This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2011 proceedings

Queuing on Energy-Efficient Wireless Transmissions with Adaptive Modulation and Coding

Jie Gong, Sheng Zhou, Zhisheng Niu
Tsinghua National Laboratory for Information Science and Technology
Department of Electronic Engineering
Tsinghua University, Beijing, 100084, China
{gongj04, zhouc}@mails.tsinghua.edu.cn, niuzhs@tsinghua.edu.cn

***Abstract*—Adaptive modulation and coding (AMC) has been widely used to improve the spectral efficiency. In this paper, we take a different look at it from energy saving point of view. Specifically, we analyze the queuing behavior of AMC systems jointly with sleep mode where the wake-up process incurs time and energy cost. We formulate the optimization problem by jointly considering energy-efficiency, queuing delay and packet loss rate, and find the solution with cross-layer adjustment of the transmit power and the sleep threshold. Numerical results show that at low traffic range, when the power consumption of idle (no data transmission) mode is un-negligible, introducing sleep mode to the AMC system significantly improves the energy efficiency compared with non-sleep system. To achieve the energy-efficiency gain, the system tends to use higher-order modulation by increasing transmit power, which also reduces the number of dropped packets.**

I. Introduction

Nowadays, reduction of energy consumption has become a key problem from both environmental and economic perspective. The rapid development of information and communication technology (ICT) not only greatly improves the capacity of wireless networks, but also leads to a large amount of energy consumption. However, because of traffic variation and channel fading, network capacity is usually not fully utilized so that the energy consumption required to support the peak capacity is wasted. Therefore, algorithms to improve energy efficiency with respect to certain Quality of Service (QoS) are urgently required.

Adaptive modulation and coding (AMC) has been widely used to enhance the spectral efficiency by matching transmission parameters to time-varying channel conditions [1] [2]. On the other hand, AMC also provides opportunity for energy-efficiency enhancement by increasing transmit power to support higher data rate, which improves the time ratio of putting devices into sleep mode. Ref. [3] studies the modulation optimization problem by taking both transmit energy and processing circuit energy into account. Although random traffic and finite buffer is not included since it considers the wireless sensor networks with periodical data transmission, the result gives an insight on the tradeoff between the energy consumption for higher-order modulation and the energy saving from sleep. Consequently, jointly considering AMC and sleep schemes can lead to energy-efficient network protocol design.

Although introducing sleep mode can dramatically reduce the energy consumption [4] [5], it is facing various challenges such as the additional time and energy cost of switching between sleep mode and active mode, queuing delay and buffer overflow. Therefore, related work in literature has focused on these problems to optimize the sleep policy. Queuing analysis of sleep policies with switching delay and energy cost is presented in [6] and [7] assuming infinite buffer length. Very recently, Ref. [8] formulates the sleep problem as a Markov decision process (MDP) problem considering the switching energy cost and the queuing delay penalty, and finds that the optimal sleep policy has a simple double-threshold structure. Most of the existing sleeping schemes assume that the server consumes the same power in active mode, which corresponds to a fixed transmission power and modulation and coding scheme. However, since the wireless channel is time-varying and sometimes suffers from deep fading, fixed transmission scheme is neither spectral efficient nor energy efficient.

In this paper, we study the queuing behavior of AMC systems jointly with sleep mode under the condition of random traffic and finite buffer length. The optimization problem is formulated to minimize the cost that combines energy consumption and queuing delay, under the transmit power and the packet loss rate constraints. We analyze the energy consumption per packet, the average delay and the packet loss rate, and propose a cross-layer optimization algorithm by adjusting the transmit power and the sleep threshold. It is shown from the numerical results that when the power consumption in idle frame with no data transmission is comparable with that of active frame, higher-order modulation combined with sleeping schemes substantially improves the energy efficiency, and noticeable energy-efficiency gain of AMC with sleep is obtained compared with non-sleep schemes.

The rest of this paper is organized as follows. We introduce the system mode in Section II. In Section III, queuing analysis, performance metrics calculation and cross-layer design are presented. Numerical results are provided in Section IV. Finally, Section V concludes the paper.

II. System Model

We consider a wireless communication link with a single-transmit and a single-receive antenna. Assume the transmitter has a finite buffer of size K and works in two modes, active and sleep. It goes to sleep as long as the buffer is empty and starts the wake-up process when the queue length exceeds a

978-1-61284-231-8/11/$26.00 ©2011 IEEE

threshold L. Assume the wake-up process is both time and energy consuming. In active mode, we deal with frame by frame transmissions at the physical layer, where each frame has a fixed duration (T_f seconds). The buffered packets are merged into each frame through modulation and coding. As a consequence, the number of packets per frame depends on the selected modulation and coding scheme. At the receiver side, the modulation-coding pairs are determined by the AMC selector according to the estimated channel state information (CSI) and sent back to the transmitter. When the transceiver is active, AMC is performed to maximize the data rate by adjusting transmission parameters to the available CSI, while maintaining a target packet error (caused by packet reception error) rate P_e. Let N denote the total number of transmission modes available. The entire SNR range is divided into N+1 nonoverlapping consecutive intervals, with boundary points denoted by $\{\gamma_n\}_{n=0}^{N+1}$. Specifically, mode n is chosen when $\gamma \in [\gamma_n, \gamma_{n+1})$. Data rate for the mode $n = 0$ is $R_0 = 0$ (bits/symbol) to avoid deep channel fades. The packet error rate (PER) is approximated as [2]

$$\mathrm{PER}_n(\gamma) \approx \begin{cases} 1, & \text{if } 0 < \gamma < \gamma_{pn}, \\ a_n \exp(-g_n \gamma), & \text{if } \gamma \geq \gamma_{pn}, \end{cases} \quad (1)$$

where n is the mode index and γ is the received signal-to-noise ratio (SNR), which is assumed Gamma distributed [9] with mean $\bar{\gamma} = P_t G \beta d^{-\kappa}/N_0$, where P_t is the transmit power, G is the antenna power gain, d is the transmit distance, β is the pathloss constant, κ is the pathloss exponent, and N_0 is the noise power. The transmission modes are adopted from the HIPERLAN/2 or the IEEE 802.11a standards [10], and the parameters a_n, g_n and γ_{pn} are obtained by fitting with packet length $N_b = 1080$. The average PER of AMC can be calculated as [2]

$$\overline{\mathrm{PER}} = \frac{\sum_{n=1}^{N} R_n \Pr(n) \overline{\mathrm{PER}}_n}{\sum_{n=1}^{N} R_n \Pr(n)}, \quad (2)$$

where $\Pr(n)$ denotes the probability with which mode n is chosen and $\overline{\mathrm{PER}}_n$ denotes the average PER of mode n. The closed-form of $\Pr(n)$ and $\overline{\mathrm{PER}}_n$ can be found in [2]. Then the boundary points $\{\gamma_n\}_{n=0}^{N+1}$ are selected so that $\overline{\mathrm{PER}}_n = P_e$ for each mode n, which naturally leads to $\overline{\mathrm{PER}} = P_e$. The scheme to determine the boundaries are detailed in [2].

The channel is assumed frequency flat and suffers block fading, i.e., it remains constant in each frame, but is variable from frame to frame. Therefore, AMC is adjusted frame by frame. A finite state Markov chain (FSMC) channel model is adopted as in [11]. The SNR region $[\gamma_n, \gamma_{n+1})$ corresponding to transmission mode n constitutes the channel state indexed by n. Denote $\mathbf{P}_c$ as the $(N+1) \times (N+1)$ state transition matrix. We assume slow fading so that transition happens only between adjacent states, i.e., $P_{i,j} = 0, |i-j| \geq 2$. The expression of the non-zero elements $P_{i,j}$ is described in [11].

The power consumption is modeled as follows. The system in sleep mode consumes P_s to maintain the basic operation such as backhaul connection and queue data storage. When the server is in wake-up process, the power consumption is P_w. In active mode, the power consumption is a function of modulation and coding mode n and transmit power P_t, denoted as $P_n(P_t)$. Define $\Pr(s)$ as the time ratio of sleep mode, $\Pr(w)$ as the time ratio of wake-up process, $\Pr(n)$ as the time ratio of modulation and coding mode n in active mode, and $\bar{N}_p$ as the average number of packets transmitted in each frame. Then the average energy consumption per packet

$$E_p = \frac{T_f}{\bar{N}_p}(\Pr(s)P_s + \Pr(w)P_w + \sum_n \Pr(n)P_n(P_t)). \quad (3)$$

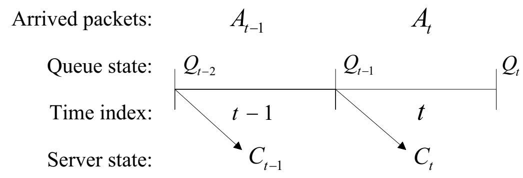

Fig. 1. Queuing model.

The objective of AMC strategy with sleep is to minimize the energy consumption plus weighted average delay of each packet while guaranteeing a target packet loss (including both packet reception error and buffer overflow) rate. The optimization problem can be modeled as

$$\begin{aligned} \min \quad & E_p + \omega \bar{W}, \\ \text{s.t.} \quad & P_l \leq P_l^{\max}, \\ & P_t \leq P_t^{\max}, \end{aligned} \quad (4)$$

where $\bar{W}$ is the average queuing delay, ω is the weight factor of delay, P_l denotes the packet loss rate with constraint $P_l^{\max}$, and $P_t^{\max}$ is the transmit power budget. The parameter ω captures the tradeoff between energy and delay, which indicates the application's sensitivity to the delay. The problem can be optimized by adjusting the wake-up threshold L that trades off between the queuing delay and the sleep time ratio, and the transmit power P_t the influences both AMC boundaries and energy consumption. We derive the performance metrics in (4) according to the queuing analysis presented in the following.

III. Performance Analysis

A. Queuing Analysis

In this section, we model and analyze the queuing arrival process, the service process and the queue state transition. The system time diagram is shown in Fig. 1, where each time-unit t represents a frame duration. During t, A_t packets arrive in the queue. The queue state Q_t is described as the number of packets in the queue at the end of each time-unit, and the server state C_t is described as the maximum number of packets that can be transmitted per time-unit.

1) Arrival Process: The process A_t is stationary with mean $E\{A_t\} = \lambda T_f$, where λ denotes the number of packets arrived per second, and independent of the queue state as well as the service process. We assume that A_t is Poisson distributed

$$P(A_t = a) = \frac{(\lambda T_f)^a \exp(-\lambda T_f)}{a!}, \quad (5)$$

where $A_t \in \mathcal{A} := \{0, 1, \cdots, \infty\}$.

2) Service Process: When the server is active, the service process is dynamic with a variable number of packets transmitted per time unit due to AMC scheme. Corresponding to each transmission mode n, let c_n (packets/time-unit) denote the number of packets transmitted per time-unit. Then we have

$$C_t \in \mathcal{C}, \quad \mathcal{C} := \{c_0, c_1, \dots, c_N\} \tag{6}$$

where $c_n = bR_n$ takes non-negative integer values, and b depends on the system resource allocated per user. Since the AMC mode n is chosen when the channel enters the state n, and the server works in state c_n, the service process in active state is also modeled as a FSMC with state transition matrix the same as $\mathbf{P}_c$.

When the server goes to sleep, the service rate remains 0 until the wake-up process is finished. Note that although the energy consumption of sleep mode and wake-up process is different, the service rate keeps the same. Therefore, we denote $c_{-1} = 0$ as the server in sleep mode or wake-up process to reduce the number of server states. While the two states are distinguished by the queue length Q_t, i.e., if $Q_t < L$, the server is in sleep mode, otherwise it is in wake-up process. To sum up, the set of server states is

$$C_t \in \mathcal{C}'. \quad \mathcal{C}' := \mathcal{C} \bigcup \{c_{-1}\} = \{c_{-1}, c_0, \dots, c_N\} \tag{7}$$

Denote w as the wake-up time of the server. In the slotted system, w is assumed a positive multiple of the frame duration. In current work, w is initially assumed geometric distributed to make use of the memoryless property, i.e. $P_w(w = kT_f) = (1-p_w)p_w^{k-1}, k \geq 1$. Actually, any random distributions or fixed wake-up time depending on real system settings can be analyzed without much difficulty, which is left for future work.

3) Queue State Transition: Recall that Q_t denotes the queue state (the number of packets in the queue) at the end of time-unit t and $Q_t \in \mathcal{Q} := \{0, 1, 2, \dots, K\}$. We assume the packets in the queue are moved out first at the beginning of time-unit t based on the server state C_t. As a result, the number of free slots in the queue is $F_t = K - \max\{0, Q_{t-1} - C_t\}$. Then, arriving packets are placed in the queue throughout time-unit t. Because of the finite buffer, packets will be dropped when the queue is full, i.e., if $A_t > F_t$, $A_t - F_t$ packets are dropped. So the queue state transition can be summarized as follows:

$$Q_t = \min\{K, \max\{0, Q_{t-1} - C_t\} + A_t\}. \tag{8}$$

Since A_t is independent of Q_{t-1} and C_t, we need only analyze the transition of joint state pair (Q_{t-1}, C_t). Since the server goes to sleep as long as the queue is empty, the state pairs $(0, c_n)_{n\geq 0}$ do not exist. Hence the joint state pair set is $\mathcal{QC} = \mathcal{Q} \times \mathcal{C}' \setminus \{(0, c_n)\}_{n=0}^{N}$. Let $P_{(q,c)(p,d)}$ denote the transition probability from $(Q_{t-1} = q, C_t = c)$ to $(Q_t = p, C_{t+1} = d)$, where $(q,c) \in \mathcal{QC}$ and $(p,d) \in \mathcal{QC}$. Then the state transition probability matrix is organized in a block form

$$\mathbf{P} = [\mathbf{A}_{q,p}], 0 \leq p, q \leq K \tag{9}$$

where

$$\mathbf{A}_{q,p} = \begin{cases} P_{(0,c_{-1})(0,c_{-1})}, & \text{if } p = 0, q = 0 \\ \left[P_{(0,c_{-1})(p,c_{-1})}, \mathbf{0}_{1\times(N+1)}\right], & \text{if } p > 0, q = 0 \\ \left[P_{(q,c)(0,c_{-1})}\right]_{c\in\mathcal{C}'}, & \text{if } p = 0, q > 0 \\ \left[P_{(q,c)(p,d)}\right]_{c,d\in\mathcal{C}'}, & \text{else} \end{cases} \tag{10}$$

where $\left[P_{(q,c)(0,c_{-1})}\right]_{c\in\mathcal{C}'} = \left[P_{(q,c_{-1})(0,c_{-1})}, P_{(q,c_0)(0,c_{-1})}, \dots, P_{(q,c_N)(0,c_{-1})}\right]^T$ is an $(N+2)\times 1$ vector where $(\cdot)^T$ denotes the transpose of a vector, and $\left[P_{(q,c)(p,d)}\right]_{c,d\in\mathcal{C}'}$ is an $(N+2)\times(N+2)$ matrix. We next calculate the state transition probability as

$$P_{(q,c)(p,d)} = \begin{cases} P(Q_t = p|Q_{t-1} = q, C_t = 0), \\ \qquad\qquad \text{if } q < L, c = c_{-1}, d = c_{-1} \\ P(Q_t = p|Q_{t-1} = q, C_t = 0)P_w(w > T_f), \\ \qquad\qquad \text{if } q \geq L, c = c_{-1}, d = c_{-1} \\ P(Q_t = p|Q_{t-1} = q, C_t = 0)P_w(w = T_f)P(C_{t+1} = d), \\ \qquad\qquad \text{if } q \geq L, c = c_{-1}, d \in \mathcal{C} \\ P(Q_t = 0|Q_{t-1} = q, C_t = c), \\ \qquad\qquad \text{if } p = 0, q > 0, c \in \mathcal{C}, d = c_{-1} \\ P(C_{t+1} = d|C_t = c)P(Q_t = p|Q_{t-1} = q, C_t = c), \\ \qquad\qquad \text{if } p > 0, q > 0, c \in \mathcal{C}, d \in \mathcal{C} \\ 0, \qquad\qquad \text{else} \end{cases} \tag{11}$$

where the transition probability $P(C_{t+1} = d|C_t = c)$ can be found from the entries of $\mathbf{P}_c$, and the stationary distribution $P(C_{t+1} = d)$ is obtained by solving $\pi_c = \pi_c \mathbf{P}_c$ and $\pi_c \mathbf{e} = 1$, where $\mathbf{e}$ is an $(N+1)\times 1$ vector with all elements equaling to 1. Here we make the following approximation that when the server finishes wake-up process and enters active mode, the channel state is randomly generated according to the stationary distribution. Based on (8), it is easy to verify that

$$P(Q_t = p|Q_{t-1} = q, C_t = c) = \begin{cases} P(A_t = p - \max\{0, q - c\}), & 0 \leq p < K \\ 1 - \sum_{0\leq p<K} P(Q_t = p|Q_{t-1} = q, C_t = c). & p = K \end{cases} \tag{12}$$

It can be proved (similar to the Appendix of [2]) that the stationary distribution $\pi = [\pi_{(0,c_{-1})}, \pi_{(1,c_{-1})}, \dots, \pi_{(1,c_N)}, \dots, \pi_{(K,c_{-1})}, \dots, \pi_{(K,c_N)}]$ exists and is unique, where $\pi_{(q,c)} := \lim_{t\to\infty} P(Q_{t-1} = q, C_t = c)$. Based on the stationary distribution π, one can calculate the average number of transmitted packets per frame, the average delay and the packet loss rate for performance evaluation.

B. System Performance

1) Energy Consumption per Packet: Define $(Q, C) := \lim_{t\to\infty}(Q_{t-1}, C_t)$. Based on the analytical results of Section III-A3, we have $\Pr(s) = \sum_{q<L} \pi_{(q,c_{-1})}, \Pr(w) = \sum_{q\geq L} \pi_{(q,c_{-1})}$ and $\Pr(n) = \sum_{q>0} \pi_{(q,c_n)}, n = 0, \dots, N$.

The average number of packets transmitted in each frame is calculated as follows. Notice that when $Q_{t-1} < C_t$, no sufficient packets are available for transmission. Hence the number of transmitted packets in the frame of state $(Q = q, C = c)$ is $N_p = \min\{q, c\}$. Consequently,

$$\bar{N}_p = \sum_{q>0,c\in\mathcal{C}} \min\{q,c\}\pi_{(q,c)}. \quad (13)$$

Then the energy consumption per packet is calculated as (3).

2) Delay: According to Little's Law [12], we have

$$\bar{W} = \frac{\bar{Q}}{(1-P_d)\lambda T_f}, \quad (14)$$

where $\bar{Q} = \sum_{q\in\mathcal{Q},c\in\mathcal{C}'} q\pi_{(q,c)}$ is the average queue length and P_d is the packet dropping (overflow or blocking) rate which is detailed in the following subsection.

3) Packet Loss Rate: The number of packets dropped at time t can be expressed as $D_t = \max\{0, A_t - K + \max\{0, Q_{t-1} - C_t\}\}$. Since A_t and state pair (Q_{t-1}, C_t) are both stationary, the ensemble-average number of packets dropped per time-unit can then be calculated as

$$\bar{D} = \sum_{a\in\mathcal{A},q\in\mathcal{Q},c\in\mathcal{C}'} \max\{0, a-K+\max\{0,q-c\}\} \cdot P(A=a)\cdot\pi_{(q,c)}. \quad (15)$$

Based on (15), the packet dropping rate is given by

$$P_d = \lim_{T\to\infty} \frac{\sum_{t=1}^{T} D_t}{\sum_{t=1}^{T} A_t} = \frac{\bar{D}}{\lambda T_f}, \quad (16)$$

where the second equality is verified in the Appendix of [2]. Then the average packet loss rate P_l can be computed as

$$P_l = 1-(1-P_d)(1-P_e). \quad (17)$$

C. Cross-Layer Design

Given P_e, λ, and the channel condition parameters, the problem (4) can be optimized by adjusting P_t at the physical layer and L at the link layer. It can be verified that given L, the packet loss rate is a nonincreasing function of P_t and the cost function in (4) has a single optimal point. Hence a simple cross-layer approach is proposed to numerically find the optimal P_t and L and is detailed as follows.

Step 1 Set $\mathcal{L} = \{1, \ldots, K\}$.
Step 2 For $L = 1$ to K
  Find $P_t^{\min}(L)$ so that $P_l \le P_l^{\max}$ for all $P_t \ge P_t^{\min}(L)$.
  If $P_t^{\min}(L) > P_t^{\max}$, set $\mathcal{L} = \mathcal{L}\setminus\{L\}$,
  else, find the optimal $P_t(L)$ as

$$P_t(L) = \arg\min_{P_t^{\min}(L)\le P_t\le P_t^{\max}} E_p + \omega\bar{W}. \quad (18)$$

Step 3 If $\mathcal{L} = \emptyset$, set $(P_t^{\mathrm{opt}}, L^{\mathrm{opt}}) = (P_t^{\max}, 1)$,
  else, find the optimal (P_t, L) as

$$(P_t^{\mathrm{opt}}, L^{\mathrm{opt}}) = \arg\min_{\substack{(P_t(L),L):\\ L\in\mathcal{L}}} E_p + \omega\bar{W}. \quad (19)$$

To enhance the searching speed, $P_t^{\min}(L)$ can be found using binary search, and the optimization (18) can be done by the golden section method [13]. While (19) can be solved by exhaustive search since $\mathcal{L}$ is finite. In the next section, we present numerical results based on our analytical results and also adopt Monte-Carlo simulations to validate our analysis.

TABLE I
OPTIMAL VALUE OF SLEEP THRESHOLD

λT_f (packets/frame)	0.5	1	1.5	2	2.5	3
Optimal L	8	14	18	20	21	20
λT_f (packets/frame)	3.5	4	4.5	5	5.5	6
Optimal L	15	9	6	1	1	1

IV. NUMERICAL RESULTS

We test the performance in single cell downlink scenario. Assume the buffer length $K = 50$, the frame length $T_f = 2$ ms, the target packet error rate $P_e = 1\times10^{-3}$, $b = 2$ and the geometric distribution parameter $p_w = 36.8\%$ for all the test. The link parameters are set according to ITU micro-cell test environment [14]. The maximum transmit power is $P_t^{\max} = 41$ dBm. The antenna power gain is assumed $G = 9$ dB. The pathloss model is $PL^{\mathrm{dB}}(d^{km}) = 33.05 + 36.7\log_{10}(d^{km})$. We choose $d = 30$ m as the transmit distance. The noise power N_0 is calculated by setting the reference SNR at distance 100 m to be 5 dB with maximum transmit power. The target packet loss rate is $P_l^{\max} = 1\times10^{-2}$.

We adopt the typical BS structure from Ref. [15] which consists of a power supplier, a central controller, a cooling equipment and 12 transmit blocks. Assume the power supplier and the central controller is always powered on, and the cooling power consumption is proportional to the number of active transmit blocks. In wake-up process and active mode with rate $R_0 = 0$, the transmit block consumes the same power as idle control. When the server is active with non-zero data rate, additional power consumption comes from the converter, the duplexer, the combiner and the power amplifier (PA) $P_{\mathrm{amp}} = \xi/\eta P_t$, where η is the drain efficiency and ξ is the modulation-dependent peak-to-average ratio (PAR). Based on the assumptions and the power budget in [15], we obtain for each transmit block, $P_s \approx 75$ W, $P_w = P_0 \approx 165$ W, $P_n(P_t) \approx 215 + \xi/\eta P_t$ W, $n = 1, \ldots, N$, where $\xi = 3(\sqrt{M} - 1/\sqrt{M} + 1)$ for M-QAM modulation and $\eta = 0.35$. In the initial phase $\omega = 5\times10^{-3}$ J/frame, which means that delaying a packet an entire frame can save 5×10^{-3} J.

The variation of optimal L along with the arrival rate is listed in Table I. It is shown to be a concave function of the arrival rate. At low traffic range, L is set to be small to minimize the queuing delay. At high traffic range on the other hand, small L is selected mainly to help reduce packet dropping. While in between, the system selects large L so that the time ratio of sleeping is maximized. In other words, the energy consumption is minimized.

Fig. 2 depicts the energy consumption per packet versus packet arrival rate. AMC w/ sleep denotes the proposed AMC system with sleep mode. AMC w/o sleep is the traditional scheme [2] in which the transmit block is always active, and only the transmit power is optimized. It is shown that AMC w/ sleep achieves high energy efficiency at low traffic range

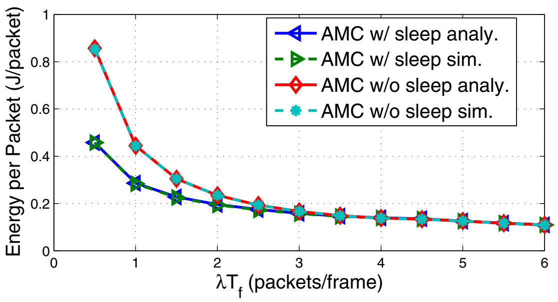

Fig. 2. Energy efficiency of AMC with sleep and without sleep.

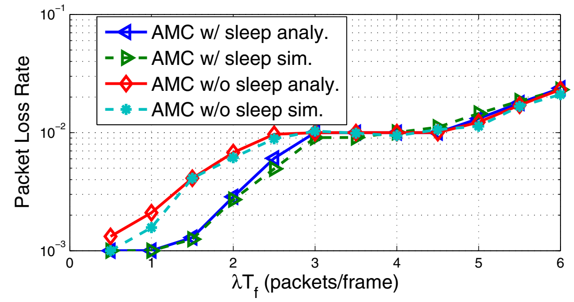

Fig. 4. Packet loss rate of AMC with sleep and without sleep.

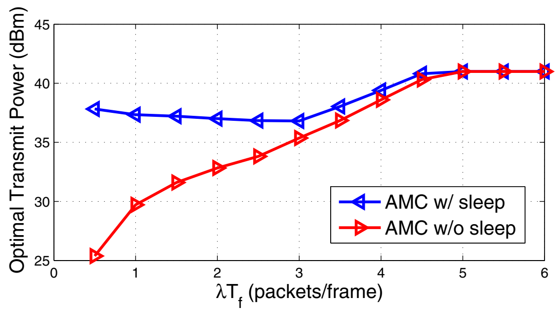

Fig. 3. Optimal transmit power of AMC with sleep and without sleep.

(46.7% gain achieved when $\lambda T_f = 0.5$ packets/frame). But the gain diminishes as the traffic becomes higher. The reason can be found in Fig. 3. At low traffic range, AMC w/ sleep, which uses higher transmit power than that without sleep (ex. 37 dBm for AMC w/ sleep and 32 dBm for AMC w/o sleep when $\lambda T_f = 2$ packets/frame), has more chance to transmit at high data rate. Consequently, the packets in the buffer are moved out quickly and the transmit block has more time to sleep, which greatly reduces the unnecessary server idle power consumption. At high traffic region, the transmit block can barely go to sleep mode even if we use the maximum transmit power, which results in the similar performance.

In Fig. 4, surprisingly, the packet loss rate of AMC with sleep is also lower than that of AMC without sleep when the traffic load is low. In fact, it is a natural outcome since the buffered packets are moved out very quickly with higher data rate, and hence fewer packets are dropped. But when the arrival rate is large ($\lambda T_f > 4.5$ packets/frame), even the maximum transmit power cannot support a required service rate, which causes heavy buffer overflow. And the packet loss rate performance is almost the same as the additional overflow brought by the short time sleep ($L = 1$) is negligible.

V. Conclusion

In this paper, we have proposed a queuing analysis of AMC systems jointly with sleep mode. Based on the queuing analysis, a cross-layer algorithm that minimize the energy consumption combined with queuing delay is proposed. According to the numerical results, in a typical BS system with sleep mode where idle mode consumes a lot of power, higher-order modulation is preferable. Also, energy-efficiency gain is much more remarkable at low traffic range. Future work includes the analysis of different wake-up settings. Extending the results to multi-cell scenario is also a valuable direction.

Acknowledgment

The work is sponsored by the National Basic Research Program of China (973 Program: 2007CB310607) and by the National Science Fund for Distinguished Young Scholars (60925002).

References

[1] M. S. Alouini and A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels," *J. Wireless Commun*, vol. 13, no. 1-2, pp. 119–143, May 2000.

[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] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sept. 2005.

[4] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy saving in cellular access networks," *Proc. IEEE ICC'09 workshop, GreenComm.*, Jun. 2009.

[5] S. Zhou, J. Gong, Z. Yang, Z. Niu, and P. Yang, "Green mobile access network with dynamic base station energy saving," *MobiCom'09 poster*, Sept. 2009.

[6] M. Yadin and P. Naor, "Queueing systems with a removable service station," *Operations Research*, vol. 14, no. 4, pp. 393–405, Dec. 1963.

[7] D. P. Heyman, "Optimal operating policies for M/G/1 queueing systems," *Operations Research*, vol. 16, no. 2, pp. 362–382, Mar. 1968.

[8] I. Kamitsos, L. Andrew, H. Kim, and M. Chiang, "Optimal sleep patterns for serving delay-tolerant jobs," *Proc. ACM e-Energy*, Apr. 2010.

[9] G. L. Stuber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA: Kluwer, 2001.

[10] A. Doufexi, S. Armour, M. Butler, A. Nix, D. Bull, J. McGeehan, and P. Karlsson, "A comparison of the HIPERLAN/2 and IEEE 802.11a wireless LAN standards," *IEEE Commun. Mag.*, vol. 40, no. 5, pp. 172–180, May 2002.

[11] J. Razavilar, K. J. R. Liu, and S. I. Marcus, "Jointly optimized bit-rate/delay control policy for wireless packet networks with fading channels," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 484–494, Mar. 2002.

[12] L. Kleinrock, *Queueing Systems*. John Wiley & Sons, 1976.

[13] S. Jacoby, J. Kowalik, and J. Pizzo, *Iterative methods for nonlinear optimization problems*. Prentice-Hall, 1972.

[14] Ericsson, "Radio characteristics of the ITU test environments and deployment scenarios." [Online]. Available: http://ftp.3gpp.org/tsg_ran/WG1_RL1/TSGR1_56b/Docs/R1-091320.zip

[15] H. Karl, "An overview of energy-efficiency techniques for mobile communication systems," Telecommunication networks Group, Technical University Berlin, Tech. Rep., Sept. 2003. [Online]. Available: http://www.tkn.tu-berlin.de/publications/papers/TechReport_03_017.pdf