral efficiency.

Recently, [8], [9], [10] have found scheduling frameworks to minimize the energy consumption in the wireless system with ergodic fading while providing the required rate to the users. The required rate is not provided in every slot but over, a possibly, long time period. In particular, Neely has shown the delay-energy tradeoff in these schemes [9]. These schemes determine the rate to be provided in each slot by solving an optimization problem. The optimization problem may be non-linear depending on coding/signaling strategy used, and hence may become computationally expensive in practice.

Here, we design a parametrized scheduling policy called *"Opportunistic Super-position Coding" (OSPC)* that exploits the channel diversity by scheduling a set users with high channel gains only, and among these users it uses super-position coding and successive decoding to exploit multiuser diversity. One of the main challenges in designing such schemes is the quantification of its performance. The quantification allows for the optimal and guaranteed control. We explicitly quantify the per user delay and the total energy requirement 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 a small delay tolerance can be exploited to achieve the 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 II, we present our system model. In Section III, we describe the OSPC policy, and in Section IV obtain analytical guarantees. In Section V, we discuss extensions of OSPC to multiband multi-access channel and to provide delay differentiation. In Section VI, we compare the performance of OSPC with PFS and delay-limited schemes using numerical computations. Finally, in Section VII, we conclude.

II. SYSTEM MODEL

We consider a multi-access system with K users that are placed uniformly at random in a cell. Time is slotted. Each user i requires a certain fraction of the total rate provided in a system, i.e., the required rate $R = \frac{\Gamma}{K}$, where Γ denotes the spectral efficiency. Alternatively, $R_i(t) = \frac{\Gamma}{K}\nu_i(t)$ denotes the arrivals for user i in slot t, where $\nu_i(t)$ is a random variable. 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. 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 infinite buffer before served. We primarily discuss the uplink communication (multi-access channel), but the results can be generalized in a straightforward fashion for the downlink case (broadcast channel) using uplink-downlink duality [11].

Now, we describe the fading model for the multi-access channel. Each user i experiences fading $d_i(t)$ in slot t. The fading $d_i(t)$ arises due to two independent effects, namely, *path loss* (denoted by s_i) and *frequency selective fading* (denoted by $f_i(t)$). 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 given the distance of a user from the transmitter. 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. Now, $d_i(t) = s_i f_i(t)$. This is referred as block fading model [12]. Let $E_i^R(t)$ ($E_i(t)$, resp.) denote the received (transmitted, resp.) energy from user i in slot t. Then, $E_i^R(t) = d_i(t)E_i(t)$. Let $\vec{d}(t) = \{d_1(t), \ldots, d_K(t)\}$. Note that the fading for users is *not* symmetric. Let N_0 denote the noise power spectral density.

Definition 1 (Scheduling Policy): A *scheduling policy* Δ is an algorithm that in each slot t determines the rate vector $\vec{\rho}^{\Delta}(t) = \{\rho_1^{\Delta}(t), \ldots, \rho_K^{\Delta}(t)\}$, and serves each user i with rate $\rho_i^{\Delta}(t)$.

We assume that s_i and $f_i(t)$ are known in each slot. Thus, a scheduling policy may adapt $\vec{\rho}^{\Delta}(t)$ to the channel state.

In general, given the rate vector any coding strategy can be used to realize it. But, it is well known that the superposition coding and successive decoding minimizes the sum energy required to realize the given rate vector for any $\vec{d}(t)$ [1]. Thus, in view of energy efficiency, we assume that superposition coding and successive decoding is used. Now, we state some relevant properties for the optimal signaling using this coding. Let $\vec{\pi}$ denote the permutation of $\vec{d}$ such that $d_{\pi_1} \leq d_{\pi_2} \leq \cdots \leq d_{\pi_K}$. Then, for optimal signaling, the successive decoding order depends only on channel gains, but not on rates. Thus,

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

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

Definition 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 3 (Stability): Let $\mathcal{B}_{u,i}$ denote the length of the u^{th} busy period of user i. The system is said to be *strongly stable* if for every user i

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

Definition 4 (Energy Efficiency): We define the energy efficiency in slot t as

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

Then, the energy efficiency is defined as

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

Definition 5 (Delay): Delay for the u^{th} arrival of user i (denoted by $D_{u,i}$) is the number slots between its departure and arrival. The delay for user i is then defined as

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

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

III. Opportunistic Super-position Coding (Δ^*)

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

- Select all users i such that $f_i(t) > \kappa$.
- 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 $\Delta^*(\kappa)$ 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 $d_i(t)$ has the distribution biased towards the users that are closer to the receiver.

We refer to κ as *opportunism threshold* as it dictates how opportunistic OSPC 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, few users with 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 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 the fixed noise power, the required energy increases exponentially with increase in the rate. Summarizing, the opportunism threshold allows us to save energy by eliminating the users in deep fade, but requires higher energy to serve scheduled users. Thus, it not clear if the opportunism threshold should improve the system performance. It, however, turns out that the energy consumption 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 when κ is large. So, clearly, the delay increases monotonically with increase in κ. Thus, the opportunism threshold provides a way to achieve delay-energy trade-off. Now, the main challenge is to quantify the delay and energy as a function of κ, so as to choose the optimal κ that guarantees the required delay while minimizing the energy.

IV. Analytical Guarantees

In this section, we obtain analytical guarantees for $\Delta^*(\kappa)$. In Theorem 1, we show that $\Delta^*(\kappa)$ strongly stable. In Theorem 2, we quantify per user delay under $\Delta^*(\kappa)$. In Theorem 3, we quantify the energy efficiency of $\Delta^*(\kappa)$. Finally, in Theorems 4 and 5 we show that the energy efficiency of $\Delta^*(\kappa)$ decreases monotonically with κ. In Subsection IV-A, we discuss the implications of these results. First, let $\gamma \overset{\text{def}}{=} \mathbb{P}\{f_i > \kappa\}$.

Theorem 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 (SLLN) for $\gamma > 0$

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

Now, result follows from Definition 3. ■

Theorem 2 (User Delay): The delay for any user i under $\Delta^*(\kappa)$ is

$$D_i^{\Delta^*(\kappa)} = \frac{1}{\gamma} \quad \text{w.p. 1.} \tag{2}$$

Proof: Let $\{T_n\}_{n\geq 1}$ denote the sequence of time slots such that $f_i(T_n) \geq \kappa$ and $f_i(t) < \kappa$ for every other t. In words, T_n denotes the n^{th} time slot in which session i is chosen by $\Delta^*(\kappa)$. Since $\{f_i(t)\}_{t\geq 1}$ is a sequence of i.i.d. random variables, $\{(T_{n+1}-T_n)\}_{n\geq 1}$ is also a sequence of i.i.d. random variables with geometric distribution and mean $1/\gamma$. Here, $T_0 = 0$. Let U_n denote the total number of arrivals of user i until T_n. Note that $(T_{n+1} - T_n)$ and arrivals in each slot are i.i.d., and fading and arrivals are mutually independent. Hence, we conclude that $\{(U_n - U_{n-1})\}_{n\geq 1}$ is a sequence of i.i.d. random variables. Here, $U_0 = 0$. By SLLN it follows that

$$\begin{aligned}\lim_{n\to\infty} \frac{U_n}{n} &= \mathbb{E}[U_1] \quad \text{w.p. 1}\\ &= \mathbb{E}[T_1]\mathbb{E}[R_i(1)] \quad \text{w.p. 1}\\ &= \frac{\Gamma}{\gamma K} \quad \text{w.p. 1.}\end{aligned} \tag{3}$$

Let $\overline{D}_{n,i}$ denote the sum of the delays experienced by all the arrivals in duration $(T_{n-1}, T_n]$. Again notice that $\{\overline{D}_{n,i}\}_{n\geq 1}$ is an i.i.d. sequence. Furthermore, the arrivals in slot $t \in (T_{n-1}, T_n]$ experience delay of $(T_n - t + 1)$ slots. Thus,

$$\overline{D}_{n,i} = \sum_{t=T_{n-1}+1}^{T_n} (T_n - t + 1)R_i(t).$$

Note that because of the independence between i.i.d. sequences $\{T_n\}_{n\geq 1}$ and $\{R_i(t)\}_{t\geq 1}$,

$$\mathbb{E}[\overline{D}_{n,i} \mid T_{n-1}, T_n] = \frac{\Gamma}{K}\frac{(T_n - T_{n-1})(T_n - T_{n-1} + 1)}{2}.$$

Thus, we obtain

$$\begin{aligned}\mathbb{E}[\overline{D}_{n,i}] &= \frac{\Gamma}{2K}\mathbb{E}[T_1(T_1+1)]\\ &= \frac{\Gamma}{K\gamma^2}.\end{aligned}$$

Moreover, by SLLN,

$$\lim_{n\to\infty} \frac{\overline{D}_{n,i}}{n} = \frac{\Gamma}{K\gamma^2} \quad \text{w.p. 1.} \tag{4}$$

Now, we evaluate $\limsup_{U\to\infty} \frac{1}{U}\sum_{u=1}^{U} D_{u,i}$. First consider the subsequence $\left\{\frac{1}{U_n}\sum_{u=1}^{U_n} D_{u,i}\right\}_{n\geq 1}$.

$$\frac{1}{U_N}\sum_{u=1}^{U_N} D_{u,i} = \frac{1}{U_N}\sum_{n=1}^{N} \overline{D}_{n,i}.$$

Thus,

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{U_N}\sum_{u=1}^{U_N} D_{u,i} &= \lim_{N\to\infty} \frac{1}{U_N}\sum_{n=1}^{N} \overline{D}_{n,i}\\ &= \lim_{N\to\infty} \frac{N}{U_N} \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \overline{D}_{n,i}\\ &= \frac{1}{\gamma} \quad \text{w.p. 1.}\end{aligned} \tag{5}$$

The last equality follows from (3) and (4).

Now, consider any $U \in \{U_{n-1}+1, \ldots, U_n\}$ and note that

$$\frac{1}{U_N}\sum_{n=1}^{N-1} \overline{D}_{n,i} \leq \frac{1}{U}\sum_{u=1}^{U} D_{u,i} \leq \frac{1}{U_{N-1}}\sum_{n=1}^{N} \overline{D}_{n,i}. \tag{6}$$

Note that as $U \to \infty$, $N \to \infty$, and $\lim_{N\to\infty} \frac{U_N}{U_{N-1}} = 1$. Now, by taking limit $U \to \infty$ in (6), we conclude from (5) that

$$\lim_{U\to\infty} \frac{1}{U}\sum_{u=1}^{U} D_{u,i} = \frac{1}{\gamma} \quad \text{w.p. 1.}$$

Thus, the result follows from Definition 5. ■

Note: 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 placed uniformly at random in a cell given that $f_i > \kappa$.

Theorem 3 (Energy Efficiency): In the mean field, i.e. as $K \to \infty$, the $(E_b/N_0)_{\text{sys}}$ under policy $\Delta^*(\kappa)$ is given by

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}}^{\Delta^*(\kappa)} = \int_0^{\infty} \frac{1}{x} \exp(\Gamma \overline{F}_{\kappa}(x)) \mathbf{d}\overline{F}_{\kappa}(x) \quad \text{w.p. 1.} \tag{7}$$

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

Let K' of these users are chosen by $\Delta^*(\kappa)$. Clearly,

$$\lim_{K \to \infty} \frac{K'}{K} = \mathbb{P}\{f_i(t) > \kappa\} = \gamma \quad \text{w.p. 1.} \tag{8}$$

Now, from (1), it follows that

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}}^{\Delta^*(\kappa)} = \sum_{i=1}^{K'} \frac{1}{\Gamma d_{\pi_i}} e^{\sum_{k<i} \rho_{\pi_k}^{\Delta^*(\kappa)}} (e^{\rho_{\pi_i}^{\Delta^*(\kappa)}} - 1). \tag{9}$$

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

$$\rho_i^{\Delta^*(\kappa)} = \frac{\Gamma}{\kappa} \overline{\nu}_i. \tag{10}$$

Let $\mathcal{A}(x)$ denote the set of chosen users i such that $f_i(t) < x$. Then (9) is equivalent to

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}}^{\Delta^*(\kappa)} = \sum_{i=1}^{K'} \frac{1}{\Gamma d_i} e^{\sum_{k \in \mathcal{A}(d_i)} \rho_k^{\Delta^*(\kappa)}} (e^{\rho_i^{\Delta^*(\kappa)}} - 1). \tag{11}$$

First, we evaluate the term $\sum_{k \in \mathcal{A}(d_i)} \rho_k^{\Delta^*(\kappa)}$ as $K \to \infty$. Let $|\mathcal{A}(x)|$ denote the cardinality of $\mathcal{A}(x)$. Note that $|\mathcal{A}(x)| \to \infty$ as $K \to \infty$ for every $x > \kappa$.

$$\begin{aligned}
&\lim_{K \to \infty} \sum_{k \in \mathcal{A}(d_i)} \rho_k^{\Delta^*(\kappa)} \\
&= \lim_{K \to \infty} \sum_{k \in \mathcal{A}(d_i)} \frac{\Gamma}{K} \overline{\nu}_i \quad \text{(from (10))} \\
&= \Gamma \lim_{K \to \infty} \frac{K'}{K} \lim_{K' \to \infty} \frac{|\mathcal{A}(d_i)|}{K'} \lim_{|\mathcal{A}(d_i)| \to \infty} \frac{\sum_{k \in \mathcal{A}(d_i)} \overline{\nu}_i}{|\mathcal{A}(d_i)|} \\
&= \Gamma \overline{F}_{\kappa}(d_i) \quad \text{w.p. 1.}
\end{aligned} \tag{12}$$

The relation (12) follows as fading and arrivals are independent.

Since $\exp(\cdot)$ is a continuous function, (12) implies that there exists a sequence of random variables $\{\epsilon_K\}_{K \geq 1}$ such that $\lim_{K \to \infty} \epsilon_K = 0$ w.p. 1, and

$$e^{\sum_{k \in \mathcal{A}(d_i)} \rho_k^{\Delta^*(\kappa)}} = e^{\epsilon_K} e^{\Gamma \overline{F}_{\kappa}(d_i)}. \tag{13}$$

Thus from (11) and (13),

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}}^{\Delta^*(\kappa)} = \frac{e^{\epsilon_K}}{\Gamma} \sum_{i=1}^{K'} \frac{1}{d_i} e^{\Gamma \overline{F}_{\kappa}(d_i)} (e^{\rho_i^{\Delta^*(\kappa)}} - 1). \tag{14}$$

Now, by Taylor's series for exponential function, (10) and (14) we have

$$(e^{\rho_i^{\Delta^*(\kappa)}} - 1) = \sum_{u=1}^{\infty} \frac{1}{u!} \left(\frac{\Gamma \overline{\nu}_i}{K}\right)^u.$$

Thus,

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}}^{\Delta^*(\kappa)} = \frac{e^{\epsilon_K}}{\Gamma} \sum_{u=1}^{\infty} \sum_{i=1}^{K'} \frac{e^{\Gamma \overline{F}_{\kappa}(d_i)}}{d_i} \frac{1}{u!} \left(\frac{\Gamma \overline{\nu}_i}{K}\right)^u. \tag{15}$$

Let us consider $u = 1$ term in (15),

$$\begin{aligned}
&\lim_{K \to \infty} e^{\epsilon_K} \sum_{i=1}^{K'} \frac{e^{\Gamma \overline{F}_{\kappa}(d_i)}}{d_i} \frac{\overline{\nu}_i}{K} \\
&= \lim_{K \to \infty} e^{\epsilon_K} \lim_{K \to \infty} \frac{K'}{K} \lim_{K' \to \infty} \frac{1}{K'} \sum_{i=1}^{K'} \frac{e^{\Gamma \overline{F}_{\kappa}(d_i)}}{d_i} \overline{\nu}_i \\
&= \gamma \mathbb{E}\left[\frac{e^{\Gamma \overline{F}_{\kappa}(d_i)}}{d_i}\right] \mathbb{E}[\overline{\nu}_i] \text{ w.p. 1.}
\end{aligned} \tag{16}$$

Now, we discuss why (16) holds. We know that fading and arrivals are independent. Hence, for (16) to hold, it suffices to argue that d_i's are i.i.d. random variables. Note that if $\Delta^*(\kappa)$ had scheduled users based on d_i rather than on f_i, then d_i 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_i'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 uniformly at random, s_i is a deterministic function of distance, and f_i and s_i are independent, we conclude that d_i's for the chosen users are i.i.d. and each d_i is distributed as $\overline{F}_\kappa(\cdot)$. Thus, $\left\{\frac{e^{\Gamma\overline{F}_\kappa(d_i)}}{d_i}\right\}_{i=1,\ldots,K'}$ is an i.i.d. sequence. Thus, from (16)

$$\begin{aligned}&\lim_{K\to\infty} e^{\epsilon_K}\sum_{i=1}^{K'}\frac{e^{\Gamma\overline{F}_\kappa(d_i)}}{d_i}\frac{\overline{\nu}_i}{K}\\ &= \int_0^\infty \frac{1}{x}\exp(\Gamma\overline{F}_\kappa(x))\mathbf{d}\overline{F}_\kappa(x) \quad \text{w.p. 1.} \qquad (17)\end{aligned}$$

Now, let us consider $u>1$ term in (15),

$$\begin{aligned}&\lim_{K\to\infty}\frac{e^{\epsilon_K}}{u!}\sum_{i=1}^{K'}\frac{e^{\Gamma\overline{F}_\kappa(d_i)}}{d_i}\frac{(\overline{\nu}_i)^u}{K^u}\\ &= \frac{1}{u!}\lim_{K\to\infty}\frac{K'}{K^u}\lim_{K'\to\infty}\frac{1}{K'}\sum_{i=1}^{K'}\frac{e^{\Gamma\overline{F}_\kappa(d_i)}}{d_i}(\overline{\nu}_i)^u\\ &= \frac{1}{u!}\lim_{K\to\infty}\frac{K'}{K^u}\mathbb{E}\left[\frac{e^{\Gamma\overline{F}_\kappa(d_i)}}{d_i}\right]\mathbb{E}[(\overline{\nu}_i)^u] \text{ w.p. 1}\\ &= 0 \text{ w.p. 1.} \qquad (18)\end{aligned}$$

The relation (18) holds because $\mathbb{E}[(\overline{\nu}_i)^u]$ is assumed to be finite for every u. Thus the required follows from Definition 4, (15), (17) and (18). ■

Note: 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.

Theorem 4 (Monotonicity of Enenrgy Efficiency): In the mean field for every $\overline{\kappa}<\kappa$,

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}}^{\Delta^*(\kappa)} \le \left(\frac{E_b}{N_0}\right)_{\text{sys}}^{\Delta^*(\overline{\kappa})} \quad \text{w.p. 1.}$$

Because of the space constraints a detailed proof is given in [13]. Here we provide a brief outline of the proof. First, note the following property of the $\log(\cdot)$ function. Let $\{a_i\}_{i\ge1}$ denote any sequence of the non-negative real numbers and let Z be any constant. Then,

$$\sum_{i=1}^{\infty}\log\left(\frac{a_i}{Z+\sum_{u<i}a_u}\right) = \log\left(\frac{\sum_{i=1}^{\infty}a_i}{Z}\right). \qquad (19)$$

Thus, under super position coding and successive decoding, we obtain

$$\sum_{i=1}^{K'}\rho_i(t) = \log\left(\frac{\sum_{i=1}^{K'}E_i^R(t)}{N_0}\right), \qquad (20)$$

where $\rho_i(t)$ denote the rate requirement of user i in slot t and K' denote the numeber of scheduled users.

Thus, from (20) 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 equal to Γ in the mean field for every κ. This can be seen as follows.

$$\begin{aligned}\lim_{K\to\infty}\sum_{i=1}^{K'}\rho_i(t) &= \lim_{K\to\infty}\sum_{i=1}^{K'}\frac{\Gamma}{K}\overline{\nu}_i(t)\\ &= \Gamma\lim_{K\to\infty}\frac{K'}{K}\lim_{K'\to\infty}\frac{1}{K'}\sum_{i=1}^{K'}\overline{\nu}_i(t)\\ &= \Gamma \text{ w.p. 1.}\end{aligned}$$

Thus, it follows that 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 as 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.

In Theorems 1 to 4, we have not assumed anything about the distribution of f_i. In the following theorem, however, we consider an important special case where f_i has infinite support. This assumption holds for many distributions used to model multi-path effect, e.g., Rayleigh, Rician and Nakagami fadings.

Theorem 5: If f_i has infinite support, i.e. $\mathbb{P}\{f_i > u\} > 0$ for every $u<\infty$, then in the mean field

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

Because of the space constraints a detailed proof is given in [13]. We briefly provide the intuition behind the result. As observed before, the required sum received energy in every slot is the same under $\Delta^*(\kappa)$ for every κ. Moreover, as the users with a higher channel gains are selected the sum transmit energy decreases monotonically. Since f_i has infinite support, we can choose κ arbitrarily large while maintaining $\gamma>0$. Thus, we expect to reduce the sum transmit energy arbitrarily close to zero by choosing sufficiently large κ.

A. Discussion on Analytical Results

Theorem 2 states that if the delay of $D\ge1$ has to be provided, then the policy $\Delta^*(\kappa)$ can achieve it with any κ satisfying $\mathbb{P}\{f_i>\kappa\}\ge 1/D$. Now, Theorem 4 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 3 quantifies the energy efficiency of the system for the given value κ. Note that Theorems 2, 3 and 4 together quantify the delay-energy tradeoff in the multi-access channel. Finally, Theorem 5 deals with the important special case. Here, we show that given any value of energy $E > 0$, there exists κ such that $\Delta^*(\kappa)$ provides the desired rate to each user while maintaining the required sum energy below E.

V. Generalizations

Now, we discuss two important generalizations. First, we consider the system with multiple non-overlapping bands. The required rate can be split on these bands. In Section V-A, we discuss how the results in Section IV 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 multi-access channel, or when the multi-access channel serves as an intermediate hop on the multiple hops traveled by the application in the network. In Section V-B, we discuss how OSPC can support this.

A. Multiband Multi-access Channel

We consider multi-access channel with M bands. We assume that the fading on these bands is statistically indistinguishable and independent. Let f_i^m denote the fast fading for user i on m^{th} sub-band. 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 multi-access 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 $f_i(t)$, i.e., $\Delta^*(\kappa)$ selects all the users with $f_i^*(t) > \kappa$ and provides the required rate on the best channel for every user. Now, Theorems 1 and 2 hold with $\gamma \stackrel{\text{def}}{=} \mathbb{P}\{f_i^* > \kappa\}$. In Theorem 3, the energy efficiency becomes

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

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, Theorems 4 and 5 hold with f_i replaced by f_i^*.

B. Delay Differentiation

The users are divided into L classes based on their delay requirements. Let $\alpha_1,\ldots,\alpha_L$ fraction of users that want delays $D_1,\ldots,D_L$ respectively. Let κ_l be the largest real number that satisfies $\mathbb{P}\{f_i > \kappa_l\} = \frac{1}{D_l}$ for every $l \le L$. Now, OSPC can use κ_l instead of κ for users of class l. Clearly, Theorems 1 and 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 3. Now, the energy efficiency for the system is the weighted sum (with respect to α_l's) of the energy efficiency of each class. Finally, Theorems 4 and 5 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, Theorems 4 and 5 follow for the whole system.

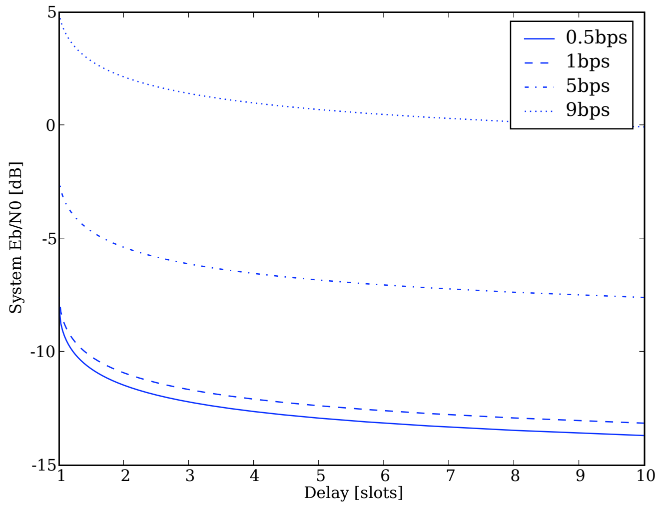

Fig. 1. System E_b/N_0 as a function of delay for various values of spectral efficiency

VI. Numerical Results

We consider a system where users are placed uniformly at random in a cell except for a forbidden region around the access point of radius $\delta = 0.01$. 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, i.e., $\mathbb{P}\{f_i(t) \le x\} = 1 - \exp(-x)$ for every t. The explicit mathematical formulations for the channel models can be found in [13]. 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 we report.

Fig. 1 demonstrates the energy delay tradeoff exhibited by OSPC. For an increase in delay from one to three slots, an energy saving of over 3dB is gained. Thus, even the small delay tolerance of the application can be

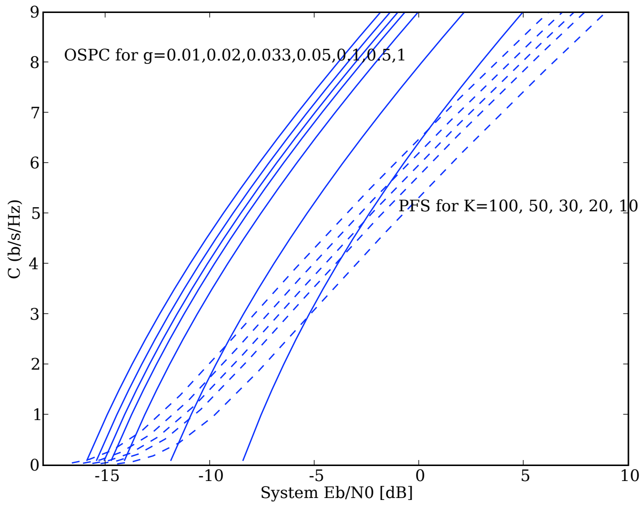

Fig. 2. Comparison between PFS (dashed) and OSPC (solid line).

exploited to obtain significant improvement in the energy efficiency of the system. Moreover, the energy required to support the given rate decreases monotonically as delay increases. This verifies Theorem 4. Note that the gain exhibits very similar behavior accross various spectral efficiencies.

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

VII. 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 OSPC 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.

References

[1] D. Tse and S. Hanly, "Multi-access fading channels-part i: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

[2] S. Hanly and D. Tse, "Multi-access fading channels-part ii: Delay-limited capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 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 Commun. Mag.*, vol. 38, no. 7, pp. 70–77, Jul. 2000.

[4] D. Tse, "Multiuser diversity through proportional fair scheduling," in *Communication theory workshop*, May 2001.

[5] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, Jun. 2002.

[6] T. Klein, K. Leung, and H. Zheng, "Enhanced scheduling algorithms for improved tcp performance in wireless ip networks," in *IEEE Globecom'04*, Dallas, TX, Nov. 2004, pp. 2744–2759.

[7] G. Caire, R. Müller, and R. Knopp, "Multiuser diversity in wireless systems: Delay-limited vs. scheduling," *to appear in IEEE Transactions on Information Theory*, 2007.

[8] P. Chaporkar and S. Sarkar, "Stable scheduling policies for maximizing throughput in generalized constrained queueing," in *IEEE INFOCOM'06*, Barcelona, Spain, Apr. 2006.

[9] M. Neely, "Energy optimal control for time varying wireless networks," in *IEEE INFOCOM'05*, Mar. 2005.

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

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

[12] 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, Oct. 1998.

[13] P. Chaporkar, K. Kansanen, and R. Müller, "Channel and multiuser diversities in wireless systems: Delay-energy tradeoff," Norwegian University of Science and Technology, Trondheim, Norway, Tech. Rep., Dec. 2006, http://www.iet.ntnu.no/~ralf/Publications.html.