Quality-of-Service Driven Power and Rate Adaptation over Wireless Links

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

***Abstract*— We propose a quality-of-service (QoS) driven power and rate adaptation scheme over wireless links in mobile wireless networks. Specifically, our proposed scheme aims at maximizing the system throughput subject to a given delay QoS constraint. First, we derive an optimal adaptation policy by integrating information theory with the concept of *effective capacity* for a block fading channel model. Our analyses reveal an important fact that there exists a fundamental tradeoff between throughput and QoS provisioning. In particular, when the QoS constraint becomes loose, the optimal power-control policy converges to the well-known *water-filling* scheme, where Shannon (ergodic) capacity can be achieved. On the other hand, when the QoS constraint gets stringent, the optimal policy converges to the *total channel inversion* scheme under which the system operates at a constant rate. Inspired by the above observations, we then consider a more practical scenario where variable-power adaptive modulation is employed over both block fading and Markov correlated fading channels. In both cases, we derive the associated power and rate adaptation policies. The obtained results suggest that the channel correlation has a significant impact on QoS-driven power and rate adaptations. The higher the correlation is, the faster the power-control policy converges to the total channel inversion when the QoS constraint becomes more stringent. Finally, we conduct simulations to verify that the adaptation policy proposed for Markov channel models can also be applied to the more general channel models.**

***Index Terms*— Mobile wireless networks, quality-of-service (QoS), effective capacity, power control, adaptive modulation, information theory, cross-layer design and optimization.**

I. Introduction

QUALITY-OF-SERVICE (QoS) guarantees play a critically important role in future mobile wireless networks. Depending on their distinct QoS requirements, differentiated mobile users are expected to tolerate different levels of delay for their service satisfactions. For instance, non-real-time services such as data disseminations aim at maximizing the throughput with a loose delay constraint. In contrast, for real-time services like multimedia video conference, the key QoS metric is to ensure a stringent delay-bound, rather than to achieve high spectral efficiency. There also exist some services falling in between, e.g., paging and interactive web surfing, which are delay-sensitive but whose delay QoS requirements are not as stringent as those of real-time applications. The diverse mobile users impose totally different and sometimes even conflicting delay QoS constraints, which impose great challenges to the design of future mobile wireless networks.

Unlike its wired counterparts, supporting diverse delay QoS in wireless environment is much more challenging since the wireless channel has a significant impact on network performance. In particular, a deterministic delay-bound QoS guarantee over wireless networks is practically infeasible due to the time-varying nature of fading channels. Alternatively, a more practical solution is to provide the *statistical QoS guarantees* [1], where we guarantee the given delay-bound with a small violation probability.

Furthermore, for wireless communications, the most scarce radio resources are power and spectral bandwidth [2]. As a result, a great deal of research has been devoted to the techniques that can enhance the spectral efficiency of wireless systems [3]. The framework used to evaluate these techniques is mainly based on information theory, using the concept of Shannon capacity [4]. Among a large number of promising schemes, power and rate adaptation has been widely considered as one of the key solutions to improve the spectral efficiency. In [5], [6], the authors showed that the optimal power and rate control policy which maximizes spectral efficiency is the so-called *water-filling* algorithm. The water-filling scheme assigns more power when the channel is in good condition and less power when the channel becomes worse. In the case that the channel quality is below a certain threshold, no information is transmitted. On the other hand, a different idea of power and rate adaptation is the scheme referred as *total channel inversion* [5], [6], where the system assigns more power to combat with deep fading and less power for the good channel in order to maintain a constant signal-to-noise ratio (SNR), such that a constant rate service process can be obtained. Clearly, from the information-theoretic viewpoint, water-filling is better than total channel inversion since the former provides higher spectral efficiency. However, a natural question that follows is whether the former is also better than the latter in terms of QoS guarantees?

It is important to note that Shannon theory does not place any restrictions on complexity and delay [6]. Consequently, in order to answer the above question, it is necessary to take the QoS metrics into account when applying the prevalent information-theoretic results. Thanks to the dual concepts of *effective bandwidth* and *effective capacity*, we obtain a powerful approach to evaluate the statistical QoS performance from the networking perspective. The effective-bandwidth theory has been extensively studied in the early 90's with the emphasis on wired asynchronous transfer mode (ATM) networks [7]–[11]. This theory enables us to analyze network statistics

Manuscript received January 5, 2006; revised May 15, 2006; accepted May 18, 2006. The associate editor coordinating the review of this paper and approving it for publication was M. Guizani. The research reported in this paper was supported in part by the National Science Foundation CAREER Award under Grant ECS-0348694. A subset of the materials in this paper was presented at the IEEE International Conference on Communications (IEEE ICC), Istanbul, Turkey, June 2006.

The authors are with the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: {jtang, xizhang}@ece.tamu.edu).

Digital Object Identifier 10.1109/TWC.2007.051075.

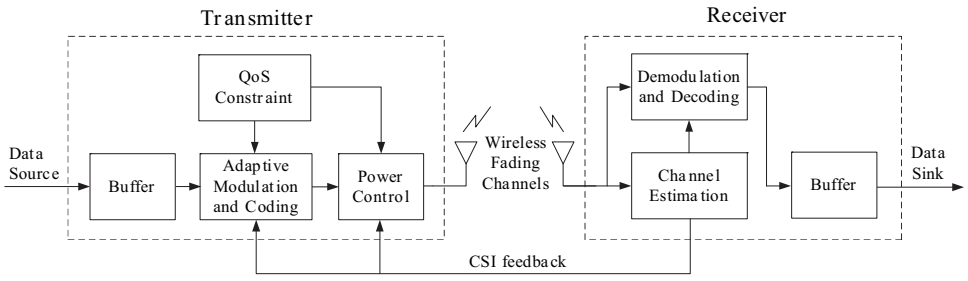


Fig. 1. The system model.

such as queue distributions, buffer overflow probabilities, and delay-bound violation probabilities, which are important for statistical QoS guarantees. In [12], Wu and Negi proposed an interesting concept termed effective capacity, which turns out to be the dual of the effective bandwidth. The effective-capacity approach is particularly convenient for analyzing the statistical QoS performance of wireless transmissions where the service process is driven by the time-varying wireless channel.

Integrating information theory with the effective capacity, in this paper we investigate the QoS-driven power and rate adaptation over wireless links in mobile wireless networks. The problem we are interested in is how to maximize the throughput subject to a given delay QoS constraint. We first focus on uncorrelated fading channels (also termed *block fading* or *quasi-static* fading channels) and investigate corresponding power and rate adaptation polices. Our analyses reveal an important fact that there exists a fundamental tradeoff between the throughput and the QoS provisioning. In particular, the higher throughput gain comes at the price of sacrificing more QoS provisioning, and vice versa. When the QoS constraint becomes loose, the optimal power-control law converges to the water-filling scheme, where Shannon (ergodic) capacity can be achieved. On the other hand, when the QoS constraint gets stringent, the optimal power-control law converges to the total channel inversion such that the system operates at a constant service rate. Motivated by the above observations, we then consider a more practical scenario where variable-power adaptive-modulation is applied over both uncorrelated and correlated fading channels. For simplicity, we use finite-state Markov chain (FSMC) to model the correlated channel processes. The FSMC-based channel model was previously proposed by Wang and Moayeri [13]. Then, this model has been extensively studied for both Rayleigh and Nakagami-m fading channel (e.g., see [14], [15] and references therein). For both block fading and FSMC-correlated fading channels, we derive the corresponding power and rate adaptation policies. Our obtained results suggest that channel correlation has a significant impact on QoS-driven power and rate allocations. The higher the correlation is, the faster the power-control policy converges to the total channel inversion as the QoS constraint becomes more stringent. Finally, we conduct simulations to verify that although the FSMC-based channel model is not perfectly accurate, the power-control law derived from it can be applied to the more general Jake's channel model [16], which has been widely used and extensively studied in literatures.

The rest of the paper is organized as follows. Section II describes our system model. Sections III introduces the statistical QoS guarantees and the concept of effective capacity. Section IV develops the optimal power and rate adaptation scheme that can maximize the effective capacity. Section V applies the above analyses to a more practical adaptive modulation-based scheme. Section VI discusses the impact of channel correlation on the power and rate adaptations. Section VII conducts simulations to evaluate the validity of our proposed adaptive schemes on the more general Jake's channel model. The paper concludes with Section VIII.

II. System Model

The system model is illustrated in Fig. 1. We concentrate on the discrete-time system over a point-to-point wireless link between the transmitter and the receiver. Let us denote the system's total spectral bandwidth by B, the mean transmit power by $\overline{P}$, and the power density of the complex additive white Gaussian noise (AWGN) by $N_0/2$ per dimension, respectively. First, the upper-protocol-layer packets are divided into *frames* at the datalink layer, which forms the "data source" as shown in Fig. 1. We assume that the frames have the same time duration, which is denoted by T_f. The frames are stored at the transmit buffer and then split into bit-streams at the physical layer. Based on the QoS constraint and the channel-state information (CSI) fed back from the receiver, the adaptive modulation and power control are employed, respectively, at the transmitter. The reverse operations are executed at the receiver side. Finally, the frames are recovered at the "data sink" for further processing.

The discrete-time channel fading process is assumed to be stationary and ergodic, which is invariant within a frame's time-duration T_f, but varies from one frame to another. Moreover, the wireless channel is flat-fading with its envelope following Nakagami-m distribution.[1] The Nakagami-m channel model is very general and often best fits land-mobile and indoor-mobile multipath propagations [17], [18]. As the parameter m varies, where $m \in [1/2, +\infty)$, the model spans a wide range of fading environments, including one-sided Gaussian fading channel ($m = 1/2$, the worst fading case), Rayleigh fading channel ($m = 1$), approximations of Rician and lognormal fading channels ($m > 1$), and additive white Gaussian noise (AWGN) channel ($m = \infty$, no fading).

Denote the channel envelope process by $\{\alpha[i], i = 1, 2, ...\}$, where i is the time index of the frame. If we use constant

[1]The power and rate adaptation scheme discussed in this paper can be applied to any other continuous channel distributions. We use Nakagami-m distribution in this paper as a general example.

power assignment, then the instantaneous transmit power, denoted by $P[i]$, is equal to $P[i] = \overline{P}$. The instantaneous received SNR, denoted by $\gamma[i]$, can be expressed as $\gamma[i] = \overline{P}\alpha^2[i]/(N_0B)$, with its mean $\overline{\gamma} = \overline{P}\mathbb{E}\{\alpha^2[i]\}/(N_0B)$, where $\mathbb{E}\{\cdot\}$ denotes the expectation. The probability density function (pdf) of $\gamma[i]$, denoted by $p_\Gamma(\gamma)$, can be expressed as [17]

$$p_\Gamma(\gamma) = \frac{\gamma^{m-1}}{\Gamma(m)}\left(\frac{m}{\overline{\gamma}}\right)^m \exp\left(-\frac{m}{\overline{\gamma}}\gamma\right),\ \gamma \geq 0 \tag{1}$$

where $\Gamma(\cdot)$ represents the Gamma function and m denotes the fading parameter of Nakagami-m distribution. Throughout this paper, we assume that the CSI is perfectly estimated at the receiver and reliably fed back to the transmitter without delay. The discussions of the imperfect CSI are not the focus of this paper. We also assume that the datalink-layer buffer size is infinite.

III. Statistical QoS Guarantees and Effective Capacity

A. The Concept of Statistical QoS Guarantees

During the early 90's, statistical QoS guarantees have been extensively studied in the contexts of effective bandwidth theory [7]–[11]. The literature on effective bandwidth is abundant. The readers are referred to Chang [7] and Kelly *et. al.* [8] for a comprehensive review.

Based on large deviation principle (LDP), Chang showed that [7], for a dynamic queueing system with stationary ergodic arrival and service processes, under sufficient conditions, the queue length process $Q(t)$ converges in distribution to a random variable (r.v.) $Q(\infty)$ such that

$$-\lim_{x\to\infty}\frac{\log\left(\Pr\{Q(\infty) > x\}\right)}{x} = \theta. \tag{2}$$

To be more specific, the above theorem states that the probability of the queue length exceeding a certain threshold x decays exponentially fast as the threshold x increases. Note that in (2), the parameter θ $(\theta > 0)$ plays a critically important role for statistical QoS guarantees, which indicates the exponential decay rate of the QoS violation probabilities. A smaller θ corresponds to a slower decay rate, which implies that the system can only provide a *looser* QoS guarantee, while a larger θ leads to a faster decay rate, which means that a more *stringent* QoS requirement can be supported. In particular, when $\theta \to 0$, the system can tolerate an arbitrarily long delay, which corresponds to the scenario studied in information theory. On the other hand, when $\theta \to \infty$, the system cannot tolerate any delay, which corresponds to an extremely stringent delay-bound. Due to its close relationship with statistical QoS provisioning, θ is called the *QoS exponent* [12]. Based on the concept of QoS exponent, the effective bandwidth is defined as the minimum constant service rate required by a given arrival process for which the QoS exponent θ is fulfilled.

B. The Effective Capacity

Inspired by the effective bandwidth, Wu and Negi proposed *effective capacity* [12], which is the dual of the original effective bandwidth. The effective capacity is defined as the maximum constant arrival rate that a given service process can support in order to guarantee a QoS requirement specified by θ. Analytically, the effective capacity can be formally defined as follows.

Let the sequence $\{R[i],\ i = 1, 2, ...\}$ denote a discrete-time stationary and ergodic stochastic service process and $S[t] \triangleq \sum_{i=1}^{t} R[i]$ be the partial sum of the service process over time sequence of $i = 1, 2, ..., t$. Assume that the Gärtner-Ellis limit of $S[t]$, expressed as $\Lambda_{\rm C}(\theta) = \lim_{t\to\infty}(1/t)\log\left(\mathbb{E}\left\{e^{\theta S[t]}\right\}\right)$ exists and is a convex function differentiable for all real θ [7, pp. 921]. Then, the effective capacity of the service process, denoted by $\mathrm{E_C}(\theta)$, where $\theta > 0$, is defined as [12, eq. (12)]

$$\mathrm{E_C}(\theta) \triangleq -\frac{\Lambda_{\rm C}(-\theta)}{\theta} = -\lim_{t\to\infty}\frac{1}{\theta t}\log\Big(\mathbb{E}\left\{e^{-\theta S[t]}\right\}\Big). \tag{3}$$

When the sequence $\{R[i],\ i = 1, 2, ...\}$ is an uncorrelated process, it is clear that the effective capacity $\mathrm{E_C}(\theta)$ reduces to

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

In this paper, our original problem is maximizing the throughput subject to a given delay QoS constraint. Notice that the effective capacity can be considered as the maximal throughput under the constraint of QoS exponent θ. Therefore, by interpreting θ as the QoS constraint in our original problem, we can formulate an equivalent new problem, which is to maximize the effective capacity for a given θ. In the following sections, we will focus on this new problem and design the corresponding resource allocation algorithms.

IV. Optimal Power and Rate Adaptation for QoS Guarantees

Conventionally, the power-control law can be expressed as a function of the instantaneous SNR $\gamma[i]$. However, our power-adaptation policy, denoted by $\mu\left(\theta, \gamma[i]\right)$, is a function of not only the instantaneous SNR $\gamma[i]$, but also the QoS exponent θ. Applying the power adaptation, the instantaneous transmit power becomes $P[i] = \mu\left(\theta, \gamma[i]\right)\overline{P}$. Note that the mean transmit power is upper-bounded by $\overline{P}$. Therefore, the power-control law needs to satisfy the mean power constraint:

$$\int_0^\infty \mu(\theta, \gamma)p_\Gamma(\gamma)d\gamma \leq 1,\ \text{for all } \theta > 0. \tag{5}$$

In this section, we also make the following two assumptions:

A1: We first assume that the channel is block fading. We make such an assumption due to the following reasons. First, the effective capacity expression (4) in uncorrelated case only depends on marginal statistics of a service process, which is much simpler than the general expression given by (3), where the higher order statistics of the service process are required. Second, we will show in Section VI that the resource allocation policy derived for block fading channel can be applied to correlated fading channel with a certain modifications.

A2: We further assume that given the instantaneous SNR $\gamma[i]$ and the corresponding power-control law $\mu\left(\theta, \gamma[i]\right)$, the adaptive modulation and coding scheme can achieve the

instantaneous capacity. Thus, the instantaneous service rate $R[i]$ of the frame i can be expressed as[2]

$$R[i] = T_f B \log_2\Big(1 + \mu\left(\theta, \gamma[i]\right)\gamma[i]\Big). \tag{6}$$

In the following discussions, we omit the discrete time-index i for simplicity. Using (4), (5), and (6), we can formally formulate our maximization problem as follows:

$$\begin{aligned} \mathrm{E}_{\mathrm{C}}^{\mathrm{opt}}(\theta) \quad &= \max_{\mu(\theta,\gamma):\int_0^\infty \mu(\theta,\gamma)p_\Gamma(\gamma)d\gamma=1} \left\{ -\frac{1}{\theta} \right. \\ &\quad \left. \cdot \log\left(\int_0^\infty e^{-\theta T_f B \log_2\left(1+\mu(\theta,\gamma)\gamma\right)} p_\Gamma(\gamma)d\gamma\right)\right\}. \end{aligned} \tag{7}$$

where $\mathrm{E}_{\mathrm{C}}^{\mathrm{opt}}(\theta)$ denotes the maximum effective capacity achieved by the optimal policy. We derive the following theorem to characterize the optimal power and rate adaptation policy.

Theorem 1: The optimal power-control policy, denoted by $\mu_{\mathrm{opt}}(\theta,\gamma)$, which maximizes the effective capacity given in (7), is determined by

$$\mu_{\mathrm{opt}}(\theta,\gamma) = \begin{cases} \dfrac{1}{\gamma_0^{\frac{1}{\beta+1}}\gamma^{\frac{\beta}{\beta+1}}} - \dfrac{1}{\gamma}, & \gamma \geq \gamma_0 \\ 0, & \gamma < \gamma_0 \end{cases} \tag{8}$$

where we define $\beta \triangleq \theta T_f B/\log 2$ as the *normalized* QoS exponent and γ_0 as the cutoff SNR threshold, which can be numerically obtained by meeting the mean power constraint:

$$\int_{\gamma_0}^{\infty} \left(\frac{1}{\gamma_0^{\frac{1}{\beta+1}}\gamma^{\frac{\beta}{\beta+1}}} - \frac{1}{\gamma}\right) p_\Gamma(\gamma)d\gamma = 1. \tag{9}$$

Proof: The proof is provided in Appendix I. ■

Theorem 1 gives the optimal power-control policy which maximizes the effective capacity. We can observe from (8) that as $\theta \to 0$, the optimal policy $\mu_{\mathrm{opt}}(\theta,\gamma)$ converges to

$$\lim_{\theta\to 0}\mu_{\mathrm{opt}}(\theta,\gamma) = \begin{cases} \dfrac{1}{\gamma_0} - \dfrac{1}{\gamma}, & \gamma \geq \gamma_0 \\ 0, & \gamma < \gamma_0 \end{cases} \tag{10}$$

which is just the water-filling formula in [6, eq. (5)]. Thus, our QoS-driven power and rate adaptation scheme reduces to the water-filling algorithm when the system can tolerate an arbitrarily long delay, which is expected since water-filling is well-known the optimal power allocation strategy without delay constraint. On the other hand, as the QoS exponent $\theta \to \infty$, the cutoff threshold $\gamma_0 \to 0$ (note that $\gamma_0 = \lambda/\beta$ as detailed in Appendix I). Therefore, the system does not enter the outage state almost surely. The optimal power control $\mu_{\mathrm{opt}}(\theta,\gamma)$ converges to

$$\lim_{\theta\to\infty}\mu_{\mathrm{opt}}(\theta,\gamma) = \frac{\sigma}{\gamma} \tag{11}$$

where $\sigma = (m-1)\overline{\gamma}/m$ for $m \geq 1$, which becomes the policy of the total channel inversion. Thus, for stringent delay QoS constraints, the optimal power control becomes the total channel inversion. Note that if the fading parameter $m < 1$, implying that the fading is severer than Rayleigh, then no total channel inversion scheme exists since the transmit power is not enough to totally invert the channel. In this case, the cutoff threshold γ_0 will converge to a small positive number as $\theta \to \infty$. Thus, the optimal power and rate adaptation policy becomes *truncated channel inversion* [6]. It is also worth noting that the optimal power-adaptation policy $\mu_{\mathrm{opt}}(\theta,\gamma)$ depends on frame duration T_f and spectral bandwidth B through the parameter β, where a system with larger value of $T_f B$ can support more stringent QoS requirements.

[2]Note that in our model, the unit for the service rate $R[i]$ and the effective capacity $\mathrm{E}_{\mathrm{C}}(\theta)$ is "bits per frame".

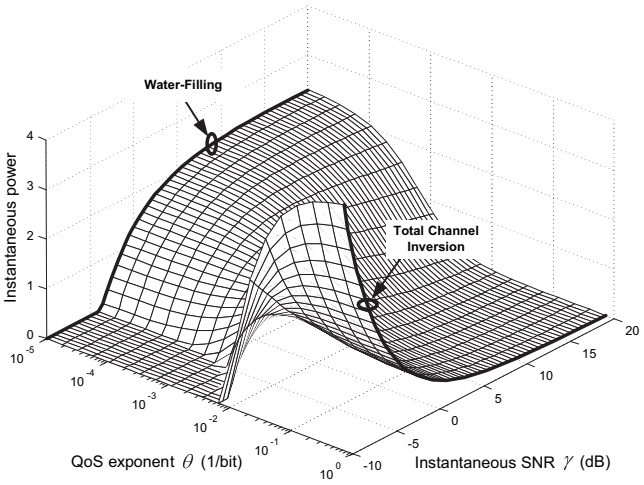


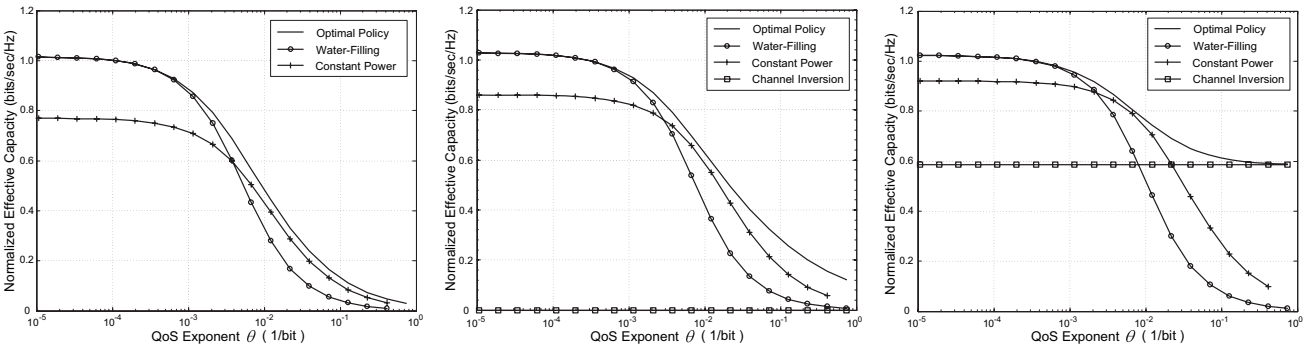
Fig. 2. The optimal power-adaptation policy. The fading parameter $m = 2$ and the average SNR $\overline{\gamma} = 0$ dB.

In all the following numerical solutions or simulation results, which are presented in Fig. 2 through Fig. 9, we set the frame duration $T_f = 2$ ms and the spectral bandwidth $B = 10^5$ Hz. The other system parameters are detailed respectively in each of these figures. Using (8), we plot the instantaneous power assignments of the optimal power-adaptation policy in Fig. 2. We can observe from Fig. 2 that for small θ, the power control assigns more power to the better channel and less power to the worse channel. In contrast, for large θ, the power control assigns less power to the better channel, but more power to the worse channel. As the QoS exponent θ varies between $(0,\infty)$, reflecting different delay QoS constraints, the corresponding optimal power-adaptation policy swings between the water-filling and the total channel inversion schemes.

Given the optimal power and rate adaptation policy, we can derive the closed-form expression for the maximum effective capacity $\mathrm{E}_{\mathrm{C}}^{\mathrm{opt}}(\theta)$ as follows:

$$\begin{aligned} \mathrm{E}_{\mathrm{C}}^{\mathrm{opt}}(\theta) &= -\frac{1}{\theta}\log\left(\int_0^\infty e^{-\beta\log\left(1+\mu_{\mathrm{opt}}(\theta,\gamma)\gamma\right)} p_\Gamma(\gamma)d\gamma\right) \\ &= -\frac{1}{\theta}\left\{\log\left(\left[\frac{m\gamma_0}{\overline{\gamma}}\right]^{\frac{\beta}{\beta+1}}\Gamma\left(m-\frac{\beta}{\beta+1},\frac{m\gamma_0}{\overline{\gamma}}\right)\right.\right. \\ &\quad \left.\left. + \gamma\left(m,\frac{m}{\overline{\gamma}}\gamma_0\right)\right) - \log\Big(\Gamma(m)\Big)\right\}. \end{aligned} \tag{12}$$

where $\gamma(\cdot,\cdot)$ and $\Gamma(\cdot,\cdot)$ denote the lower and upper incomplete Gamma functions, respectively.



(a) $m = 0.5$ (One-sided Gaussian). (b) $m = 1$ (Rayleigh). (c) $m = 2$.

Fig. 3. The Shannon-capacity-based effective capacity under different power and rate adaptation policies. The average SNR $\overline{\gamma} = 0$ dB.

For comparison purposes, we also derive the closed-form expressions of the effective capacity for other commonly used power-control policies, including the water-filling scheme, the constant power approach, and the total channel inversion. Omitting the derivation details, we obtain the closed-form expressions of the the effective capacity for water-filling, denoted by $\mathrm{E}_{\mathrm{C}}^{\mathrm{WF}}(\theta)$, as follows:

$$\mathrm{E}_{\mathrm{C}}^{\mathrm{WF}}(\theta) = -\frac{1}{\theta}\left\{\log\left(\left[\frac{m\gamma_0}{\overline{\gamma}}\right]^{\beta}\Gamma\left(m-\beta, \frac{m\gamma_0}{\overline{\gamma}}\right) + \gamma\left(m, \frac{m}{\overline{\gamma}}\gamma_0\right)\right) - \log\Big(\Gamma(m)\Big)\right\}, \quad (13)$$

and the effective capacity for the constant power approach, denoted by $\mathrm{E}_{\mathrm{C}}^{\mathrm{const}}(\theta)$, as follows:

$$\mathrm{E}_{\mathrm{C}}^{\mathrm{const}}(\theta) = -\frac{1}{\theta}\log\left(\frac{\Gamma(\beta-m)}{\Gamma(\beta)}\left(\frac{m}{\overline{\gamma}}\right)^{m} \cdot {}_1F_1\left(m; m-\beta+1; \frac{m}{\overline{\gamma}}\right) + \frac{\Gamma(m-\beta)}{\Gamma(m)} \cdot \left(\frac{m}{\overline{\gamma}}\right)^{\beta} {}_1F_1\left(\beta; \beta-m+1; \frac{m}{\overline{\gamma}}\right)\right) \quad (14)$$

respectively, where ${}_1F_1(\cdot;\cdot;\cdot)$ denotes the confluent Hypergeometric function [19]. Finally, the effective capacity of total channel inversion is simply a constant equal to $T_f B \log_2(1+\sigma)$.

The normalized effective capacity (which is defined as the effective capacity divided by B and T_f, and thus has the unit of "bits/sec/Hz") comparisons between different power and rate adaptation schemes are shown in Fig. 3. As expected, our proposed optimal power and rate adaptation always achieves the maximum effective capacity among all control policies. The optimal scheme converges to the water-filling for small θ and to the total channel inversion for large θ (when the total channel inversion exists). Note that for one-sided Gaussian channel ($m = 0.5$) and Rayleigh channel ($m = 1$), even using the optimal policy, the effective capacity also converges to zero as the QoS exponent $\theta \to \infty$. However, this is the best that the power control can do to maximize the effective capacity. This implies that no matter how much power and spectral bandwidth resource are assigned and no matter how elegant coding/modulation is employed, if no other technique (e.g., diversity or multiplexing) helps to compensate for the fading effect, Nakagami-m channels with $m \leq 1$ cannot support stringent delay QoS requirement when θ is large, which is also coincident with the fact that the zero-outage capacity of Rayleigh fading channel is zero [20].

V. QoS-Driven Power and Rate Adaptation for Adaptive MQAM

Based on Shannon theory and the concept of effective capacity, Section IV discusses the resource allocation when using ideal channel codes. In this section, we study the scenario where the transmitter employs adaptive MQAM modulation.

A. Continuous Constellation MQAM

We first assume that there is no restriction on the constellation size of adaptive MQAM, which implies that the rate of the service process can be adapted continuously. In [6], Goldsmith and Chua showed that the continuous rate adaptive MQAM has a constant power loss as compared to the Shannon capacity, where the constant only depends on bit-error rate (BER) requirement. Specifically, for each given received SNR γ and power-control policy $\mu(\theta,\gamma)$, the corresponding constellation size, denoted by $M(\gamma)$, is determined by [6, eq. (20)]

$$M(\gamma) = 1 + K\mu(\theta,\gamma)\gamma \quad (15)$$

where K is defined as $K \triangleq -1.5/\log(5\mathrm{BER})$, with BER denoting the required bit-error rate. Note that continuous rate MQAM is originally proposed to investigate the insight relationship between the Shannon capacity and the achievable spectral efficiency of MQAM modulation [6]. In practice, the constellation size $M(\gamma)$ can only be selected from a finite discrete set, which will be detailed in Section V-B.

Using (15), the service rate of continuous rate MQAM, denoted by R_{M}, can be expressed as

$$R_{\mathrm{M}} = T_f B \log_2\Big(1 + K\mu(\theta,\gamma)\gamma\Big). \quad (16)$$

Comparing (16) with (6), we can find that the only difference between these two is a constant power loss of K. Thus, the problem of maximizing the effective capacity for continuous MQAM can be solved in a similar manner to that for deriving the maximum effective-capacity in Section IV. Skipping the detailed derivations, we obtain the optimal power and rate

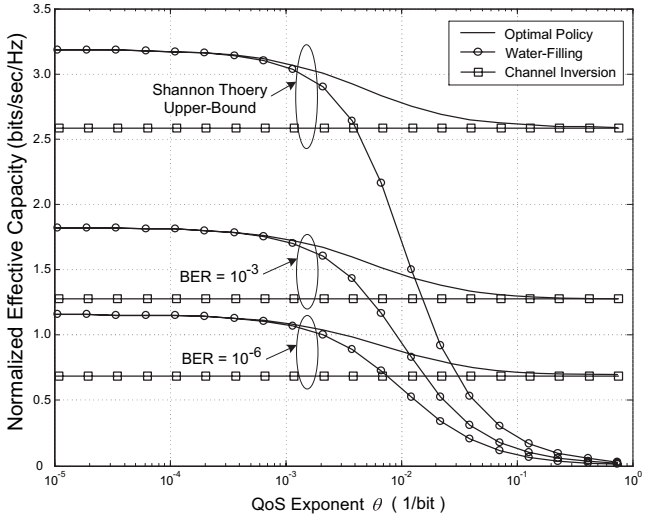


Fig. 4. The effective capacity with adaptive MQAM. The average SNR $\overline{\gamma} = 10$ dB and the fading parameter $m = 2$.

adaptation policy for continuous rate adaptive MQAM, denoted by $\mu_{\rm opt}^{\rm M}(\theta,\gamma)$, as follows:

$$\mu_{\rm opt}^{\rm M}(\theta,\gamma) = \begin{cases} \dfrac{1}{K\gamma_K^{\frac{1}{\beta+1}}\gamma^{\frac{\beta}{\beta+1}}} - \dfrac{1}{K\gamma}, & \gamma \geq \gamma_K \\ 0, & \gamma < \gamma_K \end{cases} \tag{17}$$

where γ_K is the new cutoff threshold, which needs to meet the mean power constraint:

$$\int_{\gamma_K}^{\infty} \left(\frac{1}{K\gamma_K^{\frac{1}{\beta+1}}\gamma^{\frac{\beta}{\beta+1}}} - \frac{1}{K\gamma} \right) p_\Gamma(\gamma) d\gamma = 1. \tag{18}$$

Once γ_K is obtained, we can show the resulting expression of the effective capacity is the same as (12), except that γ_0 in (12) should be replaced by γ_K. It is clear that the power adaptation law of (17) also follows the same trends as (8), which adjusts the power assignment between the water-filling and the total channel inversion, depending on the specific value of θ. Similarly, for the other non-optimal power and rate adaptation policies, we can also derive their corresponding effective capacity expressions, which are omitted for lack of space, but are evaluated by the numerical solutions as shown in Fig. 4.

Fig. 4 illustrates the normalized effective capacity comparisons between the Shannon theory-based upper-bound and the continuous rate adaptive MQAM. We can observe that as the BER requirement becomes more stringent, the effective capacities of both optimal and non-optimal schemes decrease accordingly. However, our proposed power and rate adaptations are always the optimal schemes in each group with the same BER requirement. Agreeing with our observations, the effective capacities of the optimal scheme converge to the water-filling for small θ and to the total channel inversion for large θ, respectively.

B. Discrete Constellation MQAM

The continuous rate assumption for the adaptive MQAM is not too practical. In this section, we relax this assumption by requesting that there be only N possible constellation sizes available. Specifically, we partition the entire SNR range by N non-overlapping consecutive intervals, resulting in $N+1$ boundary points denoted by $\{\Gamma_n\}_{n=0}^{N}$, where $\Gamma_0 < \Gamma_1 < \cdots < \Gamma_N$ with $\Gamma_0 = 0$ and $\Gamma_N = \infty$. Correspondingly, the adaptive modulation is selected to be in mode n if the SNR γ falls into the range $\Gamma_n \leq \gamma < \Gamma_{n+1}$. The constellation used for the zero-th mode is $M_0 = 0$ and for the nth mode is M_n-QAM, where $M_n = 2^n$ with $n = 1, 2, ..., N-1$. Thus, the spectral efficiency by using the nth mode is n bits/sec/Hz. The service rate of the nth mode, denoted by ν_n, is given by

$$\nu_n = T_f B n, \quad \text{for } n = 0, 1, ..., N-1. \tag{19}$$

To find the optimal power and rate adaptation policy for discrete rate adaptive MQAM, we first need to know how to choose the boundary points $\{\Gamma_n\}_{n=1}^{N-1}$ to maximize the effective capacity. Substituting (17) into (15), we get

$$M(\gamma) = \left(\frac{\gamma}{\gamma_K}\right)^{\frac{1}{\beta+1}} \Longrightarrow \gamma = [M(\gamma)]^{\beta+1}\gamma_K. \tag{20}$$

Although (20) is originally derived from continuous rate adaptive MQAM, it provides the guideline in choosing the boundaries for the discrete rate MQAM. Based on (20), we obtain the SNR boundaries $\{\Gamma_n\}_{n=1}^{N-1}$ for discrete rate MQAM as follows:

$$\Gamma_n = M_n^{\beta+1}\gamma_K^* \tag{21}$$

where γ_K^* denotes the new cutoff threshold for discrete rate MQAM. For each given γ_K^*, the boundaries $\{\Gamma_n\}_{n=1}^{N-1}$ are determined by (21). Then, the power-control policy is to retain a constant power for each mode $n > 0$ such that the BER requirements are satisfied. Thus, we obtain the optimal policy for the nth mode, denoted by $\mu_{\rm opt}^{n}(\theta,\gamma)$, as follows:

$$\mu_{\rm opt}^{n}(\theta,\gamma) = \begin{cases} \dfrac{(M_n - 1)}{K\gamma}, & 1 \leq n \leq N-1 \\ 0, & n = 0. \end{cases} \tag{22}$$

Let us further define $M_N \triangleq \infty$. Then, the cutoff threshold γ_K^* is determined by the mean power constraint

$$\sum_{n=0}^{N-1} \int_{M_n^{\beta+1}\gamma_K^*}^{M_{n+1}^{\beta+1}\gamma_K^*} \mu_n(\theta,\gamma) p_\Gamma(\gamma) d\gamma = 1 \tag{23}$$

which can be solved numerically.

We can observe from (21) that as $\theta \to 0$, the boundary selection policy becomes:

$$\lim_{\theta \to 0} \Gamma_n = M_n \gamma_K^* \tag{24}$$

which is same as the selection policy for the discrete rate water-filling algorithm [6, eq. (29)]. On the other hand, as $\theta \to \infty$, the threshold γ_K^* vanishes to zero, making mode 0 infinitely small. At the same time, one of the other $(N-1)$ modes dominates the entire SNR range. Again, the power-control policy converges to the total channel inversion in this case.

Using (22), we plot the instantaneous power assignments of the optimal power-adaptation policy in Fig. 5. We can observe from Fig. 5 that the power control curve has the zigzag shape

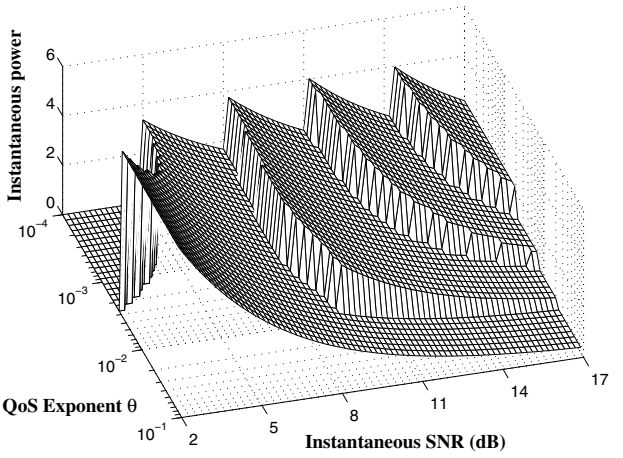


Fig. 5. The power-adaptation strategy for discrete rate adaptive MQAM. The average SNR $\overline{\gamma} = 10$ dB, the fading parameter $m = 2$, BER$= 10^{-3}$, and the number of modes $N = 5$.

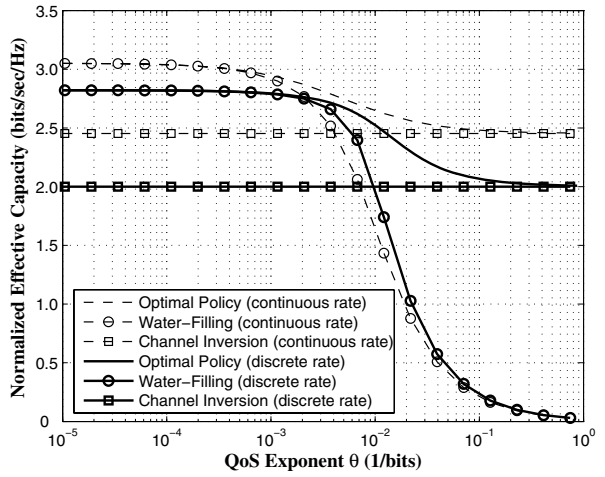


Fig. 6. The effective capacity with continuous rate and discrete rate using different power-control policies. The average SNR $\overline{\gamma} = 15$ dB, the fading parameter $m = 2$, BER$= 10^{-3}$, and the number of modes $N = 4$.

due to the constellation constraint. However, the power control policy still varies between the discrete rate water-filling at small θ and the total channel inversion at large θ, which is consistent with that of the continuous rate MQAM.

Using (4), we derive the effective capacity under the optimal policy, denoted by $\widehat{\mathrm{E}}_{\mathrm{C}}^{\mathrm{opt}}(\theta)$, as follows:

$$\widehat{\mathrm{E}}_{\mathrm{C}}^{\mathrm{opt}}(\theta) = -\frac{1}{\theta}\log\left(\sum_{n=0}^{N-1}\pi_n e^{-\theta\nu_n}\right) \tag{25}$$

where ν_n is given by (19) and

$$\begin{aligned}\pi_n &= \int_{\Gamma_n}^{\Gamma_{n+1}} p_\Gamma(\gamma)d\gamma \\ &= \frac{1}{\Gamma(m)}\left[\gamma\left(m,\frac{m}{\overline{\gamma}}\Gamma_{n+1}\right)-\gamma\left(m,\frac{m}{\overline{\gamma}}\Gamma_n\right)\right]\end{aligned} \tag{26}$$

with $\{\Gamma_n\}_{n=0}^{N}$ given by (21).

Fig. 6 compares the normalized effective capacities between continuous rate MQAM and discrete rate MQAM using both the optimal and non-optimal policies. As shown by Fig. 6, the discrete rate MQAM under the optimal policy suffers from a certain loss in performance as compared to the continuous rate MQAM due to the discrete constellation constraint. However, such a performance loss is not significant. An interesting phenomenon is that the effective capacity of discrete rate water-filling is even larger than that of the continuous rate water-filling for large θ, which is because the service rate of discrete rate scheme has smaller variance than that of the continuous rate scheme, where a service process with smaller variance can support more stringent delay QoS requirement.

VI. The Impact of Channel Correlation

A. Effective Capacity for FSMC-Based Channel Process

We derive the above analytical results by using a block fading channel model. However, this model is not always valid. In most scenarios, it is more practical to consider the correlated wireless channel models. There exist a number of models characterizing the correlated channel fading processes. For instance, the Jake's model [16] has been widely accepted as an accurate modeling approach. Based on the Jake's model, the autocorrelation of the channel gain, denoted by $A_g(\tau)$, can be expressed as $A_g(\tau) = J_0^2(2\pi f_d\tau)$ [16], where $J_0(\cdot)$ denotes the zero-th order Bessel function of the first kind and f_d is the maximum Doppler frequency. However, if using the Jake's model in our systems, it is hard to derive the effective capacity expression from (3). Then, it is even harder to find the power and rate adaptation policies. Therefore, we apply FSMC to model the correlated service process for simplicity.

Integrating the FSMC model with our discrete rate adaptive MQAM, the state of FSMC corresponds to the mode of adaptive modulation. Let $p_{i,j}$ denote the transition probability from state i to state j. We assume a slow-fading channel model such that transition only happens between adjacent states [13], [14]. Under this assumption, we have $p_{ij} = 0$, if $|i-j| > 1, \forall i,j \in \{0,1,...,N-1\}$. The adjacent transition probabilities can be approximated as follows [13]:

$$\begin{cases} p_{n,n+1} \approx \frac{N_\Gamma(\Gamma_{n+1})T_f}{\pi_n}, & \text{for } n = 0,1,...,N-2, \\ p_{n,n-1} \approx \frac{N_\Gamma(\Gamma_n)T_f}{\pi_n}, & \text{for } n = 1,2,...,N-1 \end{cases} \tag{27}$$

where $\{\Gamma_n\}_{n=0}^{N-1}$ and $\{\pi_n\}_{n=0}^{N-1}$ are given by (21) and (26), respectively, and $N_\Gamma(\gamma)$ is the level-crossing rate (LCR) calculated at SNR value of γ, which is given by [17]

$$N_\Gamma(\gamma) = \frac{\sqrt{2\pi}f_d}{\Gamma(m)}\left(\frac{m\gamma}{\overline{\gamma}}\right)^{m-\frac{1}{2}}\exp\left(-\frac{m\gamma}{\overline{\gamma}}\right). \tag{28}$$

Then, the remaining transition probabilities can be derived by using (27) as follows:

$$\begin{cases} p_{0,0} = 1 - p_{0,1} \\ p_{N-1,N-1} = 1 - p_{N-1,N-2} \\ p_{n,n} = 1 - p_{n,n-1} - p_{n,n+1},\ n = 1,...,N-2. \end{cases} \tag{29}$$

Thus, applying (27) and (29), we obtain the transition probability matrix of the FSMC, which is denoted by $\mathbf{P} = [p_{ij}]_{N\times N}$. Based on FSMC-modeled service process, we derive the following proposition to obtain the effective capacity of the service process:

Proposition 1: **If** we define the diagonal matrix $\mathbf{\Phi}(\theta) \triangleq \text{diag}\left\{e^{-\nu_0\theta}, e^{-\nu_1\theta}, ..., e^{-\nu_{N-1}\theta}\right\}$, where $\{\nu_n\}_{n=0}^{N-1}$ is given by (19), **then** the effective capacity $\mathrm{E_C}(\theta)$ of an FSMC-modeled service process is determined by

$$\mathrm{E_C}(\theta) = -\frac{1}{\theta}\log\Big(\rho\{\mathbf{P}\,\mathbf{\Phi}(\theta)\}\Big) \tag{30}$$

where $\mathbf{P}$ is the transition probability matrix of the FSMC mentioned above, and $\rho\{\cdot\}$ denotes the spectral radius of the matrix. Note that the spectral radius of a matrix is defined as the maximum of the absolute values of the eigenvalues for that matrix.

Proof: The proof is similar to [7, Example 3.3], which is omitted for lack of space. ■

B. Power and Rate Adaptation for FSMC-Based Channel Service Process

Although we employ the FSMC-model-based service process, it is still difficult to *directly* derive the power and rate adaptation policy to maximize the effective capacity described by (30). Fortunately, the wireless channel offers a unique feature that allows us to obtain the simple but *near-optimal* solution. We observe that the FSMC-modeled service process satisfy the properties described by the following proposition.

Proposition 2: **If** we denote the effective capacity functions of two FSMC-based service processes by $\mathrm{E_{C_1}}(\theta)$ and $\mathrm{E_{C_2}}(\theta)$, and they have the same marginal statistics, but differ in Doppler frequencies denoted by f_{d_1} and f_{d_2}, respectively, **then** the following equation holds:

$$\mathrm{E_{C_1}}(\theta) \approx \mathrm{E_{C_2}}\left(\frac{f_{d_2}}{f_{d_1}}\theta\right) \tag{31}$$

Proof: The proof is provided in Appendix II. ■

Remarks: Proposition 2 says that $\mathrm{E_{C_1}}(\theta)$ is approximately a horizontal-shifted version of $\mathrm{E_{C_2}}(\theta)$ along θ-axis (when θ-axis uses the logarithmic scale), where the difference between these two functionals is $10\log_{10}(f_{d_2}/f_{d_1})$ dB. Specifically, if $(f_{d_2}/f_{d_1}) > 1$, then $\mathrm{E_{C_1}}(\theta)$ is a left-shifted version of $\mathrm{E_{C_2}}(\theta)$; otherwise, $\mathrm{E_{C_1}}(\theta)$ is a right-shifted version of $\mathrm{E_{C_2}}(\theta)$.

It is well known that the Doppler frequency f_d characterizes the time correlation of channel fading processes. The larger the Doppler frequency f_d, the lower the correlation of the service process. When the Doppler frequency is large enough, the channel process can be approximately considered as uncorrelated, just like the block fading channel. For example, based on our system parameters and the standard Jake's channel model, the autocorrelation $A_g(T_f)$ passes through its first zero-point at the Doppler frequency of 191.25 Hz, which we denote by f_d^{Jake}. However, due to the inaccuracy of the FSMC-based channel model, such a Doppler frequency, denoted by f_d^{FSMC}, is about $f_d^{\mathrm{FSMC}} = 300$ Hz with the same system parameters. In the following discussions, when it is unnecessary to distinguish between these two, we denote both f_d^{Jake} and f_d^{FSMC} by f_d^*, which characterizes the Doppler frequency where the channel process can be approximately considered as uncorrelated, like the block fading channel.

Let $\mathrm{E_C^*}(\theta)$ denote the effective capacity in a block fading channel model. Then, based on Proposition 2, for an FSMC-correlated channel with the same marginal statistics and a Doppler frequency f_d ($f_d \ll f_d^*$), its effective capacity, denoted by $\mathrm{E_C^{(f_d)}}(\theta)$, can be approximated as $\mathrm{E_C^{(f_d)}}(\theta) \approx \mathrm{E_C^*}(\kappa\theta)$, where

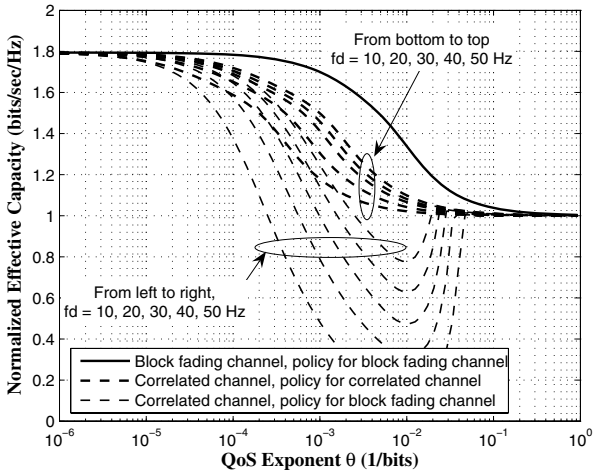


Fig. 7. The effective capacity with different power and rate control policies. The average SNR $\overline{\gamma} = 10$ dB, the fading parameter $m = 2$, BER$= 10^{-3}$, $f_d^{\mathrm{FSMC}} = 300$ Hz, and the number of modes for adaptive MQAM $N = 5$.

$$\kappa = \frac{f_d^*}{f_d}. \tag{32}$$

Likewise, for each power-adaptation policy $\mu(\theta,\gamma)$ that is used for block fading channels, the new policy $\mu(\kappa\theta,\gamma)$ can be applied to the correlated channels with Doppler frequency f_d. This policy generates a new effective-capacity functional, which is approximately a left-shifted version of the original one for the block fading channel, with a difference of $10\log_{10}(\kappa)$ dB along θ-axis. Thus, given the optimal power-adaptation policy $\mu_{\mathrm{opt}}(\theta,\gamma)$ for block fading channel, the optimal power-adaptation policy for correlated channel is approximately $\mu_{\mathrm{opt}}(\kappa\theta,\gamma)$. Note that $\kappa > 1$ due to $f_d < f_d^*$ in (32), as θ increases, the policy $\mu_{\mathrm{opt}}(\kappa\theta,\gamma)$ makes the power-control policy converge faster to the total channel inversion than the case under the block fading channel model. The higher the correlation is, the faster the power-control policy converges to the total channel inversion. Specifically, for our FSMC-based channel model with Doppler frequency f_d, the policy of choosing the boundary points $\{\Gamma_n\}_{n=1}^{N-1}$ becomes:

$$\Gamma_n = M_n^{\kappa\beta+1}\gamma_K^*(\kappa\theta) \tag{33}$$

where $\gamma_K^*(\kappa\theta)$ denotes the cutoff threshold obtained by (21) at the QoS exponent of $\kappa\theta$.

Proposition 2 plays an important role in deriving the power-control policy for the correlated channel. Applying Proposition 2, we can simply shift the existing optimal power-control policy for $10\log_{10}(\kappa)$ dB to obtain the new policies. However, since (31) given in Proposition 2 is only an approximation result, our obtained new power and rate adaptation policy is just a *near*-optimal solution.

Fig. 7 shows the normalized effective capacities of both block fading channel and FSMC-based correlated channel under different power-adaptation policies. The optimal policy

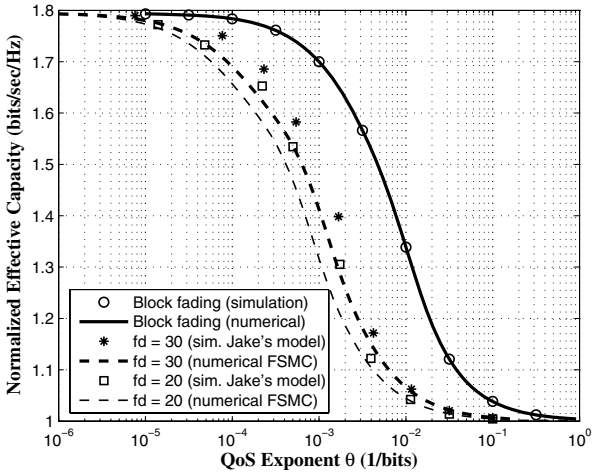


Fig. 8. The effective capacity comparisons with optimal power and rate adaptations between the FSMC-based process (numerical results) and the Jake's model (simulations). The average SNR $\overline{\gamma} = 10$ dB, the fading parameter $m = 2$, BER$= 10^{-3}$, the number of modes for adaptive MQAM is $N = 5$, $f_d^{\mathrm{FSMC}} = 300$ Hz, and $f_d^{\mathrm{Jake}} = 191.25$ Hz

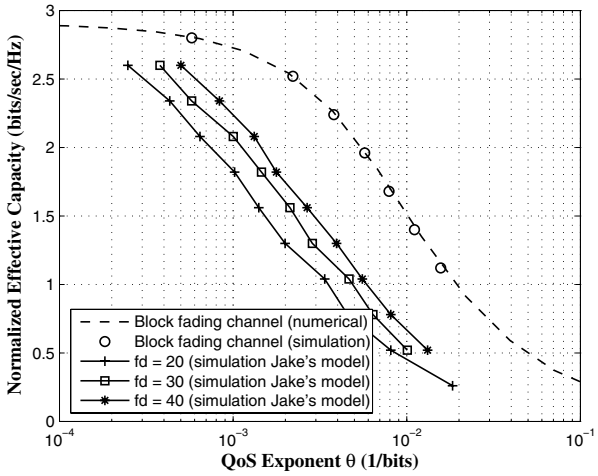


Fig. 9. The effective capacity of constant power approach. The average SNR $\overline{\gamma} = 10$ dB, the fading parameter $m = 1$ (Rayleigh channel).

$\mu_{\mathrm{opt}}(\theta, \gamma)$ for the block fading channel is derived from Section V, which generates the highest effective-capacity curve as shown by the solid line in Fig. 7. Then, according to the analyses in this section, we apply the policy $\mu_{\mathrm{opt}}(\kappa\theta, \gamma)$ to the correlated channel, which numerically generates a group of effective-capacity curves as shown by a set of dashed lines in Fig. 7. These dashed lines are virtually "parallel to" the solid line of the original block fading channel effective capacity. This is consistent with *Remarks* on Proposition 2. However, if we apply the block fading channel policy $\mu_{\mathrm{opt}}(\theta, \gamma)$ directly to the correlated channel, the resulting effective capacities decrease significantly, as shown by a group of dotted lines in Fig. 7.

VII. Simulation Results

Using FSMC-based channel model, we obtain the analytical expression for the effective-capacity and the near-optimal power and rate adaptation policies. However, it is important to verify that the policy derived from the FSMC model can also be applied to the more general scenarios, e.g., the Jake's model, without losing the performance satisfactions. Thus, in this section we simulate the Jake's channel process and compare its outcomes with the analytical results obtained in previous sections.

Applying the optimal power and rate adaptation policy given by (33), Fig. 8 shows the normalized effective capacity comparisons between the Jake's channel model and the FSMC-based channel model. We can observe from Fig. 8 that for the block fading channel, the simulation results perfectly match with the numerical results. On the other hand, when considering the channel correlation, the outcomes from simulations and numerical solutions for these two models share the same trends but differ very slightly. Such a difference can be explained as follows. As stated in Section VI, due to the inaccuracy of the channel model, the Jake's model and the FSMC model have different f_d^*'s, where $f_d^{\mathrm{FSMC}} \approx 300$ Hz and $f_d^{\mathrm{Jake}} = 191.25$ Hz, respectively. Thus, for the same Doppler frequency f_d, the resulting κ in (32) is different. Theoretically, the difference of the effective-capacity curves between these two models is $10 \log_{10}\left(f_d^{\mathrm{FSMC}}/f_d^{\mathrm{Jake}}\right) = 1.96$ dB, which is consistent with the effective capacity difference observed in Fig. 8.

The above analyses described in Fig. 8 verify that the power-adaptation policy derived by using the FSMC model can be well applied to the general Jake's channel model, where the system employs the *discrete rate MQAM*. In the following, we further show that the policy can also be applied to the general Jake's channel model where the system uses *continuous rate transmission*. Fig. 9 plots the normalized effective capacity of the constant-power approach, where we assume that the Shannon capacity can be achieved for each channel realization. For block fading channel, the simulated effective capacity agrees well with the analytical results. On the other hand, for the correlated fading channel, as the Doppler frequency f_d increases, the effective capacity also increases, with the resulting effective-capacity curves roughly "parallel to" each other. This observation implies that the effective capacity of general channel process also follows the similar trends as described in Proposition 2. Therefore, for continuous rate transmissions, the near-optimal power and rate adaptation law also has the form similar to $\mu_{\mathrm{opt}}(\kappa\theta, \gamma)$ for a certain coefficient κ, where $\mu_{\mathrm{opt}}(\theta, \gamma)$ is given by (8).

VIII. Conclusion

In this paper, we proposed and analyzed the QoS-driven power and rate control policies by applying the concept of effective capacity. Our analyses in block fading channel identified the key fact that there exists a fundamental tradeoff between spectral efficiency and QoS provisioning. Depending on the specific QoS requirements, the optimal power-adaptation policy dynamically changes between water-filling and channel inversion. For the more practical adaptive MQAM modulation-based systems, we also developed the corresponding optimal power and rate adaptation scheme. When taking the channel correlation into consideration, we proposed the simple, but

efficient, power-control scheme for Markov modeled fading channels. The simulation results verified that such an approach can also be applied to the more general channel models.

APPENDIX I
PROOF OF THEOREM 1

Proof: Since $\log(\cdot)$ is a monotonically increasing function, for each given $\theta > 0$, the maximization problem of (7) can be converted into a minimization problem as follows:

$$\min_{\mu(\theta,\gamma):\int_0^\infty \mu(\theta,\gamma)p_\Gamma(\gamma)d\gamma=1} \left\{ \int_0^\infty e^{-\theta T_f B \log_2\left(1+\mu(\theta,\gamma)\gamma\right)} \cdot p_\Gamma(\gamma)d\gamma \right\}. \quad (34)$$

It is clear from [21, Sec. 3.2] that in (34), the objective function is strictly convex and the constraint is linear with respect to $\mu(\theta,\gamma)$. Thus, the minimization problem has a unique optimal solution. Then, we can form the Lagrangian function, denoted by $\mathcal{J}$, as follows:

$$\mathcal{J} = \int_0^\infty e^{-\beta \log\left(1+\mu(\theta,\gamma)\gamma\right)} p_\Gamma(\gamma)d\gamma + \lambda\left(\int_0^\infty \mu(\theta,\gamma)p_\Gamma(\gamma)d\gamma - 1\right). \quad (35)$$

where $\beta \triangleq \theta T_f B/\log 2$ is defined as the normalized QoS exponent. Differentiating the Lagrangian function given by (35) and setting the derivative equal to zero [22, Sec. 4.2.4], we get

$$\frac{\partial \mathcal{J}}{\partial \mu(\theta,\gamma)} = \left\{\lambda - \beta\gamma\left[1+\mu(\theta,\gamma)\gamma\right]^{-\beta-1}\right\} p_\Gamma(\gamma) = 0. \quad (36)$$

Defining $\gamma_0 \triangleq \lambda/\beta$ and solving (36), we can obtain the optimal power and rate adaptation policy as shown by (8), where γ_0 is determined by the mean power constraint of (9). The proof follows. ∎

APPENDIX II
PROOF OF PROPOSITION 2

Proof: In order to prove Proposition 2, we first introduce the following lemma.

Lemma 1: Let a channel service process be modeled as a continuous-time FSMC with N states, the service rate of the nth state be denoted by $\widetilde{\nu}_n$, $(n \in \{0,1,...,N-1\})$, and the corresponding generating matrix of the continuous-time FSMC be represented by $\mathbf{Q}$, respectively. If we define $\mathbf{R} \triangleq \mathrm{diag}\{\widetilde{\nu}_0, \widetilde{\nu}_1, ..., \widetilde{\nu}_{N-1}\}$, then the effective capacity of this process, denoted by $\widetilde{\mathrm{E}}_\mathrm{C}(\theta)$, is determined by[3]

$$\widetilde{\mathrm{E}}_\mathrm{C}(\theta) = -\frac{1}{\theta}\delta\left\{\mathbf{Q} - \theta\mathbf{R}\right\} \quad (37)$$

where $\delta\{\cdot\}$ denotes the maximum real eigenvalue of the matrix.

Proof: The proof is similar to [10, Appendix], which is omitted for lack of space. ∎

[3]Note that for continuous-time FSMC, the unit for the service rate $\widetilde{\nu}_n$ and the effective capacity $\widetilde{\mathrm{E}}_\mathrm{C}(\theta)$ is "bits per second".

There exists the close relationship between continuous-time FSMC and discrete-time FSMC. Under appropriate conditions, the discrete-time FSMC can be considered as the "samples" of the embedded continuous-time FSMC. Based on our system model, the sample interval is T_f and the service rate $\nu_n = \widetilde{\nu}_n T_f$.

The relationship between transition probability matrix $\mathbf{P}$ of a discrete-time FSMC and generating matrix $\mathbf{Q}$ of a continuous-time FSMC can be expressed as

$$\mathbf{P}(T_f) = e^{\mathbf{Q}T_f} = \mathbf{I} + \mathbf{Q}T_f + o(T_f^2) \quad (38)$$

where we rewrite $\mathbf{P}$ by $\mathbf{P}(T_f)$ in (38) to emphasize that the sample interval is T_f. Given the transition probability matrix $\mathbf{P}$, the first-order approximation of the generator matrix $\mathbf{Q}$ is determined by

$$\mathbf{Q} \approx \frac{\mathbf{P}(T_f) - \mathbf{I}}{T_f}. \quad (39)$$

It is clear that the generating matrix $\mathbf{Q}$ can be expressed as $\mathbf{Q} = f_d\mathbf{A}$, where $\mathbf{A}$ only depends on the marginal statistics of the channel. Thus, we can approximate the effective capacity of a discrete-time FSMC by the effective capacity of a continuous-time FSMC as follows:

$$\mathrm{E}_\mathrm{C}(\theta) \approx \widetilde{\mathrm{E}}_\mathrm{C}(\theta)T_f. \quad (40)$$

Based on the continuous-time FSMC approximation given by (40), we prove Proposition 2 as follows. From Lemma 1, we have

$$\begin{aligned}
\widetilde{\mathrm{E}}_{\mathrm{C}_1}(\theta) &= -\frac{1}{\theta}\delta\left\{\mathbf{Q}_1 - \theta\mathbf{R}\right\} \\
&= -\frac{1}{\theta}\delta\left\{f_{d_1}\mathbf{A} - \theta\mathbf{R}\right\} \\
&= -\frac{1}{\theta}\delta\left\{\frac{f_{d_1}}{f_{d_2}}\left(f_{d_2}\mathbf{A}\right) - \frac{f_{d_1}}{f_{d_2}}\left(\frac{f_{d_2}}{f_{d_1}}\theta\mathbf{R}\right)\right\} \\
&= -\left(\frac{f_{d_1}}{f_{d_2}}\right)\frac{1}{\theta}\delta\left\{f_{d_2}\mathbf{A} - \frac{f_{d_2}}{f_{d_1}}\theta\mathbf{R}\right\} \\
&= -\frac{1}{\left(\frac{f_{d_2}}{f_{d_1}}\theta\right)}\delta\left\{\mathbf{Q}_2 - \left(\frac{f_{d_2}}{f_{d_1}}\theta\right)\mathbf{R}\right\} \\
&= \widetilde{\mathrm{E}}_{\mathrm{C}_2}\left(\frac{f_{d_2}}{f_{d_1}}\theta\right). \quad (41)
\end{aligned}$$

Plugging the approximate relationship given by (40) into (41), the proof for Proposition 2 follows. ∎

REFERENCES

[1] C.-S. Chang, *Performance Guarantees in Communication Networks*. Berlin: Springer-Verlag, 2000.

[2] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall PTR, 2001.

[3] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

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

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

[6] A. J. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[7] 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.

[8] 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.
[9] 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.
[10] G. Kesidis, J. Walrand, and C.-S. Chang, "Effective bandwidths for multiclass Markov fluids and other ATM sources," *IEEE/ACM Trans. Networking*, vol. 1, no. 4, pp. 424–428, Aug. 1993.
[11] A. I. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," *IEEE/ACM Trans. Networking*, vol. 1, no. 3, pp. 329–343, June 1993.
[12] 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.
[13] H. S. Wang and N. Moayeri, "Finite-state Markov channel — a useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.
[14] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.
[15] C.-D. Iskander and P. T. Mathiopoulos, "Analytical level crossing rates and average fading durations for diversity techniques in Nakagami fading channels," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1301–1309, Aug. 2002.
[16] W. C. Jake, Jr., *Microwave Mobile Communications*. New York: Wiley, 1974.
[17] M. K. Simon and M.S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*. New York: Wiley, 2nd Ed., 2005.
[18] J. Tang and X. Zhang, "Transmit selection diversity with maximal-ratio combining for multicarrier DS-CDMA wireless networks over Nakagami-m Fading Channels," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 1, pp. 104–112, Jan. 2006.
[19] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integral, Series, and Products*. Academic Press, 1992.
[20] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.
[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[22] A. J. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

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.

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 Assistant 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 worked as a Summer Intern 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 100 research papers in the areas of wireless networks and communications, mobile computing, cross-layer optimizations for QoS guarantees over mobile wireless networks, effective capacity and effective bandwidth theories for wireless networks, DS-CDMA, MIMO-OFDM and space-time coding, adaptive modulations and coding (AMC), wireless diversity techniques and resource allocations, wireless sensor and Ad Hoc networks, cognitive radio and cooperative communications/relay networks, vehicular Ad Hoc networks, multi-channel MAC protocols, wireless and wired network security, wireless and wired multicast networks, channel coding for mobile wireless multimedia multicast, network protocols design and modeling, statistical communications theory, information theory, random signal processing, 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 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 currently serving as an Editor for the *IEEE Transactions on Wireless Communications*, an Associate Editor for the *IEEE Transactions on Vehicular Technology*, an Associate Editor for the *IEEE Communications Letters*, an Editor for the *Wiley's Journal on Wireless Communications and Mobile Computing*, and an Editor for the *Journal of Computer Systems, Networking, and Communications*, and is also serving as the Guest Editor for the *IEEE Wireless Communications Magazine* for the special issue on "next generation of CDMA versus OFDMA for 4G wireless applications." He has frequently served as the Panelist on the U.S. National Science Foundation Research-Proposal Review Panels. He is serving or has served as the Co-Chair for the IEEE Globecom 2008 – Wireless Communications Symposium and the Co-Chair for the IEEE ICC 2008 – Information and Network Security Symposium, respectively, the Symposium Chair for the IEEE/ACM International Cross-Layer Optimized Wireless Networks Symposium 2006, 2007, and 2008, respectively, the TPC Chair for the IEEE/ACM IWCMC 2006, 2007, and 2008, respectively, the Poster Chair for the IEEE INFOCOM 2008, the Student Travel Grants Co-Chair for the IEEE INFOCOM 2007, the Panel Co-Chair for the IEEE ICCCN 2007, the Poster Chair for the IEEE/ACM MSWiM 2007 and the IEEE QShine 2006, the Publicity Chair for the IEEE/ACM QShine 2007 and the IEEE WirelessCom 2005, and the Panelist on the WiFi-Hotspots/WLAN and QoS Panel at the IEEE QShine 2004. He has served as the TPC members for more than 50 IEEE/ACM conferences, including the IEEE INFOCOM, IEEE Globecom, IEEE ICC, IEEE WCNC, IEEE VTC, IEEE/ACM QShine, IEEE WoWMoM, IEEE ICCCN, etc.

Prof. Zhang is a Senior Member of the IEEE and a Member of the Association for Computing Machinery (ACM).