Optimal power-delay trade-offs in fading channels: small delay asymptotics

Randall A. Berry
Dept. of EECS, Northwestern University
2145 Sheridan Rd., Evanston IL 60208
Email: rberry@ece.northwestern.edu

***Abstract*— When transmitting data over fading channels there is an inherent trade-off between the average transmission power and the average queueing delay of the data. This trade-off can be exploited by appropriately scheduling data transmissions over time. We study the behavior of the optimal power-delay trade-off for a single user in the regime of asymptotically small delays. In this regime, we lower bound the average power required as a function of the average delay. The rate at which this bound increases as the delay becomes asymptotically small is shown to depend on the behavior of the fading distribution near zero and the arrival statistics. We characterize this rate for two broad classes of fading distributions. For both classes, it is shown that this rate can essentially be achieved by a sequence of simple "channel threshold" policies.**

I. Introduction

In many wireless communication scenarios energy management is an important issue for reasons such as extending a device's usable life-time. Often, transmission power is one of the main energy consumers in wireless devices; consequently, there has been much interest in approaches for efficiently utilizing this resource. A basic technique for accomplishing this is through controlling the transmission rate and power over time, for example by using adaptive modulation and coding. Such approaches exploit the well-known fact that the energy per bit needed for reliable communication is decreasing in the number of degrees of freedom used to send each bit; for fixed bandwidth and packet-size, the available degrees of freedom per bit increase with a lower transmission rate. In a fading channel, another benefit of adapting the transmission rate and power is that it enables the transmitter to be "opportunistic" and send more data during good channel conditions, which again reduces the average energy per bit.

A number of energy-efficient transmission scheduling approaches have been studied in which transmission rate and/or power are adjusted over time based in part on the offered traffic as well as any available channel state information, including [1]–[7]. In each case, the goal is to effectively trade-off some cost related to packet delay (e.g. average queueing delay or a deadline by which all packets must be transmitted) with some cost related to power or energy (e.g. total energy over a finite horizon or long-term average power).

In this paper, we re-visit the model for energy-efficient transmission scheduling over a fading channel from [7]. In this

This research was supported by NSF CAREER award CCR-0238382.

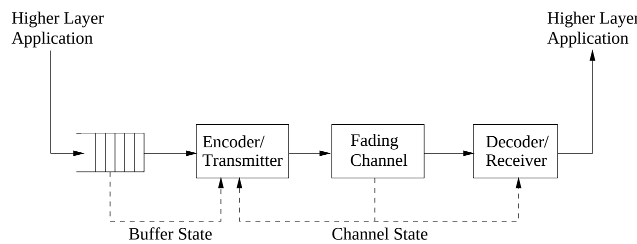

Fig. 1. System Model.

model, randomly arriving data is placed into a transmission buffer as shown in Fig. 1. Periodically, some data is removed from the buffer, encoded and transmitted over the fading channel. The transmitter can vary the transmission power and rate based on both the channel state and the buffer occupancy. As in [7], we consider the optimal power/delay trade-off, $P^*(D)$. This characterizes the minimum long-term average power under any scheduling policy as a function of the average queueing delay, for a given arrival process and channel fading process. When the channel and arrival processes are not both constant, $P^*(D)$, will be a strictly decreasing, convex function of D. In [7], the behavior of $P^*(D)$ was studied in the asymptotic regime of large delays (low power). In this regime, $P^*(D)$ approaches a limiting value of $\mathcal{P}(\bar{A})$ at rate of $\Theta\left(\frac{1}{D^2}\right)$. In [7] it was shown that this rate can be achieved with a sequence of "buffer threshold policies" whose only dependence on the buffer occupancy is via a simple threshold rule. Moreover, this weak dependence on the buffer occupancy is required for a sequence of policies to be order optimal.

Here, we focus on the behavior of $P^*(D)$ in the asymptotic regime of small delays (high power). Specifically, we study the optimal rate at which the average delay decreases to its minimum value as the average power increases. The analysis of the large delay asymptotics in [7] is based, in part, on using large deviation bounds on the buffer occupancy, which are asymptotically tight for large buffer sizes. In the small delay regime, these bounds are not very useful and we instead take a different approach. In this regime, $P^*(D)$ will be shown to behave quite differently from the large delay regime. In particular, the convergence rate depends strongly on the behavior of the fading distribution near zero. We focus on two broad classes of channels: one class ("type A") that requires infinite power to minimize the queueing delay, and one class ("type B") for which the queueing delay can be minimized with finite power. These classes include most common fading models, such as Rayleigh, Ricean and Nagakami fading. For each class, we lower bound the convergence rate in the small delay regime that can be achieved by any transmission policy. We then show that this bound is achievable for both classes when using a sequence of "channel threshold policies," under which the transmission rate only depends on the channel gain through a simple threshold rule. This demonstrates an interesting duality with the large delay regime, where instead buffer threshold policies were order optimal. For type A channels, we then show that an even simpler fixed-rate channel threshold policy is also order optimal. However, such policies are not order optimal for type B channels. Due to space considerations, we omit the proofs; these can be found in [11].

II. Problem Formulation

We begin by describing our model for the system in Figure 1. The channel is modeled as a discrete-time, block-fading channel with additive white Gaussian noise and frequency-flat fading. Over each block of N channel uses, the channel gain stays fixed. Let $\sqrt{H_n}$ denote the magnitude of the base-band channel gain during the nth block. Let $\mathbf{X}_n = (X_{n,1}, \ldots, X_{n,N})$ and $\mathbf{Y}_n = (Y_{n,1}, \ldots, Y_{n,N})$ be vectors in $\mathbb{C}^N$ which denote, respectively, the channel inputs and outputs over the nth block. These are related by:

$$\mathbf{Y}_n = \sqrt{H_n}\mathbf{X}_n + \mathbf{Z}_n. \tag{1}$$

Here the additive noise $\mathbf{Z}_n$ is a complex, circularly symmetric Gaussian random vector with zero mean and covariance matrix $\sigma^2 I$, where I denotes a $N \times N$ identity matrix. The sequence $\{\mathbf{Z}_n\}$ is independent and identically distributed (i.i.d.). The sequence of channel gains, $\{H_n\}$, is also a assumed to be i.i.d., i.e. the time for one block of N channel uses is on the order of one coherence-time. For all n, H_n takes values in $\mathcal{H} = \mathbb{R}^+$ and have a continuous probability density function $f_H(h)$ and cumulative distribution $F_H(h)$. We assume that $f_H(h) > 0$ for all $h > 0$, which is true for most channel models of interest. This implies that $F_H(h)$ is strictly increasing over $\mathcal{H}$. Both the transmitter and receiver have perfect channel state information (CSI), *i.e.*, during the nth block, the transmitter and receiver know the value of H_n.

We consider a discrete-time "fluid" model for the buffer in which time is slotted and the length of each time-slot corresponds to a block of N channel uses.[1] Let A_n be the number of bits that arrive between time n and $n-1$, and let S_n be the buffer size at the start of the nth time-slot. Denote by U_n the number of bits removed from the buffer at the start of each time-slot and transmitted over the fading channel during the time-slot. The resulting buffer dynamics are

$$S_{n+1} = \max\{S_n + A_{n+1} - U_n, A_{n+1}\}, \tag{2}$$

[1]This is a fluid model because we do not restrict the amount of "bits" that arrive to or depart from the buffer during a time-slot to be an integer.

which ensures that the arriving data (A_{n+1}) waits in the buffer for at least one time-unit. The buffer size is assumed to be infinite and we denote the buffer state space by $\mathcal{S} = \mathbb{R}^+$. We consider the case where the arrival process $\{A_n\}$ is a sequence of i.i.d. random variables taking values in a compact set $\mathcal{A} = [a_{min}, a_{max}] \subset \mathbb{R}^+$ with probability distribution $F_A(a)$ and mean $\bar{A}$. Here, a_{max} and a_{min} are, respectively, upper and lower bounds on the amount of data that can arrive per time-unit. This process is independent of both the channel fading process and the noise process.

We assume a function $S(r)$ specifies the received signal-to-noise ratio (SNR) required for the transmitter to reliably transmit at a rate of r bits per channel use during a given time-slot. The main example we consider is

$$S(r) = 2^r - 1, \tag{3}$$

which is the received SNR required for the Gaussian channel in (1) to have a capacity of r bits per channel use. More generally, the following analysis will hold for any function $S(r)$ that satisfies the following regularity property:

Definition 1: A SNR function $S(r)$ is *regular* if $S(r)$ is increasing, differentiable, and strictly convex with $S(0) = 0$, $S'(0) > 0$, and $\lim_{r\to\infty} S'(r) = \infty$.

In addition to (3), most practical modulation and coding schemes will satisfy this definition. In a time-slot when the channel gain is h, the received SNR is $\frac{hP}{\sigma^2}$, where P is the transmission power. Thus, the required transmission power to send u bits during this time-slot is $P(h,u) := \frac{\sigma^2}{h} S(u/N)$. In the case of (3),

$$P(h,u) = \frac{\sigma^2}{h}\left(2^{u/N} - 1\right), \tag{4}$$

which is the power required so that the mutual information rate in the given block is equal to u/N. Provided that N is large enough, this choice will give a reasonable indication of the power needed to reliably transmit at rate u/N. One may question the reasonableness of modeling the required power using (4) when we are analyzing the performance of a system in the regime of small delays, since to communicate reliably at rates near capacity typically requires the use of long codes and subsequently long delays. The main justification for this is that we are measuring delays on the time-scale of the queue dynamics in (2); within each time-step of this model, we assume that there are sufficient degrees of freedom available to use sophisticated coding.

Let $\mu : \mathcal{S} \times \mathcal{H} \mapsto \mathbb{R}^+$ denote a stationary (Markov) transmission policy that indicates U_n as a function of S_n and H_n. Under such a policy, $\{S_n\}$ will be a Markov chain. Under policy μ, the time-average transmission power is

$$\bar{P}^\mu := \limsup_{m\to\infty} \frac{1}{m}\sum_{n=1}^{m} \mathbb{E}(P(H_n, \mu(S_n, H_n))),$$

and the time-average delay is

$$\bar{D}^\mu := \limsup_{m\to\infty} \frac{1}{m}\sum_{n=1}^{m} \frac{\mathbb{E}(S_n)}{\bar{A}}.$$

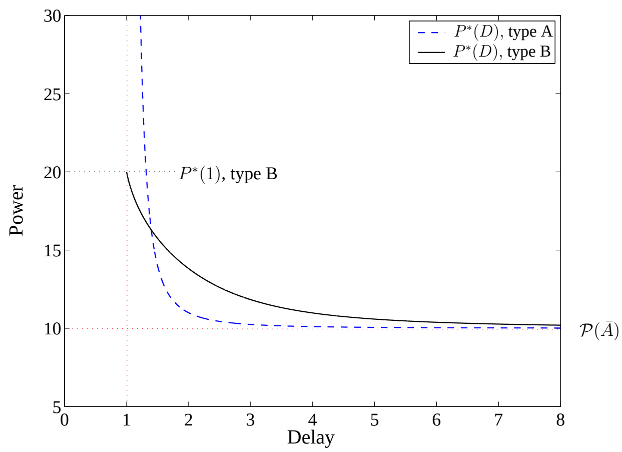

Fig. 2. Examples of $P^*(D)$ for two different channels. In the type A channel, the optimal power grows without bound as the delay approaches 1. In the type B channel the optimal power converges to the limit of $P^*(1)$.

Assuming that $\{S_n\}$ is ergodic, it follows that that $\bar{P}^\mu = \mathbb{E}_{S,H} P(S,H)$ and $\bar{D}^\mu = \frac{\mathbb{E}S}{\bar{A}}$, which is equal to the average queueing delay by Little's law.[2] For a given channel and arrival process, the optimal power/delay trade-off is

$$P^*(D) := \inf\{\bar{P}^\mu : \mu \text{ such that } \bar{D}^\mu \leq D\}.$$

This will be a decreasing, convex function as shown in Fig. 2. As $D \to \infty$, $P^*(D)$ converges to $\mathcal{P}(\bar{A})$ at a rate of $\Theta\left(\frac{1}{D^2}\right)$ [7]. The asymptotic value, $\mathcal{P}(\bar{A})$ corresponds to the minimum power required to send at average rate $\bar{A}$, ignoring any delay constraints. When $P(h,u)$ is given by (4), this is the minimum power so that the channel has a *throughput capacity* of $\bar{A}/N$ bits per channel use [8].

It will also be useful to define the *optimal delay/power trade-off*,

$$D^*(P) := \inf\{\bar{D}^\mu : \mu \text{ such that } \bar{P}^\mu \leq P\}.$$

Clearly, if $P^*(D)$ is strictly decreasing, then $D^*(P)$ will simply be its inverse. From (2), all data must spend at least one time unit in the buffer, hence $D^*(P) \geq 1$ for all P. The only way that $D^*(P) = 1$ is if $\mu(S_n, H_n) \geq A_n$ for all n, i.e. every bit is transmitted the time-slot after it arrives. The minimum power required by such a policy is $P^*(1) = \mathbb{E}_{A,H} P(H,A)$. For a constant arrival rate of $\bar{A}$ and when $P(h,u)$ is given by (4), $P^*(1)$ represents the minimum power needed for the channel to have a *delay-limited capacity* of $\bar{A}/N$ bits per channel use [9].

Depending on $f_H(h)$, $P^*(1)$ may or may not be finite. Since the arrival and channel processes are independent, $P^*(1) = \sigma^2 \mathbb{E}_H\left(\frac{1}{H}\right) \mathbb{E}_A S(A/N)$. Since $\{A_n\}$ is bounded, $P^*(1)$ will be finite if and only if $\mathbb{E}_H\left(\frac{1}{H}\right) < \infty$. Therefore, every fading distribution can be classified as either having a *positive delay-limited capacity*, if $\mathbb{E}_H\left(\frac{1}{H}\right) < \infty$, or otherwise having a *zero delay limited capacity*. We focus mainly on the following two types of distributions:

Definition 2: A channel is defined to be *type A* if H_n has a finite mean and $f_H(0) > 0$.

Definition 3: A channel is defined to be *type B* if H_n has a finite mean and $f_H(h) = \Theta(h^\gamma)$ as $h \to 0$ for some $\gamma > 0$.

It can be seen that type A channels have zero delay-limited capacity and type B have positive delay-limited capacity. Examples of type A channels include Rayleigh and Ricean fading channels. A Nakagami fading channel will also be type A if the Nakagami fading figure, m is less than 1. It will be type B when $m \geq 1$; in this case $\gamma = m - 1$. A Rayleigh channel with $m > 1$ independent diversity branches will also be type B with $\gamma = m - 1$, when either selection diversity or maximal ratio combining are used.

For a given fading density, let $G(h) := \int_h^\infty \frac{1}{\tilde{h}} f_H(\tilde{h})\, d\tilde{h}$. As $h \to 0$, $G(h) \to \mathbb{E}\left(\frac{1}{H}\right)$, which is infinite for any channel with zero delay-limited capacity. Let $G^{-1}(x)$ denote the inverse of the function $G(h)$. Since, by assumption $f_H(h) > 0$ for all $h > 0$, $G(h)$ will be strictly decreasing and approach zero as $h \to \infty$ and hence $G^{-1}(x)$ is defined for $x \in [0, G(0)]$.

III. Lower bounds

Next, we give a lower bound on $D^*(P)$ that applies for any channel and arrival processes satisfy the previous assumptions and becomes tight as $P \to P^*(1)$. We then use this to bound the rate at which $D^*(P) \to 1$ for type A and B channels.

Proposition 1: Consider a system with a regular SNR function $S(r)$. For any $P \leq P^*(1)$,

$$\begin{aligned} & D^*(P) - 1 \\ & \geq F_H\left[\left(\frac{S'(0)}{S'(a_{max}/N)}\right) G^{-1}\left(\frac{P}{\sigma^2 \mathbb{E}_A(S(A/N))}\right)\right]. \end{aligned}$$

Note that $\sigma^2 \mathbb{E}_A S(A/N)$ is the value of $P^*(1)$ for a channel in which $H_n = 1$ for all n. In a channel with positive delay-limited capacity, this satisfies $\frac{P^*(1)}{\sigma^2 \mathbb{E}_A S(A/N)} = G(0)$, and so for $P = P^*(1)$ this bound equals zero and is tight. Likewise, for a channel with zero delay limited capacity, $P^*(1) = \infty$. Hence, as $P \to \infty$, the bound approaches zero and is once again tight.

The proof of this proposition is based on considering a "fictitious system," which is identical to the original system except that all arriving data can be transmitted after waiting for 2 time-units without requiring any power. However, to transmit the data after one time-unit still requires the same power as in the original system (recall that all data must wait at least one time-unit). Therefore, the maximum delay in the fictitious system will be no more than 2 time-units. Let $\hat{D}(P)$ be the minimum average delay in this fictitious system under any transmission policy with average power no greater than P. Clearly, for the same arrival and channel processes, $\hat{D}(P) \leq D^*(P)$ for all P. Using this, we then lower bound $D^*(P)$ by lower bounding $\hat{D}(P)$.

In the fictitious system, clearly, an optimal policy will set $U_n \leq A_n$, since any other data in the buffer will leave the

[2] Here and in the following, given an ergodic process, $\{X_n\}$, we denote by X (without an index), a random variable with the corresponding steady-state distribution.

system anyway without requiring any power. It follows that an optimal policy for the fictitious system can be expressed as function of the current channel state, H_n and the number of arrivals A_n, and must be a solution to the convex program:

$$\begin{aligned}&\underset{\zeta:\mathcal{H}\times\mathcal{A}\mapsto\mathbb{R}^+}{\text{maximize}}\ \mathbb{E}_{H,A}\zeta(H,A)\\&\text{subject to: } \mathbb{E}_{H,A}\frac{\sigma^2}{H}S(\zeta(H,A)/N)\le P \qquad (5)\\&\qquad\qquad 0\le\zeta(h,a)\le a,\quad \forall h\in\mathcal{H}, a\in\mathcal{A}.\end{aligned}$$

Using standard Lagrangian techniques, the solution to this can be characterized.[3] Using this policy, we can derive the bound in Proposition 1.

The next two corollaries to Proposition 1 bound the rate at which $D^*(P)\to 1$ for both type A and type B channels.

Corollary 1: For a type A channel, as $P\to\infty$, $D^*(P)-1=\Omega(e^{-\alpha P})$, for any $\alpha>(\sigma^2 f_H(0)\mathbb{E}_A(S(A/N)))^{-1}$.

The result of this corollary can equivalently be expressed in terms of the power/delay trade-off, i.e. for a type A channel, $P^*(D)=\Omega(\ln(\frac{1}{D-1}))$ as $D\to 1$. Note that in this case the constant α is not needed.

Corollary 2: For a type B channel with parameter $\gamma>0$, as $P\to P^*(1)$ (from below), $D^*(P)-1=\Omega((P^*(1)-P)^{\frac{\gamma+1}{\gamma}})$.

The exponent $\frac{\gamma+1}{\gamma}$ is decreasing in γ, and so this bound will approach 1 slower in channels with larger values of γ.

IV. Optimal and Sub-optimal Sequences of Transmission Policies

In the previous section, we bounded $D^*(P)$ by studying an optimal transmission policy for a fictitious system. Next, we study several transmission policies for the actual system and compare their performance to the bounds.

A. Channel Threshold Policies

The first type of policies we consider are *channel threshold* policies in which transmission occurs only when the channel gain is greater than a given threshold; when this occurs the transmitter empties the buffer. More precisely, $\mu_k:\mathcal{H}\times\mathcal{S}\to\mathbb{R}^+$ is a channel threshold policy with threshold, $h_k\ge 0$, if $\mu_{h_k}(h,s)=s1_{\{h>h_k\}}$, where 1_A denotes the indicator function of the event A.

We define a *decreasing sequence* of channel threshold policies $\{\mu_k\}$ to be a sequence where the associated thresholds h_k are decreasing with $\lim_{k\to\infty}h_k=0$. Clearly, the average delay will decrease with $\bar{D}^{\mu_k}\to 1$ as $k\to\infty$. Next, we characterize the rate at which this converges as a function of the average power for type A and B channels. For this we make one additional assumption on the SNR function S.

Definition 4: A regular SNR function $S(r)$ has *exponentially bounded growth* if there exists non-negative constants M and κ such that for all $r\ge 0$, $S(r)\le M\kappa^r$.

For example, when $S(r)$ is given by (3), it satisfies this definition with $M=1$ and $\kappa=2$. This ensures that the expected SNR required to transmit all of the data which arrived in n time-slots grows at most exponentially in n.

Proposition 2: For a type A channel, if the SNR function has exponentially bounded growth, then for any decreasing sequence of channel threshold policies $\{\mu_k\}$, as $K\to\infty$, $\bar{P}^{\mu_k}\to\infty$ and $\bar{D}^{\mu_k}-1=O(\exp(-\alpha\bar{P}^{\mu_k}))$, for any $\alpha<(\sigma^2 f_H(0)\mathbb{E}_A S(A/N))^{-1}$.

This proposition implies that any decreasing sequence of channel threshold policies is nearly order optimal in the sense that the exponent α can be arbitrarily close to the bound in Corollary 1. If instead we consider the power/delay trade-off, then Proposition 2 implies that for any decreasing sequence of channel threshold policies, $\bar{P}^{\mu_k}=O(\ln(\frac{1}{\bar{D}^{\mu_k}-1}))$. In other words, in terms of the power/delay trade-off, these policies are order optimal. Therefore, in the small delay regime, the optimal convergence rate of $P^*(D)$ for type A channels is $\Theta(\ln(\frac{1}{D-1}))$. Note that this is a much faster rate of change than the $\frac{1}{D^2}$ behavior in the large delay regime.

Proposition 3: For a type B channel with parameter $\gamma>0$, if the SNR function has exponentially bounded growth, then for any decreasing sequence of channel threshold policies $\{\mu_k\}$, as $K\to\infty$, $\bar{P}^{\mu_k}\to P^*(1)$ and $\bar{D}^{\mu_k}-1=O(\left(P^*(1)-\bar{P}^{\mu_k}\right)^{\frac{\gamma+1}{\gamma}})$.

Comparing to Corollary 2, it follows that for a type B channel, decreasing sequences of channel threshold policies are order optimal. Therefore, $D^*(P)-1=\Theta((P^*(1)-P)^{\frac{\gamma+1}{\gamma}})$ as $P\to P^*(1)$ from below. Equivalently, $P^*(1)-P^*(D)=\Theta((D-1)^{\frac{\gamma}{\gamma+1}})$ as $D\to 1$. As noted previously, $\frac{\gamma+1}{\gamma}$ is decreasing to 1 as γ increases. This implies that in a Rayleigh fading channel as the number of independent diversity branches are increased, $D^*(P)$ will approach 1 at a slower rate. For a large number of diversity branches, the rate will be approximately linear in $P^*(1)-P$. Of course, $P^*(1)$ also decreases with additional diversity branches.

B. Policies with bounded transmission rates

Under a channel threshold policy, since the transmitter empties the buffer whenever it transmits, the required transmission rate can be arbitrarily large. Next, we look at a sequence of policies with bounded transmission rates. For example, such limits may be due constraints on the available coding and modulation schemes. We consider *bounded rate channel threshold policies* where the transmitter once again only transmits when the channel gain is larger than a threshold h_k. When transmitting, these policies set $U_n=A_n+\delta$, i.e. they transmit at most $a_{max}+\delta$ bits. We denote such a policy by $\phi_k(h,a)=(a+\delta)1_{\{h>h_k\}}$. Note that when $s_n/N<A_n+\delta$, there is not enough information in the buffer to transmit. In this case we assume the transmitter sends extra "dummy" bits. This is clearly a poor choice from the view of saving power, but is sufficient for our purposes.[4]

[3] Indeed, in the case where $P(h,u)$ is given by (4), the optimal policy can be related to a combination of the well-known "water-filling" power allocation and a "channel inversion" policy.

[4] Note that these policies do not base the transmission decision on only H_n and S_n, but can be viewed as using the past history of S_n and U_n. Such a dependence is not needed for an optimal policy.

We again consider a decreasing sequence of policies $\{\phi_k\}$, where the thresholds h_k decrease to 0. To study the performance of these policies, we use the following lemma which applies to any policy for which $U_n - A_n$ is an i.i.d. sequence. This is the case here, since $U_n - A_n$ depends only on H_n.

Lemma 1: For any policy where $\Delta_n = U_n - A_n$ is an i.i.d. sequence, the average buffer occupancy is bounded by

$$\frac{\mathbb{E}\{([-\Delta]^+)^2\}}{2\mathbb{E}(\Delta)} \leq \mathbb{E}S - \bar{A} \leq \frac{\sigma_\Delta^2}{2(\mathbb{E}\Delta)},$$

where $[-\Delta]^+ = \max(-\Delta, 0)$ and σ_Δ^2 is variance of Δ_n.

This Lemma follows from noting that the buffer dynamics in (2) is similar to the evolution of a Lindley process which models the delay in a continuous-time GI/G/1 queue. These bounds are essentially the same as Kingman's bounds on the average delay for such a GI/G/1 system [10].

Proposition 4: For a type A channel, let $\{\phi_k\}$ be a decreasing sequence of bounded rate channel threshold policies. Then as $k \to \infty$, $\bar{P}^{\phi_k} \to \infty$, and $\bar{D}^{\phi_k} - 1 = O(\exp(-\alpha \bar{P}^{\phi_k}))$, for any $\alpha < (\sigma^2 f_H(0)\mathbb{E}_A\{S(A/N)\})^{-1}$.

This implies that for type A channels, bounded rate channel threshold policies can achieve the same order of convergence as a channel threshold policy, which we have seen is essentially order optimal. Next, we define a *fixed-rate, channel threshold policy* $\tilde{\phi}_k(h)$ to be a policy that transmits at a fixed rate $\tilde{a}/N$, whenever the channel gain is greater than h_k, and sends nothing otherwise. This differs from the previous bounded rate policies in that the rate does not depend on A_n.

Corollary 3: For a type A channel, let $\{\tilde{\phi}_k\}$ be a decreasing sequence of fixed-rate, channel threshold policies with $\tilde{a} > a_{max}$. As $k \to \infty$, $\bar{P}^{\tilde{\phi}_k} \to \infty$ and $\bar{D}^{\tilde{\phi}_k} - 1 = O(\exp(-\alpha \bar{P}^{\tilde{\phi}_k}))$ for any $\alpha < (\sigma^2 f_H(0) S(\tilde{a}/N))^{-1}$.

In Corollary 3, the constraint on α is smaller than in Proposition 4, unless the arrival process is constant. Thus, in general this does not imply than these policies are order optimal in terms of the delay/power trade-off. However, in terms of the power/delay trade-off, we can ignore the α parameter so that these policies are order optimal in this sense. These policies do not depend on the buffer occupancy at all; this illustrates a significant difference between the small delay and large delay regimes; in the large delay regime some buffer dependence is required for any order optimal policy [7].

Next we turn to type B channels. For a type B channel or any other channel with a positive delay-limited capacity, a decreasing sequence of $\{\phi_k\}$ of bounded rate channel threshold policies, with a fixed parameter $\delta > 0$, will satisfy

$$\lim_{k\to\infty} \bar{P}^{\phi_k} = \mathbb{E}_{A,H} P(H, A+\delta) > P^*(1).$$

Therefore, any such sequence is clearly not order optimal in the small delay regime. The problem here is that in the small delay limit the power wasted on transmitting extra "dummy" bits becomes significant for type B channels, while we could ignore this in type A channels.

A better approach for such channels is as k increases to reduce both h_k as well as the parameter $\delta = \delta_k$, with $\lim_{k\to\infty} \delta_k = 0$. In this way, as $k \to \infty$, $\bar{P}^{\phi_k} \to P^*(1)$. However, as the following proposition states, such a sequence of policies still do not achieve the optimal convergence rate for type B channels. Here, for simplicity, we restrict ourselves to the case where $S(r)$ is given by (3).

Proposition 5: For a type B channel with parameter $\gamma > 0$ and $S(r)$ given by (3), let ϕ_k be a decreasing sequence of bounded rate channel threshold policies with decreasing parameters δ_k, where $\delta_k \to 0$. If as $k \to \infty$, $\bar{P}^{\phi_k} \to P^*(1)$ from below, then $\bar{D}^{\phi_k} - 1 = \Omega((P^*(1) - \bar{P}^{\phi_k})^{\frac{1}{\gamma}})$.

Note that since $\gamma > 0$, $\frac{1}{\gamma}$ is strictly less than the optimal exponent of $\frac{\gamma+1}{\gamma}$ given by Corollary 2, and so these policies are not order optimal for type B channels. This illustrates a basic difference between type A and B channels.

V. Conclusions

We have analyzed the optimal power/delay trade-off for a single user fading channel in the regime of small delays and large power. In this regime, the optimal trade-off was shown to strongly depend on the behavior of the fading distribution near zero. For two broad classes of channels, we bounded the asymptotic rate at which the average delay approaches its minimum power as the average power increases. This bound was shown to be achievable when using a sequence of simple channel threshold policies. Here we have focused on a single user communicating over a memoryless fading channel with only a long-term average power constraint. Potential directions for future work include relaxing these modeling assumptions, e.g. considering multi-user systems or channels with memory. Another possible direction is to consider models with imperfect channel knowledge, in which case outages may occur requiring data to be retransmitted.

References

[1] B. Collins and R. Cruz, "Transmission Policies for Time Varying Channels with Average Delay Constraints," in *Proc. 1999 Allerton Conf. on Commun. Control, & Comp.*, (Monticello, IL), 1999.

[2] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 124–144, Jan 2004.

[3] B. Prabhakar, E. Uysal-Biyikoglu, and A. E. Gamal, "Energy-efficient transmission over a wireless link via lazy packet scheduling," in *Proc. IEEE Infocom 2001*, 2001.

[4] V. Goyal, A. Kumar, and V. Sharma, "Power constrained and delay optimal policies for scheduling transmission over a fading channel," in *Proc. of IEEE INFOCOM 2003*, (San Francisco), Mar. 2003.

[5] A. Fu, E. Modiano, and J. Tsitsiklis, "Optimal energy allocation and admission control for communications satellites," *IEEE Trans. on Networking*, vol. 11, pp. 488–501, June 2003.

[6] E. Yeh and T. Klein, "Optimal trade-off between energy efficiency and average delay," in *Proc. of WiOpt'03*, (Sophia-Antipolis, France), 2003.

[7] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inf. Th.*, vol. 48, pp. 1135–1149, May 2002.

[8] D. Tse and S. Hanly, "Multi-access Fading Channels: Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities," *IEEE Trans. Inf. Th.*, vol. 44, pp. 2796–2815, Nov. 1998.

[9] S. Hanly and D. Tse, "Multi-access Fading Channels: Part II: Delay-Limited Capacities," *IEEE Trans. Inf. Th.*, vol. 44, pp. 2816–2831, Nov. 1998.

[10] L. Kleinrock, *Queueing Systems*, vol. II. John Wiley and Sons, 1976.

[11] R. Berry, "Optimal power-delay trade-offs in fading channels: small delay asymptotics," under submission.