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

Kunlun Wang, and Wen Chen

Department of Electronic Engineering, Shanghai Jiao Tong University, China

Emails: {kunlun1228; wenchen}@sjtu.edu.cn

Abstract—In this paper, we propose an energy efficient 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) is used at physical layer. We first investigate the stationary distribution of this system. With the stationary distribution and the AMC transmission mode, we derive the close-form expression of the delay and the energy efficiency. Then, we derive the energy efficient thresholds to choose the transmission mode. At last, we get the energy efficient transmission policy with delay constraint for given traffic. Numerical results are provided to support the theoretical development.

I. INTRODUCTION

For the rapid growth of mobile data traffic and battery powered mobile terminals, energy efficient communications are becoming increasingly important. While a large number of approaches have been proposed to reduce the energy consumption from the physical layer, very few studies focus on the cross-layer optimization.

Adaptive modulation and coding (AMC) can enhance throughput to time-varying channel conditions [1], [2], which has been widely applied in current wireless communication systems and has been incorporated in several wireless standards, i.e., IEEE 802.11a and IEEE 802.16e. 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 focus on cross-layer designs combining AMC schemes with automatic repeat request (ARQ) [3]–[5]. However, for the delay-aware traffic from upper-layer, we should consider the traffic state. In [2], [6]–[8], while the traffic and channel is known, the policy of choosing modulation constellation dynamically depending on incoming traffic state and buffer state in addition to channel state is studied. However, their works focus on the throughput and do not consider the energy efficiency. In [9], [10], the authors proposed a unified reinforcement learning solution for finding the joint optimal AMC and dynamic power management policies when the traffic arrivals and channel statistics are unknown. The performance of energy efficient transmission power has not been studied in their works either.

Taking a broader view, our work follows the energy efficient cross-layer design approach, which aims to take 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. The delay-aware traffic statistics and the queue condition are the parameters of higher layers. Our contribution can be summarized as follows:

- We obtain, via finite state Markov chain, the closed-form expression of delay of the system, which considers the queuing delay and the transmission delay.
- 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 over-flow. The average power consumption is also obtained.
- We derive the energy efficient thresholds to partition the SNR, which will improve the energy efficiency in comparison with other existing partitions from [1], [8]. And we present performance results to support the theoretical development.
- 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.

II. SYSTEM MODEL

A. Channel Model

Consider a point-to-point frame-by-frame communication system, with each frame composed of a number of packets. The channel is frequency-flat and block fading and is also corrupted with additive white Gaussian noise (AWGN) n with zero mean and variance σ^2.

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):

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

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 fading parameter ($m \geq 1/2$). We choose the Nakagami-m channel model because it applies to a large class of fading channels.

we consider adaptive modulation and coding 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 (1), the probability that the channel is in state n is given by

$$P_r(n) = \int_{\gamma_n}^{\gamma_{n+1}} f_{\bar{\gamma}}(\gamma) d\gamma. \qquad (2)$$

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

B. Traffic and Queuing Model

We assume that the user's incoming traffic is 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 a MMPP model, which is accurate and reasonable [11], and the poisson arrival is the special case of this model.

Let $\mathbb{F} = \{f_1, f_2, \cdots, f_K\}$ denote the set of states of the incoming traffic and P_{f_i, f_j} denote the probability of transition from state f_i to state f_j. Each state follows Poisson distribution with average arrival rate λ_i, $i = 1, 2, \cdots, K$. Denote the row vector $\pi^f = [\pi_1^f, \pi_2^f, \cdots, \pi_K^f]$ 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. 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 $\bar{\lambda} = \sum_{i=1}^K \pi_i^f \lambda_i$.

The user's queuing model is a single server M/G/1 queue [12], and the buffer is finite with size M. The M/G/1 model assumes Markovian or memoryless arrivals at average rate $\bar{\lambda}$, a general service distribution and a single server. Therefore, $\bar{\lambda}$ is the mean packet generation rate from the traffic $\mathbb{F}$, and c_n is the service rate at the physical layer corresponding to channel state n. The service rate c_n is in the set of $\mathbb{C} = \{c_1, \cdots, c_N\}$.

III. CROSS-LAYER QUEUING AND DELAY ANALYSIS

A. Queuing Analysis

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

can characterize the average packet successful transmission probability $\bar{P}_{s_n}$ corresponding to channel state n as

$$\bar{P}_{s_n} = 1 - \sum_{i=l}^{L} \binom{L}{i} (\bar{P}_b^n)^i (1 - \bar{P}_b^n)^{L-i}, \qquad (3)$$

where $\bar{P}_b^n$ is the average uncoded bit error rate (BER) for channel state n, which is shown in (27).

In the M/G/1 queue model, the packet service time S_{T_n} in state n has the following probability mass function:

$$P\{S_{T_n} = k\tau_n\} = \bar{P}_{s_n}(1 - \bar{P}_{s_n})^{k-1}, k = 1, 2, \cdots, N_r^{max} \qquad (4)$$

where τ_n represents the packet transmission time when the channel is in state n,

$$\tau_n(b_n) = \frac{L}{T_u b_n R_s}, \qquad (5)$$

where T_u and R_s are the time unit and the symbol rate respectively. From (4), we can get the mean service time at channel state n:

$$\begin{aligned} \mathsf{E}\{S_{T_n}\} &= \sum_{k=1}^{N_r^{max}} k\tau_n \bar{P}_{s_n}(1 - \bar{P}_{s_n})^{k-1} \\ &= \frac{\tau_n}{\bar{P}_{s_n}} \left[1 - (1 + N_r^{max} \bar{P}_{s_n})(1 - \bar{P}_{s_n})^{N_r^{max}}\right]. \end{aligned} \qquad (6)$$

From (5) and (6), the service rate c_n at state n is given by:

$$c_n = \frac{T_u b_n R_s \bar{P}_{s_n}}{L} \left[1 - (1 + N_r^{max} \bar{P}_{s_n})(1 - \bar{P}_{s_n})^{N_r^{max}}\right]^{-1} \qquad (7)$$

2) Stationary Distribution: Let index t denote the time unit 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

$$P(A_t = a | f_i) = \begin{cases} \frac{\lambda_i^a e^{-\lambda_i}}{a!}, \forall f_i \in \mathbb{F}, & \text{if } 0 \leq a \leq A \\ 0, & \text{otherwise}. \end{cases} \qquad (8)$$

Let S_t be the queue state at the start of the t-th time slot, and $S_t \in \mathbb{S} = \{s_0 = 0, s_1 = 1, \cdots, s_M = M\}$. Let $C_t \in \mathbb{C}$ denote the number of packets removed from the queue at the start of each time slot. The resulting recursion of the queue state can be summarized as

$$S_t = \min\{M, \max\{0, S_{t-1} - C_t\} + A_t\}. \qquad (9)$$

Let (F_{t-1}, C_t, S_{t-1}) denote the states of the traffic and the service and the queue, 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 $(f_i, c_x, s_q) \in \mathbb{F} \times \mathbb{C} \times \mathbb{S}$, 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

$$\mathbf{P} = \left[\mathbf{R}_{(f_i, c_x),(f_j, c_y)}\right], 0 \leq i, j \leq K, 0 \leq x, y \leq N, \qquad (10)$$

where $\mathbf{R}_{(f_i, c_x),(f_j, c_y)}$ can be shown in (11). Thus, the Markov chain has the total states $l = K \times N \times (M+1)$.

The element of $\mathbf{R}_{(f_i, c_x),(f_j, c_y)}$ is shown in (12), where P_{f_i, f_j} and P_{c_x, c_y} represent the transition probabilities of the traffic states and channel states respectively. The second

$$\mathbf{R}_{(f_i,c_x),(f_j,c_y)} = \begin{bmatrix} P_{(f_i,c_x,s_0),(f_j,c_y,s_0)} & P_{(f_i,c_x,s_0),(f_j,c_y,s_1)} & \cdots & P_{(f_i,c_x,s_0),(f_j,c_y,s_M)} \\ P_{(f_i,c_x,s_1),(f_j,c_y,s_0)} & P_{(f_i,c_x,s_1),(f_j,c_y,s_1)} & \cdots & P_{(f_i,c_x,s_1),(f_j,c_y,s_M)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{(f_i,c_x,s_M),(f_j,c_y,s_0)} & P_{(f_i,c_x,s_M),(f_j,c_y,s_1)} & \cdots & P_{(f_i,c_x,s_M),(f_j,c_y,s_M)} \end{bmatrix}. \tag{11}$$

$$\begin{aligned} P_{(f_i,c_x,s_q),(f_j,c_y,s_l)} &= P(F_t = f_j, C_{t+1} = c_y, S_t = s_l | F_{t-1} = f_i, C_t = c_x, S_{t-1} = s_q) \\ &= P_{f_i,f_j} P_{c_x,c_y} P(S_t = s_l | F_{t-1} = f_i, S_{t-1} = s_q). \end{aligned} \tag{12}$$

$$P(S_t = s_l | F_{t-1} = f_i, S_{t-1} = s_q) = \begin{cases} P(A_t = s_l - \max\{0, s_q - c_x\} | f_i), & \text{if } 0 \le s_l < M, \\ 1 - \sum_{0 \le s_l < M} P(S_t = s_l | F_{t-1} = f_i, S_{t-1} = s_q), & \text{if } s_l = M. \end{cases} \tag{13}$$

equality in (12) follows from the fact that both the channel transition and the traffic transition are independent of others.

We assume slow fading so that transition happens only between adjacent states. The nonzero elements P_{c_x,c_y} is described in [1]. At the same time, the conditional probability of (12) can be derived as (13). Therefore, based on (12) and (13), we can get the transition probability matrix $\mathbf{P}$.

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

Lemma 1. *The stationary distribution $\boldsymbol{\pi}$ of the process $\{(F_t, S_t, C_t), t \ge 0\}$ exists, and $\boldsymbol{\pi}_t \to \boldsymbol{\pi}$ as $t \to \infty$.*

Proof: Based on the theorem of [13, Theorem 4.1], the Markov chain $\{(F_t, S_t, C_t), t \ge 0\}$ exists stationary distribution only when the Markov chain is irreducible and recurrent.

The MMPP traffic has nonzero transition probability for each transition from f_i to f_j, denoted as $f_i \to f_j$. And from (13), 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

$$(f_i, c_x, s_q) \to (f_j, c_x, s_q) \tag{14}$$

has nonzero probability.

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

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

Thus, the transition from

$$(f_j, c_x, s_q) \to (f_j, c_x, s_l) \tag{15}$$

has nonzero probability. 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,

$$(f_j, c_x, s_l) \to (f_j, c_y, s_l) \tag{16}$$

also has nonzero transition probability. Based on (14), (15) and (16), we know that the $\{(F_t, S_t, C_t), t \ge 0\}$ is irreducible.

On the other hand, based on the conclusion of [13], that all states in a finite irreducible Markov chain are recurrent. In all, the stationary distribution of the Markov process $\{(F_t, S_t, C_t), t \ge 0\}$ exists. ∎

Now, the stationary distribution is obtained by solving

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}, \quad \sum_{f \in \mathbb{F}, s \in \mathbb{S}, c \in \mathbb{C}} \pi_{(f,s,c)} = 1. \tag{17}$$

The solution $\boldsymbol{\pi}$ is the left eigenvector of $\mathbf{P}$ corresponding to the eigenvalue 1.

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 packet arrivals, 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 π, we can compute P_d as [2]

$$P_d \triangleq \lim_{T \to \infty} \frac{\sum_{t=1}^{T} D_t}{\sum_{t=1}^{T} A_t} = \frac{\mathsf{E}\{D\}}{\mathsf{E}\{A_t\}} = \frac{\mathsf{E}\{D\}}{\bar{\lambda}}. \tag{18}$$

The average number of dropped packets $\mathsf{E}\{D\}$ can be found in (19). where $\theta(x, y)$ is a positive difference function, which returns the difference of x and y when $x > y$, and returns 0 when $x \le y$. With P_d available, we can get the effective average traffic rate and the system throughput.

B. Delay Analysis

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

$$c_n(S = s, C = c) = \begin{cases} c_n, & \text{if } c = c_n, s \ge c \\ \frac{s}{c} c_n, & \text{if } c = c_n, s < c. \end{cases} \tag{20}$$

Thus, the average service rate calculation corresponding to state n can be derived as

$$\bar{c}_n = \sum_{s \in \mathbb{S}, c = c_n} \frac{c_n(S = s, C = c)\pi_{(s,c)}}{\sum_{s \in \mathbb{S}, c = c_n} \pi_{(s,c)}}, \tag{21}$$

$$\mathsf{E}\{D\} = \sum_{f\in\mathbb{F},s\in\mathbb{S},c\in\mathbb{C}} \left[\theta(A_t, M-(S_{t-1}-C_t))\times P(A_t=a|f)\times\pi_{f,s,c}\right], \qquad (19)$$

where $\pi_{(s,c)}$ denotes the stationary distribution of the system state. Therefore, considering the influence of queuing, the average spectral efficiency is $\bar{b}_n = \frac{\bar{c}_n}{c_n}b_n$.

The effective traffic rate into the queue can be evaluated as $r = \bar{\lambda}(1-P_d)$. By [12], using the Pollaczek-Khintchine formula, we can get the mean queue length for state n as

$$\bar{Q}_q^n = \frac{r\mathsf{E}\{S_{T_n}^2\}}{2(1-\delta_n)}, \qquad (22)$$

where $\delta_n = r/\bar{c}_n$ is the traffic intensity or utilization, and $\mathsf{E}\{S_{T_n}^2\}$ is the second moment of the service distribution. Using (4), we can get $\mathsf{E}\{S_{T_n}^2\}$. For notational brevity, $\mathsf{E}\{S_{T_n}^2\} \triangleq f(\bar{P}_{s_n}, N_r^{max})$.

Theorem 1. *A necessary condition about 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: We omit the proof due to limited space. ∎

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 [12]. In summary, the average delay $\bar{W}_n$ for a packet corresponding to state n is given by (23), where $\tau_n(\bar{b}_n)$ is given by substituting $\bar{b}_n$ into (5). Thus, the average delay for a packet with AMC can be derived as

$$\bar{W} = \sum_{n=1}^N P_r(n)\bar{W}_n. \qquad (24)$$

Substituting (23) into (24), we can get the average delay $\bar{W}$.

IV. ENERGY EFFICIENCY ANALYSIS

In this section, we determine the energy efficient transmission policy with the joint effects of finite length queue and AMC also with the delay-aware arrival traffic.

A. Energy Efficiency

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 $\gamma = \frac{\bar{e}|h(t)|^2}{\sigma^2}$, where $|h(t)|^2$ denotes the instantaneous channel power gain. The BER for transmission mode n can be expressed as a function of the received SNR $\gamma e_{n,t}(\gamma)/\bar{e}$ as [14]

$$P_b^n \approx 0.2\exp(-\frac{1.5}{2^{b_n}-1}\frac{e_{n,t}(\gamma)}{\bar{e}}\gamma), \quad \gamma_n \leq \gamma < \gamma_{n+1}. \quad (25)$$

By considering the traffic and queueing influence, we should use the average transmission rate $\bar{b}_n$ to replace b_n. From (7),

(20) and (21), we can get $\bar{b}_n = \frac{\bar{c}_n}{c_n}b_n$. Thus, we get the transmission power for each AMC mode as follows:

$$e_{n,t}(\gamma) = \frac{\bar{e}(2^{\bar{b}_n}-1)}{1.5\gamma}\ln\frac{0.2}{P_b^n}. \qquad (26)$$

Let $\bar{P}_b^n$ denote the average BER corresponding to state n, from (2), we can derive $\bar{P}_b^n$ as

$$\bar{P}_b^n = \frac{1}{P_r(n)}\int_{\gamma_n}^{\gamma_{n+1}} 0.2\exp(-\frac{1.5}{2^{b_n}-1}\gamma)f_{\bar{\gamma}}(\gamma)d\gamma. \qquad (27)$$

Thus, the average transmission power in channel state n is

$$\bar{e}_{n,t} = \int_{\gamma_n}^{\gamma_{n+1}} \frac{\bar{e}(2^{\bar{b}_n}-1)}{1.5\gamma}\ln\frac{0.2}{\bar{P}_b^n}f_{\bar{\gamma}}(\gamma)d\gamma. \qquad (28)$$

From (2) and (28), we can approximate the actual average transmission power with AMC at time slot t as

$$\tilde{e}_t = \sum_{n=1}^N \bar{e}_{n,t}P_r(n). \qquad (29)$$

The system throughput is the average rate at which packets are successfully transmitted. Therefore, the packet dropping rate from queuing and packet violation from the channel with N_r^{max} 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

$$\bar{T} = \bar{\lambda}(1-P_d)(1-(1-\bar{P}_s)^{N_r^{max}}), \qquad (30)$$

where P_d is corresponding to (18). The average probability of successful packet transmission $\bar{P}_s$ can be calculated as the ratio of the average number of packets successfully transmitted over the total average number of transmitted packets

$$\bar{P}_s = \frac{\sum_{n=1}^N \bar{c}_n P_r(n)\bar{P}_{s_n}}{\sum_{n=1}^N \bar{c}_n P_r(n)}. \qquad (31)$$

Based on (29), (30) and (31), the energy efficiency is

$$f_{ee} \triangleq \frac{\bar{T}}{\tilde{e}_t} = \frac{\bar{\lambda}(1-P_d)(1-(1-\bar{P}_s)^{N_r^{max}})}{\sum_{n=1}^N \bar{e}_{n,t}P_r(n)}. \qquad (32)$$

It can be noted from (2) that different thresholds γ_n determine the probability distribution of different transmission rate b_n over $\mathbb{F}\times\mathbb{S}\times\mathbb{C}$ with $\bar{\gamma}$ available, which can be also called the transmission control policy $\mu(\gamma_n, \bar{\gamma})$.

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

Proof: We haven't shown the proof due to limited space. ∎

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

$$\bar{W}_n = \frac{T_u \bar{Q}_q^n}{r} + T_u \mathsf{E}\{S_{T_n}\} = \frac{T_u f(\bar{P}_{s_n}, N_r^{max})}{2(1-\delta_n)} + \frac{\tau_n(\bar{b}_n)T_u}{\bar{P}_{s_n}}\left[1 - (1 + N_r^{max}\bar{P}_{s_n})(1 - \bar{P}_{s_n})^{N_r^{max}}\right]. \qquad (23)$$

B. Energy efficient Thresholds

The choices of thresholds $\{\gamma_1, \gamma_2, \cdots, \gamma_N\}$ can be arbitrary. In [8], the equal probability method (EPM) was proposed. The partition based on minimum SNR required to acheive P_{target} (MSRE) was proposed in [1]. Since we want to get the energy efficient transmission, we set the threshold γ_n for the channel state n to be the energy efficient SNR. We consider no retransmission, f_{ee} corresponding to channel state n can be written as

$$f_{ee}(\gamma) = \frac{\bar{\lambda}(1-P_d)|h(t)|^2}{\sigma^2}\frac{f(\gamma)}{\gamma}, \qquad (33)$$

where $f(\gamma) = P_{s_n}$.

Taking derivative of (33) with respect to γ and equating it to zero, it can be shown that the energy efficient partition (EEP) γ_n^* satisfies $f(\gamma_n^*) = \gamma_n^* f'(\gamma_n^*)$. It is shown in [15] that for an S-shaped (sigmoidal) function, $f(\gamma_n^*) = \gamma_n^* f'(\gamma_n^*)$ has a unique solution, and $f(\gamma)$ is S-shaped.

Lemma 2. *The energy efficient thresholds* γ_n^* *($n = 1, \cdots, N$) is the unique solution of the following equation:* $\frac{\alpha_n L}{b_n}\sqrt{\frac{\beta_n \gamma}{2\pi}}e^{-\frac{\beta_n \gamma}{2}} + \alpha_n Q(\sqrt{\beta_n \gamma}) = 1$, *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 variable.*

Proof: The proof is Based on [15, Eq. (4),(5),(6)], ∎

C. Energy Efficient Policy with Delay Constraint

Based on the energy efficient thresholds γ_n^*, we can formulate the energy efficient optimization problem as

$$\begin{aligned} \max \quad & f_{ee}, \\ s.t. \quad & \bar{W} = \phi(\bar{\gamma}) \leqslant W_0, \\ & \bar{\gamma} \leqslant \bar{\gamma}_{max}, \\ & \bar{\gamma} \geqslant \bar{\gamma}_{min}, \end{aligned} \qquad (34)$$

where $\bar{\gamma}_{min}$ and $\bar{\gamma}_{max}$ are the minimum required SNR from theorem 1 and the maximum average received SNR, and $\bar{W}$ is corresponding to (24). Our objective is to determine the optimal prescribed average received SNR $\bar{\gamma}^{opt}(W_0)$ at the physical layer that maximizes the f_{ee}, which corresponds to the energy efficient transmission policy based on the queue state and the traffic state as well as the delay demand.

The nonlinear function of f_{ee} is complex, (34) can be numerically solved by Golden-Section method [16]. Then, we can get the energy efficient solution $\bar{\gamma}^{opt}(W_0) = \arg\max f_{ee}$ with delay constraint W_0. Hence, for given traffic $\mathbb{F}$, the energy efficient transmission policy with delay constraint can be determined as

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

The energy efficient average transmission power $\tilde{e}^{opt}$ can also be determined based on (29).
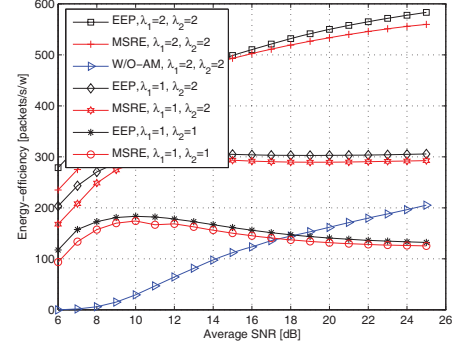

Fig. 1.   Energy efficiency versus average SNR with MSRE and EEP.

V. NUMERICAL RESULTS AND DISCUSSION

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 $L = 1080$, the maximum retransmission times of packet is $N_r^{max} = 6$. Maximum number of packet arrivals $A = 15$, average arrival rate, $\lambda_1 = 1$ packets/time-unit and $\lambda_2 = 2$ packets/time-unit. The symbol rate $R_s = 100$KHz; The Nakagami parameter $m = 1$; Doppler frequency $f_d = 10$ Hz. We assume that block (also called frame) length $T_u = 2$ ms.

B. Performance of the analysis

Fig. 1 shows the average energy efficiency, we can see that the method of EEP can offer better energy efficiency than that of MSRE, although the gap of the energy efficiency is very small at some SNRs. The energy efficiency is increasing when increasing the average arrival rate. 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. The blue curve shows the energy efficiency without adaptive modulation (W/O-AM) with average traffic arrival $\bar{\lambda} = 2$ packets/time-unit, which performs much worse than 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. 2 that the delay is decreasing when increasing the SNR. Thus, the SNR should be increased when the delay demand is more strict. Therefore, there exists an optimum energy efficient transmission policy under different delay constraints. Based on Fig. 1 and Fig. 2,
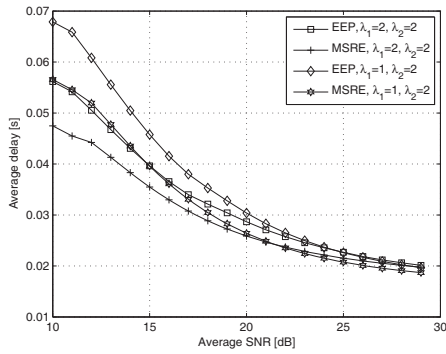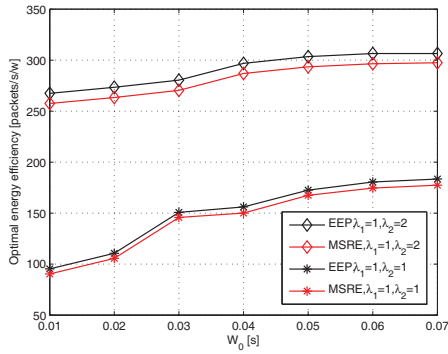
Fig. 2.   Average delay versus average SNR.

Fig. 3.   Optimal energy efficiency versus the delay constraint W_0.

the energy efficient SNR for the small average arrival rate is decreasing when the delay demand is increasing, i.e., $\bar{\lambda} = 1$ packets/time-unit. Consequently, when the delay is more tolerant, the energy efficient transmission policy for the small average arrival traffic is that the probability of choosing small transmission rate is larger. On the other hand, when the delay is more sensitive, the energy efficient transmission policy for the small average arrival traffic is that the probability of choosing large transmission rate is larger. However, the energy efficient transmission policy for the large average arrival traffic is the same no matter what the delay demand is.

Fig. 3 shows the optimal energy efficiency versus the delay W_0 for different arrival rates, 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, the optimal energy efficient transmission policy at large delay region is almost the same irrespective of delay and arrival rate variations.

VI. CONCLUSION

In this paper, we present a cross-layer framework that determines the energy efficient transmission policy based on both the physical layer and the upper-layer information. We derive the closed form expression of delay considering joint

effects of the general traffic arrival model and queuing states 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 performs better than the existing partition methods with the energy efficient partition. At last, we get the energy efficient transmission policy for different delay-aware services.

REFERENCES

[1] 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.

[2] Q. Liu, S. Zhou, and G. B. Giannakis, "Queuing 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.

[3] 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.

[4] 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.

[5] 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.

[6] X. Zhang and J. Tang, "Power-delay tradeoff over wireless networks," *IEEE Trans. Commun.*, vol. 61, no.9, pp. 3673–3684, Sep. 2013.

[7] 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.

[8] A. K. Karmokar 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.

[9] N. Mastronarde and V. D. Schaar, "Joint physical-layer and sytem-level power management for delay-sensitive wireless communications," *IEEE Trans. Mobile Computing*, vol. 12, no.4, pp. 694–709, Apr. 2013.

[10] N. Mastronarde and M. V. D. Schaar, "Fast reinforcement learning for energy-efficient wireless communication," *IEEE Trans. Sig. Proc.*, vol. 59, no.12, pp. 6262–6266, Dec. 2011.

[11] B. L. Mark and Y. Ephraim, "Explicit causal recursive estimators for continuous-time bivariate markov chains," *IEEE Trans. Sig. Proc.*, vol. 62, no.10, pp. 2709–2718, May 2014.

[12] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of Queueing Theory*. New Jersey, USA: John Wiley & Sons, 4th ed., 2008.

[13] S. M. Ross, *Introduction to Probability Models*. San Diego, CA: Academmic Press, ninth edition ed., 2007.

[14] 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.

[15] 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.

[16] E. M. T. Hendrix and B. G.-Toth, *Introduction to Nonlinear and Global Optimization*. New York, USA: Springer Optimization and Its Applications, 2010.