On Optimal Power Control for Delay-Constrained Communication over Fading Channels

Xiaochen Li, Xihua Dong, and Dapeng Wu

Abstract— In this paper, we study the problem of optimal power control for delay-constrained communication over fading channels. Our objective is to find a power control law that optimizes the link layer performance, specifically, minimizes delay bound violation probability (or equivalently, the packet drop probability), subject to constraints on average power, arrival rate, and delay bound. The transmission buffer size is assumed to be finite; hence, when the buffer is full, there will be packet drop. The fading channel under our study has a continuous state, e.g., Rayleigh fading. Since directly solving the power control problem (which optimizes the link layer performance) is particularly challenging, we decompose it into three sub-problems, and solve the three sub-problems iteratively; we call the resulting scheme Joint Queue Length Aware (JQLA) power control, which produces a local optimal solution to the three sub-problems. We prove that the solution that simultaneously solves the three sub-problems is also an optimal solution to the optimal power control problem. Simulation results show that the JQLA scheme achieves superior performance over the time domain water filling and the truncated channel inversion power control. E.g., JQLA achieves 10 dB gain at packet drop probability of 10^{-3}, over the time domain water filling power control.

Index Terms— Delay-constrained communication, power control, queuing analysis, delay bound violation probability, packet drop probability.

I. INTRODUCTION

Real-time applications such as streaming multimedia will be supported in the next generation wireless networks. Services required by these applications are different from elastic traffic in that they expect low end-to-end delay, i.e., delay-constrained communication. It is particularly challenging to provide delay guarantee or quality of service (QoS) guarantees to delay sensitive applications since a wireless channel can experience time varying channel capacity due to fading.

The studies on delay-constrained communication typically use one of the two models, i.e., the physical layer (PHY) model and the link-PHY model. The difference between the two models is whether or not a queue is included in the system. In a PHY model, there is no buffer and hence the end-to-end delay considered here only consists of channel-encoding delay and decoding delay in the physical layer; assuming a block fading channel [1], reliable communication is achieved by channel-encoding the information bits within M fading blocks

to average out the random effects caused by thermal noise and fading; then the end-to-end delay is $2M$ fading blocks[1].

For the PHY model, a delay-constrained capacity of a fading channel [1] can be used as the upper bound on the performance of a delay-constrained communication system; existing delay-constrained capacity notions include outage capacity [2], delay-limited capacity [3], and expected capacity [4]. When the channel state information (CSI) is available at the transmitter side, power control can be utilized to increase capacity. The optimal causal power control scheme which maximizes the expected capacity subject to average power constraint is time domain water filling (TDWF) [4], which can also be used to maximize the ergodic capacity [5]. An optimal non-causal power control scheme is studied in [6], where the channel gains of all M fading blocks are assumed to be known at the beginning of the transmission, which is similar to parallel channels. The optimal power control that maximizes outage capacity under non-causal CSI, is studied in [7]; in the special case of $M = 1$, the optimal power control is truncated channel inversion (TCI) [5], [7]. The optimal power control that maximizes outage capacity under causal CSI, is studied in [4].

In the PHY model, it is implicitly assumed that the arrival rate from the upper layer is equal to the information transmission rate of the physical layer, which is not valid in practice (note that there is no buffer in the PHY model). The PHY model is not practical since there will be mismatch between the arrival rate and the departure rate. To address this, we can use a link-PHY model, where a buffer is in place at the link layer to accommodate the mismatch between the arrival rate from the upper layer and the departure rate at PHY. The packets arriving from the upper layer are stored in the buffer until they are transmitted or dropped due to full buffer or delay bound violation. Both the arrival process and the channel fading process (departure process) contribute to the dynamics of the system. The end-to-end delay considered here consists of queuing delay in the buffer and encoding delay and decoding delay in the physical layer. With the knowledge of CSI at the transmitter side, we can control the departure process to achieve optimality for certain performance measures. In general, the objective of the optimization involves one of the following [8]–[13]: maximizing throughput or effective capacity [14], or minimizing energy/power, or minimizing delay (average delay or delay bound violation probability), or minimizing packet loss probability. When one

---

[1]The encoding delay is M. The decoding delay is also M. So the end-to-end delay is $2M$ fading blocks.

performance measure is chosen as the objective function for the optimization, the other performance measures will serve as constraints; hence, the problem is a constrained optimization problem.

In this paper, we study the problem of optimal power control under the link-PHY model. Our objective is to find a power control law that optimizes the link layer performance, specifically, minimizes delay bound violation probability (or equivalently, the packet drop probability), subject to constraints on average power, arrival rate, and delay bound. The buffer at the link layer is assumed to be finite; hence, when the buffer is full, there will be packet drop. The fading channel under our study has a continuous state, e.g., Rayleigh fading. Since the channel state is continuous, dynamic programming is not applicable for power control. If dynamic programming is to be used, we need to quantize the the continuous channel state; then the system will suffer from capacity loss due to quantization error; in addition, the computational complexity of dynamic programming increases quadratically with the increase of the number of discrete channel states. Since directly solving the power control problem (which optimizes the link layer performance) is particularly challenging, we take a divide-and-conquer approach, i.e., decompose the optimal power control problem into three sub-problems, and solve the three sub-problems iteratively till convergence; we call the resulting scheme Joint Queue Length Aware (JQLA) power control, which produces a local optimal solution to the three sub-problems. We prove that the solution that simultaneously solves the three sub-problems is also an optimal solution to the optimal power control problem. Simulation results show that the JQLA scheme achieves superior performance over the time domain water filling and the truncated channel inversion power control. E.g., JQLA achieves 10 dB gain at packet drop probability of 10^{-3}, over the time domain water filling power control. The algorithm to find the JQLA power control scheme is too complex; hence one may not use it in practice. Instead, the main purpose of JQLA is to explore the fundamental performance limit of power control under the link-PHY model. Although JQLA does not produce a global optimal solution, this work represents a major step toward deriving the fundamental performance limit of power control under the link-PHY model. The local optimal solution obtained by JQLA can provide a judgement on the performance of a practical power control scheme. The 10 dB gain achieved by JQLA indicates that there is much room to improve for existing power control schemes!

Power control is a heavily researched topic. The uniqueness of this work is that we study the optimal power control law that minimizes delay bound violation probability under a finite buffer and a fading channel with continuous state. This problem has not been addressed before. Existing works on delay-constrained optimal power control either minimize average delay or minimize effective capacity under an 'infinite' buffer (which is not practical) [15]–[18], or under a fading channel with finite discrete states (which suffers from capacity loss). Note that average-delay guarantee may not satisfy the requirements of delay-sensitive applications; e.g., using a handheld device to watch mobile TV over WiMax, requires

certain delay bound violation probability, which cannot be specified by average delay since average delay cannot specify the (tail) probability distribution function; e.g., for a given delay bound (say, 1 second), two systems with the same average delay of 500 ms could have quite different delay bound violation probabilities, e.g., 40% vs. 0.1%.

The rest of the paper is organized as follows. Section II introduces the system model. Section III describes our method to address the optimal power control problem. Section IV presents the simulation results. Section V concludes the paper.

II. SYSTEM MODEL

We consider a point-to-point communication model as illustrated in Fig. 1. The data packets from the upper layer enter the buffer at the link layer. Assume each packet at the link layer has a size of L bits. We assume the channel gain is perfectly known at the transmitter side. Given the channel gain and the queue length in the buffer, the power and rate control module determines the transmission power and rate (the number of packets that will be transmitted during one block). Then, the head-of-line (HOL) packets are removed from the buffer and conveyed to the physical layer. Then the packets are encoded, modulated and transmitted through a wireless channel. The information transmission rate (or code rate in unit of packets per block) in the physical layer is the same as the service rate of the buffer.

To elaborate, the wireless channel is assumed to be flat slow fading, which can be modeled as a discrete-time block-fading channel with additive white Gaussian noise, i.e., the channel gain is fixed in a block and the channel gains are independent and identically distributed (i.i.d.) [1]. We assume each block has a duration of T_b sec. The channel power gain of each block takes continuous value. Its marginal distribution is characterized by a probability density function (pdf) $f_{CH}(g)$, $g \in [0, \infty)$; without loss of generality, the noise variance is absorbed into g. We assume that a block is the smallest time interval, during which the transmitter can conduct power and rate adaptation. We assume that ideal channel code and modulation are used so that instantaneous channel capacity (determined by Shannon's capacity formula for the given channel gain in the block) can be achieved in each block. Then, the maximum error-free transmission rate in a block is uniquely determined by the transmission power through Shannon's channel capacity formula. For simplicity, we let the service rate of the buffer in a block equal to the instantaneous channel capacity in the block, and assume the resulting decoding error probability is zero due to ideal channel coding.

The queueing subsystem is modeled as a discrete-time finite-buffer queue with buffer size M packets. The data source generates packets at a constant rate μ packets per block. (In our future work, we will consider random arrival instead of constant arrival.) When newly arrived packets see a full buffer, the system drops the HOL packets in the queue and appends the incoming packets to the tail of the queue. So the maximum queueing delay is M/μ. The maximum end-to-end delay in the link-PHY model is M/μ plus the channel encoding/decoding delay, which we denote D_c. Given the end-to-end delay bound

D_{max}, if we set the buffer size $M = \mu \times (D_{max} - D_c)$, then there will be zero delay bound violation; in other words, all packets that violate the delay bound, are dropped. Hence, we say, **the delay bound violation probability is equal to the packet drop probability** in the system under our study.

Fig. 1. System model.

Fig. 2. Update of queue length.

As shown in Fig. 2, in the n-th block, the queue length $q(n)$ is observed at the observation epoch, then a batch of μ packets arrive, followed by the departure of $s(n)$ packets. Denote $d(n)$ the number of packets dropped in the n-th block. Then,

$$d(n) = \max(0, q(n) + \mu - M). \tag{1}$$

The number of packets remains in the buffer before the transmission is

$$\begin{aligned} q_r(n) &= q(n) - d(n) \\ &= \min(M, q(n) + \mu). \end{aligned} \tag{2}$$

The queue length is updated by

$$\begin{aligned} q(n+1) &= \max(0, q_r(n) - s(n)) \\ &= \max(0, \min(M, q(n) + \mu) - s(n)). \end{aligned} \tag{3}$$

The sequence $\{q(n)\}$ forms a homogeneous, irreducible, and aperiodic Markov Chain. The steady state queue length distribution can be obtained from the one step transition probability matrix $\mathbf{P}$. Since the buffer has finite capacity M, $\mathbf{P}$ is a square matrix of size $(M+1) \times (M+1)$. The i-th row j-th column of $\mathbf{P}$, $0 \leq i, j \leq M$, is

$$p_{i,j} = \text{Prob}[q(n+1) = j | q(n) = i]. \tag{4}$$

Substituting (3) into (4), we obtain

$$p_{i,j} = \text{Prob}[\max(0, \min(M, i+\mu) - s(n)) = j | q(n) = i]. \tag{5}$$

To calculate $p_{i,j}$ for each pair of $\{i, j\}$, we consider two cases as below.

Case 1) $0 \leq i < M - \mu$ and no packet-drop:

$$\begin{aligned} p_{i,j} &= \text{Prob}[\max(0, i + \mu - s(n)) = j | q(n) = i] \\ &= \begin{cases} \text{Prob}[s(n) \geq i + \mu | q(n) = i] & j = 0 \\ \text{Prob}[s(n) = i + \mu - j | q(n) = i] & 0 < j \leq i + \mu \\ 0 & i + \mu < j \leq M \end{cases} \end{aligned} \tag{6}$$

Case 2) $M - \mu \leq i \leq M$ and some packets dropped:

$$\begin{aligned} p_{i,j} &= \text{Prob}[M - s(n) = j | q(n) = i] \\ &= \text{Prob}[s(n) = M - j | q(n) = i] \quad 0 \leq j \leq M. \end{aligned} \tag{7}$$

Denote $k_x^{(i)}$ the probability that x packets can be transmitted when the queue length is i,

$$k_x^{(i)} = \text{Prob}[s(n) = x | q(n) = i], \tag{8}$$

where

$$s(n) = \left\lfloor \frac{WT_b}{L} \log_2 \left(1 + P(g(n), q(n))g(n)\right) \right\rfloor, \tag{9}$$

where W is the channel bandwidth in Hz.

Then the transition probabilities $\{p_{i,j}\}$ are given by

$$p_{i,j} = \begin{cases} \sum_{l=i+\mu}^{M} k_l^{(i)} & 0 \leq i < M - \mu, j = 0 \\ k_{i+\mu-j}^{(i)} & 0 \leq i < M - \mu, 0 < j \leq i + \mu \\ k_{M-j}^{(i)} & M - \mu \leq i \leq M, 0 \leq j \leq M \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

Denote $c_x^{(i)}$ the probability that x or more than x packets can be transmitted when the queue length is i, i.e.,

$$c_x^{(i)} = \sum_{l=x}^{M} k_l^{(i)}, \tag{11}$$

Then the transition probability matrix is given by

$$\mathbf{P} = \begin{bmatrix} c_\mu^{(0)} & \cdots & k_0^{(0)} & 0 & \cdots & 0 \\ c_{\mu+1}^{(1)} & \cdots & k_1^{(1)} & k_0^{(1)} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{M-1}^{(M-\mu-1)} & \cdots & \cdots & \cdots & \cdots & 0 \\ k_M^{(M-\mu)} & \cdots & \cdots & \cdots & \cdots & k_0^{(M-\mu)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ k_M^{(M)} & \cdots & \cdots & \cdots & \cdots & k_0^{(M)} \end{bmatrix}. \tag{12}$$

The steady state queue length distribution is given by [19]

$$\begin{cases} \boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \\ \sum_{i=0}^{M} \pi_i = 1 \end{cases} \tag{13}$$

where $\boldsymbol{\pi} = [\pi_0, \pi_1, \cdots, \pi_M]$ is a $1 \times (M+1)$ row vector. The element π_i is the probability of queue length equal to i when the queue enters the steady state, i.e.,

$$\pi_i = \lim_{n \to \infty} \text{Prob}[q(n) = i]. \tag{14}$$

Knowing the steady state queue length distribution, we can now derive the packets drop probability. The packet drop probability is defined as the ratio of the number of dropped packets to the number of total arrival packets, i.e.,

$$P_{drop} = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} d(i)}{n\mu}. \tag{15}$$

Denote $L_d(n)$ the average number of dropped packets per block in the first n blocks, i.e.,

$$L_d(n) = \frac{\sum_{i=1}^{n} d(i)}{n}. \tag{16}$$

The packet drop probability is

$$P_{drop} = \frac{L_d(\infty)}{\mu}. \tag{17}$$

When the queue enters the steady state, we have

$$\begin{aligned} L_d(\infty) &= \mathbf{E}_{q(\infty)}[d(\infty)] \\ &= \sum_{l=M-\mu}^{M} \text{Prob}[q(\infty) = l](\mu + l - M) \\ &= \sum_{l=M-\mu}^{M} \pi_l(\mu - l + M), \end{aligned} \tag{18}$$

where $d(\infty)$ and $q(\infty)$ denote the number of dropped packets, and queue length at the steady state, respectively.

Similarly, the average transmission power is,

$$\begin{aligned} \overline{P} &= \lim_{n \to \infty} \mathbf{E}_{g(n),q(n)}[P(g(n),q(n))] \\ &\overset{(a)}{=} \sum_{l=0}^{M} \pi_l \mathbf{E}_{g(n)}[P(g(n),l)]. \end{aligned} \tag{19}$$

where $P(g(n), q(n))$ is the transmission power, as a function of channel gain $g(n)$ and $q(n)$; Step (a) holds because $g(n)$ and $q(n)$ are independent. In the rest of the paper, we will omit block index n, since the channel gains are i.i.d., and the queue length distribution does not change in the steady state,

As mentioned in Section I, our objective is to find a power control law that minimizes delay bound violation probability, subject to constraints on average power, arrival rate, and delay bound. As discussed earlier, given the end-to-end delay bound D_{max} and constant arrival rate μ, we can set the buffer size $M = \mu \times (D_{max} - D_c)$ so that the delay bound violation probability is equal to the packet drop probability. Then, given average power constraint P_0, we formulate the optimal power control problem as below

$$\begin{aligned} \min_{P(g,q)} \quad & P_{drop} = \frac{1}{\mu} \sum_{l=M-\mu}^{M} \pi_l(\mu - l + M) \\ s.t. \quad & \sum_{l=0}^{M} \pi_l \mathbf{E}_g[P(g,l)] \le P_0 \\ & P(g,q) \ge 0 \\ & \boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}(P(g,q)), \\ & \sum_{i=0}^{M} \pi_i = 1 \end{aligned} \tag{20}$$

where $\mathbf{P}(P(g,q))$ implies that the transition probability matrix is a function of the power control law $P(g,q)$. The solution of (20) is the optimal power control law that we seek.

III. JQLA POWER CONTROL SCHEME

In this section, we present a method to solve (20); the resulting power control is called JQLA.

At first sight, one would use the method of Lagrange multipliers to solve (20). However, the method of Lagrange multipliers is not directly applicable; this is because without knowing the exact form of $P(g,q)$, the probability in (8) cannot be derived, and hence we do not know the explicit expression of $\{\pi_i\}$ as a function of $P(g,q)$. To address this difficulty, we decompose (20) into three sub-problems and solve the sub-problems iteratively until it converges. Each of the sub-problems is an optimization problem. By doing so, we are able to find explicit relationship between $P(g,q)$ and $\{\pi_i\}$.

The first sub-problem is intended to find the optimal transmission power for a given transition probability matrix $\mathbf{P}$. It is formulated as follows, for a given $\mathbf{P}$,

$$\min_{\{P(g,q)\}} \quad \sum_{q=0}^{M} \pi_q \mathbf{E}_g[P(g,q)] \tag{21a}$$

$$s.t. \quad \boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \tag{21b}$$

$$p_{i,j} = \text{Prob}[q(n+1) = j | q(n) = i] \tag{21c}$$

$$P(g,q) \ge 0. \tag{21d}$$

The constraint (21c) is related to the transmission power via (8), (9), and (10). Notice that the queue length q takes only discrete value; the transmission power $P(g,q)$ can be indexed by q and represented by a set of continuous functions $\{P_q(g)\}$, $q = 0, 1, \cdots, M$. Then (21a) can be rewritten as

$$\min_{\{P_q(g)\}} \quad \sum_{q=0}^{M} \pi_q \mathbf{E}_g[P_q(g)]. \tag{22}$$

Denote the optimal solution to (21) as $\{P_q^*(g, \mathbf{P})\}$.

The second sub-problem is intended to find the optimal transition probability matrix $\mathbf{P}$, for a given steady state queue length distribution $\boldsymbol{\pi}$, which minimizes the average power, i.e.,

$$\min_{\mathbf{P}} \quad \sum_{i=0}^{M} \pi_i \mathbf{E}_g[P_i^*(g, \mathbf{P})] \tag{23a}$$

$$s.t. \quad \boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \tag{23b}$$

$$p_{i,j} \ge 0, \quad \forall i, j \tag{23c}$$

$$\sum_{j=0}^{\min(i+\mu, M)} p_{i,j} = 1, \quad \forall i \tag{23d}$$

$$p_{i,j} = 0, \quad \forall i, \quad j > \min(i+\mu, M). \tag{23e}$$

(23c), (23d), (23e) are the constraints that a valid transition probability matrix needs to satisfy. Denote $\mathbf{P}^*(\boldsymbol{\pi})$ the solution to (23).

The third sub-problem is intended to find the queue length distribution that minimizes the packet drop probability, for a given $P_i^*(g, \mathbf{P}^*(\boldsymbol{\pi}))$, i.e.,

$$\min_{\boldsymbol{\pi}} \quad P_{drop} = \frac{1}{\mu} \sum_{i=M-\mu}^{M} \pi_i (\mu - i + M) \tag{24a}$$

$$s.t. \quad \sum_{i=0}^{M} \pi_i \mathbf{E}_g[P_i^*(g, \mathbf{P}^*(\boldsymbol{\pi}))] \leq P_0 \tag{24b}$$

$$\pi_i \geq 0, \quad \forall i \tag{24c}$$

$$\sum_{i=0}^{M} \pi_i = 1. \tag{24d}$$

Denote $\boldsymbol{\pi}^*$ the solution to (24). The optimal power control law is $\{P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^*))\}$. The following proposition states that $\{P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^*))\}$ is also an optimal solution to (20).

Proposition 1: Denote $P_0(g, q)$ the solution to (20). The power control law $\{P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^*))\}$ yields the same packet drop probability as $P_0(g, q)$, i.e., $\{P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^*))\}$ is also an optimal solution to (20).

For a proof of Proposition 1, see Appendix A.

The solution to the first and second sub-problems are given in Section III-A and III-B, respectively. The third sub-problem can be numerically solved by sequential quadratic programming (SQP).

A. Solution to the First Sub-problem

For a given transition probability matrix $\mathbf{P}$, the steady state queue length distribution $\{\pi_q\}$ is uniquely determined. Therefore $\{\pi_q\}$ can be viewed as fixed parameters as $\mathbf{P}$. In addition, the i-th row of $\mathbf{P}$ is uniquely determined by $P_i(g)$ and is irrelevant to $P_k(g)$, $k \neq i$. Therefore minimizing the sum in (22) is equivalent to independently minimizing $\mathbf{E}_g[P_q(g)]$ for each $q \in \{0, 1, \cdots, M\}$ as below

$$\min_{\{P_q(g)\}} \quad \mathbf{E}_g[P_q(g)]$$
$$s.t. \quad p_{i,j} = \text{Prob}[q(n+1) = j | q(n) = i] \tag{25}$$
$$P_q(g) \geq 0.$$

Denote $P_q^*(g)$ the solution to (25). Note that there is one-to-one correspondence between transmission power and maximum error-free transmission rate (packets per block). So finding $P_q^*(g)$ is equivalent to finding an optimal partition of g, and the maximum transmission rate (packet per block) of each region in the partition. By Shannon's capacity formula, the minimum power (as a function of channel gain) to convey i packets in one block is easily obtained by,

$$\tilde{P}_i(g) = \frac{A(i)}{g}, \tag{26}$$

where $A(i) = 2^{\frac{iL}{WT_b}} - 1$ ($i \geq 0$), derived from (9). $P_q^*(g)$ must be a piecewise function, formed by component functions $\tilde{P}_i(g)$, $i = \{0, 1, \cdots, \min(q + \mu, M)\}$. A set of $\tilde{P}_i(g)$ curves and an example of $P_q^*(g)$ are shown in Fig. 3.

Denote $\{R_i^{(q)}\}$ the channel gain regions where the transmission rate is i packets per block and the queue length

equals to q; and $\{R_i^{(q)}\}$ is a partition of the non-negative region of the real axis $[0, +\infty)$, i.e., $R_i^{(q)} \bigcap R_j^{(q)} = \phi$ for $i \neq j$, and $\bigcup_{i=0}^{\min(q+\mu, M)} R_i^{(q)} = [0, +\infty)$. Then, $P_q^*(g)$ can be represented by

$$P_q^*(g) = \sum_{i=0}^{\min(q+\mu, M)} \frac{A(i)}{g} \mathbf{1}(g \in R_i^{(q)}). \tag{27}$$

where $\mathbf{1}(x) = 1$ if the condition x is true, and $\mathbf{1}(x) = 0$ otherwise. Substituting (27) into $\mathbf{E}_g[P_q(g)]$, the average power is

$$\mathbf{E}_g[P_q^*(g)] = \sum_{i=0}^{\min(q+\mu, M)} \int_{g \in R_i^{(q)}} \frac{A(i)}{g} f_{CH}(g) dg. \tag{28}$$

From (3), the transition probability $p_{q,j}$ is

$$p_{q,j} = \int_{g \in R_{\min(q+\mu, M)-j}^{(q)}} f_{CH}(g) dg. \tag{29}$$

The optimization over $P_q(g)$ in (25) is converted into optimization over the regions $\{R_i^{(q)}\}$ as below

$$\min_{\{R_i^{(q)}\}} \quad \sum_{i=0}^{\min(q+\mu, M)} \int_{g \in R_i^{(q)}} \frac{A(i)}{g} f_{CH}(g) dg$$
$$\tag{30}$$
$$s.t. \quad p_{q,j} = \int_{g \in R_{\min(q+\mu, M)-j}^{(q)}} f_{CH}(g) dg.$$
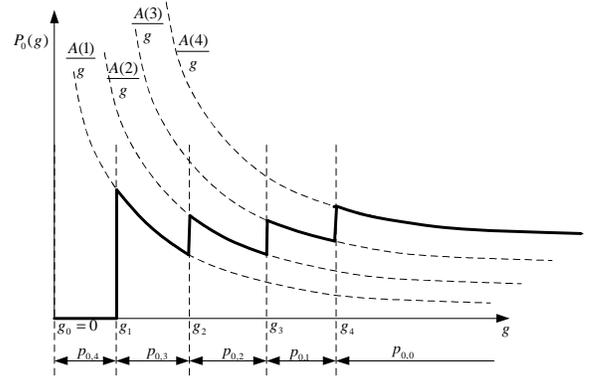
Fig. 3. An example of $P_q^*(g)$ is shown in solid curves for $q = 0$ and $\mu = 4$.

Without loss of generality, a feasible region $R_i^{(q)}$ may not be continuous. For example, it is possible that $R_i^{(q)} = [3, 5) \bigcup [7, 10)$. However, Proposition 2 states that if $\{R_i^{(q)}\}$ is the optimal solution to (30), then each region $R_i^{(q)}$ ($\forall i$) is continuous. To prove Proposition 2, we need Lemma 1.

Lemma 1: Let $0 \leq g_1 < g_2 \leq g_3 < g_4$ denote the boundary points of two regions $R_1 = [g_1, g_2)$ and $R_2 = [g_3, g_4)$. Assume that the probabilities that g falls into regions R_1 and R_2 are non-zero and identical, i.e.,

$$\int_{g_1}^{g_2} f_{CH}(x) dx = \int_{g_3}^{g_4} f_{CH}(x) dx > 0, \tag{31}$$

where $f_{CH}(g)$ is the pdf of channel gain g. Denote $\overline{P}(i,j)$ the average power of transmitting i packets when $g \in R_1$ and j packets when $g \in R_2$, i.e.,

$$\overline{P}(i,j) = \int_{g_1}^{g_2} \frac{A(i)}{x} f_{CH}(x)dx + \int_{g_3}^{g_4} \frac{A(j)}{x} f_{CH}(x)dx. \quad (32)$$

Then $\overline{P}(i,j) < \overline{P}(j,i)$ if $i < j$.

For a proof of Lemma 1, see Appendix B.

Proposition 2: Suppose the pdf of channel gain $f_{CH}(g)$ is a bounded and continuous function over $[0,\infty)$. Then the set of regions $\{R_i^{(q)}\}$ that minimizes the average power in (28), is $\{R_i^{(q)} = [g_i^{(q)}, g_i^{(q)})\}$, where the boundary points $\{g_i^{(q)}, i = 0, 1, \cdots, \min(q + \mu, M)\}$, can be obtained by solving

$$\int_{g_i^{(q)}}^{\infty} f_{CH}(g)dg = \sum_{j=i}^{\min(q+\mu,M)} p_{q,\min(q+\mu,M)-j},$$
$$\text{for} \quad i = 0, 1, \cdots, \min(q + \mu, M), \quad (33)$$

and the boundary point $g_{\min(q+\mu,M)+1} = \infty$.

For a proof of Proposition 2, see Appendix C.

To summarize, the solution to (21) can be obtained by solving (33) for $\{g_i^{(q)}\}$, and plugging $\{g_i^{(q)}\}$ into (27) as below

$$P_q^*(g, \mathbf{P}) = \sum_{i=0}^{\min(q+\mu,M)} \frac{A(i)}{g} \mathbf{1}(g \in [g_i^{(q)}, g_{i+1}^{(q)})),$$
$$\text{for} \quad q = 0, \cdots, M. \quad (34)$$

From (28), the average power is given as below

$$\mathbf{E}_g[P_q^*(g, \mathbf{P})] = \sum_{i=0}^{\min(q+\mu,M)} \int_{g_i^{(q)}}^{g_{i+1}^{(q)}} \frac{A(i)}{g} f_{CH}(g)dg. \quad (35)$$

B. Solution to the Second Sub-problem

1) Convexity of the objective function in (23a)*:* From (23d) and (23e), we have

$$p_{i,0} = 1 - \sum_{j=1}^{L(i)} p_{i,j}, \quad (37)$$

where $L(i) = \min(i + \mu, M)$ is the column index of the rightmost nonzero entry of the i-th row of $\mathbf{P}$. There are totally $\sum_{i=0}^{M} L(i)$ independent entries in matrix $\mathbf{P}$. Denote $U(j) = \max(j - \mu, 0)$ the row index of the uppermost nonzero entry of the j-th column of $\mathbf{P}$. It is easy to find that $U(L(i)) = i$ and $L(U(j)) = j$. We can rewrite (23) as (the dependent parameters $p_{i,0}$ are excluded)

$$\min_{\substack{p_{i,j} \\ i=0,\cdots,M \\ j=1,\cdots,L(i)}} \sum_{i=0}^{M} \pi_i \sum_{j=0}^{L(i)} \int_{g \in R_{L(i)-j}^{(i)}} \frac{A(L(i) - j)}{g} f_{CH}(g)dg \quad (38a)$$

$$s.t. \quad \sum_{i=U(j)}^{M} \pi_i p_{i,j} = \pi_j, \quad j = 1, 2, \cdots, M \quad (38b)$$

$$p_{i,j} \geq 0, \quad i = 0, \cdots, M, \quad j = 1, \cdots, L(i) \quad (38c)$$

$$\sum_{j=1}^{L(i)} p_{i,j} \leq 1, \quad \forall i. \quad (38d)$$

From (38b), (38c), and (38d), we know the feasible region for $p_{i,j}$ ($i = 0, \cdots, M, j = 1, \cdots, L(i)$) is convex. Next we prove that the objective function is convex in $p_{i,j}$ ($i = 0, \cdots, M$, $j = 1, \cdots, L(i)$), if $A(k) - A(k-1) > 0$ for $k \geq 1$. To prove this, we need Lemmas 2 and 3.

Lemma 2: Let $\mathbf{C}_N$ be a symmetric $N \times N$ matrix with elements $c_{i,j}$; and let $c_{i,j} = a_{\min(i,j)}$, where a_k ($k \in \{0, 1, \cdots, N\}$) are arbitrary constants and satisfy $0 < a_1 < a_2 < \cdots < a_N$. Then, $\mathbf{C}_N$ is positive definite.

For a proof of Lemma 2, see Appendix D.

Denote the i-th row (except the first element p_{i0}) of $\mathbf{P}$ by vector $\mathbf{p}_i = [p_{i,1}, p_{i,2}, \cdots, p_{i,L(i)}]^T$, and denote the average power usage when the queue length equals to i by $k_i(\mathbf{p}_i)$, which is equal to $\sum_{j=0}^{L(i)} \int_{g \in R_j^{(i)}} \frac{A(j)}{g} f_{CH}(g)dg$.

Lemma 3: If $A(k) - A(k-1) > 0$ (for $k \geq 1$), then $k_i(\mathbf{p}_i)$ is convex in $\mathbf{p}_i$.

For a proof of Lemma 3, see Appendix E.

Let $\mathbf{p} = [\mathbf{p}_0^T, \mathbf{p}_1^T, \cdots, \mathbf{p}_M^T]^T$, and

$$P_{avg}(\mathbf{p}) = \sum_{i=0}^{M} \pi_i k_i(\mathbf{p}_i). \quad (39)$$

Proposition 3: If $A(k) - A(k-1) > 0$ (for $k \geq 1$), then $P_{avg}(\mathbf{p})$ is convex in $\mathbf{p}$.

For a proof of Proposition 3, see Appendix F.

2) Solution for the constraint-relaxed optimization problem: Proposition 3 shows that the objective function (38a) is convex. The feasible region of $\mathbf{p}$ is also convex. So we can use the Lagrangian method to find the optimal solution to (38). Since the inequality constraints (38c) and (38d) make it difficult to directly solve (38), we first solve a simpler problem (40) with the equality constraint only; then use its solution to construct the solution to (38).

$$\min_{\mathbf{P}} \quad P_{avg}(\mathbf{p}) = \sum_{i=0}^{M} \pi_i k_i(\mathbf{p}_i) \quad (40a)$$

$$s.t. \quad \sum_{i=U(j)}^{M} \pi_i p_{i,j} = \pi_j, \quad j = 1, 2, \cdots, M. \quad (40b)$$

The Lagrangian function of (40) is given by

$$J(\mathbf{p}, \boldsymbol{\lambda}) = \sum_{i=0}^{M} \pi_i k_i(\mathbf{p}_i) + \sum_{j=1}^{M} \lambda_j \left(\sum_{i=U(j)}^{M} \pi_i p_{i,j} - \pi_j \right), \quad (41)$$

$$\mathcal{F}(j_0) = \max j \tag{36a}$$

$$s.t. \quad F_{CH}(B(L(i) - j_0)\hat{y}_{j_0+1}) - F_{CH}(B(L(i) - j + 1)\hat{y}_j) \geq 0, \tag{36b}$$

$$j \in \{1, 2, \cdots, j_0\}. \tag{36c}$$

---

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \cdots, \lambda_M]^T$. Differentiate $J(\mathbf{p}, \boldsymbol{\lambda})$ w.r.t. $p_{i,j}$ and set the derivative to zero, i.e.,

$$\frac{\partial}{\partial p_{i,j}} J(\mathbf{p}, \boldsymbol{\lambda}) = \pi_i \sum_{k=L(i)-j+1}^{L(i)} \frac{-B(k)}{g_k^{(i)}} + \lambda_j \pi_i = 0, \tag{42}$$

where $B(k) = A(k) - A(k-1)$. We obtain

$$\sum_{k=L(i)-j+1}^{L(i)} \frac{B(k)}{g_k^{(i)}} = \lambda_j. \tag{43}$$

Notice that when $\pi_i = 0$, $\sum_{k=L(i)-j+1}^{L(i)} \frac{B(k)}{g_k^{(i)}}$ is not necessary to be equal to λ_j. However, letting $\sum_{k=L(i)-j+1}^{L(i)} \frac{B(k)}{g_k^{(i)}} = \lambda_j$. is one of the possible solutions, as long as it leads to a valid solution of $\{p_{i,j}\}$. The proof of existence of $\{p_{i,j}\}$ is provided in Lemma 4.

For each i, solving equation groups $j = 1, 2, \cdots, L(i)$,

$$g_j^{(i)} = \frac{B(j)}{\lambda_{L(i)-j+1} - \lambda_{L(i)-j}}, \tag{44}$$

where λ_0 is assumed to be zero. From (33),

$$p_{i,j} = F_{CH}(g_{L(i)-j}^{(i)}) - F_{CH}(g_{L(i)-j+1}^{(i)}), \tag{45}$$

where $F_{CH}(x) = \int_x^\infty f_{CH}(g) \, dg$ is the complementary cumulative density function (ccdf) of channel gain. Substituting (45) into (40b), we have

$$\sum_{i=U(j)}^{M} \pi_i (F_{CH}(g_{L(i)-j}^{(i)}) - F_{CH}(g_{L(i)-j+1}^{(i)})) = \pi_j. \tag{46}$$

Summing up both sides of (46), we obtain

$$\sum_{j=k}^{M} \sum_{i=U(j)}^{M} \pi_i (F_{CH}(g_{L(i)-j}^{(i)}) - F_{CH}(g_{L(i)-j+1}^{(i)})) = \sum_{j=k}^{M} \pi_j,$$
$$k = \{1, 2, \cdots, M\}. \tag{47}$$

By changing the order of summation, the left hand side (l.h.s.) of (47) is equal to

$$\sum_{i=U(k)}^{M} \pi_i \sum_{j=k}^{L(i)} (F_{CH}(g_{L(i)-j}^{(i)}) - F_{CH}(g_{L(i)-j+1}^{(i)})) \tag{48}$$

$$= \sum_{i=U(k)}^{M} \pi_i (F_{CH}(g_0^{(i)}) - F_{CH}(g_{L(i)-k+1}^{(i)})) \tag{49}$$

$$= \sum_{i=U(k)}^{M} \pi_i (1 - F_{CH}(g_{L(i)-k+1}^{(i)})). \tag{50}$$

Substituting (44) into (50), letting (50) equal to the right hand side (r.h.s.) of (47), and moving terms to the other side, we have

$$\sum_{i=U(k)}^{M} \pi_i F_{CH}(B(L(i) - k + 1)y_k) = \sum_{i=U(k)}^{k-1} \pi_i, \tag{51}$$

where

$$y_k = (\lambda_k - \lambda_{k-1})^{-1}. \tag{52}$$

Since y_k is the only unknown variable in (51), we can solve (51) for y_k ($\forall k$). Denote $\hat{y}_k$ the solution to (51).

Lemma 4: The solution $\hat{y}_k$ to (51) exists for each $k \in \{1, \cdots, M\}$, and $\hat{y}_k \geq 0$.

For a proof of Lemma 4, see Appendix G.

Let $\hat{p}_{i,j}$ be the optimal solution to (40). Then,

$$\hat{p}_{i,j} = F_{CH}(B(L(i) - j)\hat{y}_{j+1}) - F_{CH}(B(L(i) - j + 1)\hat{y}_j). \tag{53}$$

Particularly, when $j = M$, $B(L(i) - j) = B(M - M) = 0$. We do not need to calculate $\hat{y}_{M+1}$, which does not have definition according to (51).

We have following observations about $\hat{p}_{i,j}$.

Lemma 5:
(1) $\sum_{j=1}^{L(i)} \hat{p}_{i,j} \leq 1$.
(2) $\hat{p}_{i,L(i)} \geq 0$.
(3) If the ratio $B(k+1)/B(k)$ is the same for different value of k, i.e., $B(k+1)/B(k) = c$ ($k \geq 1$), where c is a constant, then $\hat{p}_{i,j}$ has the same sign for each column j except for the upper most element $p_{U(j),j}$, $j \in \{\mu, \cdots, M\}$.
(4) $\hat{p}_{i,j} \geq 0$ for $\forall i$, $1 \leq j \leq \mu - 1$.
For a proof of Lemma 5, see Appendix H.

3) Only one negative column in the matrix of $\hat{p}_{i,j}$: In this section, we construct the optimal solution $p_{i,j}^*$ for (38) from $\hat{p}_{i,j}$. From Lemma 5, (38d) is always satisfied. Then, we only need to consider (38c), i.e., $p_{i,j} \geq 0$ for $i = 0, \cdots, M$ and $j = 1, \cdots, L(i)$. We first construct the optimal solution for a special case where only one column has $\hat{p}_{i,j} < 0$ (except for the uppermost element $\hat{p}_{U(j),j}$); then generalize the result to other cases in Section III-B.4.

Denote j_0 the index of the column with $\hat{p}_{i,j_0} < 0$ (except for the uppermost element $\hat{p}_{U(j_0),j_0}$). Define a function of j_0, denoted by $\mathcal{F}(j_0)$, as the optimal solution to (36); if (36) does not have a solution, let $\mathcal{F}(j_0) = 0$. The following proposition shows how to construct the optimal solution to (38).

Proposition 4: If the following two conditions are satisfied
1) Only one column has $\hat{p}_{i,j_0} < 0$ (except for the uppermost element $\hat{p}_{U(j_0),j_0}$), where j_0 is the index of the column,
2) $B(k+1)/B(k) = c$ for all $k \geq 1$,

$$p_{i,j}^*(j_0) = \begin{cases} \hat{p}_{i,j} & j \in \{1, \cdots, \mathcal{F}(j_0) - 1\} \bigcup \{j_0 + 1, \cdots, M\}, \quad i \in \{U(j), \cdots, M\} \\ \mathcal{G}(j_0, i, j) & j \in \{\mathcal{F}(j_0), \cdots, j_0\}, \quad i \in \{U(j), \cdots, M\} \end{cases}, \tag{54}$$

where

$$\mathcal{G}(j_0, i, j) = \begin{cases} \sum_{k=\mathcal{F}(j_0)}^{j_0} \hat{p}_{i,k} & j = \mathcal{F}(j_0), i \in \{U(j_0) + 1, \cdots, M\} \\ 0 & j \in \{\mathcal{F}(j_0) + 1, \cdots, j_0\}, i \in \{U(j_0) + 1, \cdots, M\} \\ F_{CH}(B(L(i) - j)y_{j+1,j_0}^*) - F_{CH}(B(L(i) - j + 1)y_{j,j_0}^*) & j \in \{\mathcal{F}(j_0), \cdots, j_0\}, i \in \{U(j), \cdots, U(j_0)\} \end{cases}, \tag{55}$$

where $y_{j,j_0}^* = \hat{y}_j$ if $j \in \{\mathcal{F}(j_0), j_0 + 1\}$; if $j \in \{\mathcal{F}(j_0) + 1, \cdots, j_0\}$, y_{j,j_0}^* is obtained by solving the following equation:

$$\sum_{i=U(j)}^{U(j_0)} \pi_i (1 - F_{CH}(B(L(i) - j + 1)y_{j,j_0}^*)) = \sum_{i=j}^{j_0} \pi_i. \tag{56}$$

then the solution to (38), denoted by $p_{i,j}^*(j_0)$, is given by (54). For a proof of Proposition 4, see Appendix M.

4) Arbitrary number of negative columns in the matrix of $\hat{p}_{i,j}$: Proposition 4 gives the solution for the case where only one column of $\hat{p}_{i,j}$ has negative value. The solution $p_{i,j}^*(j_0)$ differs from $\hat{p}_{i,j}$ only in columns $j \in \{\mathcal{F}(j_0), \cdots, j_0\}$; for other columns, $p_{i,j}^*(j_0)$ is the same as $\hat{p}_{i,j}$. Using this procedure iteratively, Proposition 5 states that we can construct the solution for the general case, i.e., arbitrary number of negative columns in the matrix of $\hat{p}_{i,j}$.

Proposition 5: If $B(k+1)/B(k) = c$ for all $k \geq 1$, the solution to (38) is given by Algorithm 1.

For a proof of Proposition 5, see Appendix N.

Algorithm 1: Solving the second sub-problem
  1) **Input**: $f_{CH}(g)$, $\boldsymbol{\pi}, M, \mu$
  Solve (51) for $\hat{y}_k$, $k \in \{1, \cdots, M\}$;
  Calculate $\hat{p}_{i,j}$ via (53), $i \in \{0, \cdots, M\}$, $j \in \{1, \cdots, L(i)\}$; where
  Calculate $\mathcal{F}(k)$ via (36), $k \in \{\mu, \cdots, M - 1\}$;
  Calculate $\mathcal{G}(k, i, j)$ via (55)), $k \in \{1, \cdots, M - 1\}$,
      $i \in \{U(j), \cdots, M\}$, $j \in \{\mathcal{F}(k), \cdots, k\}$;
  $k = M - 1$;
  2) If $\hat{p}_{U(k)+1,k} \geq 0$
    { For $i = U(k)$ to M
      $p_{i,k}^* = \hat{p}_{i,k}$;
    Endfor
    $k = k - 1$;
    Break}
  Else
    { For $j = \mathcal{F}(k)$ to k
      For $i = U(j)$ to M
        $p_{i,j}^* = \mathcal{G}(k, i, j)$;
      Endfor
    Endfor
    $k = \mathcal{F}(k) - 1$; }
  Endif
  If $k > 0$, go to 2);
  $p_{i,0}^* = 1 - \sum_{j=1}^{L(i)} p_{i,j}$
  3) **Output**: $\mathbf{P}^*(\boldsymbol{\pi})$.

C. Solution to the Original Problem (20)

We first sketch our idea and then present Algorithm 2 to solve the original problem (20).

Our idea is as below. Given a queue length distribution $\boldsymbol{\pi}$, we can solve (23) for $\mathbf{P}^*(\boldsymbol{\pi})$ by Algorithm 1, and then solve (21) for $P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}))$. Now, it is obvious that the first constraint (24b) in the optimization problem (24) is a function of $\boldsymbol{\pi}$, and can be denoted as $g_1(\boldsymbol{\pi}) \leq 0$, where

$$g_1(\boldsymbol{\pi}) = \sum_{i=0}^{M} \pi_i \mathbf{E}_g[P_i^*(g, \mathbf{P}^*(\boldsymbol{\pi}))] - P_0. \tag{57}$$

The optimization problem (24) can be reformulated as the following nonlinear program

$$\min_{\boldsymbol{\pi}} \quad f(\boldsymbol{\pi}) \tag{58a}$$
$$s.t. \quad \mathbf{g}(\boldsymbol{\pi}) \leq 0, \tag{58b}$$
$$h(\boldsymbol{\pi}) = 0, \tag{58c}$$

where

$$f(\boldsymbol{\pi}) = \frac{1}{\mu} \sum_{i=M-\mu}^{M} \pi_i(\mu - i + M), \tag{59}$$

$$\mathbf{g}(\boldsymbol{\pi}) = \begin{bmatrix} g_1(\boldsymbol{\pi}) \\ -\boldsymbol{\pi} \end{bmatrix}, \tag{60}$$

$$h(\boldsymbol{\pi}) = \sum_{i=0}^{M} \pi_i - 1. \tag{61}$$

(58) can be solved by Merit-function Sequential Quadratic Programming (MSQP) [20, page 583], which guarantees global convergence to a local optimal solution under some weak conditions [21].

Using the above idea, we design Algorithm 2 to seek the optimal power control law $P_0(g, q)$, which is the solution to (20). We call $P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^*))$ produced by Algorithm 2, JQLA power control law.

Algorithm 2:
  1) **Input**: $f_{CH}(g)$, μ, M, P_0, ϵ
  2) Initialization:
    Randomly select a feasible distribution $\boldsymbol{\pi}^{(0)}$, which
    satisfies (13);
    $m = 0$;
  3) Given $\boldsymbol{\pi}^{(m)}$, solve (23) for $\mathbf{P}^*(\boldsymbol{\pi}^{(m)})$ by Algorithm 1;
  4) Given $\mathbf{P}^*(\boldsymbol{\pi}^{(m)})$, solve (21) for $P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^{(m)}))$
    via (33) and (34);

5) Given $\boldsymbol{\pi}^{(m)}$ and $P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^{(m)}))$, apply one iteration of MSQP to (58) and obtain an increment $d\boldsymbol{\pi}$;

6) $\boldsymbol{\pi}^{(m+1)} = \boldsymbol{\pi}^{(m)} + d\boldsymbol{\pi}$;

7) If $|P_{drop}(\boldsymbol{\pi}^{(m+1)}) - P_{drop}(\boldsymbol{\pi}^{(m)})| < \epsilon$
   Go to 8);
   Else
   $m = m + 1$;
   Go to 3);
   Endif

8) $\boldsymbol{\pi}^* = \boldsymbol{\pi}^{(m)}$;

9) **Output**: $P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^*))$.

Algorithm 2 is actually an MSQP, which produces a local optimal solution under some weak conditions [21]. Different from a conventional MSQP, we need Step 3 and Step 4 to compute $\mathbf{P}^*(\boldsymbol{\pi}^{(m)})$ and $P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^{(m)}))$. Since functions $\mathbf{P}^*(\boldsymbol{\pi}^{(m)})$ and $P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^{(m)}))$ do not have an explicit closed form, it is particularly difficult to verify whether the sufficient conditions for global convergence of MSQP [21] are satisfied in Algorithm 2. Although this paper does not provide a convergence proof for Algorithm 2, as shown in our experiments, Algorithm 2 always converges as long as 1) the initial point is an interior point with respect to the constraints, and 2) a merit function is chosen appropriately [21].

D. Structure of the Optimal Power Control Law

The optimal power control law $\{P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^*))\}$ depends on the optimal queue length distribution $\boldsymbol{\pi}^*$, which is obtained by numerically solving the third sub-problem. However, even without knowing the closed form of $\boldsymbol{\pi}^*$, we can still obtain structural properties about the optimal power control law $\{P_q^*(g, \mathbf{P}^*(\boldsymbol{\pi}^*))\}$ from the intermediate results in solving the second sub-problem.

Since the only unknowns in (27) are channel gain regions $R_i^{(q)}$ or their boundaries $\{g_i^{(q)}\}$, hence the optimal power control law is determined by the channel gain boundaries $\{g_i^{(q)}\}$. Now we examine the structure of $\{g_i^{(q)}\}$.

We first consider the case where all $\hat{p}_{i,j} > 0$. In this case, $p_{i,j}^* = \hat{p}_{i,j}$ and the boundaries of channel gain regions $R_i^{(q)}$ are obtained by $g_i^{(q)} = B(i)\hat{y}_{L(q)-i+1}$. Hence, we have the following two properties:

- Property 1: If queue length $q < M - \mu$, we have $g_{i+1}^{(q+1)} = cg_i^{(q)}$, i.e., the values $g_i^{(q)}$, $g_{i+1}^{(q+1)}$, $g_{i+2}^{(q+2)}$, ... form a geometric sequence or an arithmetic sequence in logarithmic scale.
- Property 2: If queue length $q \geq M - \mu$, the boundaries $g_i^{(q)}$ has the same value for all values of q.

Fig. 4 shows an example of channel gain regions $R_i^{(q)}$ and their boundaries $\{g_i^{(q)}\}$ for $M = 9$ and $\mu = 3$. In Fig. 4, each horizontal axis represents channel gain g in logarithmic scale for a fixed value of queue length q. For each horizontal axis (corresponding to a queue length indexed by q), we plot the boundary points $\{g_i^{(q)}, \forall i\}$; a dark dot represents a value $g_i^{(q)}$. Due to Property 1, the points $g_i^{(q)}$, $g_{i+1}^{(q+1)}$, $g_{i+2}^{(q+2)}$, ... form a straight line (a tilted line); these points are on the axes from $q = 0$ to $q = M - \mu$. Due to Property 2, the points $g_i^{(q)}$, $g_i^{(q+1)}$, $g_i^{(q+2)}$, ... form a straight line (a vertical line); these points are on the axes from $q = M - \mu$ to $q = M$. So each tilted line is connected with a vertical line on Axis $q = M - \mu$. Hence, all boundary points $\{g_i^{(q)}, \forall q, \forall i\}$ form a regular pattern on the 2D plane, similar to plane wave-fronts in radio propagation.
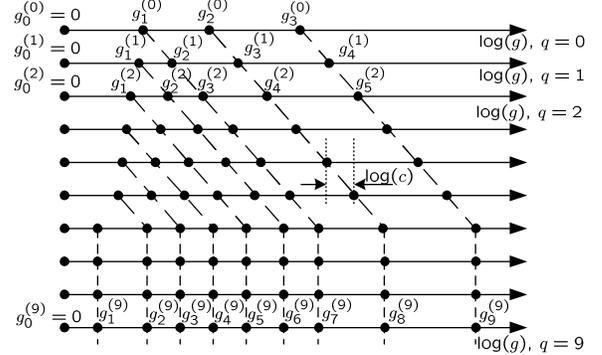
Fig. 4. An example of channel gain regions of the optimal power control law when all $\hat{p}_{i,j} \geq 0$.

For the case where some $\hat{p}_{i,j} < 0$, $p_{i,j}^*$ is obtained from $\hat{p}_{i,j}$ by Algorithm 2. Fig. 5 shows an example of channel gain regions $R_i^{(q)}$ and their boundaries $\{g_i^{(q)}\}$ for the case where not all $\hat{p}_{i,j} > 0$. In this example, $M = 9$, $\mu = 3$, $\hat{y}_7 < \hat{y}_8/c$ and $\mathcal{F}(7) = 5$; hence, some elements of the optimal transition probability matrix will be zero. As a result, Property 1 does not hold, i.e., there may exist some tilted lines not connected with a vertical line.
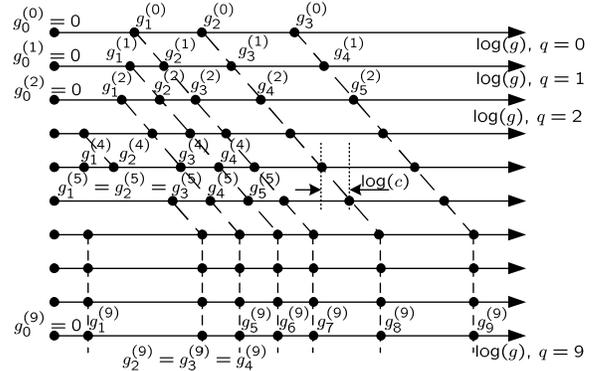
Fig. 5. An example of channel gain regions of the optimal power control law when some $\hat{p}_{i,j} < 0$.

IV. SIMULATION RESULTS

In this section, we simulate the system depicted in Fig. 1 and compare the JQLA power control law with TDWF and TCI power control.

Since we assume ideal channel coding is used, the service rate of the buffer in a block is equal to the instantaneous channel capacity in the block, and the decoding error probability in the physical layer is zero. Table I lists the simulation parameters. We conduct simulations for Case 1, Case 2, and Case 3, which represent stringent, moderate, and loose delay constraint, respectively. In Table I, the average power

constraint P_0 is different for different cases; this is because we want to keep packet drop probability $P_{drop} \leq 10^{-3}$ under JQLA power control for all the cases, and we assume that a connection has a QoS requirement of $P_{drop} \leq 10^{-3}$.

TABLE I
SIMULATION PARAMETERS

	Case 1	Case 2	Case 3
Max queueing delay $(D_{max} - D_c)$	2	5	10
Arrival rate μ	25	10	5
Average power P_0	0.94	0.14	0.046
Channel power gain g	Exponential distributed with E[g]=1		
MT_b/L	50		
Buffer size M	50		

TABLE II
SIMULATION RESULTS FOR JQLA

	Case 1	Case 2	Case 3
Packet drop probability P_{drop}	0.9×10^{-3}	0.78×10^{-3}	10^{-3}

Table II shows the packet drop probability under JQLA power control for the three cases. It can be seen that the QoS requirement of $P_{drop} \leq 10^{-3}$ is satisfied for all the cases. Fig. 6 shows JQLA power control law $P_q(g)$ obtained by Algorithm 2, for the three cases. From the figure, we have the following observations. First, if the delay constraint is stringent (Case 1), JQLA power control law $P_q(g)$ is almost invariant w.r.t. the queue length q, and for most values of g, $P_q(g)$ is a decreasing function of g, i.e., when the channel gain g is small, the transmission power is high to compensate the poor channel condition; this is similar to TCI power control. Second, JQLA power control law $P_q(g)$ under Case 2 (moderate delay constraint) is similar to that under Case 3 (loose delay constraint), in the sense that $P_q(g)$ under both cases exhibits a similar pattern: when g is very small and q is very large, $P_q(g)$ is a decreasing function of g, similar to TCI; when g is small and q is also small, $P_q(g)$ is an increasing function of g, similar to TDWF; when g is large, $P_q(g)$ is a decreasing function of g, similar to TCI. Third, for Case 3 (loose delay constraint), the cutoff threshold $g_1^{(q)}$ for small q is significantly larger than that for moderate and stringent delay constraints; (note that if channel gain $g < g_1^{(q)}$, the transmission power is zero, i.e., there is no transmission;) hence, if the delay constraint is loose, the transmitter tends to wait till the channel gets better to transmit, resulting in lower average transmission power.

Fig. 7 shows the queue length distribution under JQLA for the three cases. As shown in Fig. 7, the queue length distribution $\boldsymbol{\pi}^*$ obtained by Algorithm 2 (which is labeled as 'Analysis') agrees well with the queue length distribution obtained from simulations (which is labeled as 'Simulation'), indicating high accuracy of our analysis and Algorithm 2. It is also observed that the JQLA power control law achieves small probability for queue length $q > M - \mu$, leading to low packet drop probability. Note that $q(n) > M - \mu$ in the n-th block results in packet drops (refer to Fig. 2).

Fig. 8 shows packet drop probability P_{drop} vs. average power P_0 for JQLA, TDWF, and TCI power control. It is observed that JQLA achieves approximately 10 dB gain over TDWF when the packet drop probability is 10^{-3}. We also observe that TCI performs slightly better than TDWF; this is because the delay bound for all the three cases is not very large but TDWF needs very large delay bound to perform well.

V. CONCLUSION

In this paper, we studied the problem of optimal power control for delay-constrained communication over fading channels. Our objective is to find a power control law that optimizes the link layer performance, specifically, minimizes delay bound violation probability (or equivalently, the packet drop probability), subject to constraints on average power, arrival rate, and delay bound. The transmission buffer size is assumed to be finite; hence, when the buffer is full, there will be packet drop. The fading channel under our study has a continuous state, e.g., Rayleigh fading. Since the channel state is continuous, dynamic programming is not applicable for power control; in other words, if dynamic programming were used, the system would suffer from capacity loss due to error caused by quantizing the continuous channel state. Since directly solving the power control problem (which optimizes the link layer performance) is particularly challenging, we decomposed it into three sub-problems, and solved the three sub-problems iteratively, which produced JQLA power control scheme. We proved that the solution that simultaneously solves the three sub-problems is also an optimal solution to the optimal power control problem. Simulation results showed that the JQLA scheme achieves superior performance over the time domain water filling and the truncated channel inversion power control. E.g., JQLA achieves 10 dB gain at packet drop probability of 10^{-3}, over the time domain water filling power control. Algorithm 2 is too complex; hence one may not use it in practice. Instead, the main purpose of JQLA is to explore the fundamental performance limit of power control under the link-PHY model. Although JQLA does not produce a global optimal solution, this work represents a major step toward deriving the fundamental performance limit of power control under the link-PHY model. The local optimal solution obtained by JQLA can provide a judgement on the performance of a practical power control scheme. The 10 dB gain achieved by JQLA indicates that there is much room to improve for existing power control schemes!

APPENDIX

A. Proof of Proposition 1

Proof: Denote $\boldsymbol{\pi}_0 = [\pi_{00}, \pi_{01}, \cdots, \pi_{0M}]$ the steady state queue length distribution generated by $P_0(g, q)$. $P_0(g, q)$ is feasible to (20), therefore the average power constraint

$$\sum_{l=0}^{M} \pi_l \mathbf{E}_g[P_0(g, l)] \leq P_0, \qquad (62)$$
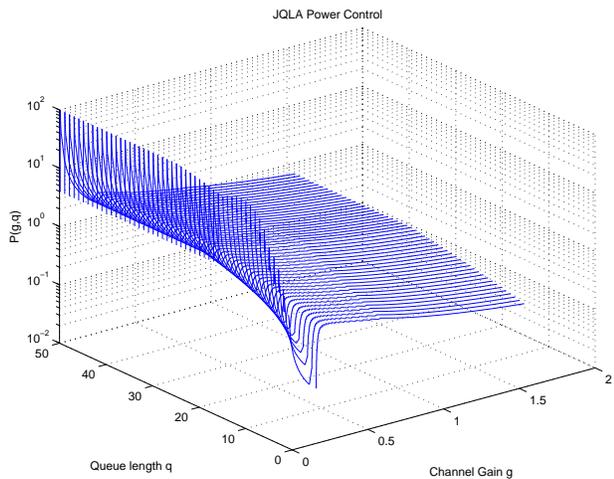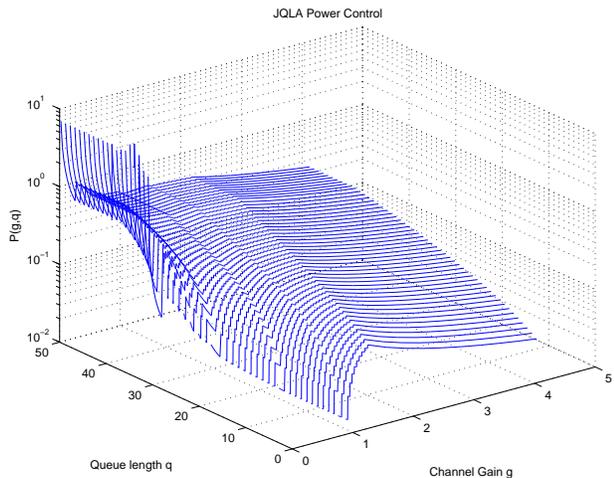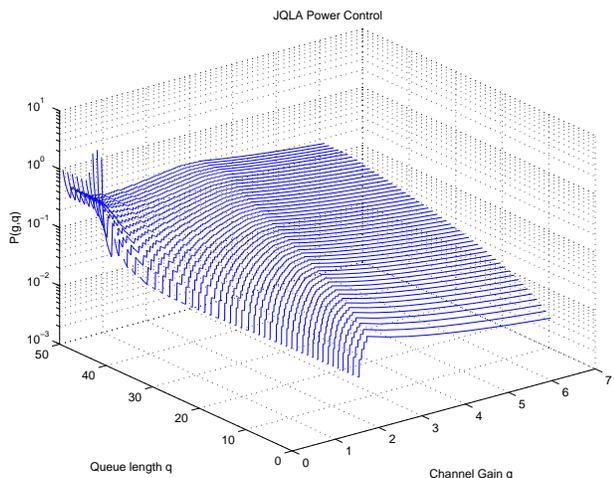
Fig. 6. $P(g, q)$ of JQLA. (a) Case 1, (b) Case 2, (c) Case 3.
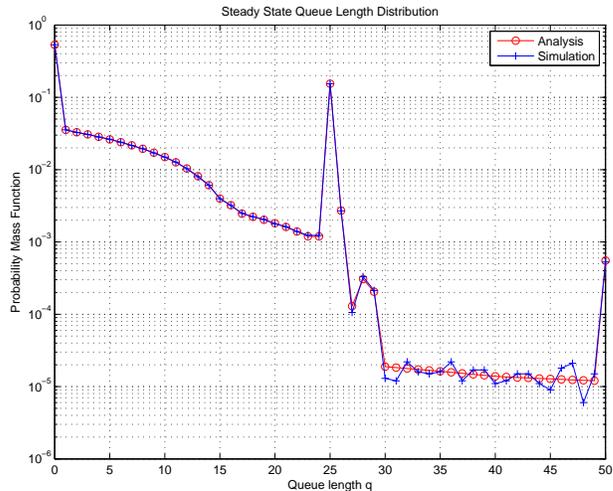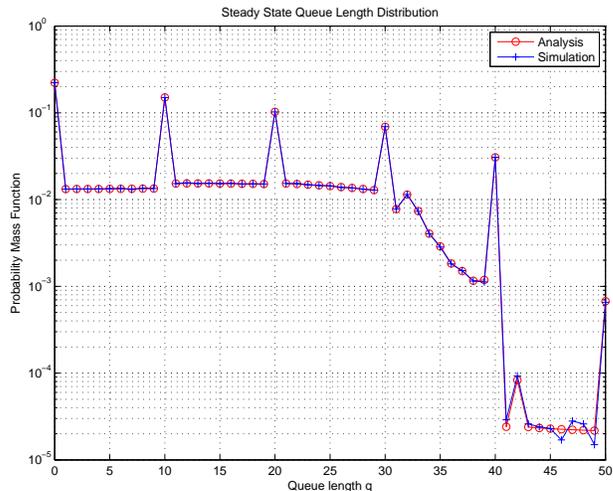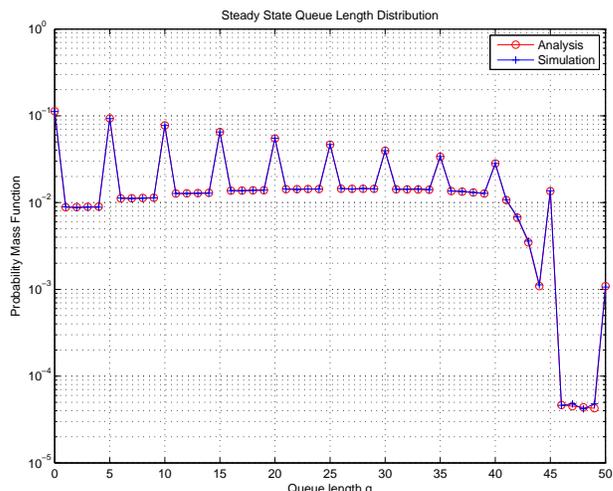
Fig. 7. Queue length distribution of JQLA. (a) Case 1, (b) Case 2, (c) Case 3.

is satisfied. From the first and second sub-problem (21) and (23), we know for the transition probability matrix $\mathbf{P}^*(\boldsymbol{\pi}_0)$,

$$\sum_{l=0}^{M} \pi_l \mathbf{E}_g[P_l^*(g, \mathbf{P}^*(\boldsymbol{\pi}_0))] \leq \sum_{l=0}^{M} \pi_l \mathbf{E}_g[P_0(g, l)] \leq P_0. \quad (63)$$

Therefore $\boldsymbol{\pi}_0$ is feasible to (24), therefore

$$\frac{1}{\mu} \sum_{l=M-\mu}^{M} \pi_l^*(\mu - l + M) \leq \frac{1}{\mu} \sum_{l=M-\mu}^{M} \pi_{0l}(\mu - l + M). \quad (64)$$

Because $P_0(g, q)$ is the optimal solution to (20), we have

$$\frac{1}{\mu} \sum_{l=M-\mu}^{M} \pi_{0l}(\mu - l + M) \leq \frac{1}{\mu} \sum_{l=M-\mu}^{M} \pi_l^*(\mu - l + M). \quad (65)$$

From (64) and (65), we have

$$\frac{1}{\mu} \sum_{l=M-\mu}^{M} \pi_{0l}(\mu - l + M) = \frac{1}{\mu} \sum_{l=M-\mu}^{M} \pi_l^*(\mu - l + M). \quad (66)$$

$\blacksquare$

B. Proof of Lemma 1

Proof: Denote $y_1 = \int_{g_1}^{g_2} \frac{1}{x} f_{CH}(x) dx$ and $y_2 = \int_{g_3}^{g_4} \frac{1}{x} f_{CH}(x) dx$. Then we have $y_1 > y_2$ since

$$
\begin{aligned}
y_1 &= \int_{g_1}^{g_2} \frac{1}{x} f_{CH}(x) dx \\
&\stackrel{(a)}{>} \frac{1}{g_2} \int_{g_1}^{g_2} f_{CH}(x) dx \\
&\stackrel{(b)}{\geq} \frac{1}{g_3} \int_{g_1}^{g_2} f_{CH}(x) dx \\
&\stackrel{(c)}{=} \frac{1}{g_3} \int_{g_3}^{g_4} f_{CH}(x) dx \\
&\stackrel{(d)}{\geq} \int_{g_3}^{g_4} \frac{1}{x} f_{CH}(x) dx \\
&= y_2 \quad (67)
\end{aligned}
$$

where (a) follows from $\frac{1}{x} > \frac{1}{g_2}$ for $x \in (g_1, g_2)$ and $g_1 \geq 0$, (b) follows from $g_2 \leq g_3$, (c) due to (31), and (d) follows from $\frac{1}{g_3} \geq \frac{1}{x}$ for $x \in [g_3, g_4]$.

Substituting y_1, y_2 into $P(i, j)$, we have

$$
\begin{aligned}
\overline{P}(i, j) &= \int_{g_1}^{g_2} \frac{A(i)}{x} f_{CH}(x) dx + \int_{g_3}^{g_4} \frac{A(j)}{x} f_{CH}(x) dx \\
&= A(i)y_1 + A(j)y_2.
\end{aligned}
$$

Similarly, we have

$$\overline{P}(j, i) = A(j)y_1 + A(i)y_2. \quad (68)$$

If $i < j$, then $A(i) < A(j)$; hence, we have

$$\overline{P}(i, j) - \overline{P}(j, i) = (A(i) - A(j))(y_1 - y_2) < 0. \quad (69)$$

$\blacksquare$

Fig. 8. Packet drop probability vs. average power. (a) Case 1, (b) Case 2, (c) Case 3.

C. Proof of Proposition 2

Proof: We first prove that if $\{R_i^{(q)}\}$ is the optimal solution to (30), then each region $R_i^{(q)}$ ($\forall i$) is continuous. We prove this by contradiction.

Assume that $\{R_i^{(q)}\}$ is the optimal solution to (30); also assume that there exists a region $R_j^{(q)}$ that contains (at least) two disconnected sub-regions, say, $[g_1, g_2)$ and $[g_5, g_6)$. Without loss of generality, assume $0 \leq g_1 < g_2 < g_5 < g_6$. Assume that a region $[g_3, g_4)$ is contained in $R_k^{(q)}$ and $g_2 < g_3 < g_4 < g_5$. Consider two cases.

- Case 1 ($k > j$): Consider two sub-cases.
  - Case 1.A: The probability of g in sub-region $[g_3, g_4)$ is greater than or equal to that in $[g_5, g_6)$, i.e.,

    $$\int_{g_3}^{g_4} f_{CH}(g)dg \geq \int_{g_5}^{g_6} f_{CH}(g)dg. \qquad (70)$$

    Then, there exists g_3' (with $g_3' \geq g_3$) such that sub-regions $[g_3', g_4)$ and $[g_5, g_6)$ have the same probability, i.e.,

    $$\int_{g_3'}^{g_4} f_{CH}(g)dg = \int_{g_5}^{g_6} f_{CH}(g)dg. \qquad (71)$$

    From Lemma 1, if we exchange the index of the two sub-regions $[g_3', g_4)$ and $[g_5, g_6)$, i.e., let $[g_3', g_4) \in R_k^{(q)}$, $[g_5, g_6) \in R_j^{(q)}$, then we can reduce the average power. Hence, the original $\{R_i^{(q)}\}$ (without exchanging the index) is not the optimal solution, which contradicts our assumption.
  - Case 1.B: The probability of g in sub-region $[g_3, g_4)$ is less than that in $[g_5, g_6)$, i.e.,

    $$\int_{g_3}^{g_4} f_{CH}(g)dg < \int_{g_5}^{g_6} f_{CH}(g)dg. \qquad (72)$$

    Then, there exists g_5' (with $g_5' > g_5$) such that sub-regions $[g_3, g_4)$ and $[g_5', g_6)$ have the same probability. From Lemma 1, if we exchange the index of the two sub-regions $[g_3, g_4)$ and $[g_5', g_6)$, then we can reduce the average power. Hence, the original $\{R_i^{(q)}\}$ is not the optimal solution, which contradicts our assumption.
- Case 2 ($k < j$): Similar to Case 1, we can show that $\{R_i^{(q)}\}$ is not the optimal solution, which contradicts our assumption.

Therefore, each region $R_i^{(q)}$ ($\forall i$) has to be continuous.

Next, we prove that if $i < j$, then $x < y$, $\forall x \in R_i^{(q)}$ and $\forall y \in R_j^{(q)}$. We prove this by contradiction.

Assume that $\{R_i^{(q)}\}$ is the optimal solution to (30); also assume that there exist $x \in R_i^{(q)}$ and $y \in R_j^{(q)}$ (with $i < j$) such that $x > y$. Without loss of generality, assume $R_i^{(q)} = [g_1, g_2)$ and $R_j^{(q)} = [g_3, g_4)$. Since $x > y$, we must have $g_4 \leq g_1$. Consider two cases.

- Case A: The probability of g in $R_i^{(q)}$ is greater than or equal to that in $R_j^{(q)}$, i.e.,

  $$\int_{g_1}^{g_2} f_{CH}(g)dg \geq \int_{g_3}^{g_4} f_{CH}(g)dg. \qquad (73)$$

  Then, there exists g_1' (with $g_1' \geq g_1$) such that sub-regions $[g_1', g_2)$ and $[g_3, g_4)$ have the same probability, i.e.,

  $$\int_{g_1'}^{g_2} f_{CH}(g)dg = \int_{g_3}^{g_4} f_{CH}(g)dg. \qquad (74)$$

  From Lemma 1, if we exchange the index of the two regions $[g_1', g_2)$ and $[g_3, g_4)$, then we can reduce the average power. Hence, the original $\{R_i^{(q)}\}$ (without exchanging the index) is not the optimal solution, which contradicts our assumption.
- Case B: The probability of g in $R_i^{(q)}$ is less than that in $R_j^{(q)}$.

  Similar to Case A, we can show that $\{R_i^{(q)}\}$ is not the optimal solution, which contradicts our assumption.

Therefore, if $i < j$, we must have $x < y$, $\forall x \in R_i^{(q)}$ and $\forall y \in R_j^{(q)}$.

Now we know that $\{R_i^{(q)}\}$ is a partitioning of $[0, \infty)$ (by definition); each region $R_i^{(q)}$ ($\forall i$) is continuous; if $i < j$, then $x < y$, $\forall x \in R_i^{(q)}$ and $\forall y \in R_j^{(q)}$. Hence, the optimal solution must have the following form (except inclusion or exclusion of the boundary points $\{g_i^{(q)}\}$): $R_i^{(q)} = [g_i^{(q)}, g_{i+1}^{(q)})$ and $0 = g_0^{(q)} \leq g_1^{(q)} \cdots \leq g_{\min(q+\mu, M)}^q \leq g_{\min(q+\mu, M)+1}^q = \infty$.

For $i = 0, 1, \cdots, \min(q + \mu, M)$, we have

$$
\int_{g_i^{(q)}}^{\infty} f_{CH}(x)dx = \sum_{j=i}^{\min(q+\mu, M)} \int_{g_j^{(q)}}^{g_{j+1}^{(q)}} f_{CH}(x)dx
$$
$$
\overset{(a)}{=} \sum_{j=i}^{\min(q+\mu, M)} p_{q, \min(q+\mu, M)-j} \qquad (75)
$$

where (a) follows from (29). ∎

D. Proof of Lemma 2

Proof: From the definition, the matrix $\mathbf{C}_N$ is given by

$$
\mathbf{C_N} = \begin{bmatrix}
a_1 & a_1 & a_1 & \cdots & a_1 \\
a_1 & a_2 & a_2 & \cdots & a_2 \\
a_1 & a_2 & a_3 & \cdots & a_3 \\
\vdots & \vdots & \vdots & \ddots & \cdots \\
a_1 & a_2 & a_3 & \vdots & a_N
\end{bmatrix}. \qquad (76)
$$

Denote $\mathcal{C}_N = \{\mathbf{C}_N\}$ the set of all possible $\mathbf{C}_N$. Consider an arbitrary matrix $C_{N+1} \in \mathcal{C}_{N+1}$. Subtracting the first column of C_{N+1} from the second column, the third column, up to the $(N+1)$-th column, we obtain

$$
\mathbf{C'_{N+1}} = \begin{bmatrix}
a_1 & 0 & 0 & \cdots & 0 \\
a_1 & a_2 - a_1 & a_2 - a_1 & \cdots & a_2 - a_1 \\
a_1 & a_2 - a_1 & a_3 - a_1 & \cdots & a_3 - a_1 \\
\vdots & \vdots & \vdots & \ddots & \cdots \\
a_1 & a_2 - a_1 & a_3 - a_1 & \vdots & a_{N+1} - a_1
\end{bmatrix}.
$$
$$ \qquad (77) $$

Then subtracting the first row of C'_{N+1} from the second row, the third row, up to the $(N+1)$-th row, we have

$$\mathbf{C}''_{N+1} = \begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 - a_1 & a_2 - a_1 & \cdots & a_2 - a_1 \\ 0 & a_2 - a_1 & a_3 - a_1 & \cdots & a_3 - a_1 \\ \vdots & \vdots & \vdots & \ddots & \cdots \\ 0 & a_2 - a_1 & a_3 - a_1 & \vdots & a_{N+1} - a_1 \end{bmatrix}. \tag{78}$$

Then, we have

$$\det(\mathbf{C}_{N+1}) = \det(\mathbf{C}'_{N+1}) = \det(\mathbf{C}''_{N+1}) = a_1 \det(\widetilde{\mathbf{C}}_N), \tag{79}$$

where $\widetilde{\mathbf{C}}_N$ is the sub-matrix by deleting the first row and first column of matrix $\mathbf{C}''_{N+1}$. It is easy to see that $\widetilde{\mathbf{C}}_N \in \mathcal{C}_N$, therefore $\det(\mathbf{C}_{N+1}) = a_1 \det(\widetilde{\mathbf{C}}_N) > 0$. Since $det(\mathbf{C}_1) > 0$ ($\forall \mathbf{C}_1 \in \mathcal{C}_1$), hence $det(\mathbf{C}_N) > 0$ holds for all $N \in \mathbf{Z}^+$.

Denote $(\mathbf{C}_N)_{i,j}$ the sub-matrix obtained by deleting the i-th row and j-th column from matrix $\mathbf{C}_N$. Since $(\mathbf{C}_N)_{i,i} \in \mathcal{C}_{N-1}$ ($\forall i \in \{1, 2, \cdots, N\}$), hence $det((\mathbf{C}_N)_{i,i}) > 0$, which holds for all $N > 1$. Therefore, $\mathbf{C}_N$ is positive definite. ∎

E. Proof of Lemma 3

Proof: From Proposition 2, $R_j^{(i)} = [g_j^{(i)}, g_{j+1}^{(i)})$. Then, we have

$$k_i(\mathbf{p}_i) = \sum_{j=0}^{L(i)} \int_{g_j^{(i)}}^{g_{j+1}^{(i)}} \frac{A(j)}{g} f_{CH}(g) dg \tag{80}$$
$$\overset{(a)}{=} \sum_{j=1}^{L(i)} A(j)[\eta(g_j^{(i)}) - \eta(g_{j+1}^{(i)})],$$

where $\eta(x) = \int_x^\infty \frac{f_{CH}(g)}{g} dg$; (a) is due to $A(0) = 0$. Since $g_{L(i)+1} = \infty$ and $\eta(\infty) = 0$, we have

$$k_i(\mathbf{p}_i) = \sum_{j=1}^{L(i)} \eta(g_j^{(i)})[A(j) - A(j-1)]. \tag{81}$$

Take partial derivative w.r.t. $p_{i,m}$, $m \in \{1, \cdots, L(i)\}$, i.e.,

$$\frac{\partial k_i(\mathbf{p}_i)}{\partial p_{i,m}} = \sum_{j=1}^{L(i)} \frac{d\eta(g_j^{(i)})}{dg_j^{(i)}} \frac{\partial g_j^{(i)}}{\partial p_{i,m}} [A(j) - A(j-1)]$$
$$= -\sum_{j=1}^{L(i)} \frac{f_{CH}(g_j^{(i)})}{g_j^{(i)}} \frac{\partial g_j^{(i)}}{\partial p_{i,m}} [A(j) - A(j-1)]. \tag{82}$$

From (33), (37)

$$g_j^{(i)} = F_{CH}^{-1}(\sum_{k=j}^{L(i)} p_{i,L(i)-k})$$
$$= F_{CH}^{-1}(\sum_{k=0}^{L(i)-j} p_{i,k})$$
$$= F_{CH}^{-1}(p_{i,0} + \sum_{k=1}^{L(i)-j} p_{i,k}) \tag{83}$$
$$= F_{CH}^{-1}(1 - \sum_{k=L(i)-j+1}^{L(i)} p_{i,k}),$$

where $F_{CH}(x) = \int_x^\infty f_{CH}(g) dg$ is the complementary cumulative density function (ccdf) of channel gain. We have

$$\frac{\partial g_j^i}{\partial p_{i,m}} = \frac{1}{F'_{CH}(F_{CH}^{-1}(1 - \sum_{k=L(i)-j+1}^{L(i)} p_{i,k}))} \times$$
$$\frac{\partial}{\partial p_{i,m}}(1 - \sum_{k=L(i)-j+1}^{L(i)} p_{i,k})$$
$$= \frac{1}{f_{CH}(g_j^{(i)})} \mathbf{1}(m \geq L(i) - j + 1). \tag{84}$$

Substituting (84) into (82), we have

$$\frac{\partial k_i(\mathbf{p}_i)}{\partial p_{i,m}} = -\sum_{j=1}^{L(i)} \frac{\mathbf{1}(m \geq L(i) - j + 1)}{g_j^{(i)}} [A(j) - A(j-1)]$$
$$= -\sum_{j=L(i)-m+1}^{L(i)} \frac{B(j)}{g_j^{(i)}}, \tag{85}$$

where $B(j) = A(j) - A(j-1)$, $j \geq 1$. We add $B(0) = 0$ to sequence $\{B(k)\}$, which will simplify the subsequent derivations. Denote $h_{m,n}^{(i)}$ the m-th row and n-th column, $m, n \in \{0, 1, \cdots, L(i)\}$, of the Hessian matrix of $k_i(\mathbf{p}_i)$, i.e.,

$$h_{m,n}^{(i)} = \frac{\partial k_i(\mathbf{p}_i)}{\partial p_{i,m} \partial p_{i,n}}$$
$$= \sum_{j=L(i)-m+1}^{L(i)} \frac{B(j)}{(g_j^{(i)})^2} \cdot \frac{\mathbf{1}(n \geq L(i) - j + 1)}{f_{CH}(g_j^{(i)})} \tag{86}$$
$$= \sum_{j=L(i)-\min(m,n)+1}^{L(i)} \frac{B(j)}{(g_j^{(i)})^2 f_{CH}(g_j^{(i)})}.$$

It is easy to see that $h_{m_1,n_1}^{(i)} = h_{m_2,n_2}^{(i)}$ if $\min(m_1, n_1) = \min(m_2, n_2)$, and $h_{m_1,n_1}^{(i)} > h_{m_2,n_2}^{(i)}$ if $\min(m_1, n_1) > \min(m_2, n_2)$. From Lemma 2, $\mathbf{H}$ is positive definite. Therefore, $k_i(\mathbf{p}_i)$ is convex in $\mathbf{p}_i$. ∎

F. Proof of Proposition 3

Proof: Let $\mathbf{p}_1 = [\mathbf{p}_{10}^T, \mathbf{p}_{11}^T, \cdots, \mathbf{p}_{1M}^T]^T$, $\mathbf{p}_2 = [\mathbf{p}_{20}^T, \mathbf{p}_{21}^T, \cdots, \mathbf{p}_{2M}^T]^T$

$$P_{avg}(\lambda\mathbf{p_1} + (1-\lambda)\mathbf{p_2}) = \sum_{i=0}^{M}\pi_i k_i(\lambda\mathbf{p}_{1i} + (1-\lambda)\mathbf{p}_{2i})$$

$$\overset{(a)}{<} \sum_{i=0}^{M}\pi_i[\lambda k_i(\mathbf{p}_{1i}) + (1-\lambda)k_i(\mathbf{p}_{2i})]$$

$$= \lambda P_{avg}(\mathbf{p_1}) + (1-\lambda)P_{avg}(\mathbf{p_2}). \tag{87}$$

where (a) follows from Lemma 3. ∎

G. Proof of Lemma 4

Proof: For $x \geq 0$, let

$$f_k(x) = \sum_{i=U(k)}^{M}\pi_i F_{CH}(B(L(i)-k+1)x). \tag{88}$$

$f_k(x)$ is a continuous and decreasing function of x. Since $0 \leq F_{CH}(x) \leq 1$, we have

$$0 \leq f_k(x) \leq \sum_{i=U(k)}^{M}\pi_i. \tag{89}$$

Because $0 \leq \sum_{i=U(k)}^{k-1}\pi_i \leq \sum_{i=U(k)}^{M}\pi_i$, there must exists $\hat{y}_k$ such that $f_k(\hat{y}_k) = \sum_{i=U(k)}^{k-1}\pi_i$, and $\hat{y}_k \geq 0$. ∎

H. Proof of Lemma 5

Proof:
(1) By (33),

$$\sum_{j=1}^{L(i)}\hat{p}_{i,j} = 1 - F_{CH}(g_{L(i)}^{(i)}) \leq 1. \tag{90}$$

(2) $\hat{p}_{i,L(i)} = 1 - F_{CH}(B(1)\hat{y}_j) \geq 0$.
(3) If $B(k+1)/B(k) = c$, for $i > U(j)$, $\mu \leq j < M$, we have

$$\hat{p}_{i,j} = F_{CH}(B(L(i)-j)\hat{y}_{j+1}) - F_{CH}(B(L(i)-j)c\hat{y}_j). \tag{91}$$

Therefore, $\hat{p}_{i,j} > 0$ if and only if $\hat{y}_{j+1}/\hat{y}_j < c$, which is irrelevant to i; $\hat{p}_{i,j} < 0$ if and only if $\hat{y}_{j+1}/\hat{y}_j > c$; $\hat{p}_{i,j} = 0$ if and only if $\hat{y}_{j+1}/\hat{y}_j = c$. I.e., $\hat{p}_{i,j}$ has the same sign for each column j except for the upper most element $p_{U(j),j}$, $j \in \{\mu, \cdots, M\}$.

(4) Assume the ratio $B(k+1)/B(k)$ takes a value of c, i.e., $B(k+1)/B(k) = c$. For $1 \leq j \leq \mu - 1$, we have $f_j(\hat{y}_j) = f_{j+1}(c\hat{y}_j) = \sum_{i=0}^{j-1}\pi_i \leq \sum_{i=0}^{j}\pi_i = f_{j+1}(\hat{y}_{j+1})$. Therefore, $c\hat{y}_j \geq \hat{y}_{j+1}$. Hence, we have

$$\hat{p}_{i,j} = F_{CH}(B(L(i)-j)\hat{y}_{j+1}) - F_{CH}(B(L(i)-j)c\hat{y}_j) \geq 0. \tag{92}$$

∎

I. Lemma 6 and Its Proof

Lemma 6: The solution to (93), denoted by y_{j,j_0}^*, exists, and $0 \leq y_{j,j_0}^* \leq \hat{y}_j$ for $j \in \{\mathcal{F}(j_0)+1, \cdots, j_0\}$.

$$\sum_{i=U(j)}^{U(j_0)}\pi_i(1 - F_{CH}(B(L(i)-j+1)y_{j,j_0}^*)) = \sum_{i=j}^{j_0}\pi_i. \tag{93}$$

Proof: For $x \geq 0$, define $f_j^*(x)$ by

$$f_j^*(x) = \sum_{i=U(j)}^{U(j_0)}\pi_i(1 - F_{CH}(B(L(i)-j+1)x). \tag{94}$$

$f_j^*(x)$ is a continuous and increasing function of x. Substituting (52) into (44), then substituting (44) into (50), and letting (50) equal to the r.h.s. of (47), we have

$$\sum_{i=U(j)}^{M}\pi_i(1 - F_{CH}(B(L(i)-j+1)y_j)) = \sum_{i=j}^{M}\pi_i. \tag{95}$$

y_j is the only unknown variable in (95). Notice that (95) is equivalent to (51), $\hat{y}_k$ is also the solution to (95). So we have

$$\sum_{i=U(j)}^{M}\pi_i(1 - F_{CH}(B(L(i)-j+1)\hat{y}_j)) = \sum_{i=j}^{M}\pi_i. \tag{96}$$

Substituting (94) into l.h.s. of (96), we have

$$\sum_{i=j}^{M}\pi_i = f_j^*(\hat{y}_j) + \sum_{i=U(j_0)+1}^{M}\pi_i(1 - F_{CH}(B(L(i)-j+1)\hat{y}_j)$$

$$\overset{(a)}{\leq} f_j^*(\hat{y}_j) + \sum_{i=U(j_0)+1}^{M}\pi_i(1 - F_{CH}(B(L(i)-j_0)\hat{y}_{j_0+1})$$

$$\overset{(b)}{=} f_j^*(\hat{y}_j) + \sum_{i=U(j_0+1)}^{M}\pi_i(1 - F_{CH}(B(L(i)-j_0)\hat{y}_{j_0+1})$$

$$\overset{(c)}{=} f_j^*(\hat{y}_j) + \sum_{i=j_0+1}^{M}\pi_i, \tag{97}$$

where (a) holds because of $\pi_i \geq 0$ and (36), i.e., for any $j \in \{\mathcal{F}(j_0)+1, \cdots, j_0\}$,

$$F_{CH}(B(L(i)-j_0)\hat{y}_{j_0+1}) < F_{CH}(B(L(i)-j+1)\hat{y}_j); \tag{98}$$

and (b) holds because from Lemma 4, $j_0 \in \{\mu, \cdots, M-1\}$; and (c) holds because of (96).

Therefore, we have

$$f_j^*(\hat{y}_j) \geq \sum_{i=j}^{j_0}\pi_i. \tag{99}$$

Since $f_j^*(0) = 0 \leq \sum_{i=j}^{j_0}\pi_i \leq f_j^*(\hat{y}_j)$, and $f_j^*(x)$ is a continuous increasing function, hence the solution to $f_j^*(x) = \sum_{i=j}^{j_0}\pi_i$ must exist, i.e., $f_j^*(y_{j,j_0}^*) = \sum_{i=j}^{j_0}\pi_i$, and we have $0 \leq y_{j,j_0}^* \leq \hat{y}_j$. ∎

J. Lemma 7 and Its Proof

Lemma 7: If $B(k+1)/B(k) = c$ for all $k \geq 1$, then $y_{j+1,j_0}^* \leq c y_{j,j_0}^*$.

Proof:

$$
\begin{aligned}
f_{j+1}^*(c y_{j,j_0}^*) &\overset{(a)}{=} f_j^*(y_{j,j_0}^*) \\
&= \sum_{i=j}^{j_0} \pi_i \\
&\geq \sum_{i=j+1}^{j_0} \pi_i \\
&= f_{j+1}^*(y_{j+1,j_0}^*).
\end{aligned}
\tag{100}
$$

(a) holds because $B(k+1)/B(k) = c$. Since $f_{j+1}^*(x)$ is a continuous increasing function, hence $y_{j+1,j_0}^* \leq c y_{j,j_0}^*$. ∎

K. Lemma 8 and Its Proof

Lemma 8: $p_{i,\mathcal{F}(j_0)}^*(j_0) \geq 0$ for $i \in \{U(j_0)+1, \cdots, M\}$, where $p_{i,\mathcal{F}(j_0)}^*(j_0)$ is defined in (54).

Proof: For $i \in \{U(j_0)+1, \cdots, M\}$, we have

$$
\begin{aligned}
p_{i,\mathcal{F}(j_0)}^*(j_0) &\overset{(a)}{=} \sum_{j=\mathcal{F}(j_0)}^{j_0} \hat{p}_{i,j} \\
&\overset{(b)}{=} F_{CH}(B(L(i) - j_0)\hat{y}_{j_0+1}) \\
&\quad - F_{CH}(B(L(i) - \mathcal{F}(j_0) + 1)\hat{y}_{\mathcal{F}(j_0)}).
\end{aligned}
\tag{101}
$$

where (a) due to (54), and (b) due to (53). Since $\mathcal{F}(j_0)$ satisfies (36b), hence the r.h.s. of (101) is greater than or equal to zero. Therefore, $p_{i,\mathcal{F}(j_0)}^*(j_0) \geq 0$. ∎

L. Lemma 9 and Its Proof

Lemma 9: If $B(k+1)/B(k) = c$ for all $k \geq 1$, then $p_{i,j}^*(j_0) \geq 0$ for $j = \{\mathcal{F}(j_0), \cdots, j_0\}$ and $i \in \{U(j), \cdots, U(j_0)\}$, where $p_{i,j}^*(j_0)$ is defined in (54).

Proof: Consider three cases as below.

- Case 1 ($j = j_0$): i can only take one value, i.e., $U(j_0)$. We have $p_{U(j_0),j_0}^* = \hat{p}_{U(j_0),j_0} = 1 - F_{CH}(B(1)y_{j_0,j_0}^*) \geq 0$.
- Case 2 ($j = \mathcal{F}(j_0)$): From (54), the definition of $p_{i,\mathcal{F}(j_0)}^*(j_0)$, and the fact that $y_{\mathcal{F}(j_0),j_0}^* = \hat{y}_{\mathcal{F}(j_0)}$, we have

$$
\begin{aligned}
p_{i,\mathcal{F}(j_0)}^*(j_0) &= F_{CH}(B(L(i) - \mathcal{F}(j_0))y_{\mathcal{F}(j_0)+1,j_0}^*) \\
&\quad - F_{CH}(B(L(i) - \mathcal{F}(j_0) + 1)\hat{y}_{\mathcal{F}(j_0)}).
\end{aligned}
\tag{102}
$$

From Lemma 6, $y_{\mathcal{F}(j_0)+1,j_0}^* \leq \hat{y}_{\mathcal{F}(j_0)+1}$; hence $F_{CH}(B(L(i) - \mathcal{F}(j_0))y_{\mathcal{F}(j_0)+1,j_0}^*) \geq F_{CH}(B(L(i) - \mathcal{F}(j_0))\hat{y}_{\mathcal{F}(j_0)+1})$. Then, from (102), we have

$$
\begin{aligned}
p_{i,\mathcal{F}(j_0)}^*(j_0) &\geq F_{CH}(B(L(i) - \mathcal{F}(j_0))\hat{y}_{\mathcal{F}(j_0)+1}) \\
&\quad - F_{CH}(B(L(i) - \mathcal{F}(j_0) + 1)\hat{y}_{\mathcal{F}(j_0)}).
\end{aligned}
\tag{103}
$$

In addition, we have

$$
\begin{aligned}
&F_{CH}(B(L(i) - \mathcal{F}(j_0))\hat{y}_{\mathcal{F}(j_0)+1}) \\
&\overset{(a)}{>} F_{CH}(B(L(i) - j_0)\hat{y}_{j_0+1}) \\
&\overset{(b)}{\geq} F_{CH}(B(L(i) - \mathcal{F}(j_0) + 1)\hat{y}_{\mathcal{F}(j_0)}),
\end{aligned}
\tag{104}
$$

where (a) because $j = \mathcal{F}(j_0)$ is the largest j that satisfies (36b), and hence $j = \mathcal{F}(j_0) + 1$ does not satisfy (36b); (b) due to (36b). Combining (103) and (104), we have $p_{i,\mathcal{F}(j_0)}^*(j_0) \geq 0$.

- Case 3 ($j \in \{\mathcal{F}(j_0)+1, \cdots, j_0-1\}$):

$$
\begin{aligned}
p_{i,j}^* &\overset{(a)}{=} F_{CH}(B(L(i) - j)y_{j+1,j_0}^*) \\
&\quad - F_{CH}(B(L(i) - j + 1)y_{j,j_0}^*) \\
&= F_{CH}(B(L(i) - j)y_{j+1,j_0}^*) \\
&\quad - F_{CH}(B(L(i) - j)c y_{j,j_0}^*) \\
&\overset{(b)}{\geq} 0.
\end{aligned}
\tag{105}
$$

where (a) is from (54) and (55), and (b) is due to Lemma 7. ∎

M. Proof of Proposition 4

To prove Proposition 4, we need Lemma 6 in Appendix I, Lemma 7 in Appendix J, Lemma 8 in Appendix K, and Lemma 9 in Appendix L. In (54), the value of $p_{i,j}^*(j_0)$ is given in four regions. For simplicity of the subsequent derivations, we decompose the last region into two regions. Denote these five regions as R_1 through R_5.

$$
\begin{cases}
R_1 = \{(i,j)|j \in \{1, \cdots, \mathcal{F}(j_0)-1\} \bigcup \{j_0+1, \cdots, M\}, \\
\quad i \in \{U(j), \cdots, M\}\} \\
R_2 = \{(i,j)|j = \mathcal{F}(j_0), i \in \{U(j_0)+1, \cdots, M\}\} \\
R_3 = \{(i,j)|j = \{\mathcal{F}(j_0)+1, \cdots, j_0\}, \\
\quad i \in \{U(j_0)+1, \cdots, M\}\} \\
R_4 = \{(i,j|j = \{\mathcal{F}(j_0)+1, \cdots, j_0\}, \\
\quad i \in \{U(j), \cdots, U(j_0)\}\} \\
R_5 = \{(i,j|j = \mathcal{F}(j_0), i \in \{U(\mathcal{F}(j_0)), \cdots, U(j_0)\}\}.
\end{cases}
\tag{106}
$$

Proof: The condition $p_{i,j} \geq 0$ is equivalent to $\pi_i p_{i,j} \geq 0$. Define a new Lagrangian function by

$$
\begin{aligned}
J(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\delta}) &= \sum_{i=0}^{M} \pi_i k_i(\mathbf{p}_i) + \sum_{j=1}^{M} \lambda_j \left(\sum_{i=U(j)}^{M} \pi_i p_{i,j} - \pi_j \right) \\
&\quad - \sum_{i=0}^{M} \sum_{j=1}^{L(i)} \pi_i \delta_{i,j} p_{i,j},
\end{aligned}
\tag{107}
$$

where $\boldsymbol{\delta}$ is a matrix with elements $\delta_{i,j}$. The optimal solution $p_{i,j}^*(j_0)$ should satisfy

$$
\frac{\partial}{\partial p_{i,j}} J(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\delta}) = 0
\tag{108a}
$$

$$
\sum_{i=U(j)}^{M} \pi_i p_{i,j} = \pi_j
\tag{108b}
$$

$$
p_{i,j} \geq 0
\tag{108c}
$$

$$
\delta_{i,j} p_{i,j} = 0
\tag{108d}
$$

$$
\delta_{i,j} \geq 0.
\tag{108e}
$$

From Lemmas 8 and 9, $p_{i,j}^*(j_0) \geq 0$, (108c) is satisfied.

Notice that $p^*_{i,j}(j_0) = 0$ in region R_3. Let $\delta_{i,j} = 0$ outside R_3, (108d) is satisfied.

Differentiate $J(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\delta})$ w.r.t. $p_{i,j}$ and let the derivative equal to zero, i.e.,

$$\frac{\partial}{\partial p_{i,j}} J(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\delta}) = \pi_i \sum_{k=L(i)-j+1}^{L(i)} \frac{-B(k)}{g_k^{(i)}} + \lambda_j \pi_i - \pi_i \delta_{i,j} = 0. \tag{109}$$

Following the same steps as in (43) to (52), we have

$$p_{i,j} = F_{CH}(B(L(i)-j)y_{j+1}^{(i)}) - F_{CH}(B(L(i)-j+1)y_j^{(i)}). \tag{110}$$

where

$$y_j^{(i)} = \frac{1}{\lambda_j - \delta_{i,j} - \lambda_{j-1} + \delta_{i,j-1}}. \tag{111}$$

And $y_j^{(i)}$ should satisfy the following equation:

$$\sum_{i=U(j)}^{M} \pi_i(1 - F_{CH}(B(L(i)-j+1)y_j^{(i)})) = \sum_{i=j}^{M} \pi_i, \tag{112}$$

(112) and (110) are derived from (108a) and (108b). Satisfying (108a) and (108b) is equivalent to satisfying (112) and (110).

Let

$$y_j^{(i)} = \begin{cases} \hat{y}_j & (i,j) \in R1 \bigcup R2 \bigcup R5 \\ y^*_{j,j_0} & (i,j) \in R4 \\ \dfrac{B(L(i)-j_0)}{B(L(i)-j+1)}\hat{y}_{j_0+1} & (i,j) \in R3 \end{cases}. \tag{113}$$

We will prove that $y_j^{(i)}$ defined in (113) satisfies (112), and $p_{i,j}$ obtained by substituting (113) into (110) is exactly $p^*_{i,j}(j_0)$ defined by (54), i.e., the optimal solution. The final step is to prove $\delta_{i,j}, \{i,j\} \in R_3$, obtained by substituting (113) into (111) satisfies (108e) ($\delta_{i,j}$ outside R_3 is set to zero to meet (108d)).

From (96), $y_j^{(i)}$ in $R1 \bigcup R2 \bigcup R5$ satisfies (112). In region $R_3 \bigcup R_4$, substitute (113) into l.h.s. of (112), i.e.,

$$\sum_{i=U(j)}^{M} \pi_i(1 - F_{CH}(B(L(i)-j+1)y_j^{(i)}))$$

$$= \sum_{i=U(j)}^{U(j_0)} \pi_i(1 - F_{CH}(B(L(i)-j+1)y^*_{j,j_0})) +$$

$$\sum_{i=U(j_0)+1}^{M} \pi_i(1 - F_{CH}(B(L(i)-j_0)\hat{y}_{j_0+1})) \tag{114}$$

$$\stackrel{(a)}{=} \sum_{i=j}^{j_0} \pi_i + \sum_{j_0+1}^{M} \pi_i$$

$$= \sum_{i=j}^{M} \pi_i$$

where (a) is due to (56) and (96), respectively. Thus, $y_j^{(i)}$ satisfies (112) in all five regions.

By substituting (113) into (110), it is easy to verify that the resulting $p_{i,j}$ is identical to $p^*_{i,j}(j_0)$.

Next we will prove that there exist $\delta_{i,j} \geq 0$, $(i,j) \in R_3$ such that (108e) is satisfied. From (111), $\delta_{i,j}$ within R_3 is determined by $y_j^{(i)}$, $j \in \{\mathcal{F}(j_0)+1, \cdots, j_0+1\}$.

From (111), $\forall (i,j) \in R_4$, we have

$$\frac{1}{\lambda_j - \lambda_{j-1}} = y_j^{(i)} = y^*_{j,j_0}. \tag{115}$$

When $j = \mathcal{F}(j_0)+1$, $i \in \{U(j_0)+1, \cdots, M\}$ (the leftmost column of R_3), substitute (111) and (115) into (113), i.e.,

$$\delta_{i,\mathcal{F}(j_0)+1} = \lambda_{\mathcal{F}(j_0)+1} - \lambda_{\mathcal{F}(j_0)} - \frac{B(L(i)-\mathcal{F}(j_0))}{B(L(i)-j_0)\hat{y}_{j_0+1}}$$

$$= \frac{1}{y^*_{\mathcal{F}(j_0)+1,j_0}} - \frac{B(L(i)-\mathcal{F}(j_0))}{B(L(i)-j_0)\hat{y}_{j_0+1}}. \tag{116}$$

From (36), and Lemma 6, $\forall j \in \{\mathcal{F}(j_0)+1, \cdots, j_0\}$, we have

$$\frac{B(L(i)-j_0)}{B(L(i)-j+1)}\hat{y}_{j_0+1} > \hat{y}_j \geq y^*_{j,j_0} \geq 0. \tag{117}$$

Therefore $\delta_{i,\mathcal{F}(j_0)+1} > 0$. Similarly, when $j \in \{\mathcal{F}(j_0)+2, \cdots, j_0\}$, we have

$$\delta_{i,j} = \frac{1}{y^*_{j,j_0}} - \frac{B(L(i)-j+1)}{B(L(i)-j_0)\hat{y}_{j_0+1}} + \delta_{i,j-1} > 0. \tag{118}$$

When $j = j_0+1$, substitute (111) into (113), i.e.,

$$\delta_{i,j_0} = \frac{1}{\hat{y}_{j_0+1}} + \lambda_{j_0} - \lambda_{j_0+1}. \tag{119}$$

(119) implies that the value of δ_{i,j_0} does not depend on i. This is consistent with δ_{i,j_0} given in (118): because $B(k+1)/B(k) = c$ and the value of $\delta_{i,\mathcal{F}(j_0)+1}$ does not depend on i. Therefore $\delta_{i,j}$ exists in region R_3, and is non-negative. This completes the proof. ∎

N. Proof of Proposition 5

Proof: $\hat{p}_{i,j}$ and $p^*_{i,j}(j_0)$ differs only in columns $j \in \{\mathcal{F}(j_0), \cdots, j_0\}$, i.e., in function $\mathcal{G}(j_0, i, j)$. Assume there are two negative columns in $\hat{p}_{i,j}$, denoted by $j_{01} < j_{02}$. From Lemma 5, we have $j_{01} \geq \mu$, and $j_{0n} < M$.

If $j_{01} < \mathcal{F}(j_{02})$, by the same method as in the proof of Proposition 4, we can prove that the optimal solution is

$$p^*_{i,j} = \begin{cases} \mathcal{G}(j_{01}, i, j) & j \in \{\mathcal{F}(j_{01}), \cdots, j_{01}\} \\ \mathcal{G}(j_{02}, i, j) & j \in \{\mathcal{F}(j_{02}), \cdots, j_{02}\} \\ \hat{p}_{i,j} & otherwise \end{cases}. \tag{120}$$

If $j_{01} \geq \mathcal{F}(j_{02})$, column j_{01} will be non-negative in $p^*_{i,j}(j_{02})$, i.e., $\mathcal{G}(j_{02}, i, j)$. Because in the proof of Proposition 4, we do not require columns $j \in \{\mathcal{F}(j_0), \cdots, j_0\}$ in $\hat{p}_{i,j}$ to be non-negative. Therefore $p^*_{i,j}(j_{02})$ is indeed the optimal solution to (23).

The above two cases can be implemented in a unique way, i.e., updating the negative columns sequentially from right to left (here updating negative column means changing the value of columns $j \in \{\mathcal{F}(j_0), j_0\}$ from $\hat{p}_{i,j}$ to $\mathcal{G}(j_0, i, j)$, where j_0 is the column index of the negative column). It is possible that the left negative column becomes non-negative during the

procedure of updating the right negative column. Then after updating the right negative column, updating the left negative column is not needed.

The above algorithm can be generalized to situations where arbitrary number of negative columns are present in $\hat{p}_{i,j}$, which is indeed Algorithm 1. ∎

REFERENCES

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

[2] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, Jan 1972.

[3] S. Hanly and D. Tse, "Multiaccess fading channels. ii. delay-limited capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2816–2831, Nov 1998.

[4] R. Negi and J. M. Cioffi, "Delay-constrained capacity with causal feedback," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2478–2494, Sept. 2002.

[5] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[6] X. Liu and A. J. Goldsmith, "Optimal power allocation over fading channels with stringent delay constraints," in *Proc. IEEE ICC 2002*, vol. 3, 28 April–2 May 2002, pp. 1413–1418.

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

[8] B. Prabhakar, E. Uysal Biyikoglu, and A. El Gamal, "Energy-efficient transmission over a wireless link via lazy packet scheduling," in *Proc. IEEE INFOCOM 2001*, vol. 1. Anchorage, AK, April 2001, pp. 386–394.

[9] A. Fu, E. Modiano, and J. Tsitsiklis, "Optimal energy allocation for delay-constrained data transmission over a time-varying channel," in *Proc. INFOCOM 2003*, vol. 2, 30 March–3 April 2003, pp. 1095–1105.

[10] M. Zafer and E. Modiano, "Optimal rate control for delay-constrained data transmission over a wireless channel," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 4020–4039, Sept. 2008.

[11] B. E. Collins and R. L. Cruz, "Transmission policies for time varying channels with average delay constraints," in *Proc. Allerton Conf. Communication, Control, Computing*, Oct. 1999, pp. 709–717.

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

[13] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 125–144, Jan. 2004.

[14] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630–643, July 2003.

[15] J. Tang and X. Zhang, "Quality-of-service driven power and rate adaptation over wireless links," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 3058–3068, August 2007.

[16] ——, "Quality-of-service driven power and rate adaptation for multi-channel communications over wireless links," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4349–4360, December 2007.

[17] X. Li, X. Dong, and D. Wu, "Hierarchical queue-length-aware power control for real-time applications over wireless networks," *accepted by Wireless Communications & Mobile Computing*, 2010.

[18] X. Li, F. Cao, X. Dong, and D. Wu, "Power control for delay constrained multichannel communications using outdated csi," *accepted by Wireless Communications & Mobile Computing*, 2010.

[19] L. Kleinrock, *Queueing systems*. Wiley New York, 1975, vol. 1.

[20] M. Bazaraa, H. Sherali, and C. Shetty, *Nonlinear programming: theory and algorithms*, 3rd ed. Wiley-interscience, 2006.

[21] P. Boggs and J. Tolle, "A strategy for global convergence in a sequential quadratic programming algorithm," *SIAM journal on Numerical Analysis*, vol. 26, no. 3, pp. 600–623, 1989.