QoS-Driven Power Allocation Over Parallel Fading Channels With Imperfect Channel Estimations in Wireless Networks

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

***Abstract*— We propose the quality-of-service (QoS) driven power allocation schemes for parallel fading channels when considering imperfect channel estimations. In particular, the parallel communication model plays a fundamental role in physical-layer evolutions of wireless networks. By integrating information theory with the concept of *effective capacity*, our proposed schemes aim at maximizing the system throughput subject to a given delay constraint. Solving the original non-convex problem by a 2-dimensional convex optimization approach, we develop the optimal allocation algorithms under different QoS and power constraints. Consistent with our previous work assuming perfect channel state information (CSI), our analyses considering imperfect CSI demonstrate that when the QoS constraint becomes more and more stringent, the optimal effective capacity decreases from the *ergodic capacity* to the *zero-outage capacity*. Moreover, our results indicate that the channel estimation error has a significant impact on QoS provisioning, especially when the delay constraint is stringent. Specifically, as long as the channel estimation is not perfect, a positive zero-outage capacity is unattainable. On the other hand, our simulations also suggest that a larger number of parallel channels can provide higher throughput and more stringent QoS, while offering better robustness against the imperfectness of CSI.**

***Index Terms*— Power control, quality-of-service (QoS), resource allocation and management, convex optimization, information theory.**

I. INTRODUCTION

THE EXPLOSIVE demand for wireless services motivates a rapid evolution of wireless wideband communications. In order to efficiently support a large number of distinct wireless applications, such as wireless Internet, mobile computing, and cellular telephoning, diverse quality-of-service (QoS) guarantees play the increasingly important role to the future wireless networks. Over the wireless environment, the most scarce radio resources are power and spectral bandwidth. In response, a great deal of research has been devoted to the techniques that can enhance the spectral efficiency of the wireless transmissions. The framework used to evaluate these techniques is mainly based on information theory [1], using the concept of either *ergodic capacity* [2] [3] or *outage capacity* [4] [5]. The ergodic capacity maximizes the average spectral efficiency with an infinitely long delay. The outage capacity, on the other hand, maintains a constant rate transmission with a certain outage probability. From the delay QoS point-of-view, such an information-theoretic framework maximizes the system throughput either without any delay constraint (i.e., ergodic capacity), or with a stringent delay constraint (i.e., outage capacity). These two extremes may not refine enough for the user's satisfactions, where a wide range of delay constraints may be requested for different applications. Consequently, to provide diverse QoS guarantees, it is necessary to take the QoS metrics into account when applying the prevalent information theory.

In our companion papers [6] [7], we proposed QoS-driven power allocation schemes for single-input-single-output (SISO) and multiple-input-multiple-output (MIMO) systems, respectively, when assuming *perfect* channel state information (CSI) available at both the transmitter and receiver. The proposed scheme aims at maximizing the system throughput subject to a given delay constraint. Our results in [6] [7] showed that the adaptive power allocation is critically important for QoS provisioning. Specifically, by integrating information theory with the concept of *effective capacity* [8] [9], we convert the original problem to the one with the target at maximizing the effective capacity through the optimal power allocation, in which the delay QoS constraint is characterized by the QoS exponent θ. Applying the effective capacity, a smaller θ corresponds to a looser QoS guarantee, while a larger θ implies a more stringent QoS requirement. In the limiting case, when $\theta \to 0$, the system can tolerate an arbitrarily long delay, which is the scenario to derive the ergodic capacity. In contrast, when $\theta \to \infty$, the system cannot tolerate any delay, which corresponds to the case to obtain the zero-outage capacity. Thus, as θ dynamically varies, the optimal power allocation builds up a bridge between the ergodic capacity and the zero-outage capacity.

As the sequel of [6] and [7], this paper focuses on QoS provisioning over parallel channels in the presence of channel estimation errors. Our study is based on the *block-fading* (also known as quasi-static) channel model. The physical validity of this model is discussed in [10]. Due to its analytical convenience, the block-fading channel model is commonly used in literatures [4]–[9], [11]–[13], which also greatly simplifies our analyses. We concentrate on communications over *parallel channels*, since this is a fundamental communication

Manuscript received Dec. 1, 2006. The research reported in this paper was supported in part by the U.S. National Science Foundation CAREER Award under Grant ECS-0348694.

Xi 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).

Jia Tang was with the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA, and is now with the CDMA Research Labs., Qualcomm, San Jose, CA XXXXX, USA (e-mail: jiat@qualcomm.com).

Digital Object Identifier 0000000000000000

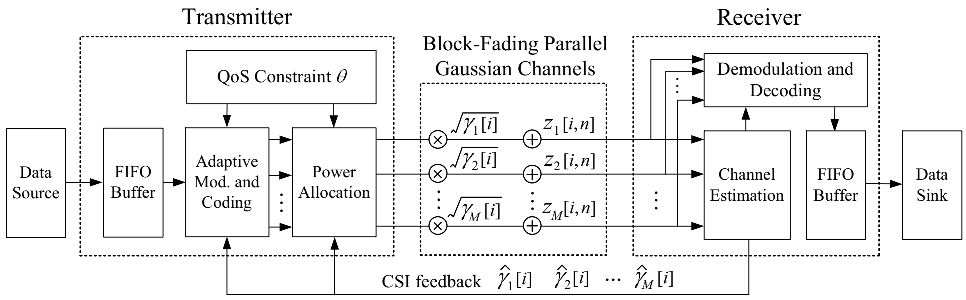

Fig. 1. The point-to-point system model.

framework, where a large number of promising techniques fall into this category. For instance, multicarrier systems employing orthogonal-frequency-division-multiplexing (OFDM) can be considered as parallel communications at the *frequency* domain [13][14]. In contrast, the MIMO system is an typical example which utilizes *spatial* domain parallel channels [2][5][11]. The emerging MIMO-OFDM architecture combines parallel channels in a joint spatial-frequency domain. On the other hand, the simple SISO system is also a special case of parallel communications, where the number of parallel channels is one.

The research of this paper is mainly motivated by a *practical concern*, where a perfect CSI is hard to obtain in real wireless networks [11]–[13], [15]. Therefore, it becomes critically important to investigate how to deal with such an imperfectness, and what its impact is on QoS provisioning. Compared to the case with perfect CSI, imperfect CSI imposes new challenges to our throughput maximization problem. In particular, the problem is *not* convex in nature. To overcome this mathematical difficulty, we decompose the original non-convex optimization problem into two orthogonal optimization sub-problems, each of which turns out to be convex and can be solved efficiently. The main contributions of this paper can be summarized as follows:

1) We derive the power allocation policy under the *total* power constraint (Theorem 1), which shows that the optimal policy is actually classic water-filling, regardless of delay requirement.
2) We propose the power allocation scheme under the *average* power constraint (Theorem 2), which shows that as the QoS exponent θ increases from zero to infinity, the optimal effective capacity decreases from the ergodic capacity to the zero-outage capacity.
3) Under stringent delay requirement, we provide necessary and sufficient conditions for the convergence of the average power (Theorem 3), which show that in the presence of channel estimation errors, the average power always diverges. Furthermore, a positive zero-outage capacity is proved to be unattainable. Alternatively, we explicitly obtain the power allocation scheme to minimize the outage probability (Theorem 4).

Our results also suggest that a larger number of parallel channels can provide higher throughput and support more stringent QoS, while offering better robustness against the wireless-channel estimation errors.

The rest of the paper is organized as follows. Section II describes our parallel system model. Sections III derives the optimal power allocation policy with different power constraints. Section IV discusses the power allocation strategy for stringent QoS provisioning. Section V conducts simulations to evaluate the performance of our proposed scheme. The paper concludes with Section VI.

Notations. We use upper- and lower-case boldface letters to denote matrices and vectors, respectively. $\mathbb{R}$ and $\mathbb{C}$ indicate the space of real and complex numbers, respectively, with possible superscript denoting the dimension of the matrices or vectors. $\mathbb{R}_+$ and $\mathbb{R}_{++}$ represent the nonnegative and positive real numbers, respectively. $(x)^+ \triangleq \max\{0, x\}$. $\mathbb{E}[\cdot]$ stands for the expectation, $\mathbb{E}_{\boldsymbol{x}}[\cdot]$ represents that the expectation is with respect to $\boldsymbol{x}$. $\boldsymbol{I}_K$ denotes a $K \times K$ identity matrix. $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{u}, \boldsymbol{\Sigma})$ means that the complex random vector $\boldsymbol{x}$ follows a jointly Gaussian distribution with mean $\boldsymbol{u}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. System Model

The system model is illustrated in Fig. 1. We concentrate on a discrete-time point-to-point link between the transmitter and the receiver in wireless networks. In particular, the transmitter and the receiver are communicating through M parallel fading channels over spectral bandwidth B. As shown in Fig. 1, a first-in-first-out (FIFO) buffer is equipped at the transmitter, which buffers the data *frames* to be transmitted to the receiver. Each frame consists of $M \times N$ symbols. 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 (i.e., $N \to \infty$) is meaningful [4][5]. The frame is then divided into M substreams, each with N symbols transmitted through one of the parallel channels. Based on a given QoS constraint θ requested by the mobile session and CSI fed back from the mobile receiver, the transmitter needs to find an optimal codeword (implemented by the adaptive modulation and coding) and a corresponding power allocation strategy, which can maximize the throughput subject to the QoS constraint θ.

The discrete-time channel process is assumed to be block-fading. Specifically, 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 the most commonly used channel distributions, such as Rayleigh, Rice, Nakagami, and Wishart, are all continuous and thus belong to this category. The transmission for the nth symbol of the ith frame can be modeled as

$$\boldsymbol{y}[i,n] = \sqrt{\boldsymbol{\Gamma}[i]}\boldsymbol{x}[i,n] + \boldsymbol{z}[i,n] \tag{1}$$

where $i = 1, 2, ...$ denotes the frame index, $n = 1, 2, ..., N$ denotes the symbol index, $\boldsymbol{x}[i,n] \in \mathbb{C}^M$ and $\boldsymbol{y}[i,n] \in \mathbb{C}^M$ are complex channel input and output symbols, respectively, $\sqrt{\boldsymbol{\Gamma}[i]} \triangleq \mathrm{diag}\{\sqrt{\gamma_1[i]}, \sqrt{\gamma_2[i]}, ..., \sqrt{\gamma_M[i]}\} \in \mathbb{R}_+^{M\times M}$ denotes the diagonal channel gain matrix, and $\boldsymbol{z}[i,n] \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$ is i.i.d. complex additive white Gaussian noise (AWGN), which, by a properly transmit power scaling, can be normalized to have the unit variance.

Let $\boldsymbol{\gamma}[i] \triangleq (\gamma_1[i], \gamma_2[i], ..., \gamma_M[i])$ denote the instantaneous CSI. When the receiver knows *perfectly* about $\boldsymbol{\gamma}[i]$, for a given power allocation $\boldsymbol{\mu}[i] \triangleq (P_1[i], P_2[i], ..., P_M[i]) \in \mathbb{R}_+^M$, the maximum instantaneous mutual information between channel inputs and outputs, denoted by $\mathcal{I}(\boldsymbol{\mu}[i], \boldsymbol{\gamma}[i])$, can be expressed as[1]

$$\mathcal{I}(\boldsymbol{\mu}[i], \boldsymbol{\gamma}[i]) \triangleq \frac{T_f B}{K} \sum_{m=1}^{M} \log_2\Big(1 + \gamma_m[i] P_m[i]\Big) \tag{2}$$

which can be achieved by the independent complex Gaussian inputs $\boldsymbol{x}[i,n] \sim \mathcal{CN}(\boldsymbol{0}, \mathrm{diag}\{\boldsymbol{\mu}[i]\})$. In Eq. (2), the parameter K with $1 \le K \le M$ is a scaling constant dependent on the specific parallel transmission scheme. For instance, when $\boldsymbol{\gamma}[i]$ corresponds to M singular-values of the spatial MIMO channel, we have $K = 1$. On the other hand, when $\boldsymbol{\gamma}[i]$ corresponds to M subchannel gains of a multicarrier system, K is equal to M.

In this paper, we are interested in the scenario where $\boldsymbol{\gamma}[i]$ is *imperfectly* known to the receiver. Let $\widehat{\gamma}[i] \in \mathbb{R}_+^M$ denote the estimation of the actual CSI $\boldsymbol{\gamma}[i]$. Given $\boldsymbol{\mu}[i]$ and $\widehat{\gamma}[i]$, the closed-form expression for the maximum instantaneous mutual information between the channel inputs and outputs turns out to be intractable, even in the simple case of $M = 1$ [12]. However, under sufficient conditions, a tight lower-bound, denoted by $\widehat{\mathcal{I}}(\boldsymbol{\mu}[i], \widehat{\gamma}[i])$, can be obtained as [11]–[13]

$$\widehat{\mathcal{I}}(\boldsymbol{\mu}[i], \widehat{\gamma}[i]) \triangleq \frac{T_f B}{K} \sum_{m=1}^{M} \log_2\left(1 + \frac{\widehat{\gamma}_m[i] P_m[i]}{1 + \sigma_e^2 \sum_{m=1}^{M} P_m[i]}\right) \tag{3}$$

where σ_e^2 denotes the variance of the channel estimation errors, which depends on the channel dynamics and channel estimation schemes employed [11], and is assumed to be known *a priori* at the both ends of the link. The mutual information lower-bound in Eq. (3) can be achieved by the independent complex Gaussian inputs and nearest neighbor decoding rule, see, e.g., [11][15] for a detailed discussion. It is also clear that when $\sigma_e^2 \to 0$, we have $\widehat{\gamma}_m[i] \to \gamma_m[i]$, and Eq. (3) reduces to Eq. (2).

In this paper, we also make the following assumptions.

[1]Throughout this paper, the unit for the mutual information and transmission rate are "bits per frame".

A1: We assume that the estimated CSI $\widehat{\gamma}[i]$ is reliably fed back to the transmitter without delay. The issues of feedback delay and unreliable feedback channels can be modeled as a *channel mean feedback* problem [18], which is not the focus of this paper. In addition, the preliminary work about the impact of feedback delay on the QoS provisioning can be found in [19].

A2: We further assume that given a power allocation $\boldsymbol{\mu}[i]$ and the estimated CSI $\widehat{\gamma}[i]$, the adaptive modulation and coding can choose an ideal channel code for each frame, such that the transmission rate, denoted by $R(\boldsymbol{\mu}[i], \widehat{\gamma}[i])$, achieves the mutual information lower-bound $\widehat{\mathcal{I}}(\boldsymbol{\mu}[i], \widehat{\gamma}[i])$ given in Eq. (3). Based on this assumption, the derived effective capacity using Eq. (3) also serves as a lower-bound for the optimal effective capacity.

A3: In practice, the channel estimation itself may cause a certain power loss. In this paper, since our focus is to study the impact of imperfect CSI on QoS provisioning, we ignore such a performance degradation factor. Based on our framework, the results can be easily extended to the case considering the cost of channel estimations.

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.

III. Power Allocation for QoS Provisioning

A. Problem Formulation

Let us define $\boldsymbol{\nu} \triangleq (\theta, \widehat{\gamma})$ as *network state information* (NSI). Then, based on Eq. (3) and assumption **A2**, the transmission rate, denoted by $R\big(\boldsymbol{\mu}(\boldsymbol{\nu}), \widehat{\gamma}\big)$, can be expressed as

$$R\big(\boldsymbol{\mu}(\boldsymbol{\nu}), \widehat{\gamma}\big) = \frac{T_f B}{K} \sum_{m=1}^{M} \log_2\left(1 + \frac{\widehat{\gamma}_m P_m(\boldsymbol{\nu})}{1 + \sigma_e^2 \sum_{m=1}^{M} P_m(\boldsymbol{\nu})}\right) \tag{4}$$

where the power allocation policy $\boldsymbol{\mu}(\theta, \widehat{\gamma}) = \boldsymbol{\mu}(\boldsymbol{\nu}) = (P_1(\boldsymbol{\nu}), P_2(\boldsymbol{\nu}), ..., P_M(\boldsymbol{\nu})) \in \mathbb{R}_+^M$ is not only the function of the estimated CSI $\widehat{\gamma}$, but also the function of the QoS exponent θ. For a given QoS constraint specified by θ, in order to find the optimal power allocation policy, denoted by $\boldsymbol{\mu}^*(\boldsymbol{\nu})$, that maximizes the effective capacity expression (see, e.g., [6, eq. (4)]), we can formulate a maximization problem as follows:

$$\boldsymbol{\mu}^*(\boldsymbol{\nu}) = \arg\max_{\boldsymbol{\mu}(\boldsymbol{\nu})} \left\{ -\frac{1}{\theta} \log\left(\mathbb{E}_{\widehat{\gamma}}\left[\mathcal{F}\big(\boldsymbol{\mu}(\boldsymbol{\nu}), \widehat{\gamma}\big) \right] \right) \right\} \tag{5}$$

where

$$\begin{aligned} \mathcal{F}\big(\boldsymbol{\mu}(\boldsymbol{\nu}), \widehat{\gamma}\big) &\triangleq e^{-\theta R\left(\boldsymbol{\mu}(\boldsymbol{\nu}), \widehat{\gamma}\right)} \\ &= \prod_{m=1}^{M} \left(1 + \frac{\widehat{\gamma}_m P_m(\boldsymbol{\nu})}{1 + \sigma_e^2 \sum_{m=1}^{M} P_m(\boldsymbol{\nu})}\right)^{-\beta} \end{aligned} \tag{6}$$

with $\beta \triangleq \theta T_f B/(K \log 2)$ defined as normalized QoS exponent. Since $\log(\cdot)$ is a monotonically increasing function, for each given QoS constraint $\theta \in \mathbb{R}_{++}$, the maximization problem above is equivalent to the following minimization problem:

$$\boldsymbol{\mu}^*(\boldsymbol{\nu}) = \arg\min_{\boldsymbol{\mu}(\boldsymbol{\nu})}\left\{\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\mathcal{F}\big(\boldsymbol{\mu}(\boldsymbol{\nu}), \widehat{\boldsymbol{\gamma}}\big)\right]\right\}. \tag{7}$$

In this paper, we mainly consider two different power constraints. A simple and practical constraint is known as *total power constraint*, also called short-term power constraint [4]. Specifically, the transmit power for each frame cannot exceed a certain threshold P_{total}, i.e.,

$$\sum_{m=1}^{M} P_m(\boldsymbol{\nu}) \leq P_{\text{total}} \tag{8}$$

for all realizations of $\widehat{\boldsymbol{\gamma}} \in \mathbb{R}_+^M$.

On the other hand, the *average power constraint*, also known as long-term power constraint [4], is often investigated from an information-theoretic point-of-view. Under the average power constraint, the mean of the transmit power cannot exceed a certain threshold P_{avg}, but no restriction is imposed on the instantaneous transmit power, i.e.,

$$\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\sum_{m=1}^{M} P_m(\boldsymbol{\nu})\right] \leq P_{\text{avg}}. \tag{9}$$

A system may be subject to the total power or/and average power constraints, which are elaborated on in the followings, respectively.

B. Power Allocation With Total Power Constraint

We first consider the problem of minimizing Eq. (7) subject to the total power constraint given by Eq. (8). It is clear that Eq. (7) achieves its minimum when the constraint in Eq. (8) is satisfied with equality. Accordingly, let us define a convex set, denoted by $\mathcal{S}$, for the power allocation policy as follows:

$$\mathcal{S} \triangleq \left\{\boldsymbol{\mu}(\boldsymbol{\nu}) : \boldsymbol{\mu}(\boldsymbol{\nu}) \in \mathbb{R}_+^M, \sum_{m=1}^{M} P_m(\boldsymbol{\nu}) = P_{\text{total}}\right\}. \tag{10}$$

Then, we have the following lemma.

Lemma 1: The objective function $\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\mathcal{F}\left(\boldsymbol{\mu}(\boldsymbol{\nu}), \widehat{\boldsymbol{\gamma}}\right)\right]$ given in Eq. (7) is strictly convex on $\mathcal{S}$.

Proof: It is easy to verify that $R(\boldsymbol{\mu}(\boldsymbol{\nu}), \widehat{\boldsymbol{\gamma}})$ given in Eq. (4) is strictly concave on $\mathcal{S}$. On the other hand, $f(x) = e^{-\theta x}$ is a strictly convex and non-increasing function for any fixed $\theta \in \mathbb{R}_{++}$. Using the property given by [20, eq. (3.10)], we know that $\mathcal{F}\left(\boldsymbol{\mu}(\boldsymbol{\nu}), \widehat{\boldsymbol{\gamma}}\right) = \exp(-\theta R(\boldsymbol{\mu}(\boldsymbol{\nu}), \widehat{\boldsymbol{\gamma}}))$ is strictly convex on $\mathcal{S}$. Finally, since the expectation is a linear operation, it preserves the strictly convexity. The proof follows. ■

Since the objective function given in Eq. (7) is strictly convex on $\mathcal{S}$, we can use the standard Lagrangian method to find the unique optimal power allocation policy, denoted by $\boldsymbol{\mu}^*_{\text{total}}(\boldsymbol{\nu}) \in \mathcal{S}$. Construct the Lagrange as follows:[2]

$$\mathcal{J}_1 = \mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\prod_{m=1}^{M}\left(1 + \frac{\widehat{\gamma}_m P_m(\boldsymbol{\nu})}{1 + \sigma_e^2 P_{\text{total}}}\right)^{-\beta}\right] + \lambda_1 \sum_{m=1}^{M} P_m(\boldsymbol{\nu}) \tag{11}$$

[2]In this paper, the explicit Lagrangian multipliers corresponding to the constraint $\boldsymbol{\mu}(\boldsymbol{\nu}) \in \mathbb{R}_+^M$ are omitted.

where λ_1 denotes the Lagrangian multiplier. By solving the Karush-Kuhn-Tucker (KKT) condition [20] of Eq. (11), we obtain the optimal power allocation policy $\boldsymbol{\mu}^*_{\text{total}}(\boldsymbol{\nu})$, which can be described by the following theorem.

Theorem 1: For each estimated fading state $\widehat{\boldsymbol{\gamma}}$, let $\pi(\cdot)$ be defined as a permutation of $\widehat{\boldsymbol{\gamma}}$ such that $\widehat{\gamma}_{\pi(1)} \geq \widehat{\gamma}_{\pi(2)} \geq \cdots \geq \widehat{\gamma}_{\pi(M)}$. For notational convenience, we also define

$$\widetilde{\gamma}_{\pi(m)} \triangleq \frac{\widehat{\gamma}_{\pi(m)}}{1 + \sigma_e^2 P_{\text{total}}} \tag{12}$$

for all $m = 1, 2, ..., M$. Then, the $\pi(m)$-th component of $\boldsymbol{\mu}^*_{\text{total}}(\boldsymbol{\nu})$, denoted by $P^*_{\pi(m)}(\boldsymbol{\nu})$, follows the classic water-filling formula and is determined by

$$P^*_{\pi(m)}(\boldsymbol{\nu}) = \left(\omega(\boldsymbol{\nu}, k) - \frac{1}{\widetilde{\gamma}_{\pi(m)}}\right)^+ \tag{13}$$

where $\omega(\boldsymbol{\nu}, k)$ denotes the time-varying water-level, which is chosen such that the total power constraint is satisfied, and is given by

$$\omega(\boldsymbol{\nu}, k) = \frac{1}{k}\left(P_{\text{total}} + \sum_{i=1}^{k} \frac{1}{\widetilde{\gamma}_{\pi(i)}}\right). \tag{14}$$

The parameter k in Eqs. (13) and (14) denotes the number of *active* channels allocated with nonzero power, which is the unique integer in $\{1, 2, ..., M\}$ such that $\omega(\boldsymbol{\nu}, k) > 1/\widetilde{\gamma}_{\pi(m)}$ for $m \leq k$ and $\omega(\boldsymbol{\nu}, k) \leq 1/\widetilde{\gamma}_{\pi(m)}$ for $m > k$.

Proof: The proof is provided in Appendix I. ■

Remark 1: The water-level $\omega(\boldsymbol{\nu}, k)$ and the active channels k are jointly determined by the channel state $\widehat{\boldsymbol{\gamma}}$. As a result, different fading states $\widehat{\boldsymbol{\gamma}}$ correspond to different $\omega(\boldsymbol{\nu}, k)$ and k. □

Remark 2: Although our objective is to maximize the throughput subject to the QoS constraint θ, Theorem 1 states that the optimal power allocation under the total power constraint is actually independent of θ. This implies that under the total power constraint, the water-filling formula is *always* the optimal power allocation policy, regardless of θ. On the other hand, since this policy does not distinguish the services with different QoS constraints, the power is not allocated in favor of the QoS provisioning. □

Substituting Eqs. (13) and (14) into Eq. (7) with some algebraic manipulations, we obtain the minimum objective function under the total power constraint as follows:

$$\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\Big[\mathcal{F}\left(\boldsymbol{\mu}^*_{\text{total}}(\boldsymbol{\nu}), \widehat{\boldsymbol{\gamma}}\right)\Big] = \mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\left\{\frac{k\Sigma_k\Pi_k\left(1 + \sigma_e^2 P_{\text{total}}\right)}{1 + (\sigma_e^2 + \Sigma_k)P_{\text{total}}}\right\}^{k\beta}\right] \tag{15}$$

where, for notational convenience, we define $\Sigma_k \triangleq 1/\left(\sum_{i=1}^{k} \widehat{\gamma}^{-1}_{\pi(i)}\right)$ and $\Pi_k \triangleq \prod_{i=1}^{k} \widehat{\gamma}^{-1/k}_{\pi(i)}$.

C. Power Allocation With Average Power Constraint

In this section, we focus on minimizing Eq. (7) subject to the average power constraint given by Eq. (9). This problem is more difficult than that under the total power constraint, since when $\sigma_e^2 > 0$, the objective function in Eq. (7) is not convex on the entire space spanned by $\boldsymbol{\mu}(\boldsymbol{\nu}) \in \mathbb{R}_+^M$. Alternatively, we obtain the optimal solution by a two-step approach.

Noticing that for each given total power P_{total}, by Theorem 1, we already know the optimal power allocation policy $\boldsymbol{\mu}^*_{\text{total}}(\boldsymbol{\nu})$. However, under the average power constraint, the instantaneous total power P_{total} changes with each fading state. In response, we rewrite P_{total} by $P_{\text{total}}(\boldsymbol{\nu})$ to emphasize such a temporal variation. To obtain the optimal power allocation under the average power constraint, we can solve the problem into two steps. The first step is to find the optimal *temporal* power allocation policy, denoted by $P^*_{\text{total}}(\boldsymbol{\nu}) \in \mathbb{R}_+$, which minimizes the objective function Eq. (7) while meeting the average power constraint:

$$\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[P^*_{\text{total}}(\boldsymbol{\nu})\right] = P_{\text{avg}}. \tag{16}$$

Once the optimal policy $P^*_{\text{total}}(\boldsymbol{\nu})$ is obtained, the second step is to assign power along the M parallel channels according to the water-filling algorithm described in Theorem 1, satisfying $\sum_{m=1}^M P^*_m(\boldsymbol{\nu}) = P^*_{\text{total}}(\boldsymbol{\nu})$.

Noting that when deriving the optimal policy $P^*_{\text{total}}(\boldsymbol{\nu})$ for the first step, an underlying assumption is that at the second step, the policy $\boldsymbol{\mu}^*_{\text{total}}(\boldsymbol{\nu})$ is applied for each $P^*_{\text{total}}(\boldsymbol{\nu})$. Therefore, the objective function for the first step can be expressed as Eq. (15), instead of the original one in Eq. (7). Based on Eq. (15), we formulate the new optimization problem as follows:

$$\begin{aligned} P^*_{\text{total}}(\boldsymbol{\nu}) &= \\ \arg\min_{P_{\text{total}}(\boldsymbol{\nu})} & \left\{ \mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\left\{ \frac{k\Sigma_k \Pi_k \left[1+\sigma_e^2 P_{\text{total}}(\boldsymbol{\nu})\right]}{1+(\sigma_e^2+\Sigma_k)P_{\text{total}}(\boldsymbol{\nu})} \right\}^{k\beta} \right] \right\} \end{aligned} \tag{17}$$

subject to the average power constraint given in Eq. (16).

Let us define a convex set, denoted by $\mathcal{S}'$, for the temporal power allocation policy as follows:

$$\mathcal{S}' \triangleq \left\{ P_{\text{total}}(\boldsymbol{\nu}) \in \mathbb{R}_+, \mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[P_{\text{total}}(\boldsymbol{\nu})\right] = P_{\text{avg}} \right\}. \tag{18}$$

Then, we have the following lemma.

Lemma 2: The objective function given in Eq. (17) is strictly convex on set $\mathcal{S}'$.

Proof: The proof is provided in Appendix II. ■

Due to the convexity of Eq. (17) on set $\mathcal{S}'$, we decompose the original non-convex problem into two sub-problems, each of which is convex. An illustration of the 2-dimensional convex optimization is shown in Fig. 2. Although the objective function is not convex, the two optimized dimensions are always convex, respectively. From Lemma 2, once again, we can use the Lagrangian technique to derive the unique optimal temporal power allocation policy $P^*_{\text{total}}(\boldsymbol{\nu}) \in \mathcal{S}'$. Construct the Lagrange as follows:

$$\begin{aligned} \mathcal{J}_2 &= \mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\left\{ \frac{k\Sigma_k \Pi_k \left[1+\sigma_e^2 P_{\text{total}}(\boldsymbol{\nu})\right]}{1+(\sigma_e^2+\Sigma_k)P_{\text{total}}(\boldsymbol{\nu})} \right\}^{k\beta} \right] \\ &\quad + \lambda_2 \mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[P_{\text{total}}(\boldsymbol{\nu})\right] \end{aligned} \tag{19}$$

where λ_2 denotes the Lagrangian multiplier. Solving the above Lagrangian problem, we obtain the optimal temporal power allocation policy $P^*_{\text{total}}(\boldsymbol{\nu})$ under the average power constraint, which can be described by the following theorem.

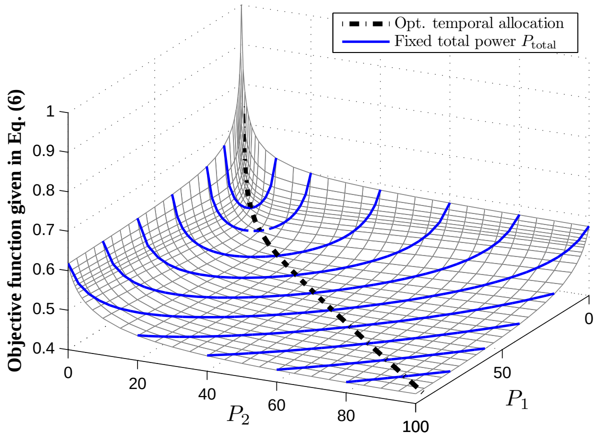

Fig. 2. The 2-dimensional convex optimization for the case of $M = 2$, where $K = 1$, $\widehat{\gamma}_1 = 5$, $\widehat{\gamma}_2 = 3$, $\beta = 0.1$, and $\sigma_e^2 = 0.1$.

Theorem 2: The optimal temporal power allocation policy $P^*_{\text{total}}(\boldsymbol{\nu}) \in \mathbb{R}_{++}$, if existing, is the unique positive solution of the following equation:

$$\frac{\left[1+(\sigma_e^2+\Sigma_k)P_{\text{total}}(\boldsymbol{\nu})\right]^{\frac{k\beta+1}{M\beta+1}}}{(k\Sigma_k)^{\frac{k\beta+1}{M\beta+1}} \Pi_k^{\frac{k\beta}{M\beta+1}} \left[1+\sigma_e^2 P_{\text{total}}(\boldsymbol{\nu})\right]^{\frac{k\beta-1}{M\beta+1}}} = \omega^* \tag{20}$$

where $\omega^* \in \mathbb{R}_+$ is a constant which is chosen such that the average power constraint is satisfied. Otherwise, if such a solution $P^*_{\text{total}}(\boldsymbol{\nu}) \in \mathbb{R}_{++}$ does not exist, then $P^*_{\text{total}}(\boldsymbol{\nu}) = 0$.

Proof: The proof is provided in Appendix III. ■

Remark 3: The constant ω^* can be called *water-level coefficient*, which is proportional to the average power constraint. The higher the average power constraint P_{avg}, the larger the water-level coefficient ω^*. Once ω^* is determined, it remains as a constant regardless of the instantaneous channel realizations. □

Unfortunately, the general closed-form solution for Eq. (20) turns out to be intractable. However, since the left-hand side of Eq. (20) is a monotonically increasing function of $P_{\text{total}}(\boldsymbol{\nu}) \in \mathbb{R}_+$, the solution, if existing, can be easily obtained numerically. Moreover, under a number of special cases, Eq. (20) can be solved in closed-form expressions.

1) $\beta \to 0$: When the normalized QoS exponent $\beta \to 0$, Eq. (20) becomes a quadratic polynomial of $P_{\text{total}}(\boldsymbol{\nu})$ and can be easily solved in closed-form. In this case, we get the following optimal temporal power allocation policy:

$$\begin{aligned} P^*_{\text{total}}(\boldsymbol{\nu})\Big|_{\beta\to 0} &= \\ & \left(\frac{-(2\sigma_e^2+\Sigma_k)+\sqrt{\Sigma_k^2+4\omega^* k\Sigma_k\sigma_e^2(\sigma_e^2+\Sigma_k)}}{2\sigma_e^2(\sigma_e^2+\Sigma_k)} \right)^+ \end{aligned} \tag{21}$$

which is the optimal temporal power allocation policy given by [11, eq. (17)] (and also [12, eq. (3)] for the case with $M = 1$) to achieve the *ergodic capacity* of the parallel channels with channel estimation errors. This is expected since when $\beta \to$

0, implying that the system can tolerate an arbitrarily long delay, the optimal effective capacity approaches the ergodic capacity [6], [7].

2) $\beta \to \infty$: When the normalized QoS exponent $\beta \to \infty$, implying stringent delay constraint, Eq. (20) becomes a linear function of $P_{\rm total}(\boldsymbol{\nu})$. The optimal temporal power allocation policy can be easily derived as follows:

$$P^*_{\rm total}(\boldsymbol{\nu})\Big|_{\beta\to\infty} = \left(\frac{\eta_k - 1}{\Sigma_k - \sigma_e^2\left(\eta_k - 1\right)}\right)^+ \tag{22}$$

where $\eta_k \triangleq (\omega^*)^{M/k} k\Sigma_k\Pi_k$. Substituting Eq. (22) into Eq. (13), the power assigned to each parallel channels can be expressed as

$$P^*_{\pi(m)}(\boldsymbol{\nu})\Big|_{\beta\to\infty} = \left(\left(\frac{1}{\Sigma_k - \sigma_e^2\left(\eta_k - 1\right)}\right)\left(\frac{\eta_k}{k} - \frac{\Sigma_k}{\widehat{\gamma}_{\pi(m)}}\right)\right)^+. \tag{23}$$

The optimal effective capacity approaches the *zero-outage capacity*[3] as $\beta \to \infty$ [6], [7]. Therefore, Eq. (23) provides the optimal power allocation policy to achieve zero-outage capacity lower-bound with channel estimation errors. The details about zero-outage capacity and outage minimization will be presented in the next section.

3) $\sigma_e^2 \to 0$: When the channel estimation is perfect, Eq. (20) also becomes a linear function of $P_{\rm total}(\boldsymbol{\nu})$. Likewise, the power assigned to each parallel channels can be obtained as

$$P^*_{\pi(m)}(\boldsymbol{\nu})\Big|_{\sigma_e^2\to 0} = \left((\omega^*)^{\frac{M\beta+1}{k\beta+1}}\Pi_k^{\frac{k\beta}{k\beta+1}} - \frac{1}{\widehat{\gamma}_{\pi(m)}}\right)^+ \tag{24}$$

which becomes the optimal power allocation for parallel channels under perfect CSI [7]. Moreover, in the limiting cases for loose QoS constraint (i.e., $\beta \to 0$), Eq. (24) reduces to the classic water-filling to achieve the ergodic capacity, which is expected, as discussed before. On the other hand, for stringent QoS constraint (i.e., $\beta \to \infty$), Eq. (24) reduces to the power allocation policy given in [4, eq. (28)] and [5, eq. (21)] to achieve the zero-outage capacity of the parallel fading channels. This is also expected since when $\beta \to \infty$, implying that the system cannot tolerate any delay, the power allocation needs to be designed to guarantee a zero-outage.

D. Power Allocation With Both Constraints

In this section, we consider the scenario where the system is subject to *both* total power constraint $P_{\rm total}$ and average power constraint $P_{\rm avg}$. The motivation of this study is the following. First, in practice, the system requires both total power and average power constraints due to hardware limitations. More importantly, as will be seen in the next section, under *only* an average power constraint, the average power does not always converge. When the average power cannot be bounded away from infinity, it is necessary to impose a total power constraint to avoid this divergence. In the following, we assume $P_{\rm total} \geq P_{\rm avg}$. Otherwise, $P_{\rm avg}$ is unattainable.

[3]The zero-outage capacity is also termed delay-limited capacity [4], [5].

To address the total power constraint, let us define another convex set, denoted by $\mathcal{S}''$, for the temporal power allocation policy as follows:

$$\mathcal{S}'' \triangleq \left\{P_{\rm total}(\boldsymbol{\nu}) : P_{\rm total}(\boldsymbol{\nu}) \in \mathcal{S}', P_{\rm total}(\boldsymbol{\nu}) \leq P_{\rm total}\right\} \tag{25}$$

where $\mathcal{S}'$ is defined in Eq. (18). It is clear that $\mathcal{S}'' \subseteq \mathcal{S}'$. Now the problem becomes maximizing the objective given in Eq. (17) on set $\mathcal{S}''$, instead of on set $\mathcal{S}'$ investigated in Section III-C. By the similar procedure used in Section III-C, we derive the optimal temporal power allocation under both constraints. The optimal temporal power assigned to each fading state, denoted by $P^*_{\rm both}(\boldsymbol{\nu})$, is simply a *truncated* version of the power derived from Theorem 2, i.e.,

$$P^*_{\rm both}(\boldsymbol{\nu}) = \min\left\{P^*_{\rm total}(\boldsymbol{\nu}), P_{\rm total}\right\} \tag{26}$$

where $P^*_{\rm total}(\boldsymbol{\nu})$ is obtained by Theorem 2. Accordingly, the water-level coefficient ω^* needs to be recalculated to meet the average power constraint.

IV. Power Allocation under Stringent Delay Constraint

When designing the QoS-driven power allocation algorithm, we are more interested in the region where the QoS constraint is stringent. Therefore, in this section, we take a close look at the power allocation performance under stringent delay constraint.

A. Convergence Analyses for the Average Power

As the delay constraint becomes stringent ($\beta \to \infty$), Eq. (22) provides the optimal temporal power allocation. However, using Eq. (22), the average power may *diverge*. In other words, for a given $P_{\rm avg}$, we probably cannot find ω^* such that $\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[P^*_{\rm total}(\boldsymbol{\nu})\right] = P_{\rm avg}$. In order to guarantee that the average power converges, we need to upper-bound the expectation of Eq. (22) away from infinity, which is equivalent to

$$\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\frac{\eta_k - 1}{\Sigma_k - \sigma_e^2\left(\eta_k - 1\right)}\right]^+ < \infty \tag{27}$$

where $\mathbb{E}[x]^+ \triangleq \mathbb{E}\left[x | x \geq 0\right]$. Explicitly characterizing the left-hand side of Eq. (27) is hard since k is time-varying depending on $\widehat{\boldsymbol{\gamma}}$. Alternatively, it is more convenient to find the necessary and sufficient conditions for the convergence. The result can be summarized in the following theorem.

Theorem 3: If $\omega^* > 1$, a necessary condition to guarantee that the average power converges to a finite number is given by

$$\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\frac{1}{\widehat{\gamma}_{\pi(1)} - \sigma_e^2(\omega^* - 1)}\right]^+ < \infty, \tag{28}$$

while a sufficient condition is given by

$$\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\frac{\Pi_M}{1 - \sigma_e^2(\omega^*)^M M\Pi_M}\right]^+ < \infty. \tag{29}$$

Otherwise, if $\omega^* \leq 1$, then $P_{\rm total}(\boldsymbol{\nu}) = 0$ always holds, and thus $\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[P_{\rm total}(\boldsymbol{\nu})\right] = 0$.

Proof: The proof is provided in Appendix IV. ■

Remark 4: When the channel estimation is perfect ($\sigma_e^2 = 0$), the sufficient condition given by Eq. (29) reduces to

$$\mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\Pi_M\right]^+ = \mathbb{E}_{\widehat{\boldsymbol{\gamma}}}\left[\Pi_M\right] < \infty \tag{30}$$

which is the condition termed *regular fading* in [5, def. 4] to achieve a positive zero-outage capacity with perfect CSI. Furthermore, when $\sigma_e^2 = 0$, it is easy to show that Eq. (30) implies Eq. (28), which is also expected since Eq. (28) is a necessary condition. □

Remark 5: For most commonly used channel distributions (e.g., Rayleigh, Nakagami, Rice, and Wishart), if $\sigma_e^2 = 0$, the *sufficient* condition given by Eq. (30) always fulfilled (additional condition of $M > 1$ may be required). Therefore, the average power always converges. However, if $\sigma_e^2 > 0$, the *necessary* condition given by Eq. (28) cannot be fulfilled. Thus, the average power always diverges. □

B. Outage Minimization

When $\beta \to \infty$, substituting Eqs. (22) and (23) into Eq. (4) with some algebraic manipulations, the instantaneous spectral efficiency $R/(T_f B)$, denoted by $\mathcal{C}$ (bits/s/Hz), can be obtained as

$$\mathcal{C} = \begin{cases} \frac{M}{K}\log_2(\omega^*), & \text{if } P^*_{\text{total}}(\boldsymbol{\nu}) > 0 \\ 0, & \text{if } P^*_{\text{total}}(\boldsymbol{\nu}) = 0 \end{cases} \tag{31}$$

which implies that the transmission is either with a *constant rate* or in an *outage*. If the outage probability is nonzero, we know from definition that the zero-outage capacity of the system is zero. The following lemma describes the impact of channel estimation error on system outage probability.

Proposition 1: If $\sigma_e^2 > 0$, then the outage probability is nonzero.

Proof: The proof is provided in Appendix V. ■

Remark 6: As long as $\sigma_e^2 > 0$, from Proposition 1 we know $\Pr\{P^*_{\text{total}}(\boldsymbol{\nu}) = 0\} > 0$. Any outage probability smaller than $\Pr\{P^*_{\text{total}}(\boldsymbol{\nu}) = 0\}$ is unattainable. In other words, the probability $\Pr\{P^*_{\text{total}}(\boldsymbol{\nu}) = 0\}$ indicates an *outage floor*. □

Corollary 1: If $\sigma_e^2 > 0$, then the system zero-outage capacity is always zero, regardless of the channel fading distributions.

Proof: The proof follows from Proposition 1. ■

Remark 7: As compared to the case with perfect CSI, where the zero-outage capacity is always *positive* when the channel is regular fading [5], the zero-outage capacity of the system with imperfect CSI is always zero, due to the presence of nonzero σ_e^2. In this case, it makes more sense to study the *outage capacity*, instead of zero-outage capacity. □

To transmit at a constant code rate $\mathcal{R}$ (bits/s/Hz), the following theorem provides the optimal allocation policy that minimizes the outage probability under an average power constraint P_{avg}.

Theorem 4: The optimal temporal power allocation policy, denoted by $P^*_{\text{out}}(\boldsymbol{\nu})$, that minimizes the outage probability while transmitting at a constant code rate $\mathcal{R}$, can be expressed as

$$P^*_{\text{out}}(\boldsymbol{\nu}) = \begin{cases} P^*_{\text{total}}(\boldsymbol{\nu}), & \text{if } P^*_{\text{total}}(\boldsymbol{\nu}) \le s^* \\ 0, & \text{otherwise} \end{cases} \tag{32}$$

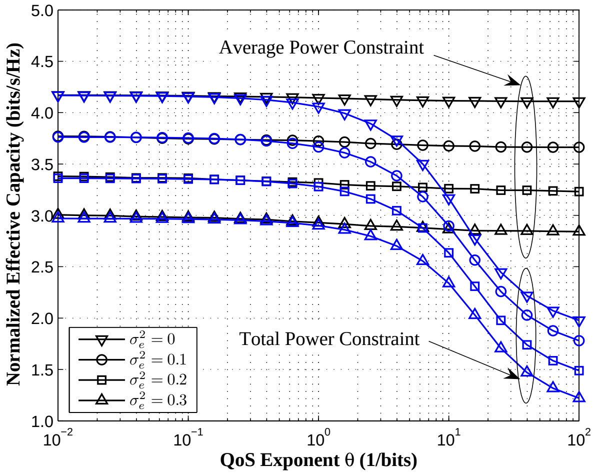

Fig. 3. The optimal effective capacity for a 4×4 MIMO system with different power constraints. The average SNR is 0 dB for both cases.

where $P^*_{\text{total}}(\boldsymbol{\nu})$ is the solution of Eq. (22) with $\omega^* = 2^{\mathcal{R}K/M}$, and $s^* \in \mathbb{R}_+$ is a constant chosen such that the average power constraint is satisfied. Based on this policy, the resulting minimum outage probability, denoted by p_{out}, is determined by

$$p_{\text{out}} = \Pr\{P^*_{\text{total}}(\boldsymbol{\nu}) = 0\} + \Pr\{P^*_{\text{total}}(\boldsymbol{\nu}) > s^*\}. \tag{33}$$

Proof: It can be easily observed from Eq. (31) that ω^* should be chosen as $\omega^* = 2^{\mathcal{R}K/M}$. The rest of the proof is based on the result of [4], which is omitted for lack of space. ■

Remark 8: The parameter s^* has the same role as the total power constraint P_{total} in previous sections. However, the power allocation policies are different. In an effective-capacity maximization problem, when the instantaneous power exceeds P_{total}, the system still use the maximum available power to transmit data [see Eq. (26)], avoiding the outage. In contrast, in an outage minimization problem, when the instantaneous power exceeds s^*, the system stops transmitting data to save the transmit power [see Eq. (32)], making the system fall into an outage. □

V. Performance Evaluations

In this section, we evaluate the performance of our proposed QoS-driven power allocation by simulations. As a typical application over parallel Gaussian channels, we simulate the MIMO system with N_t transmit antennas and N_r receive antennas. The channels between all transmit and receive antenna pairs are assumed to be i.i.d. complex Gaussian with $\mathcal{CN}(0,1)$. In this case, the parameters $K = 1$ and $M = \min\{N_t, N_r\}$. By using the minimum mean squared error (MMSE) estimator at the receiver, the range of the error variance is $0 \le \sigma_e^2 \le 1$, and the estimated channels are i.i.d. with $\mathcal{CN}(0, 1 - \sigma_e^2)$. Furthermore, we set the product $T_f B = \log 2$ such that $\theta = \beta$ for convenience. The other system parameters are detailed, respectively, in each of the figures.

Fig. 3 plots the optimal effective capacity of a 4×4 MIMO system with different power constraints. When the

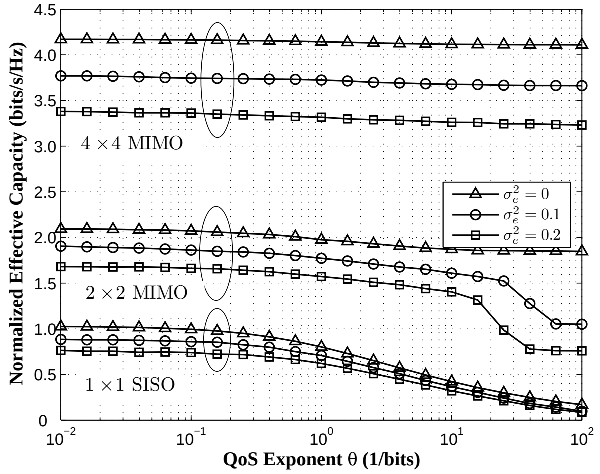

Fig. 4. The optimal effective capacity with different numbers of antennas under both total and average power constraints. The average power constraint $P_{\text{avg}} = 0$ dB and the total power constraint $P_{\text{total}} = 20$ dB.

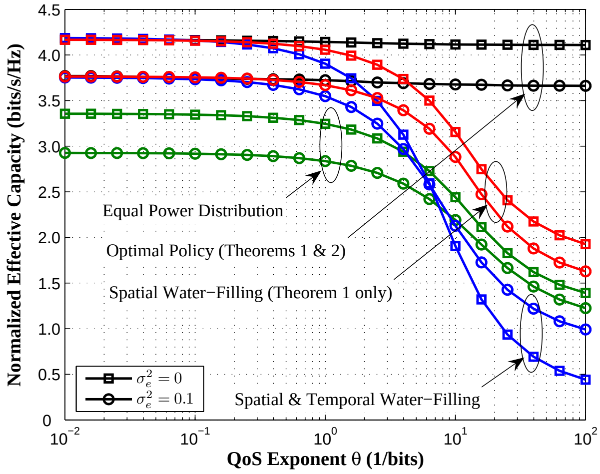

Fig. 5. The optimal effective capacity for a 4×4 MIMO system with different power allocation strategies. The average SNR is 0 dB for all the cases.

QoS constraint is loose, we can observe from Fig. 3 that total power constraint and average power constraint have neglectable performance difference. However, as the QoS constraint becomes more stringent, the average power constraint shows significant performance advantages over the total power constraint. In particular, the effective capacity under the average power constraint virtually does not decrease as θ increases, while the effective capacity under the total power constraint drops quickly as θ increases, which verifies the importance of temporal power allocation on QoS provisioning. On the other hand, the impact of channel estimation error on effective capacity is also significant.

Fig. 4 plots the optimal effective capacity under both average power constraint and total power constraint, when σ_e^2 ranges from 0 to 0.2. As shown by Fig. 4, for the 4×4 MIMO system, the effective capacities are all virtually independent of θ. On the other hand, for the 2×2 MIMO system, the QoS constraint θ does not significantly affect the effective capacity when the channel estimation is perfect ($\sigma_e^2 = 0$). However, when $\sigma_e^2 = 0.1$ or $\sigma_e^2 = 0.2$, the effective capacities significantly decrease as the QoS constraint becomes stringent. Finally, for the 1×1 SISO system, all effective capacities converges to zero as θ increases, even when the channel estimation is perfect. Thus, Fig. 4 verifies that a larger number of antennas not only provides the higher throughput, but also offers better robustness against the channel estimation error, in terms of supporting stringent QoS requirements.

Fig. 5 plots the effective capacity under different power allocation strategies. Besides our proposed power allocation with total power constraint (referred as "spatial water-filling") and with average power constraint (referred as "optimal policy"), we also simulate equal power distribution strategy, and joint spatial-temporal water-filling strategy. Note that equal power distribution is the optimal power allocation without CSI at the transmitter, and joint spacial-temporal water-filling is the optimal power allocation to achieve the ergodic capacity of the MIMO system. We can observe from Fig. 5 that for a given

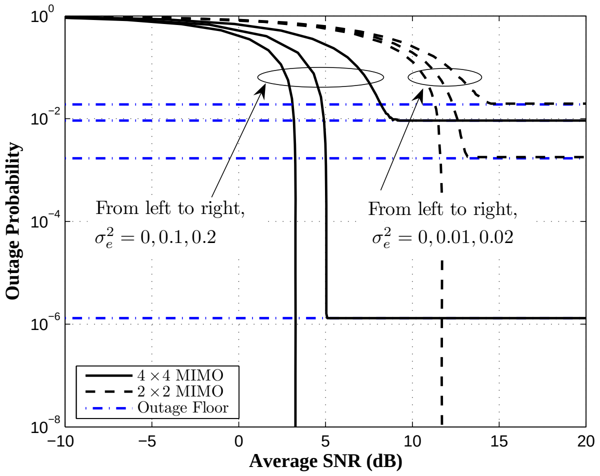

Fig. 6. The outage probability for MIMO system with different channel estimation errors. The code rate $\mathcal{R} = 6$ bits/s/Hz.

σ_e^2, our proposed optimal policy always achieves the highest effective capacity among all power allocation strategies. The advantage is more significant when the QoS constraint is stringent.

Finally, Fig. 6 plots the outage probability for MIMO systems with different channel estimation errors. Specifically, when the channel estimation is perfect ($\sigma_e^2 = 0$), the outage probability approaches zero when the average SNR is sufficiently high, which means a positive zero-outage capacity is achievable. However, when the channel estimation is imperfect ($\sigma_e^2 > 0$), the outage floor prevents the outage probability from further decreasing, no matter how much power is assigned. Fig. 6 also demonstrates that for a given code rate, the system with a larger number of antennas may tolerate severer channel estimation errors, while still maintaining better outage performance.

VI. Conclusions

We proposed and analyzed QoS-driven power allocation over parallel fading channels by taking the imperfect channel estimations into consideration. Solving the original non-convex problem by a 2-dimensional convex optimization approach, we developed power allocation algorithms for different QoS and power constraints in a general system setting. As the QoS exponent θ increases from zero to infinity, the optimal effective capacity function connects the ergodic capacity with the zero-outage capacity, which is consistent with our previous work in the case of perfect CSI. Our analyses indicate that the imperfect channel estimations have a significant impact on QoS provisioning, especially when the delay constraint is stringent. In particular, a positive zero-outage capacity is unattainable in the presence of channel estimation errors. On the other hand, our simulation results for the MIMO systems also suggest that a larger number of parallel channels can provide higher throughput and more stringent QoS, while offering better robustness against the channel estimation errors.

Appendix I
Proof of Theorem 1

Proof: Assume that there are exactly k channels out of M channels being assigned with nonzero power, where $1 \leq k \leq M$. It can be easily shown by contradiction that these k channels are $\widehat{\gamma}_{\pi(1)}, \widehat{\gamma}_{\pi(2)}, ..., \widehat{\gamma}_{\pi(k)}$. Thus, the Lagrangian $\mathcal{J}_1$ can be simplified to a new Lagrangian function, denoted by $\mathcal{J}_1'$, as follows:

$$\mathcal{J}_1' = \mathbb{E}_{\widehat{\gamma}}\left[\prod_{m=1}^{k}\left(1+\widetilde{\gamma}_{\pi(m)}P_{\pi(m)}(\boldsymbol{\nu})\right)^{-\beta}\right] + \lambda_1\sum_{m=1}^{k}P_{\pi(m)}(\boldsymbol{\nu}) \quad (34)$$

where $\widetilde{\gamma}_{\pi(m)}$ is defined in Eq. (12). Differentiating the simplified Lagrangian $\mathcal{J}_1'$ with respect to $P_{\pi(m)}(\boldsymbol{\nu})$ and setting the derivative equal to zero, we can get a set of k equations:

$$\left[1+\widetilde{\gamma}_{\pi(m)}P_{\pi(m)}(\boldsymbol{\nu})\right]^{-(\beta+1)}\prod_{i=1,\, i\neq m}^{k}\left[1+\widetilde{\gamma}_{\pi(i)}P_{\pi(i)}(\boldsymbol{\nu})\right]^{-\beta} = \frac{\lambda_1}{\beta\widetilde{\gamma}_{\pi(m)}},\ \text{for all } 1 \leq m \leq k. \quad (35)$$

Solving Eq. (35) and considering the boundary conditions, we obtain Eq. (13), where

$$\omega(\boldsymbol{\nu}, k) = \left(\frac{\beta}{\lambda_1}\right)^{\frac{1}{k\beta+1}}\prod_{i=1}^{k}\widetilde{\gamma}_{\pi(i)}^{-\frac{\beta}{k\beta+1}}. \quad (36)$$

By choosing a proper λ_1 in Eq. (36) to meet the total power constraint, $\omega(\boldsymbol{\nu}, k)$ can be simplified to Eq. (14). The proof follows. ■

Appendix II
Proof of Lemma 2

Proof: Due to the linearity of the expectation, it is sufficient to show that the objective function inside the expectation is convex on $P_{\text{total}}(\boldsymbol{\nu}) \in \mathbb{R}_+$. Following the notation of Eq. (15), we differentiate $\mathcal{F}\left(\boldsymbol{\mu}^*_{\text{total}}(\boldsymbol{\nu}), \widehat{\boldsymbol{\gamma}}\right)$ with respect to $P_{\text{total}}(\boldsymbol{\nu})$ and get the following:

$$\frac{\partial\mathcal{F}\left(\boldsymbol{\mu}^*_{\text{total}}(\boldsymbol{\nu}), \widehat{\boldsymbol{\gamma}}\right)}{\partial P_{\text{total}}(\boldsymbol{\nu})} = -\frac{\beta(k\Sigma_k)^{k\beta+1}\left[1+\sigma_e^2P_{\text{total}}(\boldsymbol{\nu})\right]^{k\beta-1}}{\Pi_k^{-k\beta}\left[1+(\sigma_e^2+\Sigma_k)P_{\text{total}}(\boldsymbol{\nu})\right]^{k\beta+1}}. \quad (37)$$

In particular, by Theorem 1 we can show that, at the critical point where the number of active channels k increases from ℓ to $\ell+1$ with $1 \leq \ell < M$, the total power is equal to

$$P_{\text{total}}(\boldsymbol{\nu}) = \frac{\ell\Sigma_\ell - \widehat{\gamma}_{\pi(\ell+1)}}{\Sigma_\ell\widehat{\gamma}_{\pi(\ell+1)} - \sigma_e^2(\ell\Sigma_\ell - \widehat{\gamma}_{\pi(\ell+1)})} \quad (38)$$

Substituting Eq. (38) into Eq. (37) and letting *either* $k=\ell$ *or* $k=\ell+1$, the derivative in Eq. (37) yields the *same* solution:

$$-\beta\Pi_\ell^{\ell\beta}\widehat{\gamma}_{\pi(\ell+1)}^{\ell\beta-1}\Sigma_\ell^{-2}\left[\Sigma_\ell\widehat{\gamma}_{\pi(\ell+1)} - \sigma_e^2(\ell\Sigma_\ell - \widehat{\gamma}_{\pi(\ell+1)})\right]^2 \quad (39)$$

which implies that the derivative of the objective function is *continuous* on $P_{\text{total}}(\boldsymbol{\nu}) \in \mathbb{R}_+$, even though the number of active channels *discretely* increases. Once verified the continuity, the twice differentiation for each given k can be easily obtained as[4]

$$\frac{\partial^2\mathcal{F}\left(\boldsymbol{\mu}^*_{\text{total}}(\boldsymbol{\nu}), \widehat{\boldsymbol{\gamma}}\right)}{\partial P^2_{\text{total}}(\boldsymbol{\nu})} = \frac{\beta(k\Sigma_k)^{k\beta+1}\left[1+\sigma_e^2P_{\text{total}}(\boldsymbol{\nu})\right]^{k\beta-2}}{\Pi_k^{-k\beta}\left[1+(\sigma_e^2+\Sigma_k)P_{\text{total}}(\boldsymbol{\nu})\right]^{k\beta+2}} \cdot\left\{\Sigma_k(k\beta+1) + 2\sigma_e^2\left[1+(\sigma_e^2+\Sigma_k)P_{\text{total}}(\boldsymbol{\nu})\right]\right\} > 0 \quad (40)$$

which demonstrates that Eq. (37) is a continuous and monotonically increasing function of $P_{\text{total}}(\boldsymbol{\nu})$. Thus, the objective function is strictly convex, and then the proof follows. ■

Appendix III
Proof of Theorem 2

Proof: Differentiating the Lagrangian function $\mathcal{J}_2$ given by Eq. (19) and setting the derivative equal to zero, we get

$$\frac{\partial\mathcal{F}\left(\boldsymbol{\mu}^*_{\text{total}}(\boldsymbol{\nu}), \widehat{\boldsymbol{\gamma}}\right)}{\partial P_{\text{total}}(\boldsymbol{\nu})} + \lambda_2 = 0. \quad (41)$$

Plugging Eq. (37) into Eq. (41) with simple algebraic manipulations, we obtain Eq. (20), where $\omega^* \triangleq (\beta/\lambda_2)^{\frac{1}{M\beta+1}}$. Similar to the proof of Lemma 2, we can show that the left-hand side of Eq. (20) is continuous and monotonically increasing function at $P_{\text{total}}(\boldsymbol{\nu}) \in \mathbb{R}_+$, even though k changes discretely. Therefore, the positive solution $P^*_{\text{total}}(\boldsymbol{\nu}) \in \mathbb{R}_{++}$ and the corresponding number of active channels k in $\{1, ..., M\}$, if exist, are unique, respectively.

Otherwise, if we cannot find such k and $P^*_{\text{total}}(\boldsymbol{\nu})$ that satisfy the two-step power allocation, from the KKT conditions and the constraint $P_{\text{total}}(\boldsymbol{\nu}) \in \mathbb{R}_+$, we know $P^*_{\text{total}}(\boldsymbol{\nu}) = 0$. Finally, the parameter $\omega^* \in \mathbb{R}_+$ should be chosen such that the average power constraint is satisfied. The proof follows. ■

[4] The twice differentiation is not continuous at the critical point when k changes. Moreover, since $\widehat{\boldsymbol{\gamma}} \in \mathbb{R}_+^M$ follows a certain continuous distribution, we have $\Sigma_k > 0$ and $\Pi_k > 0$ with probability 1.

APPENDIX IV
PROOF OF THEOREM 3

Proof: The proof is based on the following lemma.

Lemma 3: For all k in $\{1, 2, ..., M\}$, the following inequality always holds:

$$k\Sigma_k\Pi_k \leq 1. \tag{42}$$

Proof: It can be shown by definition that $k\Sigma_k$ is the *harmonic mean* of $\{\widehat{\gamma}_{\pi(i)}\}_{i=1}^k$, while Π_k^{-1} is the *geometric mean* of $\{\widehat{\gamma}_{\pi(i)}\}_{i=1}^k$. Using the well known result that the harmonic mean is always less than or equal to the geometric mean, we know the ratio $k\Sigma_k\Pi_k \leq 1$ always holds, with equality if and only if $\widehat{\gamma}_{\pi(1)} = \widehat{\gamma}_{\pi(2)} = \cdots = \widehat{\gamma}_{\pi(k)}$.[5] The proof of Lemma 3 follows. ∎

Now, we prove Theorem 3. If $\omega^* \leq 1$, from Lemma 3 we know $\eta_k = (\omega^*)^{M/k} k\Sigma_k\Pi_k \leq 1$. Then, $P_{\text{total}}(\boldsymbol{\nu}) = 0$ always holds, and $\mathbb{E}_{\widehat{\gamma}}[P_{\text{total}}(\boldsymbol{\nu})] = 0$.

Otherwise, if $\omega^* > 1$, the convergence for the average power is equivalent to the convergence of the average water-level $\omega(\boldsymbol{\nu}, k)$ given in Eq. (14), since the power assigned to each channel cannot exceed the water-level. Substituting Eq. (22) into Eq. (14) and removing the irrelevant terms, we get the following condition:

$$\mathbb{E}_{\widehat{\gamma}}\left[\frac{\Sigma_k\Pi_k}{\Sigma_k - \sigma_e^2(\eta_k - 1)}\right]^+ < \infty. \tag{43}$$

To prove the necessary condition, we need to find a lower-bound of Eq. (43), which is given by

$$\begin{aligned}
&\mathbb{E}_{\widehat{\gamma}}\left[\frac{\Sigma_k\Pi_k}{\Sigma_k - \sigma_e^2(\eta_k - 1)}\right]^+ \\
&= \mathbb{E}_{\widehat{\gamma}}\left[\frac{\Pi_k}{1 - \sigma_e^2(\omega^*)^{M/k}k\Pi_k + \sigma_e^2/\Sigma_k}\right]^+ \\
&\overset{(a)}{\geq} \mathbb{E}_{\widehat{\gamma}}\left[\frac{\Pi_1}{1 - \sigma_e^2\omega^*\Pi_1 + \sigma_e^2/\Sigma_1}\right]^+
\end{aligned} \tag{44}$$

where (a) holds since Π_k is a monotonically increasing function of k, while $(\omega^*)^{M/k}$ and Σ_k are monotonically decreasing functions of k, respectively. Plugging $\Sigma_1 = \widehat{\gamma}_{\pi(1)}$ and $\Pi_1 = 1/\widehat{\gamma}_{\pi(1)}$ into Eq. (44), and upper-bounding it away from infinity, we get the necessary condition given in Eq. (28).

Similarly, to prove the sufficient condition, we need to find an upper-bound of Eq. (43), which is given by

$$\begin{aligned}
&\mathbb{E}_{\widehat{\gamma}}\left[\frac{\Sigma_k\Pi_k}{\Sigma_k - \sigma_e^2(\eta_k - 1)}\right]^+ \\
&\leq \mathbb{E}_{\widehat{\gamma}}\left[\frac{\Pi_k}{1 - \sigma_e^2(\omega^*)^{M/k}k\Pi_k}\right]^+ \\
&\leq \mathbb{E}_{\widehat{\gamma}}\left[\frac{\Pi_M}{1 - \sigma_e^2(\omega^*)^M M\Pi_M}\right]^+.
\end{aligned} \tag{45}$$

Upper-bounding Eq. (45) away from infinity, we get the sufficient condition given in Eq. (29). The proof of Theorem 3 follows. ∎

[5] Since $\Pr\{\widehat{\gamma}_{\pi(1)} = \widehat{\gamma}_{\pi(2)} = \cdots = \widehat{\gamma}_{\pi(k)}\} = 0$ for $k \geq 2$. Therefore, the maximum $k\Sigma_k\Pi_k = 1$ is achieved by $k = 1$ with probability 1.

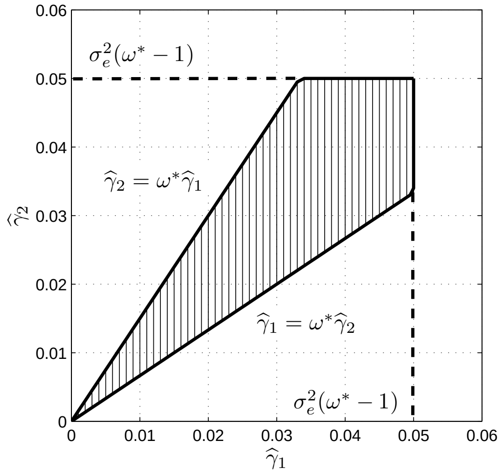

Fig. 7. The constructed outage region (shadowed area) for $M = 2$. The variance $\sigma_e^2 = 0.1$ and water-level coefficient $\omega^* = 1.5$.

APPENDIX V
PROOF OF PROPOSITION 1

Proof: If $\omega^* \leq 1$, from Theorem 3 we know that $P^*_{\text{total}}(\boldsymbol{\nu}) = 0$ always holds. Thus, the spectral efficiency $\mathcal{C}$ given in Eq. (31) is always equal to zero. The outage probability is equal to one. Otherwise, if $\omega^* > 1$, based on Eq. (22), it is *sufficient* to construct a nonempty region such that

$$0 \leq \max_k\{\Sigma_k\} \leq \min_k\left\{\sigma_e^2(\eta_k - 1)\right\}. \tag{46}$$

Inside this region, there is no positive solution for Eq. (22) for all $k \in \{1, 2, ..., M\}$, and thus $P^*_{\text{total}}(\boldsymbol{\nu}) = 0$ always holds. It is clear that in Eq. (46), $\max_k\{\Sigma_k\} = \Sigma_1 = \widehat{\gamma}_{\pi(1)}$. On the other hand,

$$\min_k\left\{\sigma_e^2(\eta_k - 1)\right\} \geq \sigma_e^2\left(\omega^*\widehat{\gamma}_{\pi(M)}/\widehat{\gamma}_{\pi(1)} - 1\right) \tag{47}$$

where the inequality holds since $\min_k\{(\omega^*)^{M/k}\} = \omega^*$ due to $\omega^* > 1$, $\min_k\{k\Sigma_k\} \geq \widehat{\gamma}_{\pi(M)}$ from the definition of the harmonic mean, and $\min_k\{\Pi_k\} = \Pi_1 = 1/\widehat{\gamma}_{\pi(1)}$. Combining Eqs. (46) and (47), we can observe that Eq. (46) always holds as long as the following stronger condition is satisfied:

$$\widehat{\gamma}_{\pi(1)} \leq \sigma_e^2\left(\omega^*\widehat{\gamma}_{\pi(M)}/\widehat{\gamma}_{\pi(1)} - 1\right). \tag{48}$$

Solving the inequalities given in Eq. (48) and noting that $\widehat{\gamma}_{\pi(1)} \geq \widehat{\gamma}_{\pi(M)} \geq 0$, we get the boundary conditions for this region as follows:

$$\begin{cases} 0 \leq \widehat{\gamma}_{\pi(M)} \leq \sigma_e^2(\omega^* - 1) \\ \widehat{\gamma}_{\pi(M)} \leq \widehat{\gamma}_{\pi(1)} \leq \min\left\{\omega^*\widehat{\gamma}_{\pi(M)}, \sigma_e^2(\omega^* - 1)\right\} \end{cases} \tag{49}$$

As long as $\omega^* > 1$ and $\sigma_e^2 > 0$, the probability measure of the region indicated by Eq. (49) is nonzero, which is a *lower-bound* for the outage probability. Therefore, the outage probability is nonzero. The proof follows. As an example, Fig. 7 plots the constructed outage region for the case of $M = 2$.

∎

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Inform. Theory*, New York: Wiley, 1991.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs Tech. Rep. BL0112170-950615-07TM, 1995.

[3] A. J. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, Nov. 1997.

[4] G. Caire, G. Taricco, and E. Biglieri, Optimum power control over fading channels, *IEEE Trans. Inform. Theory*, vol. 45, no. 5, July 1999.

[5] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with mulitple antennas," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

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

[7] J. Tang and X. Zhang, "Quality-of-service driven power and rate adaptation for multichannel communications over wireless links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 12, pp. 4349–4360, December 2007.

[8] 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, Jul. 2003.

[9] D. Wu and R. Negi, "Downlink scheduling in a cellular network for quality-of-service assurance," *IEEE Trans. Veh. Technol*., vol. 53, no. 5, pp. 1547–1557, Sep. 2004.

[10] L. OZarow, S. Shamai, and A. D. Wyner, "Infomration theoretic consideration for cellular mobile radio," *IEEE Trans. Veh. Technol*., vol. 43, ppp. 359–378, May 1994.

[11] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[12] T. E. Klein and R. G. Gallager, "Power control for the additive white Gaussian noise channel under channel estimation errors," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, June 2001, pp. 304.

[13] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, Sep. 2004.

[14] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multicarrier OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Select. Areas Commun*., vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[15] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side information" be?" *IEEE Trans. Inform. Theory*, vol. 48, pp. 1118–1134, May 2002.

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

[17] F. Kelly, S. Zachary, and I. Ziedins, "Stochastic Networks: Theory and Applications", *Royal Statistical Society Lecture Notes Series*, vol. 4, Oxford University Press, pp. 141–168, 1996.

[18] S. Zhou and G. B. Giannakis, "Adaptive modulation for multi-antenna transmissions with channel mean feedback," *IEEE Trans. Wireless Commun*., vol. 3, no. 5, pp. 1626–1636, Sept. 2004.

[19] 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.

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

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 an Assistant Professor and the Founding Director of the Division of Computer Systems Engineering, Department of Electrical Engineering and Computer Science, Beijing Information Technology Engineering Institute, Beijing, China, from 1984 to 1989. 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, Queensland, Australia, under a Fellowship from the Chinese National Commission of Education. He was with the Networks and Distributed Systems Research Department, AT&T Bell Laboratories, Murray Hills, NJ, and with AT&T Laboratories Research, Florham Park, NJ, in 1997. He has published more than 200 research papers in the areas of wireless networks and communications systems, mobile computing, network protocol design and modeling, statistical communications, random signal processing, information theory, and control theory and systems.

Prof. Zhang 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 received the Best Paper Awards in IEEE GLOBECOM 2007, IEEE GLOBECOM 2009, and IEEE WCNC 2010, respectively. He also received the 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. He is serving or has served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, a Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS for the special issue on "Broadband Wireless Communications for High Speed Vehicles", a Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS for the special issue on "Wireless Video Transmissions", an Associate Editor for IEEE COMMUNICATIONS LETTERS, a Guest Editor for IEEE COMMUNICATIONS MAGAZINE for the special issue on "Advances in Cooperative Wireless Networking", and a Guest Editor for IEEE WIRELESS COMMUNICATIONS MAGAZINE for the special issue on "Next Generation of CDMA versus OFDMA for 4G Wireless Applications". He is serving or has served as the TPC Chair for IEEE GLOBECOM 2011, TPC Vice-Chair for IEEE INFOCOM 2010, TPC Co-Chair for IEEE INFOCOM 2009 - Mini-Conference,TPC Co-Chair for IEEE GLOBECOM 2008 - Wireless Communications Symposium,TPC Co-Chair for IEEE ICC 2008 - Information and Network Security Symposium, Poster Chair for IEEE INFOCOM 2008, Student Travel Grants Co-Chair for IEEE INFOCOM 2007, etc.

Jia Tang (S'03) received the B.S. degree in Electrical Engineering from Xi'an Jiaotong University, Xi'an, China, in 2001. He is currently a research assistant working toward the Ph.D. degree in Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA.

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.

Mr. Tang received Fouraker Graduate Research Fellowship Award from Department of Electrical and Computer Engineering, Texas A&M University in 2005.