

Delay-Aware Energy-Efficient Communications Over Nakagami- m Fading Channel With MMPP Traffic

Kunlun Wang, Meixia Tao, *Senior Member, IEEE*, Wen Chen, *Senior Member, IEEE*, and Quansheng Guan, *Member, IEEE*

Abstract—In this paper, we propose a cross-layer design framework for transmitting Markov modulated Poisson process (MMPP) traffic over Nakagami- m fading channel with delay demands. The adaptive modulation and coding (AMC) technique is used at the physical layer. The energy efficiency is described as the average throughput over the average transmission power, where both of throughput and transmit power have full consideration of the queuing system. We first derive the closed-form expressions of the delay and the energy efficiency with the stationary distribution of the system. We then derive the energy efficient thresholds to choose the AMC transmission modes. At last, we derive the transmission policy to maximize the energy efficiency with delay constraints. Numerical results are provided to support the theoretical development.

Index Terms—Cross-layer design, energy efficiency, adaptive modulation and coding, queuing theory, finite state Markov channel, Markov modulated Poisson process, delay.

I. INTRODUCTION

RECENTLY, energy efficient communications in wireless networks have attracted much research attention due to the rapid growth of mobile data traffic and battery powered mobile terminals. In traditional communication systems, throughput and power are the two common performance measures to characterize the system benefit and cost, respectively. In energy efficient communications, energy efficiency as a new performance measure is used to characterize how much throughput can be achieved per unit power consumption. While a large number of approaches have been studied to reduce the power consumption in wireless communication systems from the physical layer, very few studies focus on the cross-layer optimization. Cross-layer design considers the joint effects of

the time varying channel and the dynamic data queue as well as the service traffic from the upper layers, which have significant influence to the energy efficiency of a communication system [1], [2].

Adaptive modulation and coding (AMC) scheme can enhance throughput in time-varying channel conditions [3], [4], and it has been widely incorporated in several wireless standards, i.e., Long Term Evolution (LTE), IEEE 802.11a and IEEE 802.16e. The key idea of traditional AMC scheme is to use higher order modulation for better channel condition to transmit more bits. There is a tradeoff between the reliability and the spectral efficiency in the design of AMC schemes. In [3], the authors have derived a closed-form expression of the average spectral efficiency for given reliability of packets transmission over Nakagami- m block fading channels. Although AMC is mainly used to improve the spectral efficiency of a link for a given set of quality of service (QoS) requirements, its unique nature for enhancing upper layer protocol design has spurred the development of cross-layer approaches. These approaches can integrate the QoS provisioning protocols at higher layers with energy efficient AMC implemented at the physical layer.

Many recent works on cross-layer design focus on combining AMC scheme with automatic repeat request (ARQ) [5]–[9]. The schemes of AMC and hybrid automatic repeat request (HARQ) with limited channel state information (CSI) feedback for reliable transmission have been studied for the performance of energy efficiency in [5]. Considering the queuing system for the arriving traffic, many recent works focus on the maximum throughput scheduling scheme by implementing AMC [1], [4], [10]. In [1], the objective is to maximize the system throughput under the bit error rate (BER) constraint, and the variables are the transmit power and the data rate. In [11], minimizing the average joint packet loss rate due to both erroneous transmission and buffer overflow is studied, which is equivalent to maximizing the throughput.

However, all these works do not consider the delay requirement of the traffic, which is the main QoS requirement for future wireless networks. Although the opportunistic distributed multiuser scheduling in the presence of a fixed packet deadline constraint is studied in [12], the service delay at physical channel is not considered. Considering the service delay, the authors in [2] studied the policy of choosing modulation constellation dynamically depending on incoming traffic state and buffer state in addition to channel state when the traffic and channel conditions are known. However, the energy efficiency has not been considered. In [13], [14], the authors proposed a unified

Manuscript received November 30, 2014; revised April 5, 2015 and June 26, 2015; accepted June 28, 2015. Date of publication July 8, 2015; date of current version August 7, 2015. This work was supported by the NSF of China under grants 61322102 and 61329101; by National 863 Project under grant 2015AA01A710; and by STCSM Science and Technology Innovation Program under grant 13510711200. Part of this paper was presented at the IEEE International Conference on Communications (ICC), London, U.K., June 8–12, 2015. The associate editor coordinating the review of this paper and approving it for publication was P. Popovski.

K. Wang, M. Tao, and W. Chen are with Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: kunlun1228@sjtu.edu.cn; mxtao@sjtu.edu.cn; wenchen@sjtu.edu.cn).

Q. Guan is with the School of Electronic and Information Engineering, South China University of Technology, Guangdong 510640, China (e-mail: eeqshguan@scut.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCOMM.2015.2453268

TABLE I
FREQUENTLY USED NOTATION

Quantity	Symbol	Quantity	Symbol
Number of incoming traffic	K	Buffer Size	M
Effective Average Traffic Rate into the Buffer	r	Packet Size	L
Total number of AMC transmission modes	N	Transmission Rate for Channel State n	b_n
Number of Transmission times	N_F^{max}	Average Transmission Rate for Channel State n	\bar{b}_n
Average Packet Successful Transmission Probability for Channel State n	\bar{P}_{s_n}	Service Rate for Channel State n	c_n
The Probability for Channel in State n	$P_r(n)$	BER for Channel State n	P_b^n
Received SNR Threshold	γ_n	Average BER for Channel State n	\bar{P}_b^n
Average Packet Successful Transmission Probability of AMC	\bar{P}_s	Average Service Rate Considering Queuing	\bar{c}_n
Queue State at the t -th Time-slot	S_t	Service State at the t -th Time-slot	C_t
Amount of Packets Generated by the Source	A_t	Traffic State at the t -th Time-slot	F_t
Packet Transmission Time at Channel State n	τ_n	Mean Queue Length for Channel State n	\bar{Q}_q^n
Average Packet Transmission Time for AMC	$\bar{\tau}$	Mean Queuing Time	\bar{D}_q
Total Average Delay for a Packet in Channel State n	\bar{W}_n	Total Average Delay for AMC	\bar{W}
Time Unit for a Frame	T_u	Average End-to-end Throughput	\bar{T}
Equivalent Transmit Power at Time-slot t with Rate b_n	$e_{n,t}(\gamma)$	Average Transmit Power	\bar{e}
Average Transmit Power in Channel State n Considering Queuing	\bar{e}_n	Equivalent Transmit Power for AMC	\tilde{e}
Average Packet Generate Rate	λ	Packet Dropping Rate from the Buffer	P_d

reinforcement learning solution for finding the joint optimal AMC and dynamic power management policies when the traffic arrival and channel statistics are unknown. The performance of energy efficient transmission power has not been studied in their works either. In [15], a new approach for on-line implementation of the optimal packet scheduling algorithm is proposed. The authors consider the problem of energy efficient scheduling under average delay constraint for single user fading channel. However, the average delay only includes queuing delay, excluding the transmission delay. For some important service and poor channel condition, the transmission delay also occupies an important position.

Taking a broader view, our work follows the cross-layer design approach, which takes the system variations and statistics at multiple layers of the protocol stack into account. In particular, the transmission decisions are part of the physical layer. The retransmission of packet is controlled by the data link layer, and the traffic arrival statistics also with the queue condition are the parameters of higher layers. In this context, our work is closely related to the works in [2], [4], [16]–[18], in which a similar system model with stochastic data arrival, a finite-length queue, and a time-varying channel is considered. The power-delay tradeoff have already been studied in [16], [17]. However, our work is different from [2], [4], [16]–[18] in several significant ways. First, while the objective of our work is to maximize the system energy efficiency with delay constraint, [2] and [4] concentrate more on the throughput, and [18] concentrates on the energy consumption. Second, to support delay-aware service motivates us to consider the effect of delay, and we have derived the closed-form expression of the delay including the queuing delay and the transmission delay, which is not considered in [2], [4], [16], [17]. Third, we analyze the performance of energy efficiency by modeling the channel using Nakagami-*m* fading channel, and by using a more practical and general model for incoming traffic, namely, Markov modulated Poisson process (MMPP) model. At last, we use the energy efficient thresholds to select the transmission mode, which is different from their works. We also assume that the average arrival rate is smaller than the average transmission rate so that the overflow rate can be controlled without incurring

excessive overflow. Thus, the queue is impossible to grow infinitely and drop large number of packets.

Compared with the existing literature, our contribution can be summarized as follows:

- 1) We obtain, via finite state Markov chain, the closed-form expression of the delay of the AMC transmission system, which considers the general traffic arrival and queuing model as well as the general channel model.
- 2) We derive the closed-form expression of the system throughput by taking into account the packet drop caused by both the channel transmission error and buffer overflow. Correspondingly, the average power consumption is also obtained.
- 3) By using the AMC transmission and the MMPP traffic model, we show that there is a tradeoff between delay and energy efficiency. We get the intrinsic relationship between the optimal energy efficiency and different delay requirements when the traffic arrival rate is different.
- 4) We derive the energy efficient thresholds to partition the SNR, which will improve the energy efficiency in comparison with other existing partitions from [2], [3]. In addition, we present performance results to support the theoretical development.
- 5) We get the energy efficient transmission policy with delay constraint. For the large average arrival traffic, the energy efficient transmission policy is the same no matter what the delay demand is. For the small average arrival traffic, the energy efficient transmission policy is different with different delay demands.

The rest of the paper is organized as follows. Section II describes the system model and the problem statement, including the channel model and the queuing model also with the traffic model. In Section III, we analyze the queuing and delay performance for the system to derive the delay of packet transmission. In Section IV, we derive the closed-form expression for the energy efficiency, which is relevant to the average transmission power and the system throughput. We also determine the energy efficient transmission policy under different delay constraints. Section V shows the numerical results, and the conclusions are made in Section VI. Table I lists the frequently used notations.

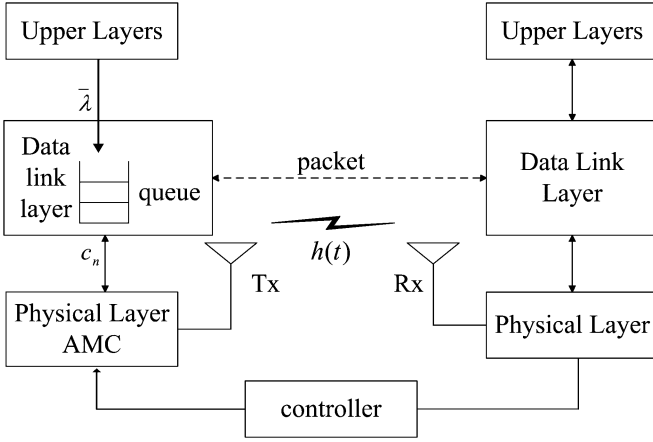


Fig. 1. System model.

II. SYSTEM MODEL AND PROBLEM STATEMENT

A. Traffic and Queuing Models

We assume that the incoming traffic is random and modeled as a Markov modulated Poisson process (MMPP), where in any state the incoming traffic is Poisson distributed and the transitions between the states are governed by an underlying Markov chain. A wide range of multimedia traffic can be represented with an MMPP model, which is accurate and reasonable [19], and the Poisson arrival is the special case of this model. Amounts of multimedia services are existing in the present and future wireless networks, which can be categorized into heterogeneous classes with different delay demands.

Let $\mathbb{F} = \{f_1, f_2, \dots, f_K\}$ denote the set of states of the incoming traffic, and P_{f_i, f_j} denotes the probability of transition from state f_i to state f_j . Each state follows Poisson distribution with average arrival rate λ_i packets/time-unit, $i = 1, 2, \dots, K$. Denote the row vector $\pi^f = [\pi_1^f, \pi_2^f, \dots, \pi_K^f]$ as the stationary distribution of the incoming traffic, and it satisfies $\pi^f = \pi^f \mathbf{P}^f$, where \mathbf{P}^f is the transition probability matrix for the underlying Markov chain governing transitions between traffic states and is given by

$$\mathbf{P}^f = \begin{bmatrix} P_{f_1, f_1} & P_{f_1, f_2} & \cdots & P_{f_1, f_K} \\ P_{f_2, f_1} & P_{f_2, f_2} & \cdots & P_{f_2, f_K} \\ \vdots & \vdots & \ddots & \vdots \\ P_{f_K, f_1} & P_{f_K, f_2} & \cdots & P_{f_K, f_K} \end{bmatrix}. \quad (1)$$

The arrival transition matrix \mathbf{P}^f is a right stochastic matrix, we can write $\sum_{j=1}^K P_{f_i, f_j} = 1$. By the stationary distribution π^f , we can get the average arrival rate as $\bar{\lambda} = \sum_{i=1}^K \pi_i^f \lambda_i$.

Consider that at the data link layer, packet streams with the packet size of L bits from upper layers arrive in a first-in-first-out (FIFO) buffer at the average arrival rate $\bar{\lambda}$. Assume that the user's queuing model is a single server M/G/1 queue [20], shown in Fig. 1. The M/G/1 model assumes Markovian or memoryless arrivals at average rate $\bar{\lambda}$, a general service distribution and a single server. Let c_n be the service rate at the physical layer corresponding to channel state n . Thus, the service state c_n is in the set of $\mathbb{C} = \{c_1, \dots, c_N\}$.

Assume that the buffer has finite capacity, denoted as M . Note that since the arriving traffic is random in nature, the queue may be empty sometimes and it may be full other times. Therefore, the packets may be dropped because of the burst arriving service and the finite buffer size. The dropped packets are considered lost.

B. Channel Model

Consider a frame-by-frame communication system, shown in Fig. 1, with each frame composed of a number of packets. The transmission time duration for each frame is defined as one time unit or time slot. The channel is frequency-flat and block fading and is also corrupted with additive white Gaussian noise (AWGN) with zero mean and variance σ^2 .

Let \bar{e} denote the average transmit signal power. For transmit power constant at \bar{e} , the channel quality can be captured by a single parameter, namely the received signal to noise ratio (SNR) γ . Due to block fading, we assume γ remains invariant within each transmission frame but can vary from frame to frame. We consider the general Nakagami- m fading model. Then the received SNR γ per frame follows a Gamma distribution with probability density function (PDF):

$$f_{\bar{\gamma}}(\gamma) = \frac{m^m \gamma^{m-1}}{\bar{\gamma}^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\bar{\gamma}}\right), \quad (2)$$

where $\bar{\gamma} \triangleq E\{\gamma\}$ is the average received SNR, $\Gamma(m) \triangleq \int_0^\infty t^{m-1} e^{-t} dt$ is the Gamma function, and m is the Nakagami- m fading parameter ($m \geq 1/2$). We choose the Nakagami- m channel model because it applies to a large class of fading channels. It includes the Rayleigh channel as the special case $m = 1$.

As shown in Fig. 1, we consider AMC for transmission at the physical layer. Let N denote the total number of AMC transmission modes available. The entire SNR range is divided into $N + 1$ nonoverlapping consecutive intervals, with boundary points given by $\{\gamma_n\}_{n=0}^{N+1}$, where $\gamma_0 = 0$ and $\gamma_{N+1} = +\infty$. The channel is said to be in state n when $\gamma \in [\gamma_n, \gamma_{n+1})$. Then the fading process can be represented by a finite-state Markov chain (FSMC). With (2), the probability that the channel is in state n is given by

$$P_r(n) = \int_{\gamma_n}^{\gamma_{n+1}} f_{\bar{\gamma}}(\gamma) d\gamma. \quad (3)$$

Let the transmission rate corresponding to channel state n be denoted as b_n bits per channel use.

Automatic repeat request (ARQ) is considered in our model. If the receiver cannot correctly decode a packet, the transmitter will repeat transmitting the packet with the maximum retransmission times $N_r^{\max} - 1$. When the receiver correctly decodes a packet, it will feedback an ACK packet to the transmitter. If the receiver still cannot decode the packet correctly after $N_r^{\max} - 1$ retransmissions, the packet will be dropped and a packet loss is declared. We want the probability of packet loss is no larger than P_{loss} . Note that the service rate c_n depends on the transmission rate b_n and the maximum retransmission times $N_r^{\max} - 1$. A frame is composed of retransmitted and new packets. Hence, different modulation levels can be used for the different retransmissions of a same packet.

C. Delay Model and Problem Statement

For the incoming MMPP traffic, the packets are stored in the finite size buffer at first, then transmitted over the Nakagami- m fading channel. In this case, the average packet delay \bar{W} is defined as the queuing time \bar{D}_q in the buffer and the average service time \bar{S}_T over the time-varying channel, i.e.,

$$\bar{W} = \bar{D}_q + \bar{S}_T.$$

Let \tilde{e} denote the average transmit power using AMC technique, which measures the energy consumed per second. The resulting system throughput can be denoted as \bar{T} , which is the number of packets that are transmitted without error per second. We define the energy efficiency as the ratio of average throughput to its transmit power, i.e.,

$$f_{ee} \triangleq \frac{\bar{T}}{\tilde{e}} \text{ packets/joule}, \quad (4)$$

which measures the number of reliable packets that are transmitted per joule of energy consumed.

In this paper, we are interested in determining the AMC rate adaption policy for given traffic \mathbb{F}_0 . The objective is to maximize the energy efficiency under the packet delay constraint W_0 . In particular, the rate adaption policy represents the probability distribution of transmission rate b_n , $n = 1, \dots, N$, the rate adaptive policy can be denoted as μ . Mathematically, the problem is given by

$$\max_{\{P_r(n)\}} \{f_{ee} | \bar{W} \leq W_0, \mathbb{F} = \mathbb{F}_0\}. \quad (5)$$

Specifically, when $P_r(n)$ is large for large transmission rate b_n , the packet delay can be small. Then the throughput and the power are large. Therefore, as we shall illustrate, the rate adaption policy under delay constraint also has a nonnegligible impact on energy efficiency.

III. CROSS-LAYER QUEUING AND DELAY ANALYSIS

In this section, we present a three-dimensional Markov process to get the packet dropping rate and the delay performance of a packet. Due to finite-length queuing, arriving packets will be dropped when the buffer is full. A packet from the source is correctly received only if it is not dropped from the queue, and correctly passed the wireless channel. Thus, the queue and the channel will influence the delay.

A. Queuing Analysis

1) *Service Rate*: We assume that a packet is in error if at least l out of L bits are corrupted. Then we can characterize the

average packet successful transmission probability \bar{P}_{s_n} corresponding to transmission rate b_n as

$$\bar{P}_{s_n} = 1 - \sum_{i=l}^L \binom{L}{i} (\bar{P}_b^n)^i (1 - \bar{P}_b^n)^{L-i}, \quad (6)$$

where \bar{P}_b^n is the average uncoded bit error rate (BER) for channel state n , which can be derived in (32).

Since the retransmission of the packet may use different AMC modes, in the M/G/1 queue model, the packet service time S_{T_n} in state n has the following probability mass function:

$$P\{S_{T_n} = \tau_n + k\bar{\tau}\} = \bar{P}_{s_n} (1 - \bar{P}_{s_n})^k, \quad k = 0, \dots, N_r^{\max} - 1, \quad (7)$$

where τ_n and $\bar{\tau}$ represent the packet transmission time when the channel is in state n and the average packet transmission time for retransmission, which are given by

$$\tau_n = \frac{L}{T_u b_n R_s}, \quad \bar{\tau} = \frac{L}{T_u \bar{b} R_s}, \quad (8)$$

where T_u and R_s are the time unit and the symbol rate respectively, and $\bar{b} = \sum_{n=1}^N P_r(n) b_n$. From (7), we can get the mean service time at channel state n as (9), shown at the bottom of the page. From (8) and (9), the service rate c_n at channel state n is given by:

$$c_n = \frac{1}{\mathbb{E}\{S_{T_n}\}}. \quad (10)$$

2) *Stationary Distribution*: Let t index the transmission frame or, equivalently, time slot and A_t be the amount of packets generated by the source between time t and $t - 1$. From the MMPP model, we can get

$$P(A_t = a | f_i) = \begin{cases} \frac{\lambda_i^a e^{-\lambda_i}}{a!}, & \forall f_i \in \mathbb{F}, \quad \text{if } 0 \leq a \leq A; \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

and the queue state is independent of the channel state. Let S_t be the queue state (the number of packets in the queue) at the start of the t -th time slot, and $S_t \in \mathbb{S} = \{s_0 = 0, s_1 = 1, \dots, s_M = M\}$. Denote by $C_t \in \mathbb{C} = \{c_1, \dots, c_N\}$ the number of packets removed from the queue using AMC at the start of each time slot. The resulting recursion of the queue state can be summarized as

$$S_t = \min\{M, \max\{0, S_{t-1} - C_t\} + A_t\}. \quad (12)$$

Let (F_{t-1}, C_t, S_{t-1}) denote the triple of the traffic state, the service rate and the queue state respectively, and let $P_{(f_i, c_x, s_q), (f_j, c_y, s_l)}$ denote the transition probability from $(F_{t-1} = f_i, C_t = c_x, S_{t-1} = s_q)$ to $(F_t = f_j, C_{t+1} = c_y, S_t = s_l)$, where

$$\begin{aligned} \mathbb{E}\{S_{T_n}\} &= \tau_n \bar{P}_{s_n} + \sum_{k=1}^{N_r^{\max}-1} (\tau_n + k\bar{\tau}) \bar{P}_{s_n} (1 - \bar{P}_{s_n})^{k-1} \\ &= \bar{P}_{s_n} (\tau_n - \bar{\tau}) + \frac{\bar{\tau}}{\bar{P}_{s_n}} \left[1 - (1 + N_r^{\max} \bar{P}_{s_n}) (1 - \bar{P}_{s_n})^{N_r^{\max}} \right] \end{aligned} \quad (9)$$

$(f_i, c_x, s_q) \in \mathbb{F} \times \mathbb{C} \times \mathbb{S}$, and $(f_j, c_y, s_l) \in \mathbb{F} \times \mathbb{C} \times \mathbb{S}$. We can organize the state transition probability matrix in a block form

$$\mathbf{P} = [\mathbf{A}_{f_i, f_j}], 0 \leq i, j \leq K, \quad (13)$$

where

$$\mathbf{A}_{f_i, f_j} = \begin{bmatrix} \mathbf{Q}_{f_1, f_1} & \mathbf{Q}_{f_1, f_2} & \cdots & \mathbf{Q}_{f_1, f_K} \\ \mathbf{Q}_{f_2, f_1} & \mathbf{Q}_{f_2, f_2} & \cdots & \mathbf{Q}_{f_2, f_K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}_{f_K, f_1} & \mathbf{Q}_{f_K, f_2} & \cdots & \mathbf{Q}_{f_K, f_K} \end{bmatrix}, \quad (14)$$

and the submatrix \mathbf{Q}_{f_i, f_j} in matrix \mathbf{P} is defined as

$$\mathbf{Q}_{f_i, f_j} = \begin{bmatrix} \mathbf{R}_{(f_i, c_1), (f_j, c_1)} & \mathbf{R}_{(f_i, c_1), (f_j, c_2)} & \cdots & \mathbf{R}_{(f_i, c_1), (f_j, c_N)} \\ \mathbf{R}_{(f_i, c_2), (f_j, c_1)} & \mathbf{R}_{(f_i, c_2), (f_j, c_2)} & \cdots & \mathbf{R}_{(f_i, c_2), (f_j, c_N)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{(f_i, c_N), (f_j, c_1)} & \mathbf{R}_{(f_i, c_N), (f_j, c_2)} & \cdots & \mathbf{R}_{(f_i, c_N), (f_j, c_N)} \end{bmatrix}. \quad (15)$$

The submatrix $\mathbf{R}_{(f_i, c_x), (f_j, c_y)}$ in matrix \mathbf{Q}_{f_i, f_j} can be shown in (16), shown at the bottom of the page. Thus, the Markov chain has the total states $l = K \times N \times (M + 1)$.

We next simplify the probability of the system state transition from state (f_i, c_x, s_q) to state (f_j, c_y, s_l) as (17), shown at the bottom of the page, where P_{f_i, f_j} and P_{c_x, c_y} represent the transition probability of the traffic as shown in (1) and the state transition probability of the transmission rate respectively. The second equality in (17) follows from the fact that both the channel transition and the traffic transition are independent of others.

We assume slow fading so that the transition happens only between adjacent channel states, i.e., $P_{c_x, c_y} = 0, |c_x - c_y| \geq 2$. The nonzero elements P_{c_x, c_y} is described in [3]. On the other hand, it is easy to get the conditional probability of $P(S_t = s_l | F_{t-1} = f_i, S_{t-1} = s_q)$ as in (18), shown at the bottom of the page. Therefore, based on (17) and (18), we can get the transition probability of the system state.

We propose a lemma to prove that the stationary distribution $\boldsymbol{\pi} = [\pi_{(f_1, c_1, s_0)}, \cdots, \pi_{(f_1, c_1, s_M)}, \cdots, \pi_{(f_1, c_N, s_0)}, \cdots, \pi_{(f_1, c_N, s_M)}, \cdots, \pi_{(f_K, c_1, s_0)}, \cdots, \pi_{(f_K, c_N, s_M)}]$ exists and is unique, where

$$\pi_{(f_i, c_x, s_q)} = \lim_{t \rightarrow \infty} P(F_{t-1} = f_i, S_{t-1} = s_q, C_t = c_x).$$

Lemma 1: The stationary distribution $\boldsymbol{\pi}$ of the process $\{(F_t, S_t, C_t), t \geq 0\}$, $(F_t, S_t, C_t) \in \mathbb{F} \times \mathbb{S} \times \mathbb{C}_a$ exists, and $\boldsymbol{\pi}_t \rightarrow \boldsymbol{\pi}$ as $t \rightarrow \infty$.

Proof: The proof is given in Appendix A. \square

This lemma suggests that when the Markov chain run a long time t_0 , then the distribution of $(F_{t_0}, S_{t_0}, C_{t_0})$ is $\boldsymbol{\pi}$ at the time $t = t_0$, which is irrelevant to the initial distribution. Let (f_i, c_x, s_q) denote the system state at time t_0 . Then

$$\pi_{(f_i, c_x, s_q)} = \lim_{t_0 \rightarrow \infty} P((F_{t_0}, S_{t_0}, C_{t_0}) = (f_i, c_x, s_q)).$$

In other words, $\pi_{(f_i, c_x, s_q)}$ is the limiting or long-run probability that the (long run) proportion of time of the system in state (f_i, c_x, s_q) . Then the stationary distribution is obtained by solving

$$\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P}, \quad \sum_{f \in \mathbb{F}, s \in \mathbb{S}, c \in \mathbb{C}} \pi_{(f, s, c)} = 1. \quad (19)$$

If the Markov chain $\{(F_t, S_t, C_t)\}$ is not assumed to be irreducible, there can be more than one solution for (19). The solution $\boldsymbol{\pi}$ is the left eigenvector of \mathbf{P} corresponding to the eigenvalue 1. Based on the linear algebra, the stationary distribution $\boldsymbol{\pi}$ of the Markov chain $\{(F_t, S_t, C_t)\}$ can also be computed as

$$\boldsymbol{\pi} = \mathbf{e}(\mathbf{I} - \mathbf{P} + \mathbf{E})^{-1}, \quad (20)$$

where $\mathbf{e} = (1, \cdots, 1)$ and all entries of the $l \times l$ matrix \mathbf{E} are equal to 1. The computation is on the condition that the transition matrix \mathbf{P} is irreducible and aperiodic. The proof can be found in [21].

$$\mathbf{R}_{(f_i, c_x), (f_j, c_y)} = \begin{bmatrix} P_{(f_i, c_x, s_0), (f_j, c_y, s_0)} & P_{(f_i, c_x, s_0), (f_j, c_y, s_1)} & \cdots & P_{(f_i, c_x, s_0), (f_j, c_y, s_M)} \\ P_{(f_i, c_x, s_1), (f_j, c_y, s_0)} & P_{(f_i, c_x, s_1), (f_j, c_y, s_1)} & \cdots & P_{(f_i, c_x, s_1), (f_j, c_y, s_M)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{(f_i, c_x, s_M), (f_j, c_y, s_0)} & P_{(f_i, c_x, s_M), (f_j, c_y, s_1)} & \cdots & P_{(f_i, c_x, s_M), (f_j, c_y, s_M)} \end{bmatrix} \quad (16)$$

$$\begin{aligned} P_{(f_i, c_x, s_q), (f_j, c_y, s_l)} &= P(F_t = f_j, C_{t+1} = c_y, S_t = s_l | F_{t-1} = f_i, C_t = c_x, S_{t-1} = s_q) \\ &= P(F_t = f_j | F_{t-1} = f_i) P(C_{t+1} = c_y | C_t = c_x) P(S_t = s_l | F_{t-1} = f_i, S_{t-1} = s_q) \\ &= P_{f_i, f_j} P_{c_x, c_y} P(S_t = s_l | F_{t-1} = f_i, S_{t-1} = s_q) \end{aligned} \quad (17)$$

$$P(S_t = s_l | F_{t-1} = f_i, S_{t-1} = s_q) = \begin{cases} P(A_t = s_l - \max\{0, s_q - c_x\} | f_i), & \text{if } 0 \leq s_l < M, \\ 1 - \sum_{0 \leq s_l < M} P(S_t = s_l | F_{t-1} = f_i, S_{t-1} = s_q), & \text{if } s_l = M \end{cases} \quad (18)$$

Remark 1: When the number of states l is larger, the numerical computation of the inverse matrix $(\mathbf{I} - \mathbf{P} + \mathbf{E})^{-1}$ in (20) can cause difficulties. In this case it is often more convenient and efficient to solve the matrix equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ iteratively.

3) *Packet Dropping Rate:* Let P_d denote the packet dropping rate. When the remaining space of the queue is smaller than the number of arriving packets, packet overflow happens. With the current service rate C_t , the remaining space is $r_t = M - (S_{t-1} - C_t)$. Thus the queue can accommodate r_t arriving packets in the current time slot. Now, if the number of arriving packets A_t is larger than r_t , $A_t - r_t$ packets will be dropped. Therefore, based on the stationary distribution of $\boldsymbol{\pi}$, we will analyze and derive the packet dropping rate P_d .

Let D_t denote the number of packets dropped at time t , we can compute P_d as [4]

$$P_d \triangleq \lim_{T \rightarrow \infty} \frac{\sum_{t=1}^T D_t}{\sum_{t=1}^T A_t} = \frac{E\{D\}}{E\{A_t\}} = \frac{E\{D\}}{\bar{\lambda}}. \quad (21)$$

The average number of dropped packets $E\{D\}$ can be found in (22), shown at the bottom of the page, where $\theta(x, y)$ is a positive difference function, which returns the difference of x and y when $x > y$, and returns 0 when $x \leq y$.

With P_d available, we can get the effective average traffic rate and the system throughput. Having considered the influence of queuing, the performance measures in terms of packet successful transmission delay will be derived in the next subsection.

B. Delay Analysis

In this subsection, we investigate the average packet delay over the Nakagami- m fading channel with AMC in details. The derived P_d and the stationary distribution of three-dimensional Markov process will be used.

To facilitate the analysis of packet service process, the queuing effect needs to be taken into account. Based on the queue

state and service rate, we can get the actual service rate given ($S = s, C = c$) as

$$c_n(S = s, C = c) = \begin{cases} c_n, & \text{if } c = c_n, s \geq c \\ \frac{s}{c}c_n, & \text{if } c = c_n, s < c. \end{cases} \quad (23)$$

Thus, the average service rate calculation corresponding to state n can be written as (24), shown at the bottom of the page, where $\pi_{(s,c)}$ denotes the stationary distribution of the system state.

The effective average traffic rate into the queue can be evaluated as $r = \bar{\lambda}(1 - P_d)$. By [22], using the Pollaczek-Khintchine formula, we can get the mean queue length for state n as

$$\bar{Q}_q^n = \frac{rE\{S_{T_n}^2\}}{2(1 - \delta_n)}, \quad (25)$$

where $\delta_n = r/\bar{c}_n$ is the traffic intensity or utilization, and $E\{S_{T_n}^2\}$ is the second moment of the service distribution. Using (7), we can get $E\{S_{T_n}^2\}$. For notational brevity, we set $E\{S_{T_n}^2\} \triangleq f(\bar{P}_{s_n}, N_r^{max})$.

Theorem 1: A necessary condition on the existence of a steady state of the queue and the finite numbers of packet dropping is that the average received SNR should satisfy $\bar{\gamma} \geq \bar{\gamma}_{min}$, and $\bar{\gamma}_{min}$ is the $\bar{\gamma}$ by setting $\sum_{n=1}^N P_r(n)c_n = \bar{\lambda}$.

Proof: The proof is given in Appendix B. \square

Remark 2: For the preceding quantities to be finite, we need $\delta_n < 1$. This condition is intuitive since we know from renewal theory that if the server was always busy, then the departure rate would be \bar{c} (see [23, Section 7.3]), which must be larger than the arrival rate r to keep things finite.

It is known that the average waiting time of a packet consists of queuing time and service time for the M/G/1 queue, and the queuing delay is $\bar{D}_q = \frac{\bar{Q}_q}{r}$, which is akin to Little's formula [22]. In summary, the average delay \bar{W}_n for a packet corresponding to state n is given by (26), shown at the bottom of the page, where

$$\begin{aligned} E\{D\} &= \sum_{f \in \mathbb{F}, s \in \mathbb{S}, c \in \mathbb{C}} DP(A_t = a | F_{t-1} = f, S_{t-1} = s, C_t = c) \\ &= \sum_{f \in \mathbb{F}, s \in \mathbb{S}, c \in \mathbb{C}} [\theta(A_t, M - (S_{t-1} - C_t)) \times P(A_t = a | f) \times \pi_{f,s,c}] \end{aligned} \quad (22)$$

$$\begin{aligned} \bar{c}_n &= \lim_{t \rightarrow \infty} \sum_{s \in \mathbb{S}, c = c_n} \frac{[c_n(S_{t-1} = s, C_t = c) \times P(S_{t-1} = s, C = c)]}{\sum_{s \in \mathbb{S}, c = c_n} P(S_{t-1} = s, C = c)} \\ &= \sum_{s \in \mathbb{S}, c = c_n} \frac{c_n(S = s, C = c) \pi_{(s,c)}}{\sum_{s \in \mathbb{S}, c = c_n} \pi_{(s,c)}} \end{aligned} \quad (24)$$

$$\bar{W}_n = \frac{T_u \bar{Q}_q^n}{r} + T_u E\{S_{T_n}\} = \frac{T_u f(\bar{P}_{s_n}, N_r^{max})}{2(1 - \delta_n)} + T_u \bar{P}_{s_n}(\tau_n - \bar{\tau}) + \frac{\bar{\tau} T_u}{\bar{P}_{s_n}} \left[1 - (1 + N_r^{max} \bar{P}_{s_n})(1 - \bar{P}_{s_n})^{N_r^{max}} \right] \quad (26)$$

$\tau_n(\bar{b}_n)$ is given by substituting \bar{b}_n into (8). Thus, the average delay for a packet with AMC can be derived as

$$\bar{W} = \sum_{n=1}^N P_r(n) \bar{W}_n. \quad (27)$$

Substituting (26) into (27), we can get the average delay \bar{W} . Note that, for a given $\bar{\gamma}$, the average delay \bar{W} can be derived. Thus, $\bar{W} = \phi(\bar{\gamma})$, and $\phi(\bar{\gamma})$ is a function of $\bar{\gamma}$. With the closed-form expression of the delay available, we will optimize the system energy efficiency for given delay demand in the next section.

IV. ENERGY EFFICIENCY ANALYSIS

In this section, we determine the energy efficiency with the joint effects of finite length queue and AMC also with the arrival data traffic, then we investigate the relationship between the energy efficiency and the packet delay. With the average service rate and the effective traffic arrival rate available, we focus on the average transmission power and the resulting system throughput over the Nakagami- m fading channel.

A. Energy Efficiency

We consider a variable power and rate transmission scheme utilizing a discrete-rate M-QAM scheme. Whenever the CSI feedback to the transmitter falls within the interval $[\gamma_n, \gamma_{n+1})$, the transmission rate b_n of AMC is chosen, data is transmitted with power $e_{n,t}(\gamma)$ at time-slot t . Thus, the received SNR is $\gamma e_{n,t}(\gamma)/\bar{e}$, and

$$\gamma = \frac{\bar{e} |h(t)|^2}{\sigma^2}, \quad (28)$$

where $|h(t)|^2$ denotes the instantaneous channel power gain. The BER for channel state n can be expressed as a function of the received SNR $\gamma e_{n,t}(\gamma)/\bar{e}$ as [24]

$$P_b^n \approx 0.2 \exp\left(-\frac{1.5}{2^{b_n}-1} \frac{e_{n,t}(\gamma)}{\bar{e}} \gamma\right), \quad \gamma_n \leq \gamma < \gamma_{n+1}. \quad (29)$$

By considering the traffic and queueing influence, we should use the average transmission rate \bar{b}_n to replace b_n . From (10), (23) and (24), we can get

$$\bar{b}_n = \frac{\bar{c}_n}{c_n} b_n. \quad (30)$$

Substituting (30) into (29), we get the equivalent transmission power for each AMC mode as a function of the BER as follows:

$$e_{n,t}(\gamma) = \frac{\bar{e}(2^{\bar{b}_n}-1)}{1.5\gamma} \ln \frac{0.2}{P_b^n}. \quad (31)$$

Let \bar{P}_b^n denote the average BER corresponding to channel state n , from (3), we can derive \bar{P}_b^n as

$$\bar{P}_b^n = \frac{1}{P_r(n)} \int_{\gamma_n}^{\gamma_{n+1}} 0.2 \exp\left(-\frac{1.5}{2^{\bar{b}_n}-1} \gamma\right) f_{\bar{\gamma}}(\gamma) d\gamma. \quad (32)$$

Based on (2) and (32), we can get the average BER \bar{P}_b^n corresponding to state n . Thus, the average transmission power in channel state n is

$$\bar{e}_n = \int_{\gamma_n}^{\gamma_{n+1}} \frac{\bar{e}(2^{\bar{b}_n}-1)}{1.5\gamma} \ln \frac{0.2}{\bar{P}_b^n} f_{\bar{\gamma}}(\gamma) d\gamma. \quad (33)$$

From (3) and (33), we can approximate the average transmission power with AMC technique as

$$\tilde{e} = \sum_{n=1}^N \bar{e}_n P_r(n). \quad (34)$$

The system throughput is the long-term average rate at which packets are successfully transmitted. Therefore, the packet dropping rate from queueing and packet violation from the channel with $N_r^{max} - 1$ retransmissions are influencing the system throughput. For an average packet arrival rate $\bar{\lambda}$, a packet dropping rate P_d , and an average packet successful transmission rate \bar{P}_s , the system average throughput \bar{T} can be calculated by

$$\bar{T} = \bar{\lambda}(1 - P_d) \left(1 - (1 - \bar{P}_s)^{N_r^{max}}\right) / T_u, \quad (35)$$

where P_d is corresponding to (21). The average probability of successful packet transmission \bar{P}_s of AMC can be calculated as the ratio of the average number of packets successfully transmitted over the total average number of transmitted packets

$$\bar{P}_s = \frac{\sum_{n=1}^N \bar{c}_n P_r(n) \bar{P}_{s_n}}{\sum_{n=1}^N \bar{c}_n P_r(n)}. \quad (36)$$

Based on (4), (34), (35), and (36), the energy efficiency can be denoted as

$$f_{ee} = \frac{\bar{T}}{\tilde{e}} = \frac{\bar{\lambda}(1 - P_d) \left(1 - (1 - \bar{P}_s)^{N_r^{max}}\right) / T_u}{\sum_{n=1}^N \bar{e}_n P_r(n)}. \quad (37)$$

As we know that different thresholds γ_n ($n = 1, \dots, N$) lead to different average transmission power and system throughput, we will derive the energy efficient thresholds in the next subsection.

B. Energy Efficient Thresholds

It can be noted from (3) that different thresholds γ_n ($n = 1, \dots, N$) determine the probability distribution of different transmission rate b_n ($n = 1, \dots, N$) over $\mathbb{F} \times \mathbb{S} \times \mathbb{C}$ with $\bar{\gamma}$ available. Thus, for a particular channel state, we can get the service rate with corresponding probability of applying in that state, which is also called the transmission control policy $\mu(\gamma_n, \bar{\gamma}) = \{P_r(n)\}$. Thus, we need to derive the thresholds that can give the energy efficient transmission policy.

The choices of thresholds $\{\gamma_1, \gamma_2, \dots, \gamma_N\}$ can be arbitrary. In [2], [25], the equal probability method (EPM) was proposed, where the received SNRs are divided so that the stationary probabilities, $P_r(n)$, $n = 1, \dots, N$ of staying in all states are the same (i.e., $P_r(1) = \dots = P_r(N) = \frac{1}{N}$). The partition based on minimum SNR required to achieve P_{target} (MSRE) was proposed in [3].

Since we want to get the energy efficient transmission, we should set the threshold γ_n for the transmission mode n to be

TABLE II
SUMMARY OF THE PARAMETERS CORRESPONDING TO THE
ENERGY EFFICIENT THRESHOLD FOR AN UNCODED
SYSTEM WITH 100 BITS PER PACKET

	Modulation	Rate(bits/sym.)	α_n	β_n	γ_n^* (dB)
Mode 1	BPSK	1	0.5858	3	2.81
Mode 2	QPSK	2	1	1	8.06
Mode 3	8-QAM	3	1.2929	0.4286	11.02
Mode 4	16-QAM	4	1.5	0.2	15.67
Mode 5	32-QAM	5	1.6464	0.0968	18.22
Mode 6	64-QAM	6	1.75	0.0476	21.60
Mode 7	128-QAM	7	1.8232	0.0236	24.61

the energy efficient SNR. For deriving energy efficient γ_n , we can approximate the throughput as

$$\bar{T} \approx \bar{\lambda}(1 - P_d)P_{s_n}(\gamma)/T_u, \quad (38)$$

where $P_{s_n}(\gamma)$ is the instantaneous packet success rate. Note that the instantaneous packet success rate is given by

$$P_{s_n}(\gamma) = \left(1 - \alpha_n Q(\sqrt{\beta_n \gamma})\right)^{\frac{2L}{b_n}}, \quad (39)$$

where $\alpha_n = 2(1 - 2^{-b_n/2})$, $\beta_n = \frac{3}{2^{b_n}-1}$ and $Q(\cdot)$ is the complementary cumulative distribution function of the standard Gaussian random variable. Thus, based on (38) and (28), the function of energy efficiency corresponding to state n can be written as

$$f_{ee}(\gamma) = \frac{\bar{\lambda}(1 - P_d) |h(t)|^2 / T_u P_{s_n}(\gamma)}{\sigma^2 \gamma}. \quad (40)$$

Taking derivative of (40) with respect to γ and equating it to zero, it can be shown that the energy efficient partition (EEP) γ_n^* satisfies

$$P_{s_n}(\gamma_n^*) = \gamma_n^* P_{s_n}'(\gamma_n^*). \quad (41)$$

It is shown (see [26]) that for an S-shaped (sigmoidal) function, $P_{s_n}(\gamma_n^*) = \gamma_n^* P_{s_n}'(\gamma_n^*)$ has a unique solution, and $P_{s_n}(\gamma)$ is S-shaped.

Lemma 2: The energy efficient thresholds γ_n^* ($n = 1, \dots, N$) is the unique solution of the following equation:

$$\frac{\alpha_n L}{b_n} \sqrt{\frac{\beta_n \gamma}{2\pi}} e^{-\frac{\beta_n \gamma}{2}} + \alpha_n Q(\sqrt{\beta_n \gamma}) = 1.$$

Lemma 2 is from [26], therefore, we can compute γ_n^* numerically for the AMC transmission mode n . Table II summarizes the results for a system with $L = 100$ bits.

C. Energy Efficient Policy With Delay Constraints

Theorem 2: For a given transmission policy $\mu^0 = \mu(\gamma_n^0, \bar{\gamma}^0)$, the average energy efficiency is nondecreasing as increasing the buffer size M , and the energy efficiency converges to a supremum.

Proof: The proof is given in Appendix C. \square

Theorem 2 reveals the influence of the buffer size to energy efficient transmission. As can be seen, the buffer size should be large enough while we only consider the energy efficient transmission.

Based on the energy efficient thresholds from Table II and the problem formulation from (5), we can reformulate the energy efficient optimization problem as

$$\begin{aligned} \max \quad & f_{ee}, \\ \text{s.t.} \quad & \bar{W} = \phi(\bar{\gamma}) \leq W_0, \\ & \mathbb{F} = \mathbb{F}_0, \\ & \bar{\gamma} \leq \bar{\gamma}_{\max}, \\ & \bar{\gamma} \geq \bar{\gamma}_{\min}, \end{aligned} \quad (42)$$

where $\bar{\gamma}_{\min}$ and $\bar{\gamma}_{\max}$ are respectively the minimum required SNR from Theorem 1 and the maximum average received SNR, and \bar{W} is corresponding to (27). Our objective is to determine the optimal prescribed average received SNR $\bar{\gamma}^{opt}$ at the physical layer that maximizes the f_{ee} , which is corresponding to the energy efficient transmission policy based on the queue state and the traffic state as well as the delay demand.

The average received SNR can influence the probability of choosing different AMC modes, so as to influence the service rate and the throughput as well as the average transmission power. The performance of energy efficiency is analyzed in terms of the throughput, power, and delay. Thus, the nonlinear function of f_{ee} is too complex. The optimization problem of (42) can be numerically solved. To solve this optimization problem, all the packet arrival process of MMPP and the queueing influence and the channel Markov model based approach presented in the previous sections will be used.

At first, in order to determine an interval that contains a interval optimum given starting point $\bar{\gamma}_0$, bracketing method is used [27]. It iteratively walks further until we are certain to have an interval (barcket) $[\alpha, \beta]$ that includes an interior optimum point. Thus, we propose a bracketing algorithm to get the interval. The algorithm 1 enlarges the initial interval with endpoints $\bar{\gamma}_0$ and $\bar{\gamma}_0 \pm \epsilon$ with a step which leads each iteration factor $\rho > 1$ bigger, and the choice of $\rho = \frac{2}{\sqrt{5}-1}$ comes from [27]. The algorithm stops when finally $\bar{\gamma}_{k-1}$ has a higher function value than $\bar{\gamma}_k$ as well as $\bar{\gamma}_{k-2}$.

Algorithm 1 Bracket($f_{ee}, W_0, \epsilon, \alpha, \beta$)

- 1: set $k := 1, \rho = \frac{2}{\sqrt{5}-1}$;
 - 2: based on (26) and (27), compute the threshold $\bar{\gamma}_0 = \phi^{-1}(W_0)$;
 - 3: **if** $f_{ee}(\bar{\gamma}_0 + \epsilon) > f_{ee}(\bar{\gamma}_0)$ **then**
 - 4: $\bar{\gamma}_1 := \bar{\gamma}_0 + \epsilon$;
 - 5: **else**
 - 6: **if** $f_{ee}(\bar{\gamma}_0 + \epsilon) < f_{ee}(\bar{\gamma}_0)$ **then**
 - 7: $\bar{\gamma}_1 := \bar{\gamma}_0 - \epsilon$;
 - 8: **end if**
 - 9: **else**
 - 10: Stop; $\bar{\gamma}_0$ is optimal;
 - 11: **end if**
 - 12: **repeat**
 - 13: $k := k + 1$;
 - 14: $\bar{\gamma}_k := \bar{\gamma}_{k-1} + \rho(\bar{\gamma}_{k-1} - \bar{\gamma}_{k-2})$;
 - 15: **until** $((f_{ee}(\bar{\gamma}_k) < f_{ee}(\bar{\gamma}_{k-1})) \vee (\bar{\gamma}_k \geq \bar{\gamma}_{\max}))$
 - 16: $\alpha := \min\{\bar{\gamma}_k, \bar{\gamma}_{k-2}\}$
 - 17: $\beta := \max\{\min\{\bar{\gamma}_k, \bar{\gamma}_{\max}\}, \bar{\gamma}_{k-2}\}$
-

With the interval that contains the optimum of (42), we will shrink it to a tiny interval enclosing the maximum point. One such method is Golden Section search [27]. This algorithm uses two evaluated points l (left) and r (right) in the interval $[\alpha, \beta]$, by locating in such a way that one of the points can be used again in the next iteration. The evaluation points l and r are located with fraction ζ in such a way that $l = \alpha + (1 - \zeta)(\beta - \alpha)$ and $r = \alpha + \zeta(\beta - \alpha)$. The value of ζ is the so called Golden Section number $\zeta = \frac{\sqrt{5}-1}{2} \approx 0.618$, which also corresponds to the value ρ in the Bracketing algorithm 1. Using the outcomes of algorithm 1 as input into the Golden Section search as $[\alpha, \beta]$ gives that the point $\tilde{\gamma}_{k-1}$ (of algorithm 1) corresponds to $\tilde{\gamma}_0^g$ (in algorithm 2).

Algorithm 2 Golden Section ($[\alpha, \beta], f_{ee}, \epsilon$)

- 1: Set $k := 1, \alpha_1 := \alpha$ and $\beta_1 := \beta, \zeta := \frac{\sqrt{5}-1}{2}$;
 - 2: $l := \tilde{\gamma}_0^g := \alpha + (1 - \zeta)(\beta - \alpha), r := \tilde{\gamma}_1^g := \alpha + \zeta(\beta - \alpha)$;
 - 3: Evaluate $f_{ee}(l) := f_{ee}(\tilde{\gamma}_0^g)$
 - 4: **repeat**
 - 5: Evaluate $f_{ee}(\tilde{\gamma}_k^g)$;
 - 6: **if** ($f_{ee}(r) > f_{ee}(l)$) **then**
 - 7: $\alpha_{k+1} := l, \beta_{k+1} := \beta_k, l := r$
 - 8: $r := \tilde{\gamma}_{k+1}^g := \alpha_{k+1} + \zeta(\beta_{k+1} - \alpha_{k+1})$;
 - 9: **else**
 - 10: $\alpha_{k+1} := \alpha_k, \beta_{k+1} := r, r := l$;
 - 11: $l := \tilde{\gamma}_{k+1}^g := \alpha_{k+1} + (1 - \zeta)(\beta_{k+1} - \alpha_{k+1})$;
 - 12: **end if**
 - 13: $k := k + 1$;
 - 14: **until** ($\beta_k - \alpha_k < \epsilon$)
-

As a consequence, we can get the energy efficient solution $\tilde{\gamma}^{opt} = \arg \max f_{ee}$ under the delay constraint W_0 . Hence, for given traffic \mathbb{F}_0 , the energy efficient transmission policy under delay constraint can be determined as

$$\mu^{opt} = \mu(\gamma_n^*, \tilde{\gamma}^{opt}(W_0)).$$

The energy efficient average transmission power can also be determined.

In the next section, we will show the performance of the system based on our derivations and analysis by simulations and numerical methods.

V. NUMERICAL RESULTS AND DISCUSSION

This section presents simulation results to evaluate the previous theoretical development of delay-aware energy efficient communications with joint consideration of traffic, queuing and channel condition.

A. System Parameters

Unless specified otherwise, for all simulations, we assume that number of traffic, channel and queue states are $K = 2$, $N = 7$ and $M + 1 = 51$ respectively. The packet size $L = 100$, the maximum retransmission times of packet is $N_r^{max} = 6$. For traffic \mathbb{F}_0 , average arrival rate, $\lambda_1 = 1$ packets/time-unit and

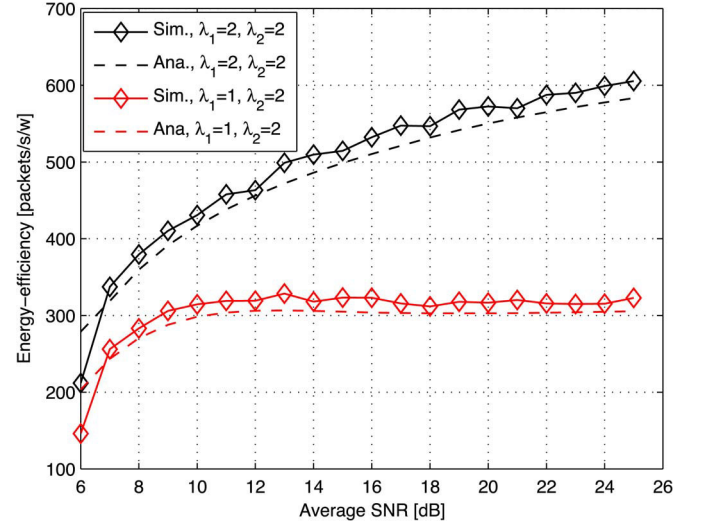


Fig. 2. The analysis and simulation results of the average energy efficiency versus average SNR for Markov modulated Poisson traffic model under Nakagami- m fading channel with $m = 1$.

$\lambda_2 = 2$ packets/time-unit, and the transition probability matrix characterizing the MMPP service as

$$\mathbf{P}^f = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}.$$

Maximum number of packet arrivals $A = 15$. The symbol rate $R_s = 100$ KHz; The Nakagami parameter $m = 1$ (which corresponds to Rayleigh fading channel with no line of sight (LOS) component); Doppler frequency $f_d = 10$ Hz. Like 3GPP LTE standard, we assume that block (also called frame) length $T_u = 0.002$ s.

B. Performance of Energy Efficient Thresholds

Under the condition of only traffic f_1 , we study the performance of energy efficient thresholds as a function of average received SNR and as a function of the packet arrival rate $\bar{\lambda}$. In order to validate the analytical energy efficiency, we simulate the cross-layer system with packet transmission over a Rayleigh fading channel with EEP method. Fig. 2 shows the results of analysis and simulated values of energy efficiency for different average traffic rates as a function of channel average SNR. It can be observed that the variations of analytical and simulation results agree reasonably well.

Fig. 3 shows the average energy efficiency, we can see that the method of EEP can offer better energy efficiency than that of MSRE at different SNRs, although the gap of the energy efficiency is very small at some SNRs. The energy efficiency is increasing when increasing the average arrival rate, that means, for large packet arrival rate, the average throughput is increasing faster than the average transmission power. However, when the average SNR is increasing greatly, there is no increasing in energy efficiency of both the MSRE and EEP for the small average arrival rate. This result comes from the fact that the throughput is equivalent to increasing very small for the small average arrival rate, but the average transmit power is linearly increasing with SNR. The blue curve shows the energy efficiency without

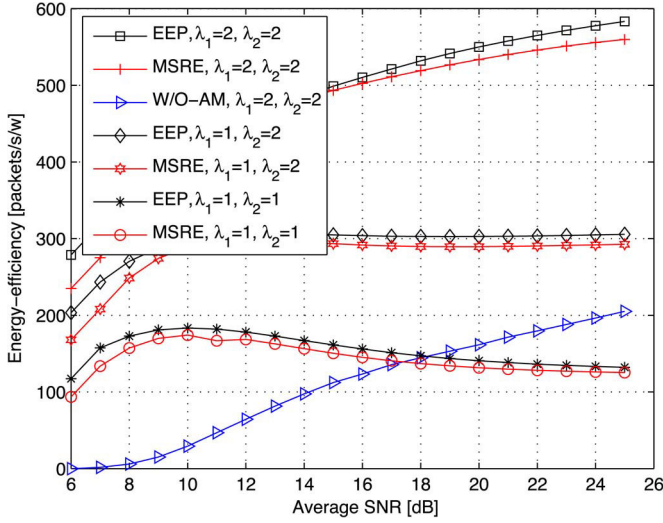


Fig. 3. Average energy efficiency versus average SNR between the minimum SNR required to achieve P_{target} (MSRE) and the energy efficient partition (EEP), where $P_{\text{target}} = 0.001 \frac{1}{N_r^{\max}+1}$.

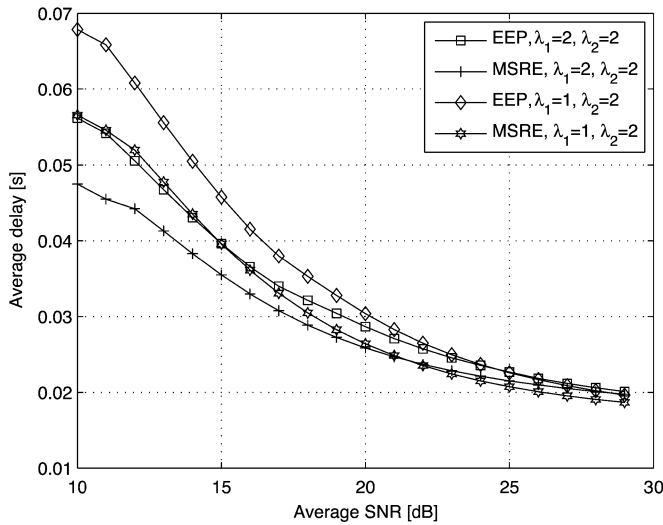


Fig. 4. Average delay versus average SNR.

adaptive modulation (W/O-AM) with average traffic rate $\bar{\lambda} = 2$ packets/time-unit, which performs much worse than that of the cross-layer policy with adaptive modulation.

Ensure the energy efficient transmission, we can also observe from the figure that the average SNR should be as large as possible when the average arrival rate is large, i.e., $\bar{\lambda} = 2$ packets/time-unit, which means that the probability of choosing large transmission rate should be increased. With regard to the delay, we can see from Fig. 4 that the delay is decreasing when increasing SNR. Thus, the SNR should be increased when the delay demand is more strict. Therefore, we can get the optimal energy efficient transmission policy under different delay constraints. Based on Fig. 3 and Fig. 4, the energy efficient SNR for the small average arrival rate is decreasing when increasing the delay demand, i.e., $\bar{\lambda} = 1$ packets/time-unit. Thus, from the numerical results of [3], when the delay is large, the result of the energy efficient transmission policy for the small average arrival traffic is that the probability of

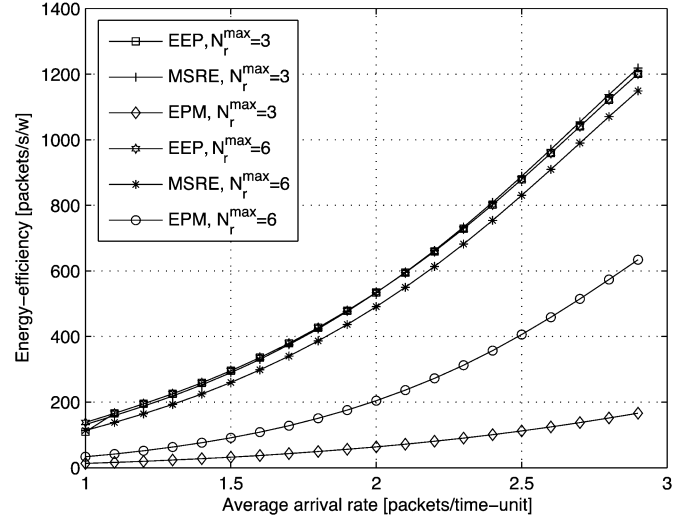


Fig. 5. Average energy efficiency versus average packet arrival rate $\bar{\lambda}$, the packet arrival rate in traffic state f_2 is kept zero, and the average SNR $\bar{\gamma} = 20$ dB.

choosing small transmission rate is larger than that of choosing large transmission rate. On the other hand, when the delay is small, the result of the energy efficient transmission policy for the small average arrival traffic is that the probability of choosing large transmission rate is larger than that of choosing small transmission rate. However, the energy efficient transmission policy for the large average arrival traffic is invariant no matter what the delay demand is.

Fig. 5 shows that the energy efficiency of EEP and MSRE is increasing when reducing the retransmission times, while the energy efficiency of EPM is increasing when increasing the retransmission times. This implies that we should choose the minimum retransmission times according to the required packet loss rate at the physical channel. We can also find that MSRE gives better performance than EEP when the average arrival rate is increasing. This comes from the fact that although the EEP is better in power consumption at physical layer transmission, the packet loss rate by queuing is larger than that of MSRE in the regime of large traffic arrival rate, since MSRE has packet loss rate constraint. Thus the throughput of EEP is much smaller than that of MSRE, as well as the system energy efficiency. It is evident from the plot that the EPM always has worst performance in energy efficiency. At the same time, we can see from Fig. 6 that the power consumption is increasing when increasing the traffic arrival rate, since large traffic arrival rate causes large transmission rate. Although the power consumption of EPM is less than that of EEP when the retransmission times is 3, the throughput of EPM is much less than that of EEP based on Fig. 5. In all, The EEP can still perform better in power consumption than the other methods.

We can see the delay performance versus the average arrival rate in Fig. 7, where the delay is increasing when increasing the average arrival rate. On the other hand, it can be seen from Fig. 7 that EPM achieves smaller delay in comparison with the other two methods for the reason of larger transmission rate. However, EPM has much larger average power consumption than those of EEP and MSRE, which causes poor performance in energy efficiency.

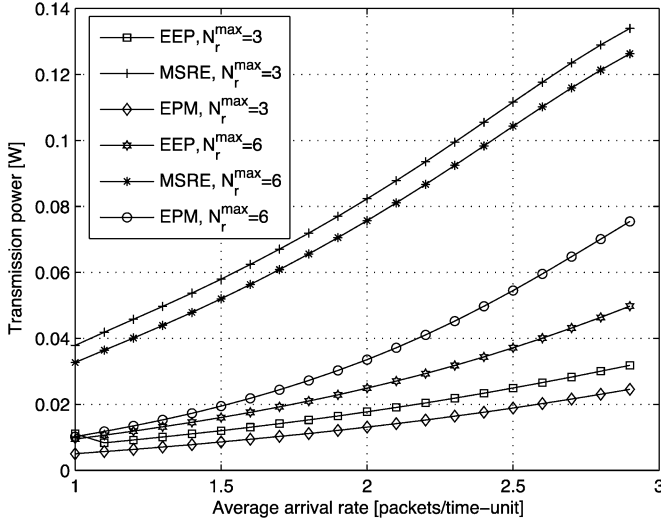
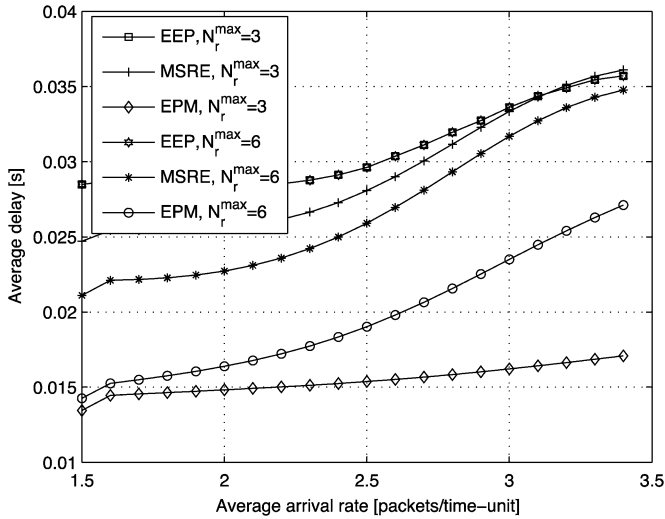


Fig. 6. Average power consumption versus average packet arrival rate.

Fig. 7. Average delay versus average packet arrival rate $\bar{\lambda}$, the packet arrival rate in traffic state f_2 is kept zero, and the average SNR $\bar{\gamma} = 20$ dB.

C. Performance of Energy Efficient Policy

Fig. 8 shows the optimal energy efficiency versus the average traffic arrival rate for two different delay demands, which are the delay tolerant and the delay sensitive. Specifically, when the delay requirement is higher than 20 ms, we call the service be delay tolerant, otherwise, the service be delay sensitive. For our simulation, the delay constraints of delay tolerant and delay sensitive are respectively $W_0 = 20$ ms and $W_0 = 70$ ms. It can be seen from the figure that EEP always performs better than MSRE for different arrival rates.

Since the energy efficiency for the large average arrival rate is increasing when increasing the average transmission power, the optimal energy efficiency can be obtained from the largest available average transmission power. For larger packet arrival rate, the queue remains almost full most of the times, that results in increasing packet overflow rate from the queue. Thus the average transmission power should be larger to keep large throughput. This means that we can get the optimal

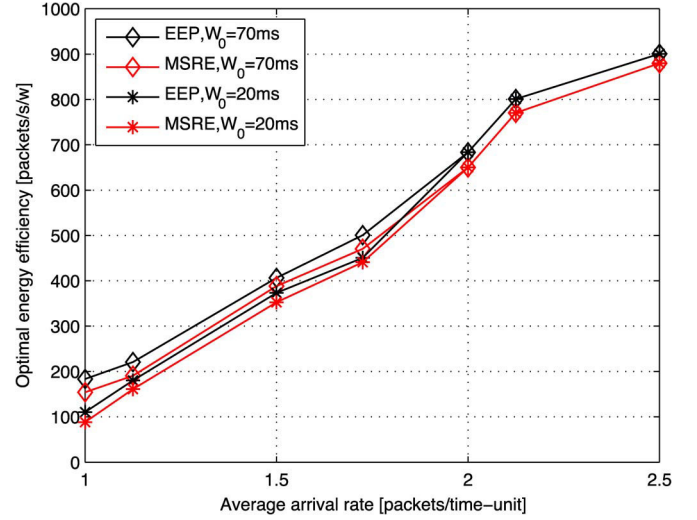
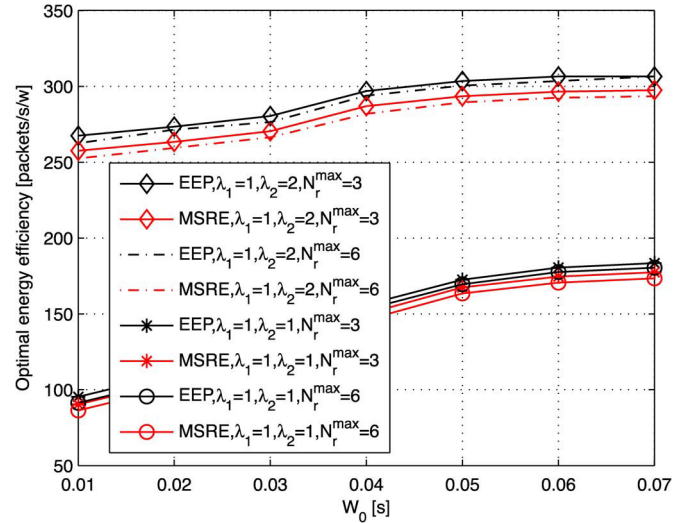


Fig. 8. Optimal energy efficiency versus the average traffic arrival rate for two different delay demands.

Fig. 9. Optimal energy efficiency versus the delay constraint W_0 for EEP and MSRE transmission.

energy efficiency for the AMC transmission by using as much transmission power as possible when the average arrival rate of the traffic is much large, i.e., $\bar{\lambda} = 2$ packets/time-unit, which is corresponding to the result of increasing the probability with using the large transmission rate. On the other hand, the delay tolerant service needs small average SNR. Consequently, we can observe from Fig. 8 that the optimal energy efficiency of delay tolerant traffic converges to that of delay sensitive traffic when the average arrival rate is increasing.

Fig. 9 shows the optimal energy efficiency versus the delay constraint W_0 for different arrival rates, where we use EEP for AMC transmission. We can observe from the curve that the energy efficiency is increasing when increasing W_0 at the regime of small W_0 . At the regime of large W_0 , the optimal energy efficiency converges to a stable value. Therefore, at large delay region, the optimal energy efficient transmission policy is almost the same irrespective of delay and arrival rate variations. We can also see from Fig. 9 that the energy efficiency is

decreasing when increasing the retransmission times for given delay constraint W_0 . This comes from the fact that delay can be larger when the retransmission times is larger. Then the power consumption should be larger to guarantee the delay constraint.

VI. CONCLUSION

In this paper, we present a cross-layer framework that determines the energy efficient transmission policy based on both the physical and the upper-layer information. We derive the closed-form expression of delay considering joint effects of the general arriving traffic and the dynamic data queue as well as the general channel model. With regard to the physical layer transmission, we propose an energy efficient partition method to achieve the AMC transmission. Our numerical results show that the energy efficiency is better than those of the existing partition methods with the energy efficient AMC partition. At last, we propose an algorithm to get the energy efficient transmission policy under different delay constraints. By the energy efficient policy, we get the result that the optimal energy efficiency increases when the average traffic arrival rate increases.

APPENDIX A

PROOF OF LEMMA 1

Based on the theorem of [23, Theorem 4.1], the Markov chain $\{(F_t, S_t, C_t), t \geq 0\}$ exists stationary distribution only when the Markov chain is irreducible and recurrent. The Markov chain is said to be irreducible if there is only one class, that is, if all states are accessible to each other. Thus, we only need to prove that any state (f_i, c_x, s_q) can access any state (f_j, c_y, s_l) , where $(f_i, c_x, s_q) \in \mathbb{F} \times \mathbb{C} \times \mathbb{S}$ and $(f_j, c_y, s_l) \in \mathbb{F} \times \mathbb{C} \times \mathbb{S}$.

The MMPP traffic has nonzero transition probability for each transition from f_i to f_j , denoted as $f_i \rightarrow f_j$. From (18), we can get

$$P\{(c_x, s_q)|(c_x, s_q)\} = P(A_t = s_q - \max\{0, s_q - c_x\}|f_i).$$

Then the transition

$$(f_i, c_x, s_q) \rightarrow (f_j, c_x, s_q) \quad (43)$$

has nonzero probability.

When the traffic stays in the state f_j , the transition probability

$$P\{(c_x, s_q)|(c_x, s_l)\} = P(A_t = s_l - \max\{0, s_l - c_x\}|f_j).$$

Thus, the transition from

$$(f_j, c_x, s_q) \rightarrow (f_j, c_x, s_l) \quad (44)$$

has nonzero probability. By the finite channel model, the channel state x can always have transition path to the state y from the neighbour state. Then state c_x can go to state c_y , and

$$P\{(c_x, s_l)|(c_y, s_l)\} = P(A_t = s_l - \max\{0, s_l - c_x\}|f_i).$$

Therefore,

$$(f_j, c_x, s_l) \rightarrow (f_j, c_y, s_l) \quad (45)$$

also has nonzero transition probability. Based on (43), (44), and (45), we can get the result that the $\{(F_t, S_t, C_t), t \geq 0\}$ is irreducible.

On the other hand, based on the conclusion of [23], that all states in a finite irreducible Markov chain are recurrent. In all, the stationary distribution of the Markov process $\{(F_t, S_t, C_t), t \geq 0\}$ exists.

APPENDIX B

PROOF OF THEOREM 1

From [22], we know that the queue size increases with bound depending on the condition $\bar{c} > r$. Since $r = \bar{\lambda}(1 - P_d)$, therefore, $\sup_{0 \leq P_d \leq 1} r = \bar{\lambda}$, causing $\bar{c} \geq \bar{\lambda}$.

In AMC transmission strategy, while increasing the average received SNR, the probability of choosing the large transmission rate is increasing based on the numerical results of [3]. We know that

$$c_n = \frac{T_u b_n R_s \bar{P}_{s_n}}{L} \left[1 - (1 + N_r^{max} \bar{P}_{s_n}) (1 - \bar{P}_{s_n})^{N_r^{max}} \right]^{-1},$$

which is an increasing function of n , and $c_n \geq \bar{c}_n$. On the other hand

$$\bar{c} = \sum_{n=1}^N P_r(n) \bar{c}_n,$$

thus,

$$\sum_{n=1}^N P_r(n) c_n \geq \sum_{n=1}^N P_r(n) \bar{c}_n = \bar{c}.$$

Therefore, the necessary condition can be written as $\sum_{n=1}^N P_r(n) c_n \geq \bar{\lambda}$ for simple calculation, and $\sum_{n=1}^N P_r(n) c_n$ is an increasing function of $\bar{\gamma}$. And when $\bar{\gamma} = \bar{\gamma}_{min}$, we have $\sum_{n=1}^N P_r(n) c_n = \bar{\lambda}$. In all, when $\bar{\gamma} \geq \bar{\gamma}_{min}$, we can get $\bar{c} > r$.

APPENDIX C

PROOF OF THEOREM 2

Based on (22), when the buffer size is increasing to infinite and S_{t-1} is small, we can get

$$\theta(A_t, M - (S_{t-1} - C_t)) = \theta(A_t, M) = 0.$$

Note that while $S_{t-1} \rightarrow M$, $\pi_{f,s=S_{t-1},c} \rightarrow 0$. As a result, we can get $P_d = 0$ as M increases.

For given $(\gamma_n^0, \bar{\gamma}^0)$, the packet successful transmission rate $\bar{P}_s(\gamma_n^0, \bar{\gamma}^0)$ can be computed from (36). Thus, the supremum of the average throughput $\sup_{\mu^0} \bar{T} = \bar{\lambda}(1 - (1 - \bar{P}_s(\gamma_n^0, \bar{\gamma}^0))^{N_r^{max}})/T_u$.

On the other hand, the average power consumption $\tilde{e}(\gamma_n^0, \bar{\gamma}^0)$ is determinate for given $(\gamma_n^0, \bar{\gamma}^0)$. In all, we can get the supremum of the average energy efficiency as

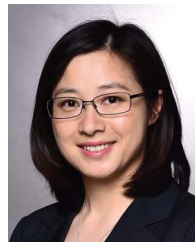
$$f_{ee} = \frac{\bar{\lambda} \left(1 - (1 - \bar{P}_s(\gamma_n^0, \bar{\gamma}^0))^{N_r^{max}} \right) / T_u}{\tilde{e}(\gamma_n^0, \bar{\gamma}^0)}.$$

REFERENCES

- [1] A. T. Hoang and M. Motani, "Cross-layer adaptive transmission: Optimal strategies in fading channels," *IEEE Trans. Commun.*, vol. 56, no. 5, pp. 799–807, May 2008.
- [2] A. K. Karmakar and V. K. Bhargava, "Performance of cross-layer optimal adaptive transmission techniques over diversity nakagami- m fading channels," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3640–3652, Dec. 2009.
- [3] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1746–1755, Sep. 2004.
- [4] Q. Liu, S. Zhou, and G. B. Giannakis, "Queueing with adaptive modulation and coding over wireless links: Cross-layer analysis and design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1142–1153, May 2005.
- [5] J. Choi and J. Ha, "On the energy efficiency of AMC and HARQ-IR with QoS constraints," *IEEE Trans. Veh. Tech.*, vol. 62, no. 7, pp. 3261–3270, Sep. 2013.
- [6] L. B. Le, E. Hossain, and A. S. Alfa, "Radio link level performance evaluation in wireless networks using multi-rate transmission with ARQ-based error control," *IEEE Trans. Wireless Commun.*, vol. 5, no. 10, pp. 2647–2653, Oct. 2006.
- [7] F. Ishizaki and G. U. Hwang, "Cross-layer design and analysis of wireless networks using the effective bandwidth function," *IEEE Trans. Wireless Commun.*, vol. 6, no. 9, pp. 3214–3219, Sep. 2007.
- [8] X. Wang, Q. Liu, and G. B. Giannakis, "Analyzing and optimizing adaptive modulation coding jointly with ARQ for QoS-guaranteed traffic," *IEEE Trans. Veh. Tech.*, vol. 56, no. 2, pp. 710–720, Mar. 2007.
- [9] M. Poggioni, L. Rugini, and P. Banelli, "QoS analysis of a scheduling policy for heterogeneous users employing AMC jointly with ARQ," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2639–2652, Sep. 2010.
- [10] T. Kim and J.-T. Lim, "Queueing analysis in a multiuser diversity system with adaptive modulation and coding," *IEEE Trans. Veh. Tech.*, vol. 60, no. 1, pp. 338–342, Jan. 2011.
- [11] X. Bai and A. Shami, "Two dimensional cross-layer optimization for packet transmission over fading channel," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3813–3822, Oct. 2008.
- [12] M. M. Butt, R. R. Muller, and K. Kansanen, "Individual packet deadline delay constrained opportunistic scheduling for multiuser systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2014, pp. 1–16, Apr. 2014.
- [13] N. Mastronarde and V. D. Schaar, "Joint physical-layer and system-level power management for delay-sensitive wireless communications," *IEEE Trans. Mobile Comput.*, vol. 12, no. 4, pp. 694–709, Apr. 2013.
- [14] N. Mastronarde and M. V. D. Schaar, "Fast reinforcement learning for energy-efficient wireless communication," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6262–6266, Dec. 2011.
- [15] N. Salodkar, A. Bhorkar, A. Karandikar, and V. S. Borkar, "An on-line learning algorithm for energy efficient delay constrained scheduling over a fading channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 732–742, May 2008.
- [16] X. Zhang and J. Tang, "Power-delay tradeoff over wireless networks," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3673–3684, Sep. 2013.
- [17] R. A. Berry, "Optimal power-delay tradeoffs in fading channels: small-delay asymptotics," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3939–3952, Jun. 2013.
- [18] J. Lee and N. Jindal, "Energy-efficient scheduling of delay constrained traffic over fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1866–1875, Apr. 2009.
- [19] B. L. Mark and Y. Ephraim, "Explicit causal recursive estimators for continuous-time bivariate Markov chains," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2709–2718, May 2014.
- [20] R. S. Atilla Eryilmaz, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," *IEEE/ACM Trans. Netw.*, vol. 15, no. 6, pp. 1333–1344, Dec. 2007.
- [21] V. Schmidt, "Markov Chains and Monte-Carlo Simulation," Inst. Stochastics, Ulm Univ., Ulm, Germany, Jul. 2010.
- [22] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of Queueing Theory*, 4th ed. Hoboken, NJ, USA: Wiley, 2008.
- [23] S. M. Ross, *Introduction to Probability Models*, 9th ed. San Diego, CA, USA: Academic, 2007.
- [24] J. S. Harsini and F. Lahouti, "Adaptive transmission policy design for delay-sensitive and bursty packet traffic over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 776–786, Feb. 2009.
- [25] C.-D. Iskander and P. T. Mathiopoulos, "Fast simulation of diversity nakagami fading channels using finite-state Markov models," *IEEE Trans. Broadcast.*, vol. 49, no. 3, pp. 269–277, Sep. 2003.
- [26] F. Meshkati, A. J. Goldsmith, H. V. Poor, and S. C. Schwartz, "A game-theoretic approach to energy-efficient modulation in CDMA networks with delay QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, pp. 1069–1078, Aug. 2007.
- [27] E. M. T. Hendrix and B. G. Toth, *Introduction to Nonlinear and Global Optimization*. New York, NY, USA: Springer-Verlag, 2010.



Kunlun Wang received the B.S. degree from Hangzhou Dianzi University, Hangzhou, China, in 2009, and the M.S. degree from South China University of Technology, Guangzhou, China, in 2012. He is currently pursuing the Ph.D. degree at Network Coding and Transmission Laboratory, Shanghai Jiao Tong University, Shanghai, China. His current research interests include energy efficient communications, cross-layer design, and MIMO systems.



Meixia Tao (S'00–M'04–SM'10) received the B.S. degree in electronic engineering from Fudan University, Shanghai, China, in 1999, and the Ph.D. degree in electrical and electronic engineering from Hong Kong University of Science and Technology in 2003. She is currently a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. Prior to that, she was a Member of Professional Staff at Hong Kong Applied Science and Technology Research Institute during 2003–2004, and a Teaching Fellow then an Assistant Professor at the Department of Electrical and Computer Engineering, National University of Singapore from 2004 to 2007. Her current research interests include cooperative communications, wireless resource allocation, MIMO techniques, and physical layer security.



Wen Chen (M'03–SM'11) received the B.S. and M.S. degrees from Wuhan University, China, in 1990 and 1993, respectively, and the Ph.D. degree from the University of Electro-Communications, Tokyo, Japan, in 1999. He was a researcher of the Japan Society for the Promotion of Sciences (JSPS) from 1999 through 2001. In 2001, he joined the University of Alberta, Canada, starting as a Postdoctoral Fellow with the Information Research Laboratory and continuing as a Research Associate in the Department of Electrical and Computer Engineering. Since 2006, he has been a Full Professor in the Department of Electronic Engineering, Shanghai Jiao Tong University, China. His interests cover network coding, cooperative communications, cognitive radio, and MIMO-OFDM systems.



Quansheng Guan (S'09–M'11) received the B.Eng. degree in electronic engineering from Nanjing University of Post and Telecommunications, China in 2006, and the Ph.D. degree from South China University of Technology (SCUT) in 2011. From 2009 to 2010, he was a visiting Ph.D. student with the University of British Columbia, Canada. From 2012 to 2013, he was a Postdoc Researcher at the Chinese University of Hong Kong. In 2013, he was a Visiting Scholar at Singapore University of Technology and Design. He is currently an Associate Professor with the School of Electronic and Information Engineering, SCUT. His research interests include wireless communications and networking, and networked interactions and economics.