# Power-Delay Tradeoff over Wireless Networks

Xi Zhang, *Senior Member, IEEE,* and Jia Tang

***Abstract*—When transmitting stochastic traffic flows over wireless networks, there exists an inherent tradeoff between average transmit power and corresponding queuing-delay bound. In this paper, we investigate such a tradeoff and show how average power increases as delay-bound requirement for wireless network traffics becomes stringent. Specifically, we propose the resource allocation schemes to minimize the power consumption subject to a delay quality-of-service (QoS) constraint, where the delay constraint is in terms of queue-length *decay rate* when an arrival traffic is transmitted through the wireless networks. We focus on orthogonal-frequency-division-multiplexing (OFDM) communications under three different network infrastructures, namely, point-to-point link, multihop amplify-and-forward (AF) network, and multiuser cellular network. We derive the optimal resource allocation policies for each scenario, and compare their performances with other existing resource-allocation policies. The obtained simulation and numerical results show that using our proposed optimal resource-allocation policies, significant power saving can be achieved. Furthermore, our OFDM-based communications systems can significantly reduce the power consumption, especially under stringent delay constraint.**

***Index Terms*—Power control, statistical delay-bounded quality-of-service (QoS) guarantees, effective capacity, wireless networks, resource allocation and management, scheduling, OFDM-based communications systems, convex optimization, information theory.**

## I. INTRODUCTION

THE EXPLOSIVE demand for wireless services motivates a rapid evolution of wireless wideband communications and networks to transmit the various types of real-time traffics. In order to efficiently support a large number of distinct wireless applications such as video/audio and interactive data transmissions over mobile networks, the diverse delay-bounded quality-of-service (QoS) guarantees play an increasingly important role in the next-generation multimedia wireless networks. This implies that there are more challenging problems emerging in wireless networks algorithm designs, resource control, and performance optimizations.

Over the wireless environments, it is well known that power is one of the most scarce radio resources. Consequently, a great deal of research has been devoted to designing resource allocation policies that can minimize the transmit power. Power minimization can not only save energy for the mobile terminals, and thus prolong the network lifetime, but also introduce less interference to the other users, and hence increase the entire system capacity. At the same time, however, to guarantee a certain delay-QoS constraint, sufficient power needs to be assigned for the transmission in order to satisfy the QoS required by the traffic flows. Therefore, there exists an inherent tradeoff between transmit power and delay constraint. In this paper, we focus on this power-delay tradeoff and study: (i) the dynamics on how the transmit power increases as the delay constraint gets more stringent; (ii) how to design the optimal resource allocation policy that can minimize the power consumption while still satisfying the required delay QoS.

In order to tackle the above problems, it is necessary to introduce a tractable delay-QoS performance metric. Thanks to the dual concepts of effective bandwidth [1] and effective capacity [2], [3], [4], we obtain a powerful approach to evaluate *statistical* QoS performance of the wireless networks. Specifically, we design the optimal resource allocation schemes to minimize the average transmit power subject to a delay constraint, where the delay QoS is in terms of queue-length *decay rate*, which can be jointly determined by the effective bandwidth of the arrival traffic and the effective capacity of the wireless channel. In this paper, we concentrate on the orthogonal-frequency-division-multiplexing (OFDM)-based communications model, since this is one of the most promising communication frameworks for future wireless networks. Moreover, noting that OFDM is a special case of parallel communications in the *frequency* domain, our proposed scheme can be readily extended to other parallel communications models, such as multiple-input multiple-output (MIMO) system, one of the most popular *spatial* domain parallel communication schemes.

Resource allocation for delay QoS guarantees has been extensively studied in the context of power control, scheduling, and admission control (see e.g., [3], [5], [6], [7], [8], [9], [4], [10], [11], and references therein). In [5], [6], the authors apply information theory to develop power control policies achieving *delay-limited* capacity. Compared to ergodic capacity, which does not impose any delay constraint, delay-limited capacity focuses on another extreme where the delay QoS is stringent. Such a pure information theoretic approach may not be fine-grained enough for the user's satisfactions, where a wide range of delay constraints may be requested for different applications. On the other hand, increasing research efforts have been paid to the tradeoff between energy/power consumption and queuing-delay QoS control over wireless channels [12], [13], [14], [15], [16], [17], [18]. A major category of the related works [14], [15], [16], [17] attempt to guarantee the

Manuscript received January 18, 2012; revised October 1, 2012. The associate editor coordinating the review of this paper and approving it for publication was A. Mackenzie.

The research reported in this paper was supported in part by the U.S. National Science Foundation under Grant CNS-205726 and the U.S. National Science Foundation CAREER Award under Grant ECCS-0348694.

X. Zhang is with the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA (e-mail: xizhang@ece.tamu.edu).

J. Tang was with the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA. He is now with Qualcomm Research Labs, Santa Clara, California, USA (e-mail: jiat@qualcomm.com).

Digital Object Identifier 10.1109/TCOMM.2013.061913.120001

hard delay-upper-bound in queuing delay for delay-sensitive services and/or to limit the maximum queue length under the finite buffer size. In [15], the authors developed a scheme to minimize the total energy expenditure over a single time-varying fading channel, while ensuring the guarantees of the deadline constraint. The problem was formulated by using a stochastic control process over a continuous-time queuing system, and the solution was obtained and characterized by a Lagrangian duality approach.

The authors of [16] tackled the delay bound assurance problem for a multi-packet scheduling scenario, where multiple packets with different packet sizes need to be delivered within their respective delay-bound requirement. Literature [17] proposed a novel calculus approach for energy minimization under diverse strict delay/buffer constraints. The authors of [14] studied distortion minimizing for multimedia video under given delay bound constraints. These efforts in fact all characterize the delay QoS requirement for a dynamic queuing system in a deterministic way, where the delay bound is targeted to be within certain threshold. These approaches well qualify the needs of real-time services. However, over wireless fading channels having essential nature of randomness, the deterministic hard delay-bound QoS guarantees are usually too costly in terms of energy consumptions. This is because the severe and time-varying deep fading states of wireless channels will drastically increase energy expenditure to compensate for the power loss during wireless transmissions.

The second category of research efforts focused more on guaranteeing the *average-delay* bound QoS guarantees for queuing-delay control [12], [13], [17]. The authors of [12] derived the scaling law trading off the average-delay against the energy consumptions. They extended the Berry-Gallager bound from a single user context to a multiuser one. They also proposed a class of algorithms to achieve the fundamental tradeoff characterized by the Berry-Gallager bound in the multiuser scenario. Among others, literature [13] proposed an on-line implementation for the optimal packet scheduling algorithm towards energy minimization subject to the average-delay constraint; authors of [17] considered the problem to minimize the average delay subject to the power constraint. The other related research works under the second category include papers [7], [8], [9], which aim at the average-delay control with power-delay tradeoff over wireless fading channels. While average-delay control can alleviate the pressure on energy consumption as compared to the deterministic hard delay-bound assurance, the average delay is a marginal statistics, which may not be sufficient to characterize the dynamics of delay distribution. In particular, for real-time multimedia applications, the key QoS requirement is not the average delay, but the *bounded delay*. So, the guarantees of average-delay does not necessarily satisfy the delay QoS requirements, especially for the bounded delay in real-time and multimedia services.

In contrast to the above two categories of characterization of delay QoS requirements, we in this paper concentrate on statistically guaranteeing the *bounded queuing delay* through a different yet more promising approach, often referred as the statistical QoS guarantees. On the one hand, we do not impose the hard delay bound QoS requirements for queuing delay control. On the other hand, we aim at statistically controlling the queue-length and delay distributions, and correspondingly, the QoS provisioning is characterized by the queue-length-/delay-bound violation probabilities, and effective capacity. The motivation of our work is to confine the delay/queue-length bound violation probability to a very small value range, as the wireless channel fading status varies dramatically. In the meantime, our approach can directly map the actual delay QoS requirement to certain parameters termed QoS exponent, which cannot be achieved by the average-delay optimization control.

In [3], [4], [10], based on effective bandwidth and effective capacity, the authors investigated resource allocation for statistical QoS guarantees. However, aforementioned previous works only focused on the point-to-point based wireless networks. In this paper, we extend the results to the scenarios for multiuser cellular and multihop wireless networks under OFDM-based communication systems. In [11], [19], the authors studied scheduling policies for multiuser cellular networks, but the power allocation is not discussed. Instead, we consider a joint scheduling, power allocation, and rate allocation problem in this paper. The major contributions of this paper are summarized as follows:

1) **Framework**: Combining information theoretic results with the statistical QoS theory, we propose a framework to investigate the power-delay tradeoff for wireless networks. In order to find the optimal tradeoff, we formulate a power minimization problem in a convex optimization setting. This framework is generic and applicable to different wireless network infrastructures.
2) **Algorithms**: Under our proposed framework, we derive resource allocation algorithms that achieve the optimal power-delay tradeoff under three different network infrastructures: (i) point-to-point link; (ii) multihop amplify-and-forward (AF) network; (iii) multiuser cellular network. Considering the implementation issues, we also derive a low-complexity suboptimal algorithm for point-to-point link, and an adaptive allocation algorithm for cellular networks, respectively. Moreover, over the multiuser cellular networks, joint power allocation and time-slot allocation provides guidelines on how to design optimal scheduling policies for efficient QoS guarantees.
3) **Performance**: We evaluate the performance of the optimal power-delay tradeoff by extensive simulations. For comparison purposes, we also derive the closed-form effective capacity expressions for a number of existing policies, under which the corresponding power-delay tradeoff can be evaluated numerically. Simulation and numerical results show that our proposed optimal resource allocations can reduce the transmit power significantly, especially when the delay constraint is stringent. Furthermore, OFDM-based communications system can greatly reduce the power consumption under stringent delay QoS.

The rest of the paper is organized as follows. Sections II formulates the problem and sets up the system model. Section III derives the power allocation policy for point-to-point link. Sections IV and V develop the resource allocation strategies for multihop AF networks and multiuser cellular networks, respectively. Section VI conducts simulations and numerical analyses to evaluate the performance of our proposed schemes. The paper concludes with Section VII.

*Notations.* Throughout this paper, we use the following notations. We use lower-case boldface letters to denote column vectors. $(\cdot)^T$ represents the transpose. $[x]^+ \triangleq \max\{0,x\}$. $\mathbb{E}[\cdot]$ stands for the expectation. All logarithm functions are based on $e$, unless otherwise stated.

## II. Problem Formulations

### A. Statistical QoS Guarantees

Since the early 90's, statistical QoS guarantees have been extensively studied in the contexts of *effective bandwidth* theory [1], [20]. Basically, the effective bandwidth, denoted by $\mathcal{A}(\theta)$, is defined as the minimum *constant* service rate for a given arrival process, such that a required QoS constraint $\theta$ can be guaranteed. In the above definition, the QoS constraint $\theta$ characterizes the queue-length decay rate. Specifically, for a dynamic queueing system with stationary ergodic arrivals and service processes, under sufficient conditions, the queue length process $Q(t)$ converges in distribution to a random variable $Q(\infty)$ such that [1]

$$-\lim_{x\to\infty}\frac{\log\left(\Pr\{Q(\infty)>x\}\right)}{x}=\theta. \tag{1}$$

In other words, the probability of the queue-length exceeding a certain threshold $x$ decays exponentially fast as $x$ increases. In Eq. (1), the parameter $\theta$ ($\theta>0$), called QoS exponent [2], [3], determines this decay rate. A smaller $\theta$ corresponds to a slower decay rate, which implies that the system can only provide a *looser* QoS guarantee, while a larger $\theta$ leads to a faster decay rate, which means that a more *stringent* QoS can be supported. In particular, when $\theta\to 0$, the system can tolerate an arbitrarily long delay; when $\theta\to\infty$, the system cannot tolerate any delay.

Inspired by the effective bandwidth theory, the authors of [2] proposed a dual concept termed *effective capacity*. The effective capacity, denoted by $\mathcal{S}(\theta)$, is defined as the maximum *constant* arrival rate that a given service process can support in order to guarantee a QoS requirement specified by $\theta$. Integrating effective bandwidth with effective capacity, we can analyze the statistical QoS performance of a queuing system where both arrival and service processes are time-varying. For this general case, in order to guarantee a QoS constraint $\theta$, a necessary and sufficient condition is $\mathcal{S}(\theta)\geq\mathcal{A}(\theta)$. The above decoupling of effective bandwidth and effective capacity enables us to analyze $\mathcal{A}(\theta)$ and $\mathcal{S}(\theta)$ independently, while achieving the QoS performance of the entire system.

### B. Problem Formulation

Our problem is to find the minimum average transmit power for a certain arrival traffic flow such that a required delay-QoS constraint can be satisfied. To formulate a tractable problem in the context of statistical QoS guarantees, first, we interpret the delay-QoS constraint by a required QoS exponent $\theta$. Second, for a certain arrival traffic, we assume that the corresponding effective bandwidth $\mathcal{A}(\theta)$ of the arrival process is known. Thus, we can focus on optimizing the service process. Our problem becomes how to minimize the average transmit power such that the resulting $\mathcal{S}(\theta)$ still satisfies $\mathcal{S}(\theta)\geq\mathcal{A}(\theta)$ at a given $\theta$.[1] Mathematically, we need to solve the following general problem:

$$\min_{\boldsymbol{p}\in\mathcal{C}}\sum_i\mathbb{E}\left[p_i\right] \tag{2}$$

subject to

$$\mathcal{S}(\theta)\geq\mathcal{A}(\theta)\triangleq A \tag{3}$$

where $\mathcal{C}$ denotes a feasible set for the instantaneous transmit power vector $\boldsymbol{p}=(p_1,p_2,...p_D)^T$, and the dimension $D$ of the vector $\boldsymbol{p}$ is determined by the specific problem. As $\theta$ becomes larger and larger, indicating more and more stringent QoS constraint, the power-delay tradeoff characterizes how the required average transmit power increases with $\theta$.

In general, the analytical expression of the effective capacity for an arbitrary stationary ergodic service process is complicated and difficult to analyze (see e.g., [2, eq. (12)]). However, when the service process is an independent and identically distributed (i.i.d.) process, the effective capacity expression can be greatly simplified. Let the sequence $\{R[i]\geq 0,\ i=1,2,...\}$ denote an i.i.d. service process. Then, the effective capacity $\mathcal{S}(\theta)$ of this process is given by [3]:

$$\mathcal{S}(\theta)=-\frac{1}{\theta}\log\Big(\mathbb{E}\left[e^{-\theta R[i]}\right]\Big). \tag{4}$$

In this paper, we assume that the channel is i.i.d. block-fading, with one block independent of another block. Thus, it can significantly simplify the effective capacity derivations. Moreover, through our study in [3], we observe that there exists a simple and efficient approach to convert the power allocation policy obtained in i.i.d. block-fading channels to that over the correlated block-fading channels, making the investigation of i.i.d. block-fading channel more applicable.

Since $\log(\cdot)$ is a monotonically increasing function, for each given QoS constraint $\theta>0$, the constraint given in Eq. (3) is equivalent to the following expression:

$$\mathbb{E}\left[e^{-\theta R[i]}\right]\leq e^{-\theta A}. \tag{5}$$

In the rest of the paper, we solve different versions of this modified optimization problem expressed by Eqs. (2) and (5) for different wireless network infrastructures.

### C. System Model

To simplify the presentation, in this section we describe the system based on a point-to-point model. For multihop AF networks and multiuser cellular networks, we will further detail the system descriptions in their respective sections.

We concentrate on a discrete-time OFDM system. The transmitter and the receiver communicate through $N$ subchannels over a given spectral bandwidth $B$. A first-in-first-out

[1]The analytical results derived as the functions of $\theta$ in the rest of the paper still hold true for the heterogeneous multiusers cases where $\theta$ differs for different given users. However, to simplify the presentation, we just consider the homogeneous users cases in the paper.

(FIFO) buffer is equipped at transmitter, which buffers the data *frames* to be transmitted to the receiver. The frame duration is denoted by $T_f$, which is assumed to be less than the fading coherence time, but sufficiently long so that the information-theoretic assumption of infinite code-block length is meaningful [5]. The frame is then divided into $N$ substreams, each transmitted through one of the subchannels. Based on a given QoS constraint $\theta$ requested by the arrival traffic and channel state information (CSI) fed back from the receiver, the transmitter needs to find an optimal codeword (implemented by the adaptive modulation and coding) and a corresponding resource allocation strategy, which can minimize the average transmit power subject to the QoS constraint $\mathcal{S}(\theta) \geq \mathcal{A}(\theta)$.

The discrete-time channel process is assumed to be block-fading. The path gains are constant within a frame's duration $T_f$, but vary *independently* from one frame to another, following a certain continuous distribution. Note that our proposed scheme is independent of channel distributions, but we will focus on Rayleigh fading throughout the paper for simplicity. The probability density function (pdf) of the channel gain $\gamma_n$ at the $n$th subchannels, denoted by $p_{\Gamma_n}(\gamma)$, can be expressed as $p_{\Gamma_n}(\gamma) = \frac{1}{\overline{\gamma}_n} e^{-\frac{\gamma}{\overline{\gamma}_n}}$, where $\overline{\gamma}_n$ denotes the average channel gain for the $n$th subchannel with $n = 1, 2, ..., N$. Let $\boldsymbol{\gamma}[i] \triangleq (\gamma_1[i], \gamma_2[i], ..., \gamma_N[i])^T$ denote the instantaneous CSI at the $i$th time frame. When the receiver feeds $\boldsymbol{\gamma}[i]$ back to the transmitter, for a given power allocation $\boldsymbol{p}[i] \triangleq (p_1[i], p_2[i], ..., p_N[i])^T$, the maximum transmission rate, denoted by $R[i]$, can be expressed as

$$R[i] = \frac{T_f B}{N} \sum_{n=1}^{N} \log\left(1 + \gamma_n[i] p_n[i]\right) \tag{6}$$

which can be achieved by independent complex Gaussian code-books. We assume that $\boldsymbol{\gamma}[i]$ can be perfectly estimated at the receiver and reliably fed back to the transmitter without delay. Moreover, given a power allocation $\boldsymbol{p}[i]$ and CSI $\boldsymbol{\gamma}[i]$, we assume that the adaptive modulation and coding can choose an ideal channel code for each frame, such that the transmission rate achieves its maximum $R[i]$ given in Eq. (6). Based on this assumption, the derived minimum average transmit power using Eq. (6) serves as a lower-bound for real systems using practical coding schemes.

In the following discussions, since the block-fading channel process is i.i.d., its instantaneous marginal statistics is independent of the frame index $i$, and thus we may omit the frame index $i$ for simplicity, unless otherwise stated.

## III. Power-Delay Tradeoff Over a Single Link

### A. Optimal Resource Allocation

In this section, we investigate the power-delay tradeoff for a point-to-point OFDM link. In order to derive the optimal resource allocation policy that minimizes the average transmit power, we formulate the problem $(P1)$ as follows:

$$(P1) \quad \min_{\boldsymbol{p} \in \mathcal{C}_1} \sum_{n=1}^{N} \mathbb{E}\left[p_n\right] \tag{7}$$

subject to Eq. (5), where $R[i]$ in Eq. (5) is given by Eq. (6), and

$$\mathcal{C}_1 \triangleq \left\{ \boldsymbol{p} | \boldsymbol{p} = (p_1, p_2, ..., p_N)^T, \;\; p_n \geq 0, \; \forall n \in \{1, 2, ..., N\} \right\}. \tag{8}$$

It is easy to verify that both the objective function Eq. (7) and the constraint Eq. (5) are convex over the convex set $\mathcal{C}_1$. Thus, this problem can be solved by the convex optimization approach [21]. We derive the optimal resource allocation policy, which is summarized by the following theorem.

*Theorem 1:* For each fading state $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \cdots, \gamma_N)^T$, let $\pi(\cdot)$ denote a permutation of $\boldsymbol{\gamma}$ such that $\gamma_{\pi(1)} \geq \gamma_{\pi(2)} \geq \cdots \geq \gamma_{\pi(N)}$. Then, the $\pi(n)$-th component of $\boldsymbol{p}$, denoted by $p_{\pi(n)}$, follows the water-filling formula and is determined by

$$p_{\pi(n)} = \left[\omega - \gamma_{\pi(n)}^{-1}\right]^{+} \tag{9}$$

where

$$\omega = \nu^{\frac{N\beta+1}{M\beta+1}} \prod_{m=1}^{M} \gamma_{\pi(m)}^{-\frac{\beta}{M\beta+1}} \tag{10}$$

denotes the time-varying water-level, with $\nu$ chosen to meet the QoS constraint Eq. (5), and $\beta = \theta T_f B/N$ represents the normalized QoS exponent, and parameter $M$ denotes the number of *active* channels allocated with nonzero power, which, if exists, is the unique integer in $\{1, 2, \cdots, N\}$ such that $\omega > 1/\gamma_{\pi(m)}$ for $m \leq M$ and $\omega \leq 1/\gamma_{\pi(m)}$ for $m > M$. Otherwise, if such an $M$ does not exist, we have $p_n = 0$ and the system is in an outage state.

*Proof:* The proof is provided in Appendix A. ■

Theorem 1 tells us that the optimal policy allocates the transmit power across both frequency and time domains. Over the frequency domain, the optimal allocation is always water-filling, regardless of the QoS constraint [see Eq (9)]. On the other hand, over the time-domain, the optimal allocation heavily depends on QoS constraint [see time-varying water-level $\omega$]. In particular, when the QoS constraint is loose ($\beta \to 0$), $\omega$ converges to a constant. Thus, the temporal power allocation becomes water-filling as well. On the other hand, when the QoS constraint is stringent ($\beta \to \infty$), the temporal power allocation tries to "invert" the channel, under which the delay-limited capacity can be achieved [5].

### B. Suboptimal Allocation Algorithm

As mentioned above, the optimal frequency domain power allocation is always water-filling, no matter what the QoS constraint is. However, it is well known that with much lower computational complexity, *equal power allocation* across frequency domain is near optimal, especially at high SNR regime. This fact motivates us to study low-complexity suboptimal power allocation algorithm for implementation considerations. Specifically, instead of water-filling at frequency domain, our suboptimal algorithm simply applies uniform power distribution. At the time domain, we derive the optimal temporal allocation policy in the following.

The temporal power allocation problem $(P1')$ can be formulated as

$$(P1') \quad \min_{P_{tot} \geq 0} \mathbb{E}\left[P_{tot}\right] \tag{11}$$

subject to Eq. (5) with $R[i] = R$, where

$$R = \frac{T_f B}{N} \sum_{n=1}^{N} \log\left(1 + \frac{P_{tot}\gamma_n}{N}\right) \tag{12}$$

where $P_{tot} = \sum_{n=1}^{N} p_n = N p_n$ which denotes the total power allocated to a fading state $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \cdots, \gamma_N)^T$. It can be verified again that problem $(P1')$ is a convex optimization problem. We can readily derive this suboptimal policy by using the standard Lagrangian technique as described by the following proposition.

*Proposition 1:* The optimal temporal power allocation $P_{tot}$ is the unique solution of following equation:

$$\nu \prod_{m=1}^{N}\left(1 + \frac{P_{tot}\gamma_m}{N}\right)^{-\beta}\left\{\sum_{n=1}^{N}\gamma_n\left(1 + \frac{P_{tot}\gamma_n}{N}\right)^{-1}\right\} - 1 = 0, \tag{13}$$

if $\nu\sum_{n=1}^{N}\gamma_n > 1$, where $\nu$ is chosen to meet the QoS constraint Eq. (5). Otherwise, if $\nu\sum_{n=1}^{N}\gamma_n \le 1$, then we have $P_{tot} = 0$.

*Proof:* The proof follows by the standard Lagrangian approach, which is omitted for lack of space. ■

There is in general no closed-form solution for Eq. (13). However, since the left-hand side of Eq. (13) is monotonically decreasing convex function, numerical techniques such as bisection search or Newton's method can be efficiently applied to find the solution. For instance, using Newton's method, let $f(P_{tot})$ denote the left-hand side of Eq. (13), we can update $P_{tot}$ as follows:

$$P_{tot} = P_{tot} - \frac{f(P_{tot})}{f'(P_{tot})} \tag{14}$$

where

$$f'(P_{tot}) = -\frac{\nu}{N}\left\{\beta\left[\sum_{n=1}^{N}\gamma_n\left(1+\frac{P_{tot}\gamma_n}{N}\right)^{-1}\right]^2 + \left[\sum_{n=1}^{N}\gamma_n^2\left(1+\frac{P_{tot}\gamma_n}{N}\right)^{-2}\right]\right\} \times \prod_{m=1}^{N}\left(1+\frac{P_{tot}\gamma_m}{N}\right)^{-\beta}. \tag{15}$$

It is also worth noting that when the number of subchannels $N = 1$, the suboptimal algorithm that we developed above reduces to the optimal allocation policy derived in Theorem 1.

### C. *Performance of Constant-Power Rate-Control Policy*

For comparison purpose, we also investigate power-delay tradeoff performance of constant-power rate-control policy. Under such a policy, there is no transmit power adaptation over time and frequency domains, but the transmission rate is adapted according to the fading states (by using adaptive modulation and coding). Note that this policy is widely applied in current wireless communication systems and under active investigation for future wireless standards.

Following the similar idea to the previous sections, we can formulate the power-delay tradeoff problem $(P1'')$ as

$$(P1'') \quad \min_{\mathbb{E}[e^{-\theta R}] \le e^{-\theta A}} \overline{P} \tag{16}$$

where

$$R = \frac{T_f B}{N}\sum_{n=1}^{N}\log\left(1 + \frac{\overline{P}\gamma_n}{N}\right). \tag{17}$$

If we further assume that all the subchannels are i.i.d., then the QoS constraint can be derived as

$$\mathcal{G}_1\left(\theta, \overline{P}\right) \le e^{-\theta A} \tag{18}$$

where

$$\mathcal{G}_1\left(\theta, \overline{P}\right) \triangleq \mathbb{E}\left[e^{-\theta R}\right] = \left[\int_0^{+\infty}\left(1 + \frac{\overline{P}\gamma}{N}\right)^{-\beta} p_{\Gamma_n}(\gamma)d\gamma\right]^N \tag{19}$$

The integral in Eq. (19) can be calculated by using the techniques described in [22]. After some algebraic manipulations, we derive the closed-form expression for $\mathcal{G}_1\left(\theta, \overline{P}\right)$ as follows:

$$\mathcal{G}_1\left(\theta, \overline{P}\right) = \left[\left(\frac{N}{\overline{P}\overline{\gamma}}\right)\exp\left(\frac{N}{\overline{P}\overline{\gamma}}\right)E_\beta\left(\frac{N}{\overline{P}\overline{\gamma}}\right)\right]^N \tag{20}$$

where $E_\nu(\cdot)$ denotes the $\nu$th order exponential integral function [22]. Thus, we obtain the minimum average transmit power $\overline{P}$ by letting $\mathcal{G}_1\left(\theta, \overline{P}\right) = e^{-\theta A}$.

## IV. Power-Delay Tradeoff Over Multihop AF (Amplify-and-Forward) Relay Networks

### A. *Resource Allocation for Multihop AF Relay Networks*

In this section, we turn to studying the power-delay tradeoff over multihop AF relay OFDM networks. We focus on AF relay due to its low complexity compared to decode-and-forward (DF) relay. In AF protocol, the relay nodes simply amplify and then forward what they receive to the next hop. Let $L$ denote the number of hops in the relay networks, and $p_{\ell,n}$ with $\ell = 1, ..., L$ and $n = 1, ..., N$ denote the transmit power assigned to the $\ell$th hop at $n$th subchannel, respectively. Then, the achievable rate of AF protocol, denoted by $R_{AF}$, can be expressed as [23]

$$R_{AF} = \frac{T_f B}{2N}\sum_{n=1}^{N}\log\left(1+\left[\prod_{\ell=1}^{L}\left(1 + \frac{1}{2p_{\ell,n}\gamma_{\ell,n}}\right) - 1\right]^{-1}\right). \tag{21}$$

Note that in Eq. (21), each node works at half duplex mode, which implies that they cannot transmit and receive at the same time, but only for half of the frame duration, which results in a factor of 2 shown in Eq. (21).

To characterize the power-delay tradeoff, our power minimization problem $(P2)$ for multihop AF networks becomes

$$(P2) \quad \min_{\boldsymbol{p}\in\mathcal{C}_2}\sum_{\ell=1}^{L}\sum_{n=1}^{N}\mathbb{E}\left[p_{\ell,n}\right] \tag{22}$$

subject to Eq. (5) with $R[i] = R$, where $\mathcal{C}_2 \triangleq \{\boldsymbol{p} : p_{\ell,n} \ge 0, \text{ for all } \ell, n\}$ and $R$ is given by $R = R_{AF}$ in Eq. (21).

Problem $(P2)$ is still not easy to solve since the constraint is not convex on $\mathcal{C}_2$. To simplify the problem, we can make the following approximation at the high SNR regime:

$$R_{AF} \lessapprox \widetilde{R}_{AF} \triangleq \frac{T_f B}{2N}\sum_{n=1}^{N}\log\left(1+\left[\sum_{\ell=1}^{L}\frac{1}{2p_{\ell,n}\gamma_{\ell,n}}\right]^{-1}\right). \quad (23)$$

The difference between the actual value $R_{AF}$ and its approximated expression $\widetilde{R}_{AF}$ becomes negligible as the SNR increases. The approximation $\widetilde{R}_{AF}$ in Eq. (23) takes the advantages of mathematical tractability over $R_{AF}$ in Eq. (21). Specifically, $\widetilde{R}_{AF}$ is strictly concave on $\mathcal{C}_2$, which makes the resultant optimization much easier than the original problem $(P2)$. Furthermore, $\widetilde{R}_{AF}$ serves as a tight upper-bound for $R_{AF}$ at the high SNR regime, which results in a *lower-bound* for the minimum transmit power. On the other hand, by applying the obtained policy directly to $R_{AF}$, we can get an *upper-bound* for the minimum transmit power (since it is achievable). We will see by the numerical examples later that the upper-bound and lower-bound are close to each other, even at moderate SNR regime.

Replacing $R_{AF}$ by $\widetilde{R}_{AF}$, we derive the optimal resource allocation policy for multihop AF relay networks in the following theorem.

*Theorem 2:* Under optimal power allocation, $\widetilde{R}_{AF}$ given in Eq. (23) can be simplified as

$$\widetilde{R}_{AF} = \frac{T_f B}{2N}\sum_{n=1}^{N}\log\left(1+2\widetilde{p}_n\widetilde{\gamma}_n\right) \quad (24)$$

where $\widetilde{\gamma}_n = \left(\sum_{i=1}^{L}\sum_{j=1}^{L}\frac{1}{\sqrt{\gamma_{i,n}\gamma_{j,n}}}\right)^{-1}$ and $\widetilde{p}_n$ denotes the total power allocated to the $n$th subchannel, i.e., $\widetilde{p}_n = \sum_{\ell=1}^{L} p_{\ell,n}$. Thus, under optimal power allocation, the multihop AF relay problem is converted to a single hop problem, where Theorem 1 provides the optimal power allocation policy. Furthermore, the power allocated to each hop is given by

$$p_{\ell,n} = \frac{1}{\sqrt{\gamma_{\ell,n}}}\left(\sum_{i=1}^{L}\frac{1}{\sqrt{\gamma_{i,n}}}\right)^{-1}\widetilde{p}_n \quad (25)$$

*Proof:* The proof is provided in Appendix B. ■

## B. Performance of Constant-Power Rate-Control Policy

We also study the power-delay tradeoff of constant-power rate-control policy for multihop AF relay networks. In this section, we focus on a special case where the number of hops $L=2$. Then, the power-delay tradeoff problem $(P2')$ can be formulated as follows:

$$(P2') \quad \min_{\mathbb{E}[e^{-\theta R}]\leq e^{-\theta A}} \overline{P} \quad (26)$$

where $R = [T_f B/(2N)]\sum_{n=1}^{N}\log\left(1+\widehat{\gamma}_n\right)$ with $\widehat{\gamma}_n \triangleq [\overline{P}/(LN)]\left(\sum_{\ell=1}^{L}\gamma_{\ell,n}^{-1}\right)^{-1}$. We further assume that all the subchannels are i.i.d., then the pdf of $\widehat{\gamma}_n$, denoted by $p_{\widehat{\Gamma}_n}(\gamma)$, can be derived as [24]

$$p_{\widehat{\Gamma}_n}(\gamma) = \gamma e^{-\frac{4N\gamma}{\overline{P}\overline{\gamma}}}\left(\frac{4N}{\overline{P}\overline{\gamma}}\right)^2\left[K_0\left(\frac{4N\gamma}{\overline{P}\overline{\gamma}}\right)+K_1\left(\frac{4N\gamma}{\overline{P}\overline{\gamma}}\right)\right] \quad (27)$$

where $K_\nu(\cdot)$ denotes the $\nu$th order modified Bessel function of the second kind [22]. Thus, the QoS constraint can be written as follows:

$$\begin{aligned}\mathcal{G}_2\left(\theta,\overline{P}\right) \triangleq \mathbb{E}\left[e^{-\theta R}\right] &= \left[\int_0^{+\infty}(1+\gamma)^{-\frac{\beta}{2}}p_{\widehat{\Gamma}_n}(\gamma)d\gamma\right]^N \\ &\leq e^{-\theta A}. \end{aligned} \quad (28)$$

Again, the integral in Eq. (28) can be calculated by using the techniques described in [22]. After some tedious algebraic manipulations, we derive the closed-form expression for $\mathcal{G}_2\left(\theta,\overline{P}\right)$ given in Eq. (28) as follows:

$$\begin{aligned}\mathcal{G}_2\left(\theta,\overline{P}\right) = \Bigg\{\sqrt{\pi}\left[\Gamma\left(\frac{\beta}{2}\right)\right]^{-1}\Bigg[G_{2,3}^{3,1}\left(\frac{8N}{\overline{P}\overline{\gamma}}\middle| \begin{matrix}-1, \frac{1}{2} \\ -1, 1, \frac{\beta}{2}-2\end{matrix}\right) \\ + G_{2,3}^{3,1}\left(\frac{8N}{\overline{P}\overline{\gamma}}\middle| \begin{matrix}-1, \frac{1}{2} \\ 0, 0, \frac{\beta}{2}-2\end{matrix}\right)\Bigg]\left(\frac{4N}{\overline{P}\overline{\gamma}}\right)^2\Bigg\}^N\end{aligned} \quad (29)$$

where $\Gamma(\cdot)$ denotes the Gamma function, and $G_{p,q}^{m,n}(\cdot|\cdot)$ stands for the Meijer's $G$-function (due to the space limit, the readers are referred to [22] for the detailed descriptions on the Meijer's $G$-function). Thus, we obtain the minimum average transmit power $\overline{P}$ by letting $\mathcal{G}_2\left(\theta,\overline{P}\right) = e^{-\theta A}$.

# V. Power-Delay Tradeoff Over Cellular Wireless Networks

## A. Optimal Resource Allocation

In this section, we concentrate on the power-delay tradeoff over multiuser OFDM cellular networks. Without loss of generality, we focus on downlink transmission, where the base station transmit data streams to multiple mobile users in a dynamic time-division multiple access (TDMA) mode. Let $K$ denote the number of mobile users in the networks, $p_{k,n}$ with $k=1,...,K$ and $n=1,...,N$ denote the transmit power assigned to the $k$th user at $n$th subchannel, and $\alpha_k$ with $k=1,...,K$ denote the portion of frame assigned to the $k$th user, respectively. Then, in order to characterize the power-delay tradeoff, our power minimization problem $(P3)$ for multiuser cellular networks becomes

$$(P3) \quad \min_{(\boldsymbol{\alpha},\boldsymbol{p})\in\mathcal{C}_3}\sum_{k=1}^{K}\sum_{n=1}^{N}\mathbb{E}\left[p_{k,n}\right] \quad (30)$$

subject to

$$\mathbb{E}\left[e^{-\theta_k R_k}\right] \leq e^{-\theta A_k}, \text{ for all } k=1,2,...,K \quad (31)$$

where

$$R_k = \frac{T_f B}{N}\sum_{n=1}^{N}\alpha_k\log\left(1+\frac{p_{k,n}\gamma_{k,n}}{\alpha_k}\right) \quad (32)$$

and $\mathcal{C}_3 \triangleq \left\{(\boldsymbol{\alpha},\boldsymbol{p}) : p_{k,n}\geq 0, 0\leq\alpha_k\leq 1, \sum_{k=1}^{K}\alpha_k\leq 1\right\}$.

Problem $(P3)$ states that we need to find an optimal power and time-slot allocation policy (which can also be considered as a scheduling policy) to minimize the total average transmit power, such that each user's QoS constraint is satisfied. We can verify, but omit details for lack of space, that the objective function Eq. (30) and the constraint Eq. (31) are convex over the convex set $\mathcal{C}_3$. Based on convex optimization, we derive the optimal power and time-slot allocation policy, which are detailed in following theorem.

*Theorem 3:* Let $q_{k,n} = p_{k,n}/\alpha_k$ denote the instantaneous power assigned to the $k$th user at the $n$th subchannel. For the $k$th user at each fading state, let $\mathcal{N}_k$ denote the index set of subchannels which are allocated with nonzero power, and $N_k$ denote the cardinality of $\mathcal{N}_k$, respectively. Furthermore, let $\pi_k(\cdot)$ denote a permutation of subchannel gains in a descending order such that $\gamma_{k,\pi_k(1)} \geq \gamma_{k,\pi_k(2)} \geq \cdots \geq \gamma_{k,\pi_k(N_k)} \triangleq \gamma_k^{\min}$. Then, the optimal power allocation follows the classic water-filling algorithm and is determined by

$$\begin{cases} q_{k,\pi(n)} = q_k^{\min} + \left(\frac{1}{\gamma_k^{\min}} - \frac{1}{\gamma_{k,\pi(n)}}\right) & n \leq N_k \\ q_{k,\pi(n)} = 0 & n > N_k \end{cases} \tag{33}$$

where

$$q_k^{\min} = \left[\frac{\mu - \Sigma_k}{N_k W\left(\frac{\Pi_k(\mu-\Sigma_k)}{eN_k}\right)} - \frac{1}{\gamma_k^{\min}}\right]^+ \tag{34}$$

where $\Pi_k \triangleq \left(\prod_{n=1}^{N_k} \gamma_{k,\pi_k(n)}\right)^{\frac{1}{N_k}}$, $\Sigma_k \triangleq \sum_{n=1}^{N_k} 1/\gamma_{k,\pi_k(n)}$, and $W(\cdot)$ denotes the Lambert $W$-function [22]. On the other hand, the optimal time-slot allocation $\alpha_k$ for the $k$th user is given by

$$\alpha_k = \frac{1}{\beta_k N_k}\left[\frac{\log(\lambda_k \beta_k \Pi_k)}{1 + W\left(\frac{\Pi_k(\mu-\Sigma_k)}{eN_k}\right)} - 1\right]^+ \tag{35}$$

where $\beta_k = \theta_k T_f B/N$. After obtaining $q_{k,n}$ and $\alpha_k$, the optimal power allocation is determined by $p_{k,n} = q_{k,n}\alpha_k$. In Eqs. (34) and (35), $\lambda_k$ is chosen such that each user's QoS constraint is satisfied, and $\mu$ is chosen at each fading state such that $\sum_{k=1}^{K} \alpha_k = 1$. If at some fading states, $\lambda_k \beta_k \Pi_k < 1$ for all $k$, then $\sum_{k=1}^{K} \alpha_k \equiv 0$ (there is no $\mu$ satisfying $\sum_{k=1}^{K} \alpha_k = 1$). In this case, we have $p_{k,n} = 0$ for all $k$ and $n$, and the system is in an outage state.

*Proof:* The proof is provided in Appendix C. ■

Similar to the previous sections, we also derive the power-delay tradeoff for the constant-power rate-control policy, where each user is assigned resources as $p_{k,n} = (\overline{P}/N)/K$ and $\alpha_k = 1/K$. The performance of this scheme can be evaluated by the approach similar to that we described in Section III. Another alternative policy is optimal power allocation only with fixed time-slot allocation (i.e., $\alpha_k = 1/K$). We will compare these policies to the optimal one in the next section.

### B. Adaptive Resource Allocation Algorithm

The optimal resource allocation policy derived above depends on fading statistics through the Lagrangian multipliers $\{\lambda_k\}_{k=1}^K$. When the channel fading statistics are known, multipliers $\{\lambda_k\}_{k=1}^K$ can be numerically obtained in advance by extensive searching procedure, which is in general computational complex. Furthermore, in practice, the fading statistics may not be known *a priori*, which makes the problem of finding $\{\lambda_k\}_{k=1}^K$ even more complicated. Therefore, it is practically important to estimate $\{\lambda_k\}_{k=1}^K$ from the samples of the channel fading process. In this section, we propose an adaptive algorithm to estimate $\{\lambda_k\}_{k=1}^K$.

*Definition 1:* A *supgradient* of a function $f(\cdot)$ at a point $\boldsymbol{x}$ is any vector $\boldsymbol{\xi}$ such that

$$f(\widetilde{\boldsymbol{x}}) \leq f(\boldsymbol{x}) + (\widetilde{\boldsymbol{x}} - \boldsymbol{x})^T \boldsymbol{\xi} \tag{36}$$

for all $\widetilde{\boldsymbol{x}}$.

Our fundamental approach for deriving the proposed adaptive algorithm is summarized by the following theorem.

*Theorem 4:* For a given set of multipliers $\boldsymbol{\lambda}$, let $\boldsymbol{p}^*(\boldsymbol{\lambda})$, $\boldsymbol{\alpha}^*(\boldsymbol{\lambda})$, and $\mu^*(\boldsymbol{\lambda})$ denote the optimal resource allocation obtained from Theorem 3. Also define the following dual function:

$$\mathcal{J}_3(\boldsymbol{\lambda}) \triangleq \mathcal{L}_3(\boldsymbol{\alpha}^*(\boldsymbol{\lambda}), \boldsymbol{p}^*(\boldsymbol{\lambda}), \mu^*(\boldsymbol{\lambda}), \boldsymbol{\lambda}) \tag{37}$$

where $\mathcal{L}_3(\boldsymbol{\alpha}, \boldsymbol{p}, \mu, \boldsymbol{\lambda})$ is Lagrangian function given by Eq. (52) in Appendix C. Then, a supgradient of $\mathcal{J}_3(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda}$ is given by $\boldsymbol{\xi} \triangleq (\xi_1, \xi_2, ..., \xi_K)^T$, where

$$\xi_k = \mathbb{E}\left[\prod_{n=1}^{N}\left(1 + \frac{p_{k,n}^*(\boldsymbol{\lambda})\gamma_{k,n}}{\alpha_k^*(\boldsymbol{\lambda})}\right)^{-\alpha_k^*(\boldsymbol{\lambda})\beta_k}\right] - e^{-\theta A_k} \tag{38}$$

In addition, $\mathcal{J}_3(\boldsymbol{\lambda})$ is a concave function of $\boldsymbol{\lambda}$.

*Proof:* The proof is provided in Appendix D. ■

Theorem 4 suggests the following update for $\boldsymbol{\lambda}$ by a supgradient method [21]:

$$\boldsymbol{\lambda}[i+1] = \boldsymbol{\lambda}[i] + \epsilon[i]\boldsymbol{\xi}[i] \tag{39}$$

where $\epsilon[i]$ denotes the step size at time $i$. Under the sufficient condition on step size selection, the supgradient method is guaranteed to geometrically converge to the optimal $\boldsymbol{\lambda}^*$ for concave $\mathcal{J}_3(\boldsymbol{\lambda})$. It is also worth noting that at the points where $\mathcal{J}_3(\boldsymbol{\lambda})$ is differentiable, the supgradient reduces to gradient.

In practice, since the expectation in Eq. (38) is hard to obtain in advance, we may replace it by time-average due to the ergodicity as follows:

$$\xi_k[i] \approx \frac{1}{i}\sum_{j=1}^{i}\left[\prod_{n=1}^{N}\left(1 + \frac{p_{k,n}^*(\boldsymbol{\lambda}[j])\gamma_{k,n}[j]}{\alpha_k^*(\boldsymbol{\lambda}[j])}\right)^{-\alpha_k^*(\boldsymbol{\lambda}[j])\beta_k}\right] - e^{-\theta A_k}. \tag{40}$$

In summary, for each $\boldsymbol{\lambda}[i]$, we obtain resource allocation $(\boldsymbol{\alpha}[i], \boldsymbol{p}[i])$ by Theorem 3. Then, the adaptive algorithm updates $\boldsymbol{\lambda}[i]$ by Eq. (39), where the supgradient is replaced by its time-averaged version Eq. (40). Under the sufficient condition on step size selection, the adaptive algorithm geometrically converges to the neighborhood of the optimal $\boldsymbol{\lambda}$.

## VI. Numerical and Simulation Evaluations

We evaluate the power-delay tradeoff based on our proposed resource allocation by simulations and numerical methods. As a simple example of OFDM-based communications systems, we simulate a multicarrier system with $N$ i.i.d. subchannels. Also, we set the product $T_f B = 1$ such that $\beta = \theta/N$ for convenience. The other system parameters are detailed, respectively, in each of the figures.

Figure 1 plots the optimal power-delay tradeoff over point-to-point OFDM link with different number $N$ of subchannels. We can observe from Fig. 1 that for a single channel

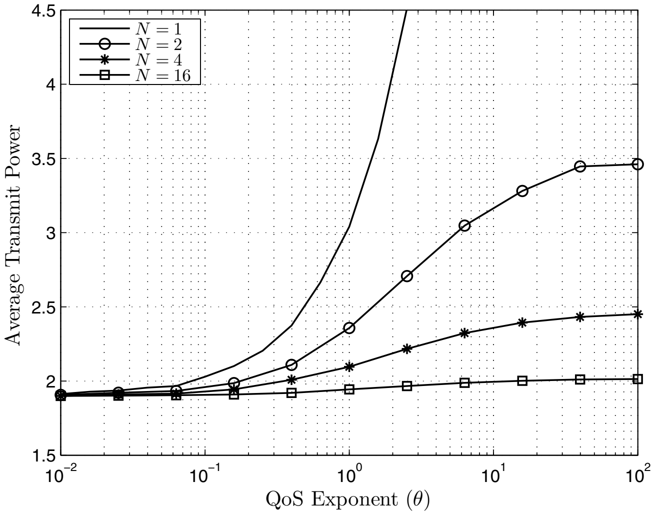



Fig. 1. The optimal power-delay tradeoff over point-to-point OFDM link, where $N$ subchannels are modeled as i.i.d. Rayleigh fading with $\overline{\gamma} = 0$ dB.

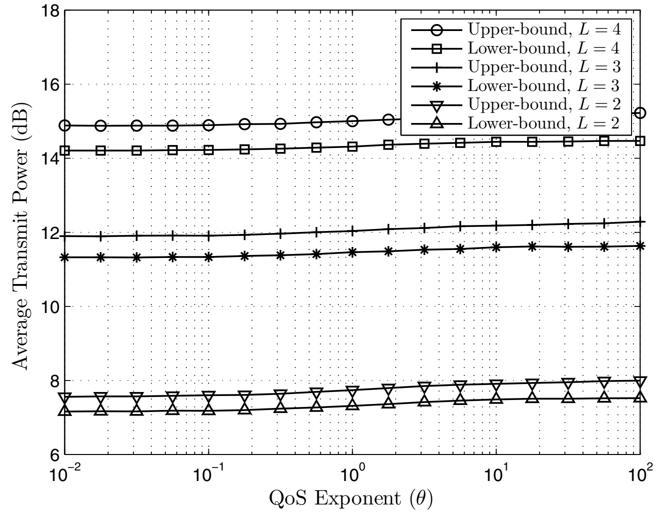



Fig. 3. The upper-bound and lower-bound for the power-delay tradeoff over multihop AF networks, where $N = 8$ subchannels are modeled as i.i.d. Rayleigh fading with $\overline{\gamma} = 0$ dB.

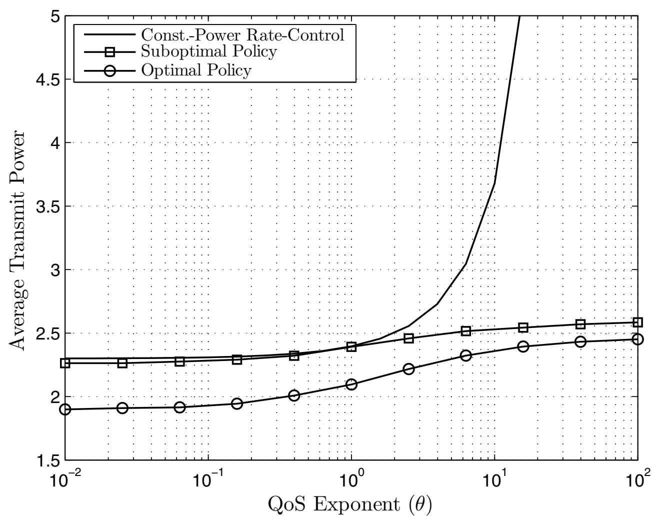



Fig. 2. The power-delay tradeoff comparisons between different policies over point-to-point OFDM link, where $N = 8$ subchannels are modeled as i.i.d. Rayleigh fading with $\overline{\gamma} = 0$ dB.

system, the average power cannot be upper-bounded as the QoS exponent increases, which is consistent with information theoretic result that over the Rayleigh fading channel, channel inversion scheme (i.e., the optimal policy as $\theta \to \infty$) requires infinite power. On the other hand, as the number $N$ of subchannels increases, the average transmit power converges to a finite value. The larger the number $N$ given, the lower the transmit power required. When $N = 16$, the average transmit power virtually invariant as $\theta$ increases, which implies the system can provide stringent QoS with almost the same power consumption as that with loose QoS. Thus, the power-control is virtually independent of delay-QoS constraint.

Figure 2 shows power-delay tradeoff comparisons between different policies over point-to-point OFDM link, where the number of subchannels $N = 8$. From Fig. 2 we can see that our optimal policy significantly outperforms the constant-power rate-control policy, especially when the delay constraint is stringent. On the other hand, our suboptimal policy performs similar to constant-power policy for small $\theta$, but similar to optimal policy for large $\theta$. The performance loss compared to optimal policy comes from uniform power allocation across subchannels, which however gains the benefit of the low computational complexity.

Figure 3 illustrates the upper- and lower-bounds for the optimal power-delay tradeoff over multihop AF networks. In the simulations, we assume that all the nodes are placed on a straight line with equal distance. The distance between each adjacent nodes are normalized to a unit such that the average channel gain $\overline{\gamma} = 0$ dB. Thus, over an $L$-hop network, the distance between original sender and final receiver is $L$. From Fig. 3 we can observe that the upper-bounds and corresponding lower-bounds are close to each other, where the differences are all within 1 dB for different values of $\theta$. Also, with $N = 8$ subchannels, the total average transmit power only increases slightly as $\theta$ increases, which demonstrates the advantages of using multichannels for relay networks.

Figure 4 compares the power-delay tradeoff using optimal policy with other resource allocation policies, including constant-power rate-control AF protocol, optimal direct transmission, and constant-power direct transmission. In the simulations, we assume that the path loss exponent is 3 for direct transmission. Fig. 4 shows that the optimal AF relay significantly outperforms the direct transmission. The advantage is about 5 dB compared with the optimal direct transmission. On the other hand, the constant-power policies cannot support stringent delay applications. The gap between the optimal AF and constant-power policies cannot be upper-bounded as $\theta$ increases.

Figure 5 shows power consumption for the multiuser networks with different number of users and different QoS constraints. We assume that the users are homogeneous, i.e., all users have the same QoS constraint and the same channel statistics. Furthermore, for a fair comparison, we assume that

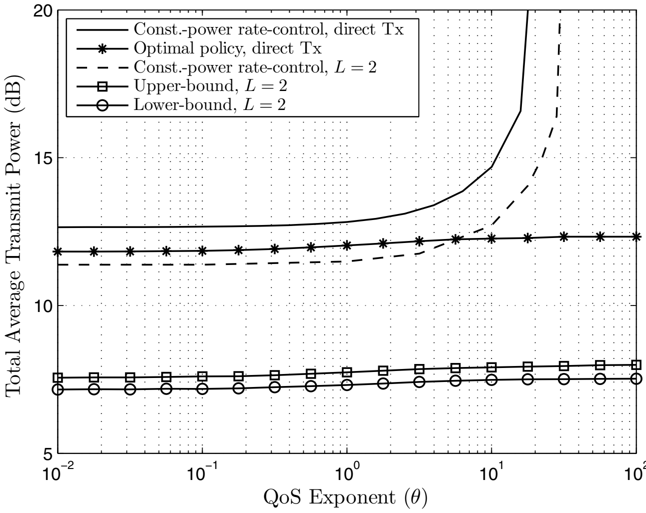



Fig. 4. The power-delay tradeoff comparisons between point-to-point link and multihop AF relay, when using different policies. The number $N = 8$ of subchannels are modeled as i.i.d. Rayleigh fading with $\overline{\gamma} = 0$ dB. The path loss exponent is set to be 3.

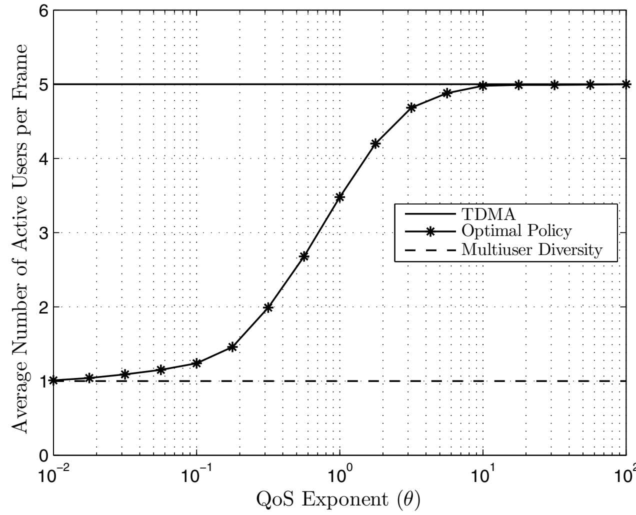



Fig. 6. The average number of scheduled users per frame under different resource allocation policies. The number of subchannels $N = 2$ and the number of users $K = 5$. The QoS constraint of the effective bandwidth is same for each user and is set to be $\mathcal{A}(\theta) = 1/K$ nats/sec/Hz.

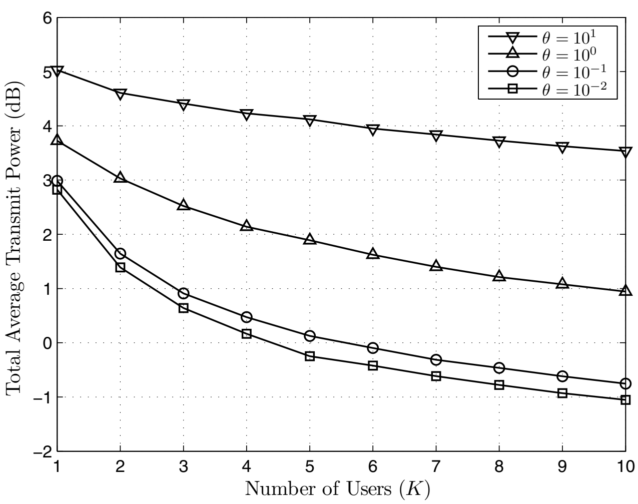



Fig. 5. The total average power consumption as the number $K$ of users increases. The number of subchannels $N = 2$. The QoS constraint of the effective bandwidth is same for each user and is set to be $\mathcal{A}(\theta) = 1/K$ nats/sec/Hz.

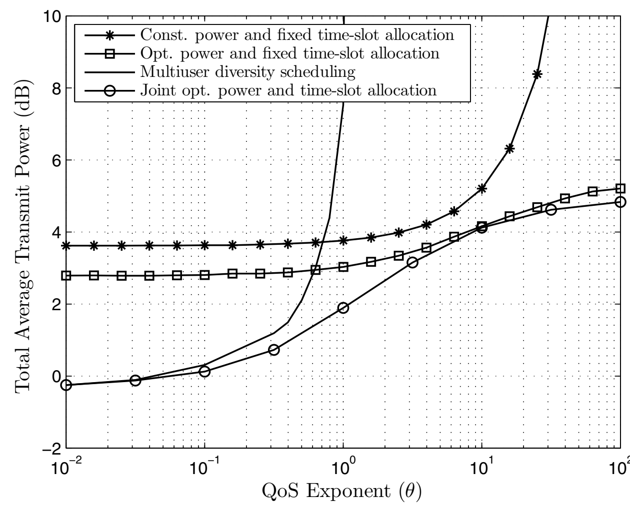



Fig. 7. The power-delay tradeoff comparisons between different resource allocation policies for multiuser cellular networks. The number of subchannels $N = 2$ and the number of users $K = 5$. The QoS constraint of the effective bandwidth is same for each user and is set to be $\mathcal{A}(\theta) = 1/K$ nats/sec/Hz.

each user requests an effective bandwidth of $\mathcal{A}(\theta) = 1/K$, thus the sum effective bandwidth constraint $\sum_{k=1}^{K} \mathcal{A}(\theta) = 1$ is the same regardless of the number of users. From Fig. 5 we observe that when the QoS constraint is loose, there exists a significant multiuser diversity gain [25] as the number $K$ of mobile users increases. For example, when $\theta = 10^{-2}$, the power consumption with $K = 10$ users is 4 dB less than that with $K = 1$ user. The reason is that the system can always schedule the user with best channel conditions. However, when the QoS is stringent, the multiuser diversity gain diminishes. When $\theta = 10^1$, the gain is only less than 1.5 dB for $K = 10$ compared to the case with $K = 1$. The reason is that with stringent delay constraint, we need to schedule all the users at each frame, no matter what the channel conditions are.

To better understand how delay constraint affects scheduling policy performance and selections, Fig. 6 shows the average number of scheduled users per frame. We can see that when the delay constraint is loose, the average number of active users converges to 1, thus multiuser diversity scheduling policy applies. On the other hand, when the delay constraint is stringent, the optimal policy always schedules all the users at each frame, thus TDMA policy is near optimal. As the delay constraint varies from loose to stringent, the optimal policy changes from multiuser diversity scheduling to TDMA.

Finally, Fig. 7 compares the power-delay tradeoff of multiuser networks under different resource allocation policies. The multiuser diversity scheduling selects the best user at each frame and performs the water-filling power allocation. The fixed time-slot allocation assigns each user with $\alpha_k = 1/K$

of the time-slot. From Fig. 7, we can observe that our proposed joint optimal power and time-slot allocation scheme outperforms all the other resource allocation policies. Fig. 7 also shows that when the delay constraint is stringent, the policy of optimal power allocation with fixed time-slot allocation is close to the joint optimal power and time-slot allocation, which again shows that TDMA scheduling is near optimal for large $\theta$. In this case, power allocation has more significant impact on the power-delay tradeoff performance than scheduling.

## VII. Conclusions

We proposed and analyzed the resource allocation to achieve the optimal power-delay tradeoff over various wireless networks. As the QoS exponent $\theta$ increases from zero to infinity, the optimal power-delay tradeoff curve characterizes how much power is necessary to guarantee a given QoS requirement. For each network scenario, we derived the optimal resource allocation policies. Our results indicate that over an OFDM communication system, it is possible to achieve stringent QoS guarantee with little power increase compared to that with loose delay constraint. Compared to existing resource allocation policies, which are widely employed in current communications systems, our proposed resource allocation policies significantly reduce the average transmit power under the stringent delay constraints.

## Appendix A
## Proof of Theorem 1

*Proof:* We construct the Lagrangian function as follows:

$$\mathcal{L}_1(\boldsymbol{p},\lambda)=\sum_{n=1}^{N}\mathbb{E}[p_n]+\lambda\left(\mathbb{E}\left[\prod_{n=1}^{N}(1+p_n\gamma_n)^{-\beta}\right]-e^{-\theta A}\right) \tag{41}$$

where $\beta=\theta T_f B/N$ and $\lambda$ denotes the Lagrangian multiplier corresponding to the QoS constraint. Taking the partial derivative of $\mathcal{L}_1(\boldsymbol{p},\lambda)$ with respect to (w.r.t.) $p_n$, we have

$$\begin{aligned}\frac{\partial\mathcal{L}_1(\boldsymbol{p},\lambda)}{\partial p_n} &= 1-\lambda\beta\gamma_n(1+p_n\gamma_n)^{-1}\prod_{m=1}^{N}(1+p_m\gamma_m)^{-\beta}\\ &= 1-\lambda\beta\gamma_n(1+p_n\gamma_n)^{-1}\Phi(\boldsymbol{p}),\end{aligned} \tag{42}$$

where

$$\Phi(\boldsymbol{p})\triangleq\prod_{m=1}^{N}(1+p_m\gamma_m)^{-\beta}, \tag{43}$$

which is a scalar function of the vector $\boldsymbol{p}$. Then, by the Karush-Kuhn-Tucker (KKT) condition [21], we have

$$\begin{cases}\dfrac{\partial\mathcal{L}_1(\boldsymbol{p},\lambda)}{\partial p_n} =1-\dfrac{\lambda\beta\gamma_n\Phi(\boldsymbol{p})}{(1+p_n\gamma_n)}=0, & \text{if } p_n>0;\\ \left.\dfrac{\partial\mathcal{L}_1(\boldsymbol{p},\lambda)}{\partial p_n}\right|_{p_n=0} =1-\lambda\beta\gamma_n\Phi(\boldsymbol{p})\geq 0, & \text{if } p_n=0;\end{cases} \tag{44}$$

Assume that there are $M\ (\leq N)$ channels satisfying $p_n>0$. Next, by contradiction we will show the claim that the SNRs of these $M$ channels are $\gamma_{\pi(1)},\gamma_{\pi(2)},\dots,\gamma_{\pi(M)}$, respectively, which are the $M$ largest SNRs. If this claim does not hold, there must exist a certain $m\in\{1,2,\dots,N\}$ and a certain $j\in\{1,2,\dots,N\}$, where $m\neq j$, such that $p_m>0$ and $p_j=0$, but

$$\gamma_m<\gamma_j. \tag{45}$$

However, based on Eq. (44), we can derive

$$\begin{cases}\gamma_m = \dfrac{1+p_m\gamma_m}{\lambda\beta\Phi(\boldsymbol{p})}>\dfrac{1}{\lambda\beta\Phi(\boldsymbol{p})}, & \text{for } p_m>0;\\ \gamma_j \leq \dfrac{1}{\lambda\beta\Phi(\boldsymbol{p})}, & \text{for } p_j=0,\end{cases} \tag{46}$$

where $\Phi(\boldsymbol{p})$ is defined in Eq. (43). Equation (46) suggests that

$$\gamma_m>\gamma_j, \tag{47}$$

contradicting the assumption given in Eq. (45). Therefore, by contradiction we can see that all the $M$ channels with non-zero powers have the SNRs equal to $\gamma_{\pi(1)},\gamma_{\pi(2)},\dots,\gamma_{\pi(M)}$, respectively.

Then, solving the $M$ equations based on Eq. (44), and considering the boundary condition $p_m\geq 0$, we can obtain Eq. (9), where $\nu=(\lambda\beta)^{\frac{1}{N\beta+1}}$. ■

## Appendix B
## Proof of Theorem 2

*Proof:* Construct the Lagrangian function as follows:

$$\begin{aligned}\mathcal{L}_2(\boldsymbol{p},\lambda) &= \lambda\mathbb{E}\left[\prod_{n=1}^{N}\left(1+\left[\sum_{\ell=1}^{L}\frac{1}{2p_{\ell,n}\gamma_{\ell,n}}\right]^{-1}\right)^{-\frac{\beta}{2}}\right]\\ &\quad -\lambda e^{-\theta A}+\sum_{\ell=1}^{L}\sum_{n=1}^{N}\mathbb{E}\left[p_{\ell,n}\right]\end{aligned} \tag{48}$$

where $\lambda$ denotes the Lagrangian multiplier corresponding to the QoS constraint. For all $p_{\ell,n}>0$, we have

$$\begin{aligned}\frac{\partial\mathcal{L}_2(\boldsymbol{p},\lambda)}{\partial p_{\ell,n}} &= 1-\lambda\beta\gamma_{\ell,n}\prod_{m=1}^{N}\left(1+\left[\sum_{\ell=1}^{L}\frac{1}{2p_{\ell,m}\gamma_{\ell,m}}\right]^{-1}\right)^{-\frac{\beta}{2}}\\ &\quad\times\left(1+\left[\sum_{\ell=1}^{L}\frac{1}{2p_{\ell,n}\gamma_{\ell,n}}\right]^{-1}\right)^{-1}\\ &\quad\times\left(1+p_{\ell,n}\gamma_{\ell,n}\sum_{i=1,i\neq\ell}^{L}\frac{1}{p_{i,n}\gamma_{i,n}}\right)^{-2}\\ &= 0.\end{aligned} \tag{49}$$

Letting $\Delta_n=\sum_{\ell=1}^{L}\left(p_{\ell,n}\gamma_{\ell,n}\right)^{-1}$, we have

$$\frac{\lambda\beta}{\gamma_{\ell,n}(p_{\ell,n}\Delta_n)^2}\prod_{m=1}^{N}\left(1+2\Delta_m^{-1}\right)^{-\frac{\beta}{2}}\left(1+2\Delta_n^{-1}\right)^{-1}=1 \tag{50}$$

for all $\ell=1,2,...,L$. Solving these $L$ equations, we obtain

$$p_{i,n}=\sqrt{\frac{\gamma_{j,n}}{\gamma_{i,n}}}p_{j,n}. \tag{51}$$

Substituting Eq. (51) into Eq. (23) with some algebraic manipulations, we can obtain Eq. (24). The proof follows. ■

## APPENDIX C
## PROOF OF THEOREM 3

*Proof:* Construct the Lagrangian function as follows:

$$\mathcal{L}_3(\boldsymbol{\alpha},\boldsymbol{p},\mu,\boldsymbol{\lambda}) = \sum_{k=1}^{K}\sum_{n=1}^{N}\mathbb{E}\left[p_{k,n}\right]+\mu\left(\sum_{k=1}^{K}\alpha_k-1\right) + \sum_{k=1}^{K}\lambda_k\left(\mathbb{E}\left[\prod_{n=1}^{N}\left(1+\frac{p_{k,n}\gamma_{k,n}}{\alpha_k}\right)^{-\alpha_k\beta_k}\right]-e^{-\theta A_k}\right) \quad (52)$$

where $\mu$ denotes the Lagrangian multiplier corresponding to time-slot constraint, and $\lambda_k$ denotes the Lagrangian multipliers corresponding to the $k$th user's QoS constraint. For all $\alpha_k > 0$, we have

$$\frac{\partial\mathcal{L}_3(\boldsymbol{\alpha},\boldsymbol{p},\mu,\boldsymbol{\lambda})}{\partial\alpha_k} = \mu-\lambda_k\beta_k\prod_{m=1}^{N}\left(1+\frac{p_{k,m}\gamma_{k,m}}{\alpha_k}\right)^{-\alpha_k\beta_k} \times\left\{\sum_{n=1}^{N}\left[\log\left(1+\frac{p_{k,n}\gamma_{k,n}}{\alpha_k}\right)-\frac{p_{k,n}\gamma_{k,n}}{\alpha_k+p_{k,n}\gamma_{k,n}}\right]\right\}=0. \quad (53)$$

Otherwise, if $\alpha_k = 0$, then $p_{k,n} = 0$ for all $n$. Similarly, for all $p_{k,n} > 0$, we have

$$\frac{\partial\mathcal{L}_3(\boldsymbol{\alpha},\boldsymbol{p},\mu,\boldsymbol{\lambda})}{\partial p_{k,n}} = 1-\lambda_k\beta_k\gamma_{k,n}\prod_{m=1}^{N}\left(1+\frac{p_{k,m}\gamma_{k,m}}{\alpha_k}\right)^{-\alpha_k\beta_k} \times\left(1+\frac{p_{k,n}\gamma_{k,n}}{\alpha_k}\right)^{-1}=0 \quad (54)$$

In general, solving $\alpha_k$ and $p_{k,n}$ from Eqs. (53) and (54) gives us the optimal allocation policy. However, it is difficult to directly solve Eqs. (53) and (54). Thus, we need to find an alternative way to solve this problem.

Let us define $q_{k,n} = p_{k,n}/\alpha_k$ as the instantaneous power assigned to the $k$th user at the $n$th subchannel. Then, substituting $q_{k,n}$ into Eqs. (53) and (54), we obtain

$$\prod_{m=1}^{N}\left(1+q_{k,m}\gamma_{k,m}\right)^{\alpha_k\beta_k}=\frac{\lambda_k\beta_k}{\mu} \times\left\{\sum_{n=1}^{N}\left[\log(1+q_{k,n}\gamma_{k,n})-\frac{q_{k,n}\gamma_{k,n}}{1+q_{k,n}\gamma_{k,n}}\right]\right\} \quad (55)$$

and

$$(1+q_{k,n}\gamma_{k,n})\prod_{m=1}^{N}\left(1+q_{k,m}\gamma_{k,m}\right)^{\alpha_k\beta_k}=\lambda_k\beta_k\gamma_{k,n}, \quad (56)$$

respectively. Dividing Eq. (56) by Eq. (55), we obtain

$$1+q_{k,n}\gamma_{k,n} = \mu\gamma_{k,n} \times\left\{\sum_{m=1}^{N}\left[\log(1+q_{k,m}\gamma_{k,m})-\frac{q_{k,m}\gamma_{k,m}}{1+q_{k,m}\gamma_{k,m}}\right]\right\}^{-1} \quad (57)$$

for all $n = 1, 2, ..., N$. Solving these $N$ equations as described by Eq. (57), we obtain the following expression:

$$q_{k,n}=q_{k,m}+\left(\frac{1}{\gamma_{k,m}}-\frac{1}{\gamma_{k,n}}\right), \quad \forall m,n\in\mathcal{N}_k \quad (58)$$

where $\mathcal{N}_k$ denotes the index set of subchannels for the $k$th user which is allocated with nonzero power. It is clear from Eq. (58) that the power allocation follows frequency-domain water-filling formula. Expressing all $q_{k,n}$ in terms of $q_k^{\min}$, we can then obtain Eq. (33). Substituting Eq. (33) into Eq. (57), we can derive

$$N_k\left(1+q_k^{\min}\gamma_k^{\min}\right)\left\{\log\left(1+q_k^{\min}\gamma_k^{\min}\right) + \left[\log\left(\frac{\Pi_k}{\gamma_k^{\min}}\right)-1\right]\right\}=\gamma_k^{\min}(\mu-\Sigma_k). \quad (59)$$

Solving Eq. (59), we get the optimal power allocation for $q_k^{\min}$ as expressed in Eq. (34). On the other hand, solving Eq. (55), we get solution of $\alpha_k$ as follows:

$$\alpha_k=\left[\left(\beta_k\sum_{n=1}^{N}\log(1+q_{k,n}\gamma_{k,n})\right)^{-1}\left\{\log\left(\frac{\lambda_k\beta_k}{\mu}\right) + \log\left(\sum_{n=1}^{N}\left[\log(1+q_{k,n}\gamma_{k,n})-\frac{q_{k,n}\gamma_{k,n}}{1+q_{k,n}\gamma_{k,n}}\right]\right)\right\}\right]^{+}. \quad (60)$$

Plugging Eq. (34) into Eq. (60) with some tedious manipulations, we obtain Eq. (35). This completes the proof. ■

## APPENDIX D
## PROOF OF THEOREM 4

*Proof:* To simplify the presentation, let us re-write the original minimization problem $(P3)$ in a compact form as follows:

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) \quad (61)$$

$$\text{s.t.}\quad \boldsymbol{g}(\boldsymbol{x})\le\boldsymbol{0} \quad (62)$$

$$h(\boldsymbol{x})\le 0 \quad (63)$$

where $\boldsymbol{x} \triangleq (\boldsymbol{\alpha}^T,\boldsymbol{p}^T)^T$, the objective $f(\boldsymbol{x})$ corresponds to power minimization Eq. (30), the first constraint $\boldsymbol{g}(\boldsymbol{x})$ corresponds to the QoS constraints in Eq. (31), and the second constraint $h(\boldsymbol{x})$ corresponds to time-slot constraint $\sum_{k=1}^{K}\alpha_k-1\le 0$. The Lagrangian function given in Eq. (52) can be rewritten as

$$\mathcal{L}_3(\boldsymbol{x},\mu,\boldsymbol{\lambda})=f(\boldsymbol{x})+\mu h(\boldsymbol{x})+\boldsymbol{\lambda}^T\boldsymbol{g}(\boldsymbol{x}). \quad (64)$$

Let $\boldsymbol{x}^*(\boldsymbol{\lambda})$ and $\mu^*(\boldsymbol{\lambda})$ denote the optimal values of $\boldsymbol{x}$ and $\mu$ for a given $\boldsymbol{\lambda}$, which can be readily obtained by Theorem 3. Then, the Lagrangian dual problem can be expressed as

$$\mathcal{J}_3(\boldsymbol{\lambda})=\inf_{\boldsymbol{x}}\mathcal{L}_3(\boldsymbol{x},\mu^*(\boldsymbol{\lambda}),\boldsymbol{\lambda})=\mathcal{L}_3(\boldsymbol{x}^*(\boldsymbol{\lambda}),\mu^*(\boldsymbol{\lambda}),\boldsymbol{\lambda}) \quad (65)$$

Then, for all $\widetilde{\boldsymbol{\lambda}}$, we have

$$\begin{aligned}\mathcal{J}_3(\widetilde{\boldsymbol{\lambda}}) &= \inf_{\boldsymbol{x}}\mathcal{L}_3(\boldsymbol{x},\mu^*(\widetilde{\boldsymbol{\lambda}}),\widetilde{\boldsymbol{\lambda}})\\ &\le \mathcal{L}_3(\boldsymbol{x}^*(\boldsymbol{\lambda}),\mu^*(\widetilde{\boldsymbol{\lambda}}),\widetilde{\boldsymbol{\lambda}})\\ &= f(\boldsymbol{x}^*(\boldsymbol{\lambda}))+\mu^*(\widetilde{\boldsymbol{\lambda}})h(\boldsymbol{x}^*(\boldsymbol{\lambda}))+\widetilde{\boldsymbol{\lambda}}^T\boldsymbol{g}(\boldsymbol{x}^*(\boldsymbol{\lambda}))\\ &\le f(\boldsymbol{x}^*(\boldsymbol{\lambda}))+\widetilde{\boldsymbol{\lambda}}^T\boldsymbol{g}(\boldsymbol{x}^*(\boldsymbol{\lambda})) \quad (66)\\ &= f(\boldsymbol{x}^*(\boldsymbol{\lambda}))+\boldsymbol{\lambda}^T\boldsymbol{g}(\boldsymbol{x}^*(\boldsymbol{\lambda}))+(\widetilde{\boldsymbol{\lambda}}-\boldsymbol{\lambda})^T\boldsymbol{g}(\boldsymbol{x}^*(\boldsymbol{\lambda}))\\ &= \mathcal{J}_3(\boldsymbol{\lambda})+(\widetilde{\boldsymbol{\lambda}}-\boldsymbol{\lambda})^T\boldsymbol{g}(\boldsymbol{x}^*(\boldsymbol{\lambda})) \quad (67)\end{aligned}$$

where Eq. (66) is due to the constraint $h(\boldsymbol{x}^*(\boldsymbol{\lambda}))\le 0$ and Eq. (67) is due to the Complementary Slackness condition:

$$\mu^*(\boldsymbol{\lambda})h(\boldsymbol{x}^*(\boldsymbol{\lambda}))\equiv 0. \quad (68)$$

Thus, from Definition 1, $\boldsymbol{g}(\boldsymbol{x}^*(\boldsymbol{\lambda}))=\boldsymbol{\xi}$ is a supgradient of $\mathcal{J}_3(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda}$. Furthermore, since $\mathcal{J}_3(\boldsymbol{\lambda})$ has supgradient at every point $\boldsymbol{\lambda}$, $\mathcal{J}_3(\boldsymbol{\lambda})$ is a concave function. The proof follows. ■

## REFERENCES


[1] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Automatic Contorl*, vol. 39, no. 5, pp. 913–931, May 1994.

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

[3] J. Tang and X. Zhang, "QoS-driven power and rate adaptation over wireless links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3058–3068, Aug. 2007.

[4] J. Tang and X. Zhang, "Cross-layer-model based adaptive resource allocation for statistical QoS guarantees in mobile wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2318–2328, June 2008.

[5] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1468–1489, July 1999.

[6] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with mulitple antennas," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1273–1289, May 2001.

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

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

[9] R. A. Berry, "Optimal power-delay trade-offs in fading channels–small delay asymptotics," submitted to *IEEE Trans. Inf. Theory*, 2006.

[10] J. G. Kim and M. Krunz, "Bandwidth allocation in wireless networks with gauranteed packet-loss performance," *IEEE/ACM Trans. Networking*, vol. 8, pp. 337–349, June 2000.

[11] D. Wu and R. Negi, "Downlink scheduling in a cellular network for quality-of-service assurance," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 1547–1557, Sep. 2004.

[12] M. J. Neely, "Optimal energy and delay tradeoffs for multi-user wireless downlinks," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3095–3113, Sep. 2007.

[13] N. Salodkar, A. Bhorkar, A. Karandikar, and V. S. Borkar, "An on-line learning algorithm for energy efficient delay constrained scheduling over a fading channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 732–742, May 2008.

[14] F. Fu and M. van der Schaar, "Decomposition principles and online learning in cross-layer optimization for delay-sensitive applications," *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1401–1415, Mar. 2010.

[15] M. Zafer and E. Modiano, "Minimum energy transmission over a wireless channel with deadline and power constraints," *IEEE Trans. Automatic Contorl*, vol. 54, no. 12, pp. 2841–2852, Dec. 2009.

[16] W. Chen, M. J. Neely, and U. Mitra, "Energy-efficient transmissions with individual packet delay constraints," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2090–2109, Sep. 2008.

[17] M. Zafer and E. Modiano, "A calculus approach to energy-efficient data transmission with quality-of-service constraints," *IEEE/ACM Trans. Networking*, vol. 17, no. 13, pp. 898–911, June 2009.

[18] M. Goyal, A. Kumar, and V. Sharma, "Power constrained and delay optimal policies for scheduling transmission over a fading channel," in *Proc. 2003 IEEE INFOCOM*, pp. 311–320.

[19] D. Wu and R. Negi, "Utilizing multiuser diversity for efficient support of quality of service over a fading channel," *IEEE Trans. Veh. Technol.*, vol. 54, no. 3, pp. 1198–1206, May 2005.

[20] F. Kelly, S. Zachary, and I. Ziedins, "Stochastic networks: theory and applications," *Royal Statistical Society Lecture Notes Series*, vol. 4, pp. 141–168, 1996.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[22] I. S. Gradshteyn and I.M. Ryzhik, *Table of Integral, Series, and Products*. Academic Press, 1992.

[23] M. O. Hasna and M.-S. Alouini, "Outage probability of multihop transmission over nakagami fading channels," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 216–218, May 2003.

[24] M. O. Hasna and M.-S. Alouini, "Performance analysis of two-hop relayed transmissions over Rayleigh fading channels," in *Proc. 2002 IEEE VTC – Fall*, pp. 1992–1996.

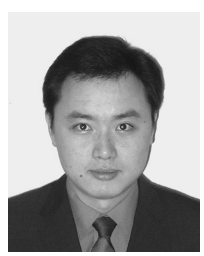
[25] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. 2002 IEEE ICC*, pp. 331–335.



**Xi Zhang** (S'89-SM'98) received the B.S. and M.S. degrees from Xidian University, Xi'an, China, the M.S. degree from Lehigh University, Bethlehem, PA, all in electrical engineering and computer science, and the Ph.D. degree in electrical engineering and computer science (Electrical Engineering-Systems) from The University of Michigan, Ann Arbor.

He is currently an Associate Professor and the Founding Director of the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station. He was a research fellow with the School of Electrical Engineering, University of Technology, Sydney, Australia, and the Department of Electrical and Computer Engineering, James Cook University, Australia. He was with the Networks and Distributed Systems Research Department, AT&T Bell Laboratories, Murray Hill, New Jersey, and AT&T Laboratories Research, Florham Park, New Jersey, in 1997. He has published more than 250 research papers on wireless networks and communications systems, network protocol design and modeling, statistical communications, random signal processing, information theory, and control theory and systems. He received the U.S. National Science Foundation CAREER Award in 2004 for his research in the areas of mobile wireless and multicast networking and systems. He is an IEEE Distinguished Lecturer for both IEEE Communications Society and IEEE Vehicular Technology Society. He received Best Paper Awards at IEEE GLOBECOM 2007, IEEE GLOBECOM 2009, and IEEE WCNC 2010, respectively. He also received a TEES Select Young Faculty Award for Excellence in Research Performance from the Dwight Look College of Engineering at Texas A&M University, College Station, in 2006.

Prof. Zhang is serving or has served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, twice as a Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS for two special issues on "Broadband Wireless Communications for High Speed Vehicles" and "Wireless Video Transmissions," an Associate Editor for IEEE COMMUNICATIONS LETTERS, a Lead Guest Editor for *IEEE Communications Magazine* for the special issue on "Advances on Cooperative Wireless Networking" and a Guest Editor for *IEEE Wireless Communications Magazine* for the special issue on "Next Generation CDMA versus OFDMA for 4G Wireless Applications", an Editor for Wiley's JOURNAL ON WIRELESS COMMUNICATIONS AND MOBILE COMPUTING, JOURNAL OF COMPUTER SYSTEMS, NETWORKING, AND COMMUNICATIONS, and Wiley's JOURNAL ON SECURITY AND COMMUNICATIONS NETWORKS, and an Area Editor for Elsevier's JOURNAL ON COMPUTER COMMUNICATIONS, among others. He is serving or has served as the TPC Chair for IEEE GLOBECOM 2011, TPC Vice-Chair IEEE INFOCOM 2010, TPC Area Chair for IEEE INFOCOM 2012, Panel/Demo/Poster Chair for ACM MobiCom 2011, General Vice-Chair for IEEE WCNC 2013, Panel/Demo/Poster Chair for ACM MobiCom 2011, and TPC/General Chair for numerous other IEEE/ACM conferences, symposia, and workshops.


**Jia Tang** (S'03) received the B.S. degree in Electrical Engineering from Xian Jiaotong University, Xian, China, in 2001, and the Ph.D degree from Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA, in 2006, respectively. He is now a staff engineer at Qualcomm Research Labs, Santa Clara, California. His research interests include mobile wireless communications and networks, with emphasis on cross-layer design and optimizations, wireless quality-of-service (QoS) provisioning for mobile multimedia networks and wireless resource allocation. Dr. Tang received Fouraker Graduate Research Fellowship Award from Department of Electrical and Computer Engineering, Texas A&M University in 2005.