Minimum Energy Transmission over a Wireless Channel with Deadline and Power Constraints

Murtaza Zafer *Member, IEEE,* and Eytan Modiano *Senior Member, IEEE*

***Abstract*—We consider optimal rate-control for energy-efficient transmission of data, over a time-varying channel, with packet-deadline constraints. Specifically, the problem scenario consists of a wireless transmitter with B units of data that must be transmitted by deadline T over a fading channel. The transmitter can control the transmission rate over time and the required instantaneous power depends on the chosen rate and the present channel condition, with limits on short-term average power consumption. The objective is to obtain the optimal rate-control policy that minimizes the total energy expenditure while ensuring that the deadline constraint is met. Using a continuous-time stochastic control formulation and a Lagrangian duality approach, we explicitly obtain the optimal policy and show that it possesses a very simple and intuitive form. Finally, we present an illustrative simulation example comparing the energy costs of the optimal policy with the full power policy.**

***Index Terms*—Energy, Delay, Rate control, Deadline, Wireless channel, Quality of Service**

I. INTRODUCTION

Real-time data communication over wireless networks, inherently, involves dealing with packet-delay constraints, time-varying stochastic channel conditions and scarcity of resources; one of the important resource constraint is the transmission energy expenditure [1], [2]. In principle, for a point-to-point wireless link, packet-deadline constraints can always be met by transmitting at high rates, however, such an approach leads to higher transmission energy cost. When the transmitter has energy limitations, one can instead utilize rate-control to minimize the energy expenditure. Clearly, minimizing the energy cost has numerous advantages in efficient battery utilization of mobile devices, increased lifetime of sensor nodes and mobile ad-hoc networks, and efficient utilization of energy sources in satellites.

In modern wireless devices, rate-control can be achieved in many ways that include adjusting the power level, symbol rate, coding rate/scheme, signal-constellation size and any combination of these approaches. Furthermore, in some technologies, the receiver can detect these changes directly from the received data without the need for an explicit rate-change control information [6]. In fact, with present technology, rate-control can be achieved very rapidly over time-slots of a few millisecond duration [3], thereby, providing a unique opportunity to utilize dynamic rate-control algorithms.

M. Zafer (mzafer@us.ibm.com) is a Research Staff Member at IBM T. J. Watson Research Center, Hawthorne, USA; and E. Modiano (modiano@mit.edu) is an Associate Professor at Massachusetts Institute of Technology, Cambridge, USA. This work was supported by NSF ITR grant CCR-0325401, by DARPA/AFOSR through the University of Illinois grant no. F49620-02-1-0325, by ONR grant number N000140610064 and by ARO Muri grant number W911NF-08-1-0238

Associated with a transmission rate, there is a corresponding power expenditure which is governed by the power-rate function. Specifically, a power-rate function is a relationship which gives the amount of transmission power that would be required to transmit at a certain rate for a given bit-error probability. Two fundamental aspects of this function, which are exhibited by most encoding/communication schemes and hence are common assumptions in the literature [7]–[12], [16], [18], are as follows. First, for a fixed bit-error probability and channel state, the required transmission power is a non-negative, increasing, convex function of the rate. This implies, from Jensen's inequality, that transmitting data at a low rate over a longer duration is more energy-efficient as compared to a high rate transmission. Second, the wireless channel is time-varying which varies the convex power-rate curves as a function of the channel state. As good channel conditions require less transmission power, one can exploit this variability over time by adapting the rate in response to the channel conditions. Thus, we see that by intelligently adapting the transmission rate over time, energy cost can be reduced.

In this paper, we consider the following setup: The transmitter has B units of data that must be transmitted by deadline T over a wireless fading channel. The channel state (which is defined in Section II-B) is stochastic and modelled as a Markov process. The transmission rate can be controlled over time and the expended power depends on both the chosen rate and the present channel condition. The transmitter has short-term average power limits and the objective is to dynamically adapt the rate over time such that the transmission energy cost is minimized and the deadline constraint is met. To address this problem, we consider a continuous-time stochastic control formulation and utilize Lagrangian duality to obtain the optimal policy. The optimal rate function takes the simple and intuitive form, given as,

*optimal rate = amount of data left * urgency of transmission*

where the urgency functions can be computed offline as the solution of a system of ordinary differential equations. The problem described above is a canonical problem with applications in a wide variety of settings. One of the primary motivations is a real-time monitoring scenario, where data is collected by the sensor nodes and must be transmitted to a central processing node within a certain fixed time interval. Energy-efficiency here translates into a higher lifetime of the sensor devices and the network as a whole. Similarly, in case of wireless data networks, mobile devices running real-time applications such as video streaming and Voice-over-IP generate data packets with deadline constraints on them.

Minimizing the transmission energy cost here directly leads to an efficient utilization of the limited battery energy. Finally, for satellite networks, where there are stringent limitations on stored energy, the above problem has significance for applications involving delay-constrained data communication.

Transmission power and rate control are an active area of research and have been studied earlier in the context of network stability [15], [16], average throughput [17], average delay [7], [12] and packet drop probability [18]. However, this literature considers "average metrics" that are measured over an infinite time horizon and hence do not apply for deadline constrained data. As data services over wireless networks are evolving and real-time applications are being introduced, there is a strong need for addressing communication issues associated with packet delays. In particular, with strict deadlines, rate adaptation simply based on steady-state probability distribution of the channel states does not suffice and one needs to take into account the system dynamics over time, thus introducing new challenges and complexity into the problem. Recent work in this direction includes [5], [8]–[11]. The work in [8] studied offline formulations under non-causal knowledge of the future channel states and devised heuristic online policies using the optimal offline solution. The authors in [9] studied several data transmission problems using discrete-time Dynamic Programming (DP). However, the problem that we consider in this work becomes intractable using this methodology, due to the large state space in the DP-formulation or the well-known "curse of dimensionality". The works in [5], [10], [11] studied formulations without channel fading and in particular, in the work in [5] we used a calculus approach to obtain minimum energy policies with general arrival curves and quality-of-service constraints. This paper generalizes our earlier work in [13] by incorporating explicit short-term average power constraints which arise in practice due to limitations on energy consumption in batteries. As compared to [13], the additional complexity arising due to power constraints is addressed here using a Lagrange duality approach in combination with a stochastic control formulation. Part of the work in this paper has been presented earlier in [14].

The rest of the paper is organized as follows. In Section II, we give a description of the problem setup, while, in Section III, we utilize techniques from stochastic optimal control and Lagrange duality to obtain the optimal policy. In Section IV, we give simulation results illustrating the gains achieved by the optimal policy, and, finally in Section V we conclude the paper.

II. Problem Setup

We consider a continuous-time model of the system and assume that the rate can be varied continuously in time. Clearly, such a model is an approximation of a communication system that operates in discrete time-slots, however, the assumption is justified since in practice the time-slot durations are on the order of 1 msec [3], and much smaller than packet delay requirements which are on the order of 100's of msec. Thus, one can view the system as operating in continuous-time. Such a model is advantageous as it makes the problem

modulation	bits/symbol	SNR/symbol
2 PAM	1	$0.25\ d^2$
4 QAM	2	$0.50\ d^2$
16 QAM	4	$1.25\ d^2$
64 QAM	6	$5.25\ d^2$

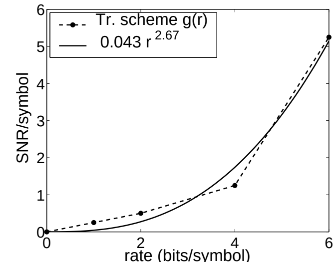


Fig. 1. Modulation scheme considered in [16] as given in the table. The corresponding plot shows the least squares monomial fit, $0.043r^{2.67}$, to the scaled piecewise linear power-rate curve.

mathematically tractable and yields simple solutions, which can be applied in practice, in a straightforward manner, by evaluating the rate functions at discrete time-slots, as done for the simulations in Section IV.

To proceed, in the next section, we describe the transmission model, followed by the Markov model for the channel evolution and finally give a detailed description of the mathematical formulation of the problem.

A. Transmission Model

Let $h(t)$ denote the channel gain between the transmitter and the receiver, $P(t)$ the transmitted signal power and $P^{rcd}(t)$ the received signal power at time t. We make the common assumption [7], [8], [10]–[12], [16] that the required received signal power for reliable communication (with a fixed bit-error probability) is convex in the rate, i.e. $P^{rcd}(t) = g(r(t))$. Since the received signal power is given by, $P^{rcd}(t) = |h(t)|^2 P(t)$, the required transmission power to achieve rate $r(t)$ is given by,

$$P(t) = \frac{g(r(t))}{c(t)} \tag{1}$$

where $c(t)$ is defined as, $c(t) \triangleq |h(t)|^2$, and $g(r)$ is a non-negative, convex, increasing function for $r \geq 0$. The quantity $c(t)$ is referred to as the *channel state* at time t. Its value at time t is assumed known through channel measurement (based on receiver feedback, pilot measurement or other sophisticated schemes) but evolves stochastically in the future. It is worth emphasizing that the power-rate relationship in (1) encompasses much more generality than discussed above. For example, $c(t)$ could represent a combination of stochastic variations in the system and (uncontrollable) interference from other transmitter-receiver pairs, as long as the power-rate relationship obeys (1).

The function $g(r)$ depends on the transmission scheme utilized (modulation, coding, etc.), and there is no single analytical expression that describes all the schemes. The Shannon formula, a generally used model [8] for which $g(r)$ takes an exponential form, applies for the ideal coding scenario. In this paper, we consider a simplification and take $g(r)$ within the class of *monomial functions*, namely, $g(r) = kr^n,\ n > 1, k > 0\ (n, k \in \mathbb{R})$. While this assumption restricts the generality of the solution to any transmission scheme, it has several justifications which make the results still meaningful. In particular, in the low SNR, low rate regime of operation, Shannon's formula is well-approximated as a linear function and thus can be approximated by the monomial class of functions considered. More generally, even for practical transmission schemes, one can obtain a good approximation of $g(r)$ to the form kr^n. As an example, consider the QAM modulation scheme considered in [16] and reproduced here in Figure 1. The table gives the rate and the normalized signal power per symbol, where d represents the minimum distance between signal points and the scheme is designed for error probabilities less than 10^{-6}. The plot gives the least squares monomial fit to the transmission scheme, and it can be seen that the monomial approximation is fairly close over the range of rates considered in this example. Lastly, the monomial approximation lends itself to mathematical analysis which yields useful insights that can be applied in practice to develop simple heuristics. As a note, without loss of generality, throughout the paper we take the constant $k = 1$, since, any other value of k simply scales the energy cost without affecting the optimal policy results.

B. Channel Model

We consider a time-homogeneous, first-order, discrete state space, Markov process for the channel state $c(t)$. Markov processes constitute a large class of stochastic processes that exhaustively model a wide set of fading scenarios and there is substantial literature on these models [19]–[22] and their applications to communication networks [22], [23].

Denote the channel stochastic process as $C(t)$ and the state space as $\mathcal{C}$. Let $c \in \mathcal{C}$ denote a particular value of the channel state and $\{c(t), t \geq 0\}$ be a sample path. Starting from state c, the channel can transition to a set of new states ($\neq c$) and this set is denoted as $\mathcal{J}_c$. Let $\lambda_{c\tilde{c}}$ denote the channel transition rate from state c to $\tilde{c}$, then, the sum transition rate at which the channel jumps out of state c is, $\lambda_c = \sum_{\tilde{c}\in\mathcal{J}_c} \lambda_{c\tilde{c}}$. Clearly, the expected time that $C(t)$ spends in state c is $1/\lambda_c$ and one can view $\frac{1}{\lambda_c}$ as the coherence time of the channel in state c.

Now, define $\lambda \triangleq \sup_c \lambda_c$ and a random variable, $Z(c)$, as,

$$Z(c) \triangleq \begin{cases} \tilde{c}/c, & \text{with probability } \lambda_{c\tilde{c}}/\lambda,\ \tilde{c} \in \mathcal{J}_c \\ 1, & \text{with probability } 1 - \lambda_c/\lambda \end{cases} \quad (2)$$

With this definition, we obtain a compact description of the process evolution as follows. *Given a channel state c, there is an exponentially distributed time duration with rate λ after which the channel state changes. The new state is a random variable which is given as $C = Z(c)c$.* Clearly, from (2) the transition rate to state $\tilde{c} \in \mathcal{J}_c$ is unchanged at $\lambda_{c\tilde{c}}$, whereas with rate $\lambda - \lambda_c$ there are indistinguishable self-transitions. This is a standard uniformization technique and there is no process generality lost with the new description as it yields a stochastically identical scenario. The representation simply helps in notational convenience.

The other technical assumptions in the model are as follows. The channel state space, $\mathcal{C}$, is a countable space (it could be infinite), and $\mathcal{C} \subseteq \mathbb{R}^+$. The states $c = 0, \infty$ are excluded from $\mathcal{C}$ since each of this state leads to a singularity in (1). The set $\mathcal{J}_c, \forall c$, is a finite subset of $\mathcal{C}$. Transition rate $\lambda_c, \forall c$ is bounded which ensures that λ defined as the supremum is finite. For all c, the support of $Z(c)$ lies in $[z_l, z_h]$, where $0 < z_l \leq z_h < \infty$. This ensures that $C(t)$ does not hit 0 or ∞, a.s. (almost surely), over a finite time interval.

Example: As an example, consider a two-state channel model with values c_b and c_g. These represent a two level quantization of the physical channel gain, where, if the measured channel gain is below a threshold the channel is considered as "bad" and $c(t)$ is assigned an average value c_b, otherwise $c(t) = c_g$ for the "good" condition. Let the transition rate from the good to the bad state be λ_{gb} and from the bad to the good state be λ_{bg}. Let $\gamma = c_b/c_g$, and using the earlier notation, $\lambda = \max(\lambda_{bg}, \lambda_{gb})$. For state c_g we have,

$$Z(c_g) = \begin{cases} \gamma, & \text{with probability } \lambda_{gb}/\lambda \\ 1, & \text{with probability } 1 - \lambda_{gb}/\lambda \end{cases} \quad (3)$$

To obtain $Z(c_b)$, replace γ with $1/\gamma$ and λ_{gb} with λ_{bg} in (3).

C. Problem Formulation

As mentioned earlier, the transmitter has B units of data and a deadline T by which the data must be sent. Let $x(t)$ denote the amount of data left in the buffer and $c(t)$ be the channel state at time t. The system state can be described as (x, c, t), where the notation means that at time t, we have $x(t) = x$ and $c(t) = c$. Let $r(x, c, t)$ denote the chosen transmission rate for the corresponding system state (x, c, t). Since the underlying process is Markov, it is sufficient to restrict attention to transmission policies that depend only on the present system state [26]. Clearly then, (x, c, t) is a Markov process.

Given a policy $r(x, c, t)$, the system evolves in time as a Piecewise-Deterministic-Process (PDP) as follows. We are given $x(0) = B$ and $c(0) = c_0$. Until t_1, where t_1 is the first time instant after $t = 0$ at which the channel changes, the buffer is reduced at the rate $r(x(t), c_0, t)$. Hence, over the interval $[0, t_1)$, $x(t)$ satisfies the ordinary differential equation,

$$\frac{dx(t)}{dt} = -r(x(t), c_0, t) \quad (4)$$

Equivalently, $x(t) = x(0) - \int_0^t r(x(s), c_0, s) ds \ \ ,\ t \in [0, t_1]$. Then, starting from the new state $(x(t_1), c(t_1), t_1)$ until the next channel transition we have, $\frac{dx(t)}{dt} = -r(x(t), c(t_1), t)$, $t \in [t_1, t_2)$; and this procedure repeats until $t = T$ is reached.

At time T, the data that missed the deadline (amount $x(T)$) is assigned a penalty cost of $\frac{\tau g(x(T)/\tau)}{c(T)}$ for some $\tau > 0$. This peculiar cost can be viewed in the following two ways. First, it simply represents a specific penalty function where τ can be adjusted and in particular made small enough[1] so that the data that misses the deadline is small. This will ensure that with good source-coding, the entire data can be recovered even if $x(T)$ misses the deadline. Second, note that $\frac{\tau g(x(T)/\tau)}{c(T)}$ is the amount of energy required to transmit $x(T)$ data in time τ with the channel state being $c(T)$. Thus, τ is the small time window in which the remaining data is completely transmitted out assuming that the channel state does not change over that period. In fact, viewing $T + \tau$ as the actual deadline, τ then

[1]For $g(\cdot)$ strictly convex, making τ smaller increases the penalty cost.

models a small buffer window in which unlimited power can be used to meet the deadline, albeit at an associated cost.

Let P denote the maximum power limit of the transmitter, a restriction imposed by the hardware and battery limitations. This limit would translate into a constraint of the form $\frac{g(r(x(t),c(t),t))}{c(t)} \leq P, \forall t \in [0,T)$. However, imposing such a strict constraint that must be satisfied at all times and on all the channel sample paths makes the problem intractable. To overcome this difficulty, we partition the interval $[0,T]$ into multiple periods and impose an average power constraint in each of the periods. Such a constraint is less restrictive and the optimization is over a much bigger class of policies. Let the interval $[0,T]$ be partitioned into L equal periods[2], where the value of L is fixed based on the hardware limitations. Then, over each partition the power constraint requires that the expected energy cost, $E\left[\int \frac{1}{c(s)} g(r(x(s),c(s),s))ds\right]$, is less than $P(T/L)$, i.e. we require,

$$E\left[\int_{\frac{(k-1)T}{L}}^{\frac{kT}{L}} \frac{g(r(x(s),c(s),s))}{c(s)} ds\right] \leq \frac{PT}{L} \ , \quad k = 1, \ldots, L \tag{5}$$

Note that T/L is the duration of each partition interval and $[\frac{(k-1)T}{L}, \frac{kT}{L})$ is the k^{th} interval, $k = 1, \ldots, L$. Clearly, by varying L, the duration of the partition interval can be varied and the power constraint can be made either more or less restrictive.

Let Φ denote the set of all transmission policies, $r(x,c,t)$, that satisfy the following,

(a) $0 \leq r(x,c,t) < \infty$, (non-negativity of rate)

(b) $r(x,c,t) = 0$, if $x = 0$ (no data left to transmit)[3].

We say that a policy $r(x,c,t)$ is *admissible*, if $r(x,c,t) \in \Phi$ and additionally if it also satisfies the power constraints as given in (5).

Denote the optimization problem as $(\mathcal{P})$, we can now summarize it as follows,

$$(\mathcal{P}) \quad \inf_{r(\cdot)\in\Phi} \ E\left[\int_0^T \frac{1}{c(s)} g(r(x(s),c(s),s))ds + \frac{\tau g(\frac{x(T)}{\tau})}{c(T)}\right]$$

$$\text{subject to} \quad E\left[\int_0^{\frac{T}{L}} \frac{1}{c(s)} g(r(x(s),c(s),s))ds\right] \leq \frac{PT}{L}$$

$$\vdots$$

$$E\left[\int_{\frac{T(L-1)}{L}}^{T} \frac{1}{c(s)} g(r(x(s),c(s),s))ds\right] \leq \frac{PT}{L}$$

All the expectations above are conditional on the starting state (x_0, c_0)[4]. For the analysis, we will keep the general notation x_0 but its value in our case is simply $x_0 = B$. Note that problem $(\mathcal{P})$ as stated above has at least one admissible solution since a policy that does not transmit any data and simply incurs the penalty cost is an admissible policy. Furthermore, as shown in Appendix C, this simple policy has a finite cost and hence the minimum value of the objective function above is finite.

III. Optimal Policy

In order to solve problem $(\mathcal{P})$, we consider a Lagrange duality approach. The basic steps involved in such an approach are as follows: (a) form the Lagrangian by incorporating the constraints into the objective function using Lagrange multipliers, (b) obtain the dual function by minimizing over the primal space, and (c) maximize the dual function with respect to the Lagrange multipliers. Finally, we need to show that there is no duality gap, that is, maximizing the dual function gives the optimal cost for the constrained problem. There are, however, important subtleties in problem $(\mathcal{P})$. First, the domain of the rate functions $r(\cdot)$ is a functional space which makes $(\mathcal{P})$ an infinite dimensional optimization, and, second, $(\mathcal{P})$ is a stochastic optimization and by this we mean that there is a probability space involved over which the expectation is taken. We now present the technical details of the various steps mentioned above.

A. Dual Function

Consider the inequality constraints in $(\mathcal{P})$ and re-write them as follows,

$$E\left[\int_{\frac{(k-1)T}{L}}^{\frac{kT}{L}} \frac{g(r(\cdot))}{c(s)} ds\right] - \frac{PT}{L} \leq 0, \quad k = 1, \ldots, L \tag{6}$$

Let $\bar{\nu} = (\nu_1, \ldots, \nu_L)$ be the Lagrange multiplier vector for these constraints and since these are inequality constraints, the vector $\bar{\nu}$ must be non-negative, i.e. $\nu_1 \geq 0, \ldots, \nu_L \geq 0$. The *Lagrangian function* is then given as,

$$\begin{aligned}\mathcal{H}(r(\cdot), \bar{\nu}) =& E\left[\int_0^T \frac{g(r(\cdot))}{c(s)} ds + \frac{\tau g(\frac{x(T)}{\tau})}{c(T)}\right] \\ &+ \sum_{k=1}^{L} \nu_k \left(E\left[\int_{\frac{(k-1)T}{L}}^{\frac{kT}{L}} \frac{g(r(\cdot))}{c(s)} ds\right] - \frac{PT}{L}\right)\end{aligned} \tag{7}$$

Re-arranging the above equation, it can be written in the form,

$$\begin{aligned}\mathcal{H}(r(\cdot), \bar{\nu}) =& E\left[\int_0^T \frac{(1+\nu(s))g(r(\cdot))}{c(s)} ds + \frac{\tau g(\frac{x(T)}{\tau})}{c(T)}\right] \\ &-(\nu_1 + \ldots + \nu_L)\frac{PT}{L}\end{aligned} \tag{8}$$

where $\nu(s)$ takes value ν_k over the k^{th} partition interval, i.e. $\nu(s) = \nu_k, s \in [\frac{(k-1)T}{L}, \frac{kT}{L})$.

The *Dual function*, denoted as $\mathcal{L}(\bar{\nu})$, is defined as the infimum of the Lagrangian function $\mathcal{H}(r(\cdot), \bar{\nu})$ over $r(x,c,t) \in \Phi$. Thus, we have,

$$\mathcal{L}(\bar{\nu}) = \inf_{r(\cdot)\in\Phi} \mathcal{H}(r(\cdot), \bar{\nu}) \tag{9}$$

A point to note here is that the policies $r(x,c,t)$ over which the above minimization is considered do not have to satisfy the power constraints, though the other requirements still apply.

[2]Extensions to arbitrary sized partitions is fairly straightforward but such a generality is omitted for mathematical simplicity.

[3]We also require that $r(x,c,t)$ be locally Lipschitz continuous in x ($x > 0$) and piecewise continuous in t. This ensures that the ODE in (4) has a unique solution.

[4]To avoid being cumbersome on notation, we will throughout represent conditional expectations without an explicit notation but rather mention the conditioning parameter whenever there is ambiguity.

This is because the short term power constraints (violation) have been added as a cost in the objective function of the dual problem. Thus, the Lagrangian in (9) is minimized over the set Φ without the power constraints.

A well-known property of the dual function is that for a given Lagrange vector $\bar{\nu} \geq 0$, the dual function $\mathcal{L}(\bar{\nu})$ gives a lower bound to the optimal cost in $(\mathcal{P})$. This standard property is referred to as *weak duality* and it applies in our case as well. Let $J(x_0, c_0)$ denote the optimal cost for problem $(\mathcal{P})$ (i.e. the minimum value of the objective function) with (x_0, c_0) being the starting state, we then have the following result.

Lemma 1: (Weak Duality) *Consider problem $(\mathcal{P})$ and let (x_0, c_0) be the starting state at $t = 0$. Then, for all $\bar{\nu} \geq 0$, we have, $\mathcal{L}(\bar{\nu}) \leq J(x_0, c_0)$.*

Proof: This is a standard result and the proof is omitted for brevity but can be found in [4]. ∎

We, next, proceed to evaluate the dual function $\mathcal{L}(\bar{\nu})$ by solving the minimization problem given in (9).

Evaluating the dual function: The approach we adopt to evaluate the dual function is to view the problem in L stages corresponding to the L partition intervals and solve for the optimal rate functions in each of the partitions with the necessary boundary conditions at the edges. An immediate observation from (8) shows that the effect of the Lagrange multipliers is to multiply the instantaneous power-rate function $\frac{g(r(\cdot))}{c(s)}$ with a time-varying function $(1+\nu(s))$. Thus, the difference over the various intervals is in a different multiplicative factor to the cost function, which for the k^{th} interval is, $1+\nu(s) = 1+\nu_k$. Intuitively, the Lagrange multipliers re-adjust the cost function which causes the data transmission to be moved among the various time-periods. For example, if $\nu_k > \nu_l$, then it becomes more costly to transmit in the k^{th} period than the l^{th} period and this has the effect of (relatively) increasing the data transmission in the l^{th} period.

Since (9) involves a minimization over $r(\cdot)$ for fixed Lagrange multipliers $\bar{\nu}$, the second term in (8), i.e. $\frac{(\nu_1+\ldots+\nu_L)PT}{L}$, is irrelevant for the minimization and we will neglect it for now. Define,

$$H^r_\nu(x, c, t) = E\left[\int_t^T \frac{(1+\nu(s))g(r(\cdot))}{c(s)}ds + \frac{\tau g(\frac{x(T)}{\tau})}{c(T)}\right] \quad (10)$$

$$H_\nu(x, c, t) = \inf_{r(\cdot)\in\Phi} H^r_\nu(x, c, t) \quad (11)$$

where the expectation in (10) is conditional on the state (x, c, t). In simple terms, $H^r_\nu(x, c, t)$ is the *cost-to-go function* starting from state (x, c, t) for policy $r(\cdot)$ and $H_\nu(x, c, t)$ is the corresponding *optimal cost-to-go function*. Relating back to (8), $H^r_\nu(x_0, c_0, 0)$ is the expectation term in (8) and $H_\nu(x_0, c_0, 0)$ is the minimization of this term over $r(\cdot) \in \Phi$. Clearly from (8) and (9), having solved for $H_\nu(x, c, t)$, we then obtain the dual function as simply,

$$\mathcal{L}(\bar{\nu}) = H_\nu(x_0, c_0, 0) - \frac{(\nu_1 + \ldots + \nu_L)PT}{L} \quad (12)$$

Finally, in the process of obtaining $H_\nu(x, c, t)$, we also obtain the optimal rate function that achieves the minimum in (11).

Now, to proceed, focus on the k^{th} partition interval so that $t \in [\frac{(k-1)T}{L}, \frac{kT}{L})$ and consider a small interval $[t, t+h)$, within this partition. Let some policy $r(\cdot)$ be followed over $[t, t+h)$ and the optimal policy thereafter, then using Bellman's principle [24] we have,

$$H_\nu(x, c, t) = \min_{r(\cdot)} \Big\{ E\left[\int_t^{t+h} \frac{(1+\nu_k)g(r(x(s), c(s), s))}{c(s)}ds\right] + EH_\nu(x_{t+h}, c_{t+h}, t+h)\Big\} \quad (13)$$

where x_{t+h} is short-hand for $x(t+h)$ and the expectation is conditional on (x, c, t). The left side in the equation above is the optimal cost if the optimal policy is followed right from the starting state (x, c, t), whereas on the right side, the expression within the minimization bracket is the total cost with policy $r(\cdot)$ being followed over $[t, t+h]$ and the optimal policy thereafter. Removing the minimization gives the following inequality,

$$H_\nu(x, c, t) \leq E\int_t^{t+h} \frac{(1+\nu_k)g(r(x(s), c(s), s))}{c(s)}ds + E\left[H_\nu(x_{t+h}, c_{t+h}, t+h)\right] \quad (14)$$

$$H_\nu(x, c, t) \leq E\int_t^{t+h} \frac{(1+\nu_k)g(r(x(s), c(s), s))}{c(s)}ds + E\left[H_\nu(x_{t+h}, c_{t+h}, t+h)\right] \quad (15)$$

Dividing by h in (15) and taking the limit $h \downarrow 0$ gives,

$$A^r H_\nu(x, c, t) + \frac{(1+\nu_k)g(r)}{c} \geq 0 \quad (16)$$

The above inequality follows since, $\frac{E\int_t^{t+h}\left(\frac{(1+\nu_k)g(r(\cdot))}{c_s}\right)ds}{h} \to \frac{(1+\nu_k)g(r)}{c}$, where r is the value of the transmission rate at time t, i.e. $r = r(x, c, t)$. The term $A^r H_\nu(x, c, t)$ is defined as $A^r H_\nu(x, c, t) \triangleq \lim_{h\downarrow 0} \frac{EH_\nu(x_{t+h}, c_{t+h}, t+h) - H_\nu(x, c, t)}{h}$ and this quantity $A^r H_\nu(x, c, t)$ is called the differential generator of the Markov process $(x(t), c(t))$ for policy $r(\cdot)$. Intuitively, it is a natural generalization of the ordinary time derivative for a function that depends on a stochastic process. An elaborate discussion on this topic can be found in [24]–[26]. For our case, using the time evolution as given in (4), the quantity $A^r H_\nu(x, c, t)$ can be evaluated as,

$$A^r H_\nu(x, c, t) = \frac{\partial H_\nu(x, c, t)}{\partial t} - r\frac{\partial H_\nu(x, c, t)}{\partial x} + \lambda(E_z[H_\nu(x, Z(c)c, t)] - H_\nu(x, c, t)) \quad (17)$$

where E_z is the expectation with respect to the $Z(c)$ variable; $Z(c)$ is as defined in (2). Now, in the above steps from (14)-(16), if policy $r(\cdot)$ is replaced with the optimal policy $r^*(\cdot)$, equation (16) holds with equality and we get,

$$A^{r^*} H_\nu(x, c, t) + \frac{(1+\nu_k)g(r^*)}{c} = 0 \quad (18)$$

Hence, for a given system state (x, c, t), the optimal transmission rate, r^*, is the value that minimizes (16) and the minimum value of the expression equals zero. Over the k^{th} partition interval with $t \in [\frac{(k-1)T}{L}, \frac{kT}{L})$, we thus get the following *Optimality Equation*,

$$\min_{r\in[0,\infty)} \left[\frac{(1+\nu_k)g(r)}{c} + A^r H_\nu(x,c,t)\right] = 0 \quad (19)$$

Substituting $A^r H_\nu(\cdot)$ from (17), we see that (19) is a partial differential equation in $H_\nu(x,c,t)$, also referred to as the Hamilton-Jacobi-Bellman (HJB) equation,

$$\min_{r\in[0,\infty)} \Big\{\frac{(1+\nu_k)g(r)}{c} + \frac{\partial H_\nu}{\partial t} - r\frac{\partial H_\nu}{\partial x} + \lambda(E_z[H_\nu(x, Z(c)c, t)] - H_\nu(x,c,t))\Big\} = 0 \quad (20)$$

The boundary conditions for $H_\nu(\cdot)$ are as follows. At $t = T$, $H_\nu(x,c,T) = \frac{\tau g(\frac{x}{\tau})}{c}$, since starting in state (x,c) at time T, the optimal cost simply equals the penalty cost. At each of the partition interval, $t = kT/L$, we require that $H_\nu(\cdot)$ be continuous at the edges, so that the functions evaluated for the various intervals are consistent.

We now solve the above optimality PDE equation to obtain the function $H_\nu(x,c,t)$, and the corresponding optimal rate function denoted as $r^*_\nu(x,c,t)$ (the subscript ν is used to indicate explicit dependence on the Lagrange vector $\bar{\nu}$). Theorem I summarizes the results while an intuitive explanation of the optimal rate function is presented later. Before proceeding further, we need some additional notations regarding the channel process. Let there be m channel states in the Markov model and denote the various states $c \in \mathcal{C}$ as $c^1, c^2, \ldots, c^m$. Given a channel state c^i, the values taken by the random variable $Z(c^i)$ are denoted as $\{z_{ij}\}$, where $z_{ij} = c^j/c^i$. The probability that $Z(c^i) = z_{ij}$ is denoted as p_{ij}. Clearly, if there is no transition from state c^i to c^j, $p_{ij} = 0$.

Theorem *I: **(General Markov Channel)** Consider the minimization in (11) with $g(r) = r^n$, $(n > 1, n \in \mathbb{R})$. For $k = 1,\ldots,L$ and $t \in \left[\frac{(k-1)T}{L}, \frac{kT}{L}\right)$ (k^{th} partition interval), we have,*

$$r^*_\nu(x,c^i,t) = \frac{x}{f_i^k(T-t)}, \qquad i = 1,\ldots,m \quad (21)$$

$$H_\nu(x,c^i,t) = \frac{(1+\nu_k)x^n}{c^i(f_i^k(T-t))^{n-1}}, \quad i = 1,\ldots,m \quad (22)$$

For a fixed k, the functions $\{f_i^k(s)\}_{i=1}^m$, $s \in \left[\frac{(L-k)T}{L}, \frac{(L-k+1)T}{L}\right]$ are the solution of the following ordinary differential equation (ODE) system,

$$(f_1^k(s))' = 1 + \frac{\lambda f_1^k(s)}{n-1} - \frac{\lambda}{n-1}\sum_{j=1}^m \frac{p_{1j}}{z_{1j}}\frac{(f_1^k(s))^n}{(f_j^k(s))^{n-1}} \quad (23)$$

$$\vdots$$

$$(f_m^k(s))' = 1 + \frac{\lambda f_m^k(s)}{n-1} - \frac{\lambda}{n-1}\sum_{j=1}^m \frac{p_{mj}}{z_{mj}}\frac{(f_m^k(s))^n}{(f_j^k(s))^{n-1}} \quad (24)$$

The following boundary conditions apply: if $k = L$, $f_i^L(0) = \tau(1+\nu_L)^{\frac{1}{n-1}}, \forall i$ and if $k = 1,..,L-1$, $f_i^k\left(\frac{(L-k)T}{L}\right) = \left(\frac{1+\nu_k}{1+\nu_{k+1}}\right)^{\frac{1}{n-1}} f_i^{k+1}\left(\frac{(L-k)T}{L}\right), \forall i$. The dual function in (12) is then given as (let $c_0 = c^j$, for some $j \in \{1,\ldots,m\}$),

$$\mathcal{L}(\bar{\nu}) = \frac{(1+\nu_1)x_0^n}{c^j(f_j^1(T))^{n-1}} - \frac{(\nu_1+\ldots+\nu_L)PT}{L} \quad (25)$$

Proof: See Appendix A. ■

The above solution can be understood as follows. For each partition interval, k, there are m functions $\{f_i^k(s)\}_{i=1}^m$ corresponding to the respective channel states. The subscript in the notation for f refers to the channel state index while the superscript refers to the partition interval. Now, given that the present time t lies in the k^{th} interval, the optimal rate function has the closed form expression $\frac{x}{f_i^k(T-t)}$ as given in (21), while $H_\nu(\cdot)$ is as given in (22). The functions $\{f_i^k(s)\}_{i=1}^m$ for the k^{th} interval are the solution of the ODE system in (23)-(24) over $s \in \left[\frac{(L-k)T}{L}, \frac{(L-k+1)T}{L}\right]$ with the initial boundary condition given as, $f_i^k\left(\frac{(L-k)T}{L}\right) = \left(\frac{1+\nu_k}{1+\nu_{k+1}}\right)^{\frac{1}{n-1}} f_i^{k+1}\left(\frac{(L-k)T}{L}\right), \forall i$. This ensures that $H_\nu(x,c,t)$ is continuous at the interval edges, $t = \frac{kT}{L}$. For the L^{th} interval the boundary condition is, $f_i^L(0) = \tau(1+\nu_L)^{\frac{1}{n-1}}$, $\forall i$; this ensures that at $t = T$, $H_\nu(x,c^i,T) = \frac{(1+\nu_L)x^n}{c^i(f_i^L(0))^{n-1}} = \frac{x^n}{c^i\tau^{n-1}}$, same as the penalty cost function for $g(r) = r^n$ (as required).

The functions $\{f_i^k(s)\}$ can be evaluated starting with $k = L$ and the initial boundary condition $f_i^L(0) = \tau(1+\nu_L)^{\frac{1}{n-1}}$, to obtain $\{f_i^L(s)\}_{i=1}^m$ over $s \in \left[0, \frac{T}{L}\right]$. Having obtained $\{f_i^L(s)\}_{i=1}^m$, then consider, $k = L-1$, and using the earlier mentioned boundary conditions obtain $\{f_i^{L-1}(s)\}_{i=1}^m$, $s \in \left[\frac{T}{L}, \frac{2T}{L}\right]$. Proceeding backwards this way, we obtain all the functions $\{f_i^k(s)\}$.

In full generality, the ODE system in (23)-(24) can be easily solved numerically using standard techniques and as shown in Lemma 3 in the Appendix, the system has a unique positive solution. Furthermore, this computation needs to be done only once before the system starts operating and $\{f_i^k(s)\}$ can be pre-determined and stored in a table. Once $\{f_i^k(s)\}$ are known, the closed form structure of the optimal policy in (21) warrants no further computation.

Constant Drift Channel: Theorem I gives the dual function and the optimal-rate results for a general Markov channel model. By considering a special structure on the channel model, which we refer to as the "*Constant Drift*" channel, the functions $\{f_i^k(s)\}$ can be evaluated in closed-form.

Under this channel model, we assume that the expected value of the random variable $1/Z(c)$ is independent of the channel state, i.e. $E[1/Z(c)] = \beta$, a constant. Thus, starting in state c, if $\tilde{c}$ denotes the next transition state we have $E\left[\frac{1}{\tilde{c}}\right] = E\left[\frac{1}{Z(c)}\right]\frac{1}{c} = \frac{\beta}{c}$. This means that if we look at the process $1/c(t)$, the expected value of the next state is a constant multiple of the present state. We refer to β as the "*drift*" parameter of the channel process. If $\beta > 1$, the process $1/c(t)$ has an upward drift; if $\beta = 1$, there is no drift and if $\beta < 1$, the drift is downwards. As a simple example of such a Markov model, suppose that the channel transitions at rate $\lambda > 0$ and at every transition the state either improves by a factor $u > 1$ with probability p_u, or worsens by a factor $1/u$ with probability p_d $(= 1 - p_u)$. Thus, given some state $c > 0$ the next channel state is either uc or c/u, and, $E[1/Z(c)] = p_u/u + up_d$. Hence, the drift parameter is $\beta = p_u/u + up_d$.

There are various situations where a constant drift channel model is applicable over the time scale of the deadline interval. For example, when a mobile device is moving in the direction of the base station, the channel has an expected drift towards improving conditions and vice-versa. Similarly, in case of satellite channels, changing weather conditions such as cloud cover makes the channel drift towards worsening conditions and vice-versa. For cases when the time scale of these drift changes is longer than the deadlines on the data, a constant drift channel model serves as an appropriate model.

Theorem *II: **(Constant Drift Channel)** Consider the minimization in (11) with $g(r) = r^n$ and the constant drift channel model with parameter β. For $k = 1, \ldots, L$, $t \in [\frac{(k-1)T}{L}, \frac{kT}{L})$,*

$$H_\nu(x, c, t) = \frac{(1+\nu_k)x^n}{c(f^k(T-t))^{n-1}} \tag{26}$$

$$r^*_\nu(x, c, t) = \frac{x}{f^k(T-t)} \tag{27}$$

Let $\eta = \frac{\lambda(\beta-1)}{n-1}$, then [5],

$$\begin{aligned} f^k(T-t) &= \tau(1+\nu_k)^{\frac{1}{n-1}} e^{-\eta(T-t)} \\ &+ \frac{1}{\eta}\Big\{ \sum_{j=0}^{L-k-1} \left(\frac{1+\nu_k}{1+\nu_{L-j}}\right)^{\frac{1}{n-1}} e^{-\eta(T-t)} \left(e^{\frac{\eta(j+1)T}{L}} - e^{\frac{\eta jT}{L}}\right) \Big\} \\ &+ \frac{1}{\eta}\left(1 - e^{-\eta\left((T-t)-(L-k)\frac{T}{L}\right)}\right) \end{aligned} \tag{28}$$

The dual function in (12) is given as,

$$\mathcal{L}(\bar{\nu}) = \frac{(1+\nu_1)x_0^n}{c_0(f^1(T))^{n-1}} - \frac{(\nu_1 + \ldots + \nu_L)PT}{L} \tag{29}$$

Proof: See Appendix B. ∎

Thus, from above, we see that the constant drift channel model admits a closed-form solution of the dual function and the optimal rate.

B. Strong Duality

In Theorems I and II, we fixed a Lagrange vector $\bar{\nu}$ and obtained the dual function $\mathcal{L}(\bar{\nu})$ and the optimal rate function that achieves the minimum in (11). Now, from Lemma 1, given a Lagrange vector $\bar{\nu} \geq 0$, the dual function is a lower bound to the optimal cost of the constrained problem, $\mathcal{P}$. Thus, intuitively, it makes sense to maximize $\mathcal{L}(\bar{\nu})$ over $\bar{\nu} \geq 0$. Theorem III below states that strong duality holds, i.e. maximizing $\mathcal{L}(\bar{\nu})$ over $\bar{\nu} \geq 0$ gives the optimal cost of $\mathcal{P}$, and furthermore, the optimal rate function $r^*(x, c, t)$ for the constrained problem $\mathcal{P}$ is the same as $r^*_\nu(x, c, t)$ obtained in Theorem I with $\bar{\nu} = \bar{\nu}^*$ (where $\bar{\nu}^*$ is the maximizing Lagrange vector).

[5]For $k = L$ the summation term in (28) is taken as zero.

As in Lemma 1, let $J(x_0, c_0)$ denote the optimal cost of $(\mathcal{P})$ with the initial state (x_0, c_0) at $t = 0$, where $x_0 = B$ and $c_0 \in \mathcal{C}$. We then have the following result.

Theorem *III: **(Strong Duality)** Consider the dual function defined in (9) for $\bar{\nu} \geq 0$, we then have,*

$$J(x_0, c_0) = \max_{\bar{\nu} \geq 0} \mathcal{L}(\bar{\nu}) \tag{30}$$

and the maximum on the right is achieved by some $\bar{\nu}^* \geq 0$. Let $r^*(x, c, t)$ denote the optimal transmission policy for problem $(\mathcal{P})$, then, $r^*(x, c, t)$ is as given in (21) for $\bar{\nu} = \bar{\nu}^*$.

Proof: See Appendix C. ∎

For the maximization in (30), the dual functions are given in Theorems I (general markov channel) and II (constant drift channel). It can also be shown using a standard argument that the dual function is concave [27] which makes the maximization much simpler since there is a unique global (and local) maxima. Using a standard gradient search algorithm the vector $\bar{\nu}^*$ can be obtained numerically and this computation needs to be done offline.

C. Optimal Policy for $(\mathcal{P})$

The optimal policy for problem $(\mathcal{P})$ can now be obtained by combining Theorems I and III and is given as follows. For $k = 1, \ldots, L$ and $t \in [\frac{(k-1)T}{L}, \frac{kT}{L})$ (k^{th} partition interval),

$$r^*(x, c^i, t) = r^*_{\nu^*}(x, c^i, t) = \frac{x}{f_i^k(T-t)}, \qquad i = 1, \ldots, m \tag{31}$$

where the functions $\{f_i^k(s)\}$ are evaluated with $\bar{\nu} = \bar{\nu}^*$. As mentioned earlier, the computation for $\bar{\nu}^*$ and $\{f_i^k(s)\}$ needs to be done offline before the data transmission begins. In practice, if the transmitter has computational capabilities, these computations can be carried out at $t = 0$ for the given problem parameters, otherwise, the $\bar{\nu}^*$ and $\{f_i^k(s)\}$ can be pre-determined and stored in a table in the transmitter memory. Having known $\{f_i^k(s)\}$, the closed form structure of the optimal policy as given in (31) warrants no further computation and is simple to implement. At time t, the transmitter looks at the amount of data in the buffer x, the channel state c, the partition interval k in which t lies, and computes the rate for that communication slot as simply $\frac{x}{f_i^k(T-t)}$.

The solution in (31) provides several interesting observations and insights as follows. At time t, the optimal rate is a linear function of x, hence, as intuitively expected, the rate is proportionately higher when there is more data left in the queue. For a given channel state c^i and time t, the slope of this function is given as $\frac{1}{f_i^k(T-t)}$, thus, in some sense we can view the quantity $\frac{1}{f_i^k(T-t)}$ as the "urgency" of transmission under the channel state c^i and with time $(T-t)$ left until the deadline. This view gives a nice separation form for the optimal rate:

*optimal rate = amount of data left * urgency of transmission*

Another interesting observation can be made if we set $\lambda = 0$ (no channel variations), $P = \infty$ (no power constraints) and set $\tau = 0$ (infinite penalty cost for missing the deadline).

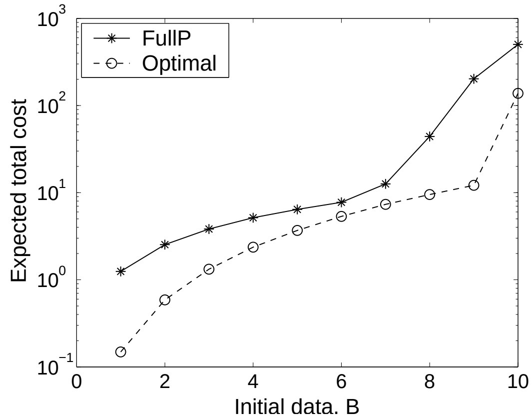


Fig. 2. Total cost comparison of the optimal and the full power policy.

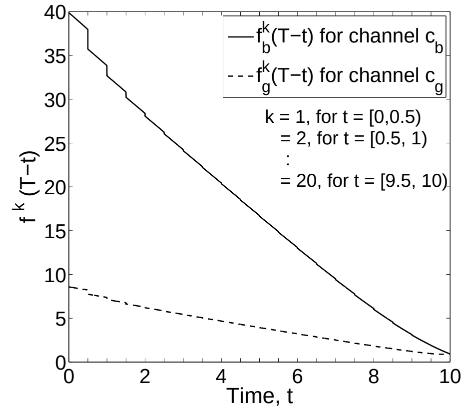


Fig. 3. Plot of $\{f_g^k(T-t), f_b^k(T-t)\}_{k=1}^L$ for channel states c_g, c_b respectively.

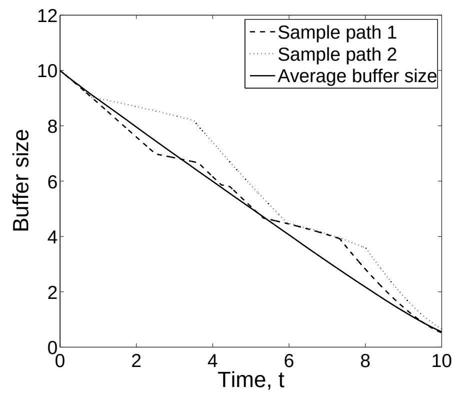


Fig. 4. Buffer state evolution over time for the optimal policy.

Clearly with $P = \infty$ the Lagrange vector $\bar{\nu}^* = (0, \dots, 0)$ and the optimal rate function takes the form $r^*(x, c^i, t) = \frac{x}{T-t}$. Thus, under no power constraints and no channel variations, the optimal policy is to transmit at a rate that just empties the buffer by the deadline. This observation is consistent with previous results in the literature for non-fading/time-invariant channels [5], [8], [10].

IV. Simulation Results

In this section, we consider an illustrative example and present energy cost comparisons for the optimal and the Full Power (FullP) policy. In FullP policy, the transmitter always transmits at full power, P, and so given the system state (x, c, t) the rate is chosen as, $r(x, c, t) = g^{-1}(cP) = (cP)^{1/n}$, for $g(r) = r^n$. The simulation setup is as follows. The channel model is the two-state model as described earlier in Section II-B, with parameters $\lambda_{bg} = 1$, $\lambda_{gb} = 3/7$, $c_g = 1$ and $c_b = 0.2$; thus, $\lambda = \max(\lambda_{bg}, \lambda_{gb}) = 1$ and $\gamma = c_b/c_g = 0.2$. It can be easily checked that with the above parameters, in steady state the fraction of time spent in the good state is 0.7 and 0.3 in the bad state. The deadline is taken as $T = 10$ and the number of partition intervals as $L = 20$. The power-rate function is, $g(r) = r^2$ and the value of τ in the penalty cost function is taken as 0.01 which is 0.1% of the deadline; thus, a time window of 0.1% is provided at T. To simulate the process, the communication slot duration is taken as $dt = 10^{-3}$ implying that there are $T/dt = 10{,}000$ slots over the deadline interval. For each slot, the transmission rate is computed as given by the corresponding policy and the total cost is obtained as the sum of the energy cost over the time-slots plus the penalty cost. Expectation is then taken as an average over 10^4 sample paths. Let $\{f_g^k(s), f_b^k(s)\}_{k=1}^L$ denote the $f(\cdot)$ functions for channel states c_g and c_b respectively; these are computed using a standard ODE solver wherein $\bar{\nu}^*$ is evaluated using a standard function maximizer in MATLAB and these computations are carried out offline before the system operation.

Figure 2 is a plot of the expected total cost of the two policies with the initial data amount B varied from 1 to 10. The value of P is chosen such that at $B = 5$, even with bad channel condition over the entire deadline interval, the entire data can be transmitted at full power. This implies, $P = \frac{1}{\gamma}(5/T)^2 = 1.25$ ($5/T$ is the rate required to serve 5 units in time T). Thus, $B \le 5$ gives the regime in which full power always meets the deadline and $B > 5$ is the regime in which data is left out which then incurs the penalty cost. It is evident from the plot that the optimal policy gives a significant gain in the total cost (note that the y-axis is on a log scale) and at around $B = 1$, FullP policy incurs almost 10 times the optimal cost. Thus, we see that dynamic rate adaptation can yield significant energy savings.

For a particular value of B, taken as $B = 10$, the functions $\{f_g^k(T-t), f_b^k(T-t)\}_{k=1}^L$ are plotted in Figure 3. The index $k = 1, \dots, L$ denotes the partition interval, and since $L = 20$, each partition is of size 0.5 time units. It can be seen from the plot that $f_g^k(T-t) \le f_b^k(T-t)$, which implies that given x units of data in the buffer the optimal rate $\frac{x}{f_g^k(T-t)}$ is higher under the good channel state c_g than the bad channel state c_b. This is intuitively expected since the optimal policy must exploit the good channel condition which has a lower energy cost and transmit more data in this state.

Figure 4 is a plot of the buffer state $x(t)$ over time for the parameter values as stated earlier with $B = 10$. The solid curve is the average value of $x(t)$ in time, while the other two curves are the $x(t)$ values for two illustrative sample paths.

From the plot, we see that the average buffer size is a smooth, monotonically decaying function with an almost linear decay. On each sample path, however, the buffer size goes through stages of faster decay which correspond to good channel state and higher rate transmission, and slower decay which correspond to the bad channel state where the transmission rate is lower.

V. Conclusion

We considered energy-efficient transmission of data over a wireless fading channel with deadline and power constraints. Specifically, we addressed the scenario of a wireless transmitter with short-term average power constraints, having B units of data that must be transmitted by deadline T over a fading channel. Using a novel continuous-time optimal-control formulation and Lagrangian duality, we obtained the optimal transmission policy that dynamically adapts the rate over time and in response to the channel variations to minimize the transmission energy cost. The optimal policy is shown to have a simple and intuitive form, given as,

*optimal rate = amount of data left * urgency of transmission*

The work in this paper and the approach adopted open up various interesting research directions involving deadline-constrained data transmission over wireless channels. One of the natural extensions is to consider a network scenario involving control of multiple transmitter-receiver pairs when there are deadline constraints on data transmission.

Acknowledgements

The authors would like to acknowledge Asuman Ozdaglar and Devavrat Shah for helpful discussions on the work.

References

[1] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry and T. Pappas, “Advances in efficient resource allocation for packet-based real-time video transmission”, *Proceedings of the IEEE*, vol. 93, no. 1, Jan. 2005.
[2] A. Ephremides, “Energy concerns in wireless networks”, *IEEE Wireless Communications*, vol. 9, issue 4, pp. 48-59, August 2002.
[3] A. Jalali, R. Padovani, R. Pankaj, “Data throughput of CDMA-HDR a high efficiency high data rate personal communication wireless system”, *IEEE Vehicular Technology Conf.*, vol. 3, 2000.
[4] M. Zafer, “Dynamic rate-control and scheduling algorithms for quality-of-service in wireless networks”, PhD Thesis, Massachusetts Institute of Technology (MIT), Sept. 2007.
[5] M. Zafer, E. Modiano, “A calculus approach to minimum energy transmission policies with quality of service guarantees”, *Proceedings of the IEEE INFOCOM 2005*, vol. 1, pp. 548-559, March 2005.
[6] L. Tsaur, D. Lee, “Closed-loop architecture and protocols for rapid dynamic spreading gain adaptation in CDMA networks”,*In Proc. of IEEE INFOCOM*, 2004.
[7] R. Berry, R. Gallager, “Communication over fading channels with delay constraints”, *IEEE Tran. on Information Theory*, vol. 48, no. 5, May 2002.
[8] E. Uysal-Biyikoglu and A. El Gamal, “On adaptive transmission for energy-efficiency in wireless data networks,” in *IEEE Tran. on Information Theory*, Dec. 2004.
[9] A. Fu, E. Modiano, J. Tsitsiklis, “Optimal energy allocation for delay constrained data transmission over a time-varying channel”, *In Proc. of IEEE INFOCOM 2003*, vol. 2, pp 1095 - 1105, April 2003.
[10] M. Khojastepour, A. Sabharwal, “Delay-constrained scheduling: power efficiency, filter design and bounds”, *In Proc. of IEEE INFOCOM 2004*, March 2004.
[11] P. Nuggehalli, V. Srinivasan, R. Rao, “Delay constrained energy efficient transmission strategies for wireless devices”, *In Proc. of IEEE INFOCOM 2002*.
[12] B. Collins, R. Cruz, “Transmission policies for time varying channels with average delay constraints”, *Allerton conf. on comm., control and computing*, Monticello, IL, 1999.
[13] M. Zafer, E. Modiano, “Optimal Rate Control for Delay-Constrained Data Transmission over a Wireless Channel”, *IEEE Transactions on Information Theory*, vol. 54, issue 9, pp. 4020-4039, Sept. 2008.
[14] M. Zafer and E. Modiano, “Optimal adaptive data transmission over a fading channel with deadline and power constraints”, *Conference on Information Sciences and Systems*, Princeton, New Jersey, March 2006.
[15] L. Tassiulas and A. Ephremides, “Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks”, *IEEE Tran. on Automatic Control*, vol. 37, no. 12, Dec. 1992.
[16] M. Neely, E. Modiano and C. Rohrs, “Dynamic power allocation and routing for time varying wireless networks,” *IEEE INFOCOM*, 2003.
[17] X. Liu, E. Chong, N. Shroff, “A framework for opportunistic scheduling in wireless networks ” *Computer Networks*, 41, pp. 451-474, 2003.
[18] B. Ata, “Dynamic power control in a wireless static channel subject to a quality of service constraint”, *Operations Research 53* (2005), no 5, 842-851.
[19] F. Babich and G. Lombardi, “A Markov model for the mobile propagation channel”, *IEEE Transactions on Vehicular Technology*, vol. 49, no. 1, pp. 63-73, Jan. 2000.
[20] H. S. Wang and P. Chang, “On verifying the first-order Markovian assumption for a Rayleigh fading channel model”, *IEEE Transactions on Vehicular Technology*, vol. VT-45, pp. 353-357, May 1996.
[21] Q. Zhang, S. Kassam, “Finite-state markov model for rayleigh fading channels”, *IEEE Tran. on Communications*, vol. 47, no. 11, Nov. 1999.
[22] E. N. Gilbert, “Capacity of burst-noise channel”, *Bell Syst. Tech. J.*,vol. 39, no. 9, pp. 1253-1265, Sept. 1960.
[23] A. Goldsmith and P. Varaiya, “Capacity, mutual Information and coding for finite state markov channels”, *IEEE Trans. Information Theory*, 1996.
[24] W. Fleming and H. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer-Verlag, 1993.
[25] M. Davis, *Markov Models and Optimization*, Chapman and Hall, 1993.
[26] B. Oksendal, *Stochastic Differential Equations*, Springer, 5^{th} edn., 2000.
[27] D. Luenberger, *Optimization by Vector Space Methods*, John Wiley & sons, 1969.
[28] G. R. Grimmett, and D. Stirzaker, *Probability and Random Processes*, 3rd ed. New York, Oxford University Press, 2001.

Appendix A
Proof of Theorem I – *General Markov Channel*

Consider first the L^{th} partition interval, i.e. $k = L$. The system state space for this interval is $(x, c, t) \in [0, B] \times \mathcal{C} \times \left[\frac{(L-1)T}{L}, T\right)$ and over this period, (21) and (22) take the form,

$$r^*_\nu(x, c^i, t) = \frac{x}{f^L_i(T-t)}, \qquad i = 1, \ldots, m \tag{32}$$

$$H_\nu(x, c^i, t) = \frac{(1+\nu_L)x^n}{c^i(f^L_i(T-t))^{n-1}}, \quad i = 1, \ldots, m \tag{33}$$

The Hamilton-Jacobi-Bellman (HJB) equation over this period is given as in (20) with $\nu_k = \nu_L$. To prove optimality of the above functional forms, we need the following verification result from stochastic optimal-control theory [24]. It states that if we can find a functional form $H_\nu(x, c^i, t)$ that satisfies the HJB equation and the boundary conditions then it is the optimal solution.

Lemma 2: (Verification Result) *([24], Chap III, Theorem 8.1) Consider the L^{th} partition interval and let $H_\nu(x, c, t)$ defined on $(x, c, t) \in [0, B] \times \mathcal{C} \times \left[\frac{(L-1)T}{L}, T\right]$, solve the equation in (20) with the boundary conditions $H_\nu(0, c, t) = 0$ and $H_\nu(x, c, T) = \frac{\tau g(\frac{x}{\tau})}{c}$. Then,*

1) $H_\nu(x, c, t) \leq H^r_\nu(x, c, t), \ \forall\ r(\cdot) \in \Phi$

2) *Let $r_\nu^*(x,c,t) \in \Phi$ be such that r_ν^* is the minimizing value of r in (20), then, $r_\nu^*(x,c,t)$ is an optimal policy, $H_\nu(x,c,t)$ is the minimum cost-to-go function and,*

$$H_\nu(x,c,t) = E\left[\int_t^T \frac{g(r_\nu^*(x(s),c(s),s))}{c(s)}ds + \frac{\tau g(\frac{x(T)}{\tau})}{c(T)}\right] \quad (34)$$

By verifying the requirements in the above lemma, we now show that (32) and (33) are the optimal solution for the L^{th} interval. First note that $g(r) = r^n$ and from the boundary conditions on $f_i^L(s)$ in Theorem I, we have $f_i^L(0) = \tau(1+\nu_L)^{\frac{1}{n-1}}, \forall i$. Using this it is easy to check that the boundary conditions $H_\nu(0,c,t) = 0$ and $H_\nu(x,c,T) = \frac{\tau g(\frac{x}{\tau})}{c}$ are satisfied.

Now, substituting (32) and (33) into the PDE equation in (20) gives,

$$\frac{(1+\nu_L)x^n}{c^i(f_i^L(T-t))^n} + \frac{-(1+\nu_L)x^n(1-n)(f_i^L)'(T-t)}{c^i(f_i^L(T-t))^n} - \frac{x}{f_i^L(T-t)}\frac{n(1+\nu_L)x^{n-1}}{c^i(f_i^L(T-t))^{n-1}} + \lambda\sum_{j=1}^m \frac{p_{ij}}{z_{ij}c^i}\frac{(1+\nu_L)x^n}{(f_j^L(T-t))^{n-1}} - \lambda\frac{(1+\nu_L)x^n}{c^i(f_i^L(T-t))^{n-1}} = 0$$

Cancelling out $\frac{(1+\nu_L)x^n}{c^i}$, simplifying the above and setting $s = T-t$ gives the following ODE system (note, $t \in \left[\frac{(L-1)T}{L}, T\right]$ implies that $s = (T-t) \in [0, T/L]$),

$$(f_i^L(s))' = 1 + \frac{\lambda f_i^L(s)}{n-1} - \frac{\lambda}{n-1}\sum_{j=1}^m \frac{p_{ij}}{z_{ij}}\frac{(f_i^L(s))^n}{(f_j^L(s))^{n-1}}, \quad i = 1,\ldots,m \quad (35)$$

Thus, from above we see that $r_\nu^*(\cdot)$ and $H_\nu(\cdot)$ as given in (32) and (33) respectively, would satisfy the optimality PDE equation if the functions $\{f_i^L(s)\}_{i=1}^m$, $s \in [0,T/L]$, satisfy the above ODE system with the boundary conditions $f_i^L(0) = \tau(1+\nu_L)^{\frac{1}{n-1}}, \forall i$. The following lemma shows that indeed such a set of positive functions exists and also that they are unique.

Lemma 3: (Existence and Uniqueness of the ODE solution in (35)) *The ODE system in (35) with the boundary conditions $f_i^L(0) = \tau(1+\nu_L)^{\frac{1}{n-1}}, \forall i$, has a unique positive solution for $s \in [0,T/L]$.*

Proof: The proof is omitted for brevity but can be found in [4] ■

This completes the verification that $H_\nu(x,c,t)$ as given in (33) satisfies the optimality PDE equation. Furthermore, it is easy to check that the rate r^* as given in (32) is the minimizing value of r in (20) (take the first derivative with respect to r and set it to zero). The admissibility of $r_\nu^*(x,c,t)$ follows by noting that the functional form in (32) is continuous and locally Lipschitz in x, continuous in t and satisfies $r_\nu^*(0,c,t) = 0$. Thus, we have verified all the requirements in Lemma 2 and this completes the proof that (32) and (33) give the optimal solution over the L^{th} partition interval.

Now, consider the $(L-1)^{th}$ partition interval, i.e. $k = L-1$. The system state space for this interval is $(x,c,t) \in [0,B] \times \mathcal{C} \times \left[\frac{(L-2)T}{L}, \frac{(L-1)T}{L}\right)$ and over this period, equations (21) and (22) take the form,

$$r_\nu^*(x,c^i,t) = \frac{x}{f_i^{L-1}(T-t)}, \quad i = 1,\ldots,m \quad (36)$$

$$H_\nu(x,c^i,t) = \frac{(1+\nu_{L-1})x^n}{c^i(f_i^{L-1}(T-t))^{n-1}}, \quad i = 1,\ldots,m \quad (37)$$

Suppose that we start with a system state in the $(L-1)^{th}$ partition interval. Once we reach the L^{th} interval, i.e. $t = \frac{(L-1)T}{L}$, we know from the preceding arguments that (33) gives the minimum cost and (32) gives the optimal rate to be followed thereafter. Thus, for the optimization over the $(L-1)^{th}$ interval, we can abstract the L^{th} interval as a terminal cost of $H_\nu\left(x,c^i,(L-1)T/L\right) = \frac{(1+\nu_L)x^n}{c^i(f_i^L(T/L))^{n-1}}$, applied at $t = \frac{(L-1)T}{L}$. The minimization problem in (11) over the $(L-1)^{th}$ interval is therefore identical to that over the L^{th} interval (discussed earlier) except that we now have a different boundary condition given as, $H_\nu(x,c,(L-1)T/L) = \frac{(1+\nu_L)x^n}{c^i(f_i^L(T/L))^{n-1}}$. Using the functional form in (37), this boundary condition translates into $f_i^{L-1}\left(\frac{T}{L}\right) = \left(\frac{1+\nu_{L-1}}{1+\nu_L}\right)^{\frac{1}{n-1}} f_i^L\left(\frac{T}{L}\right), \forall i$ (as outlined in the Theorem I statement). Now, following an identical set of arguments as done for the L^{th} partition interval, it is easy to check that (36) and (37) give the optimal solution over the $(L-1)^{th}$ interval.

Finally, recursively going backwards and considering the partition intervals, $k = L-2, L-3, \ldots, 1$, it follows that (21) and (22) with the boundary conditions as in the Theorem statement, give the optimal solution. This completes the proof.

Appendix B
Proof of Theorem II – *Constant Drift Channel*

The proof for this theorem is identical to that of the general case in Theorem I except that now the functions $\{f_i^k(s)\}$ can be evaluated in closed form. Therefore, to avoid repetition we only present the details regarding the functions $\{f_i^k(s)\}$. As before, start with the L^{th} partition interval and suppose that for all the channel states the $f_i^L(s)$ function is the same, i.e. $f_i^L(s) = f^L(s)$. The ordinary differential for $f^L(s)$ then becomes,

$$(f^L)'(s) = 1 + \frac{\lambda f^L(s)}{n-1} - \frac{\lambda}{n-1}f^L(s)\left(\sum_j \frac{p_{ij}}{z_{ij}}\right) \quad (38)$$

$$= 1 - \frac{\lambda f^L(s)}{n-1}(\beta - 1) \quad (39)$$

where $\sum_j \frac{p_{ij}}{z_{ij}} = E[1/Z(c^i)] = \beta, \ \forall i$, by the constant drift channel assumption. The solution to the above ODE evaluated over $s \in [0,T/L]$ with the boundary condition $f^L(0) = \tau(1+\nu_L)^{\frac{1}{n-1}}$ is given as (let $\eta = \frac{\lambda(\beta-1)}{n-1}$),

$$f^L(s) = \tau(1+\nu_L)^{\frac{1}{n-1}}e^{-\eta s} + \frac{1}{\eta}\left(1 - e^{-\eta s}\right) \quad (40)$$

Clearly, for $k = L$, equation (28) is the same as (40) above (set $s = T-t$).

Now, consider the $(L-1)^{th}$ partition interval and following the same argument as for the L^{th} partition interval, it is easy to see that $f^{L-1}(s)$ satisfies the same ODE as given in (39). This ODE must now be evaluated over $s \in \left[\frac{T}{L}, \frac{2T}{L}\right]$ with the following boundary condition,

$$f^{L-1}\left(\frac{T}{L}\right) = \left(\frac{1+\nu_{L-1}}{1+\nu_L}\right)^{\frac{1}{n-1}} f^L\left(\frac{T}{L}\right)$$
$$= \tau(1+\nu_{L-1})^{\frac{1}{n-1}} e^{-\frac{\eta T}{L}} + \frac{1}{\eta}\left(\frac{1+\nu_{L-1}}{1+\nu_L}\right)^{\frac{1}{n-1}}\left(1-e^{-\eta\frac{T}{L}}\right)$$

Evaluating the ODE with the above boundary condition gives $f^{L-1}(s)$ as follows,

$$f^{L-1}(s) = \tau(1+\nu_{L-1})^{\frac{1}{n-1}} e^{-\eta s} + \frac{1}{\eta}\left(\frac{1+\nu_{L-1}}{1+\nu_L}\right)^{\frac{1}{n-1}} \times \left(e^{-\eta\left(s-\frac{T}{L}\right)} - e^{-\eta s}\right) + \frac{1}{\eta}\left(1-e^{-\eta\left(s-\frac{T}{L}\right)}\right) \quad (41)$$

Again for $k = L-1$, equation (28) is the same as (41) above with $s = T-t$. Recursing backwards and following the same steps as earlier, it can be seen that $f^k(s)$ can be written in the general form as given in (28).

APPENDIX C
PROOF OF THEOREM III – *Strong Duality*

The optimization problem $(\mathcal{P})$ as stated earlier is given as,

$$(\mathcal{P}) \quad \min_{r(\cdot)\in\Phi} \; E\left[\int_0^T \frac{1}{c(s)} g(r(x(s),c(s),s))ds + \frac{\tau g(\frac{x(T)}{\tau})}{c(T)}\right] \quad (42)$$
$$\text{sub. to } E\left[\int_{\frac{(k-1)T}{L}}^{\frac{kT}{L}} \frac{g(r(x(s),c(s),s))}{c(s)} ds\right] \le \frac{PT}{L}, k = 1,..,L$$

Before proceeding to show strong duality holds, we first interchange the expectations and the integrals and re-write the above problem in a standard form as in [27]. But to do that, we need the following. Let $I_{[a,b]}(s)$ be the indicator function for the interval $s \in [a,b]$; it is defined as,

$$I_{[a,b]}(s) \triangleq \begin{cases} 1, & \text{if } s \in [a,b] \\ 0, & \text{otherwise} \end{cases} \quad (43)$$

Also define,

$$K^r(s) \triangleq \frac{g(r(x(s),c(s),s))}{c(s)} I_{[0,T]}(s) + g\left(\frac{x_0 - \int_0^T r(x(t),c(t),t)dt}{\tau}\right)\frac{1}{c(T)} I_{[T,T+\tau]}(s) \quad (44)$$

Let $F(r(\cdot))$ denote the total cost for policy $r(\cdot)$ (i.e. the objective function in $(\mathcal{P})$). From (42), it is given as,

$$F(r(\cdot)) = E\left[\int_0^T \frac{1}{c(s)} g(r(x(s),c(s),s))ds + \frac{\tau g(\frac{x(T)}{\tau})}{c(T)}\right] \quad (45)$$

Using (44), we can re-write the above as,

$$F(r(\cdot)) = E\left[\int_0^{T+\tau} K^r(s)ds\right] \quad (46)$$

For any policy $r(\cdot) \in \Phi$, it is clear that $K^r(s),\ s \in [0, T+\tau]$ is a collection of non-negative random variables which depend on the underlying channel stochastic process. Hence, using Fubini's theorem [28], we can interchange the expectation and the integral which gives,

$$F(r(\cdot)) = \int_0^{T+\tau} E[K^r(s)]ds \quad (47)$$

Similarly, we can interchange the expectation and the integral for the power constraint inequalities in (42). Thus, we can now re-write the optimization problem $(\mathcal{P})$ as,

$$\min_{r(\cdot)\in\Phi} \; F(r(\cdot)) \quad (48)$$
$$\text{subject to} \int_{\frac{(k-1)T}{L}}^{\frac{kT}{L}} E\left[\frac{g(r(x(s),c(s),s))}{c(s)}\right] ds - \frac{PT}{L} \le 0, \quad k = 1,\dots,L$$

where $F(r(\cdot))$ is as given in (47). Now, having written the optimization problem $(\mathcal{P})$ in the above form, the strong duality result in [27] (Theorem 1, sec. 8.6, pp. 224) gives the results as stated in Theorem III, which then completes the proof. However, as a final step we need to verify the technical conditions required in [27]. These are presented below with a description of the technical requirement and the proof for its validity in our case.

(1) $F(r(\cdot))$ *is a convex functional over* $r(\cdot) \in \Phi$

Consider two policies $r_1(x,c,t), r_2(x,c,t) \in \Phi$ and let $0 \le \alpha \le 1$. Let $\tilde{r}(x,c,t) = \alpha r_1(x,c,t) + (1-\alpha) r_2(x,c,t)$; since $r_1(\cdot), r_2(\cdot) \in \Phi$ it is easy to check that $\tilde{r}(\cdot)$ also lies in Φ. Now,

$$\begin{aligned} K^{\tilde{r}}(s) &= \frac{g(\tilde{r}(x(s),c(s),s))}{c(s)} I_{[0,T]}(s) \\ &+ g\left(\frac{x_0 - \int_0^T \tilde{r}(x(t),c(t),t)dt}{\tau}\right)\frac{1}{c(T)} I_{[T,T+\tau]}(s) \\ &\le \alpha\Big\{\frac{g(r_1(x(s),c(s),s))}{c(s)} I_{[0,T]}(s) \\ &+ g\left(\frac{x_0 - \int_0^T r_1(x(t),c(t),t)dt}{\tau}\right)\frac{1}{c(T)} I_{[T,T+\tau]}(s)\Big\} \\ &+ (1-\alpha)\Big\{\frac{g(r_2(x(s),c(s),s))}{c(s)} I_{[0,T]}(s) \\ &+ g\left(\frac{x_0 - \int_0^T r_2(x(t),c(t),t)dt}{\tau}\right)\frac{1}{c(T)} I_{[T,T+\tau]}(s)\Big\} \\ &= \alpha K^{r_1}(s) + (1-\alpha) K^{r_2}(s) \end{aligned}$$

where the inequality above follows since $g(r)$ is a convex function of r. Thus, from above $K^r(s)$ is a convex functional over $r(\cdot) \in \Phi$ and this implies that $E[K^r(s)]$ is a convex functional. It then directly follows that $F(r(\cdot))$ is a convex functional over $r(\cdot) \in \Phi$.

(2) *Let* $G_k(r(\cdot)) = \left(\int_{\frac{(k-1)T}{L}}^{\frac{kT}{L}} E\left[\frac{g(r(x(s),c(s),s))}{c(s)}\right] ds - \frac{PT}{L}\right)$, $k = 1,\dots,L$, *then,* $G_k(r(\cdot))$ *is a convex functional over* $r(\cdot) \in \Phi$. The proof for this is identical to the previous case and omitted here for brevity.

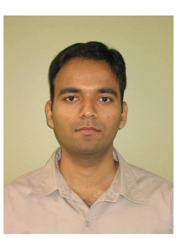
(3) *Minimum cost for problem $(\mathcal{P})$ is finite*

To see this consider the simple policy which does not transmit any data and only incurs the terminal cost. The total cost for this policy is simply the expected penalty cost and is given as,

$$\begin{aligned} \text{total cost} &= E\left[\frac{g(x_0/\tau)}{c(T)}\tau\right] = \tau g\left(\frac{x_0}{\tau}\right)\ E\left[\frac{1}{c(T)}\right] \\ &\leq \tau g\left(\frac{x_0}{\tau}\right)\sum_{j=0}^{\infty}\left(\frac{1}{c_0(z_l)^j}\right)\frac{(\lambda T)^j\, e^{-\lambda T}}{j!} \\ &= \frac{\tau}{c_0}\, g\left(\frac{x_0}{\tau}\right) e^{\frac{\lambda T}{z_l}}\, e^{-\lambda T} < \infty \end{aligned}$$

The inequality above follows by first conditioning that the channel makes j transitions over $[0, T]$, taking $c(T) = (z_l)^j c_0$, where $(z_l)^j c_0$ is the worst possible channel quality starting with state c_0 and making j transitions, and finally taking expectation with respect to j (number of transitions, j, is Poisson distributed with rate λT and $z_l > 0$ is the least value that any $Z(c)$ can take). Now, since there exists an admissible policy with a finite cost, it follows that the minimum cost over all admissible $r(\cdot)$ is finite.

(4) *Let* $G_k(r(\cdot)) = \left(\int_{\frac{(k-1)T}{L}}^{\frac{kT}{L}} E\left[\frac{g(r(x(s),c(s),s))}{c(s)}\right] ds - \frac{PT}{L}\right)$, $k = 1, \ldots, L$, *then, a policy* $r(\cdot) \in \Phi$ *exists such that* $G_k(r(\cdot)) < 0,\ \forall k$ (the interior-point policy). Take $r(\cdot)$ as the policy that does not transmit at all and only incurs the terminal cost.



Murtaza A. Zafer received the B.Tech degree in Electrical Engineering from the Indian Institute of Technology (IIT) Madras, India, in 2001, and the S. M. and Ph.D. degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (MIT), MA, USA, in 2003 and 2007 respectively. Currently, he is a Research Staff Member at IBM Thomas J. Watson Research Center, NY, USA, where his research is on the theory and algorithms for wireless communications, mobile ad-hoc and sensor networks. He spent the summer of 2004 at the Mathematical Sciences Research center, Bell Laboratories Alcatel-Lucent Inc. He is a co-author of the Best Student Paper award at the WiOpt conference, 2005, and a recipient of the Siemens and Philips award for academic excellence in 2001.

Eytan Modiano received his B.S. degree in Electrical Engineering and Computer Science from the University of Connecticut at Storrs in 1986 and his M.S. and PhD degrees, both in Electrical Engineering, from the University of Maryland, College Park, MD, in 1989 and 1992 respectively. He was a Naval Research Laboratory Fellow between 1987 and 1992 and a National Research Council Post Doctoral Fellow during 1992-1993. Between 1993 and 1999 he was with MIT Lincoln Laboratory where he was the project leader for MIT Lincoln Laboratory's Next Generation Internet (NGI) project. Since 1999 he has been on the faculty at MIT; where he is presently an Associate Professor. His research is on communication networks and protocols with emphasis on satellite, wireless, and optical networks.

He is currently an Associate Editor for IEEE Transactions on Information Theory, The International Journal of Satellite Communications, and for IEEE/ACM Transactions on Networking. He had served as a guest editor for IEEE JSAC special issue on WDM network architectures; the Computer Networks Journal special issue on Broadband Internet Access; the Journal of Communications and Networks special issue on Wireless Ad-Hoc Networks; and for IEEE Journal of Lightwave Technology special issue on Optical Networks. He was the Technical Program co-chair for Wiopt 2006, IEEE Infocom 2007, and ACM MobiHoc 2007.