Power-Delay Tradeoff Improvement with Adaptive Modulation Scheme under Practical Power Model

Yan Chen, Shunqing Zhang, and Shugong Xu
Huawei Technologies, Co. Ltd., Shanghai, China
Email: {bigbird.chenyan, zhangshunqing, shugong}@huawei.com

***Abstract*—In this paper, we focus on the power-delay tradeoff for downlink transmission systems. In particular, we target to design the adaptive modulation scheme that achieves the optimal power-delay tradeoff under practical concerns, such as the practical power consumption model of a base station and the dynamics of user traffic. The optimal modulation level selection problem is formulated in a Markov decision problem and a cross-layer approach is adopted to take both queue and channel dynamics into consideration. Simulation results show that the proposed policy achieves about 20% power saving for the same delay performance, compared with fixed modulation level schemes. In addition, the impact of the practical power consumption model under dynamic traffic does change the optimal modulation level selection philosophy.**

I. Introduction

As has been pointed out, the radio base station (BS) is the major energy killer in the wireless access networks [1]. Therefore, increasing the energy efficiency of BSs becomes an urgent task to support the rapidly increasing traffic while limiting the energy consumption. However, the reduction of energy consumption can hardly be achieved without any cost to pay. Tradeoffs, as a result, are inevitable. The key question is *when* and *where* to tradeoff *what* such that the reduction consumption reduction can be achieved without degrading user's quality-of-service and/or quality-of-experience. In this paper, we focus on delay-sensitive applications, and try to design the resource scheduling policy that could achieve the optimal power-delay tradeoff under practical system concerns.

To understand the DL-PW tradeoff, let us start with the simplest case first excluding the impact of both channel and traffic dynamics. For point-to-point transmission over AWGN channels, Shannon's formula tells us that $R = W\log_2(1+\frac{P}{WN_0})$ bit information are transmitted each second; hence, it takes $t_b = 1/R$ second to transmit a bit. Therefore, the average power per bit can be expressed as

$$P_b = N_0 W t_b \left(2^{\frac{1}{t_b W}} - 1\right) = N_0 b \left(2^{\frac{1}{b}} - 1\right), \qquad (1)$$

where b is the modulation level in an uncoded transmission system. The relation shows $P_{b,t}$ is a monotonically decreasing function of t ($t \to 0, \mathcal{P}_{b,t} \to \infty$ and $t \to \infty, \mathcal{P}_{b,t} \to N_0 \ln 2$).

There are many works extending the basic results in (1) to packet transmission in the literature. For example, [2] proposed a lazy schedule to minimize the transmission power while guaranteeing the transmission before a strict deadline. The physical layer transmission model in [2] is oversimplified and channel dynamics are not exploited. A benchmark paper [3] took both channel uncertainties and random traffic into consideration and characterized a bound for asymptotically large delay, which was extended to multi-user case in [4]. Furthermore, policies that achieve optimized delay performance have been proposed in [5] and [6]. For instance, [5] proved that longest queue highest possibility rate is optimal for multiaccess channels, and [6] found the delay-optimal precoder adaptation policy for multi-stream MIMO systems.

In this paper, we target to design the adaptive modulation scheme that achieves the optimal power-delay tradeoff for downlink transmission systems with practical concerns that are not addressed in the above literatures. To be more specific, we consider finite modulation levels and the practical power consumption model of a base station, including the circuit power, the processing power, as well as the non-ideal power efficiency of the devices, along the whole signal path in a base station. We shall adopt a cross-layer approach and design the adaptive modulation scheme based on both traffic and channel dynamics. Moreover, we shall also investigate how the practical energy consumption model of a base station impact the design philosophy under dynamic user traffic.

The rest of the paper is organized as follows. Section II introduces the system model and section III formulates the optimal modulation level selection problem into a Markov Decision Problem and finds the solutions that achieves the Pareto optimal boundaries of the tradeoff between average total power and delay. Section IV gives some numerical examples and section V concludes the whole paper.

II. System Model

The scenario of downlink transmission of delay-sensitive applications from base station (BS) to K mobile users is considered. The system diagram is shown in Fig. 1. Simple TDMA based Round-Robin multi-user scheduler is assumed and the users are scheduled on a τ-length time slot basis in Round-Robin (RR) fashion. When user k is scheduled, the scheduler decides the modulation level and the corresponding power for transmission, according to both the channel state information (CSI) and the queue state information (QSI) of the k-th user.

A. Physical Layer Model

We consider flat Rayleigh fading between the BS and users. The received signal at the k-th user is

$$y_k = h_k \sqrt{P_{t,k}} x_k + n_k, \ \ \forall 1 \le k \le K, \qquad (2)$$

978-1-4673-0990-5/12/$31.00 ©2012 IEEE

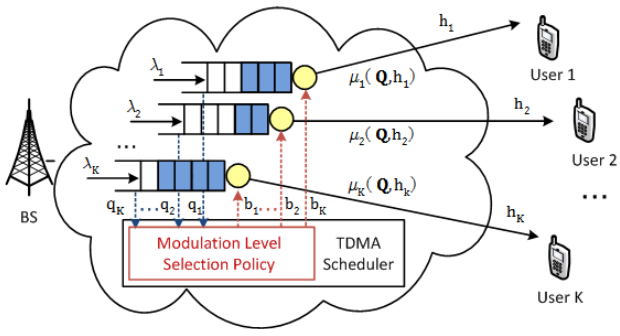

Fig. 1. System diagram of multi-user downlink transmission.

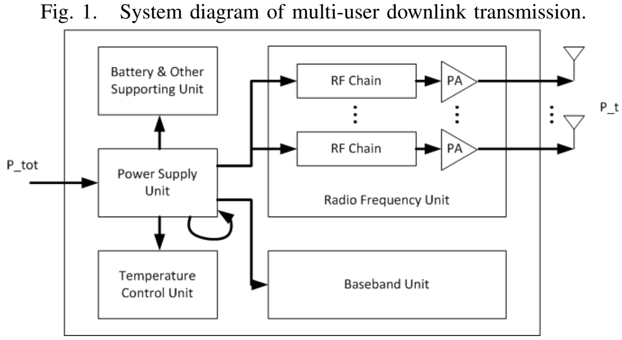

Fig. 2. Base station power consumption chain.

where y_k denotes the received signal at the UT side and n_k denotes the random noise with variance N_0. x_k denotes the transmitted signal at the BS side with normalized signal power, i.e. $\mathbb{E}[x_k x_k^*] = 1$. h_k represents all the channel attenuations including pathloss and Rayleigh fading. The variance of h_k is $\sigma_k^2 = \kappa_0 d_k^{-\alpha}$, where d is the distance between the BS and the k-th user, α is the path loss exponent and κ_0 is the corresponding coefficient. We assume blocking fading channel such that h_k remains static during a round of RR scheduling (K time slots with slot duration τ) but varies independently between users and scheduling rounds.

Uncoded adaptive M-QAM modulation is applied to the transmission of each queue. With peak power constraint, the constellation size M is finite and can only be chosen from natural numbers that are powers of 2, i.e. $M = 2^b$, where b is the modulation level and can be chosen from the set $\Phi = \{2, ..., b_{max}\}$. For given channel condition h_k, the symbol error rate (SER) of 2^b-QAM modulation is bounded by [7]

$$\epsilon_k \leq \left(1 - \frac{1}{\sqrt{2^{b_{max}}}}\right) \exp\left(-\frac{3P_{t,k}|h_k|^2}{2(2^{b_k}-1)N_0 B}\right) \quad (3)$$

B is the system bandwidth. We further assume the channel state information h_k of each user can be accurately feedback to the BS for modulation level selection.

B. Power Consumption Model

From (3), under a given SER ϵ_0, the minimum transmit power needed to support modulation level b is

$$P_t = \frac{N_0 B \left(\ln\left(1 - 1/\sqrt{2^{b_{max}}}\right) - \ln \epsilon_0\right)}{1.5|h|^2}\left(2^b - 1\right) \quad (4)$$

However, in practice, the transmit power is only a portion of the power consumption a base station. Non-ideal power efficiency of the devices make the total input power of a BS P_{in} much larger than the transmit P_t. A typical power chain for a base station is sketched in Fig. 2. Given the air-interface transmit power P_t (omitting the subscript of user index k), the power consumed in the whole system consists of the power from power amplifier (PA), denoted as P_{PA}, the power from electronic circuits in the radio frequency (RF) components, denoted as P_c, the processing power in the base band (BB), denoted as P_{BB}, and the static power consumed in other supporting subsystems such as battery unit, denoted as P_{sta}. Moreover, the total power input should also consider the power supply efficiency, denoted as η_{PS}, $\eta_{PS} < 1$, and cooling expenditure factor, denoted as η_c, $\eta_c < 1$. Assuming n_a antennas in the system, the total power consumption of a base station can be expressed as

$$P_{in} = \frac{n_a(P_{PA} + P_c) + P_{BB} + P_{sta}}{\eta_{PS}(1 - \eta_c)}. \quad (5)$$

Using the model in [8], the power consumption from the PA can be approximated by $P_{PA} = (1 + \alpha)P_t$, where $\alpha = \xi/\eta_{PA} - 1$ with η_{PA} the drain efficiency of the RF power amplifier which is device-dependent and ξ the peak-to-average ratio (PAR), which is dependent on the modulation scheme and the associated constellation size. For M-QAM modulation, we have $\xi = 3\frac{\sqrt{M}-1}{\sqrt{M}+1} = 3\frac{2^{b/2}-1}{2^{b/2}+1}$ [7].

In the dynamic traffic case, the base station may not always be transmitting. In fact, it is an efficient power saving approach to turn off the PA and part of its associated circuit power. In this case, the power consumption of the cooling system (temperature control unit) can also be lowered down. We define the base station power consumption when not transmitting data as residual power, denoted as $P_{c,off}$, and its counterpart in the case of base station transmission is $P_{c,on}$, where

$$P_{c,on} = \frac{n_a P_c + P_{BB} + P_{sta}}{\eta_{PS}(1 - \eta_c)}$$

and $P_{c,off} = \beta P_{c,on}$, $\beta < 1$.

C. Queue model, System States, and Control Policy

There are K independent queues with finite buffer length L_k corresponding to each user's data stream at the BS. Let $Q_k(t)$ denotes the buffer state (number of packets stored) of the k-th queue at the i-th slot and we have $0 \leq Q_k(t) \leq L_k$. The arrival of the K queues is assumed to follow K independent Poisson processes with mean arrival rates $(\lambda_1, .., \lambda_K)$ (unit: packets per slot). The packet length of the k-th data stream, N_k, follows exponential distribution with mean packet size $\overline{N}_k$ (unit: bits per packet). The *observed system state* at the scheduler at time slot i (suppose user k is scheduled in this slot) consists of both the CSIT between user k and the BS and the QSI of all queues, i.e. is $\chi_k(t) = (h_k(t), \mathbf{Q}(t))$, where $\mathbf{Q}(t) = \{q_1(t), q_2(t), \ldots, q_K(t)\}$.

Definition 1: (*Modulation Level Selection Policy*) For any user k, the modulation level selection policy $\mathcal{P}$:

$\{0, 1, ..., L\}^K \times \mathbb{C} \to \mathbb{R}$ is a mapping from the currently observed system state $\chi_k(t) = (h_k(t), \mathbf{Q}(t))$ to a modulation level $b_k(\chi_k(t)) = \mathcal{P}(\chi_k(t))$ chosen from the set $\Phi = \{0, 1, ..., b_{max}\}$.[1]

Given system state χ_k, in the scheduled slot of user k, $b_k(\chi(k))(\ll N_k)$ bits are transmitted. With the assumption of exponentially distributed packet length, the packet service time follows exponential distribution too with mean service rate conditioned on the system state χ_k as

$$\mu_k(\chi_k) = \frac{B}{K} b_k(\chi_k)/\overline{N}_k, \tag{6}$$

in which $1/K$ comes from the TDMA scheduling.

Definition 2: (*Average Delay of User k*) By Little's Law [10], for given policy $\mathcal{P}$, the average delay of the k-th user is

$$\overline{D}_k(\mathcal{P}) = \frac{1}{\lambda_k} \limsup_T \frac{1}{T} \mathbb{E}_{\chi_k} \left[\sum_{i=1}^{T} Q_k(t) \right] \tag{7}$$

Definition 3: (*Average Power Consumption of User k*) For given policy $\mathcal{P}$, the average delay of the k-th user is

$$\overline{P}_k(\mathcal{P}) = \limsup_T \frac{1}{T} \mathbb{E}_{\chi_k} \left[\sum_{i=1}^{T} P_k(t) \right] \tag{8}$$

III. Power-Delay Tradeoff Characterization

We are trying to find the optimal policy $\mathcal{P}^*$ that minimized the total average delay of K data streams while satisfy the long term average power constraint at the BS, namely

$$\min_{\mathcal{P}} \quad \sum_k \overline{D}_k(\mathcal{P}) \tag{9}$$

$$\text{s.t.} \quad \sum_k \overline{P}_k(\mathcal{P}) \le \overline{P}_{tot}, \ \ \forall k = 1, \dots, K \tag{10}$$

By Little's law, we may rewrite $\overline{D}_k(\mathcal{P}) = \overline{q}_k(\mathcal{P})/\lambda_k$.

For a given multiplier γ (denoting the level of punishment for power consumption), we may change the the above problem into the following non-constrained problem

$$\mathcal{J}_1(\mathcal{P}) \triangleq \sum_k \overline{q}_k(\mathcal{P})/\lambda_k + \gamma \sum_k \overline{P}_k(\mathcal{P}) \tag{11}$$

The solution to the $\mathcal{J}_1$-minimization problem corresponds to a point on the Pareto frontier of the tradeoff between total average power and sum average delay. Therefore, by traversing all values of γ, we can obtain the whole tradeoff curve.

In the following, we first formulate $\mathcal{J}_1$-minimization problem into a reduced-state Markov Decision Problem (MDP), and then find the optimal modulation level selection policy by iteratively solving the Bellman equations.

A. Markov Decision Problem Formulation

Assume that the scheduling slot duration τ is substantially smaller than the average packet inter-arrival time as well as average packet service time ($\tau \ll 1/\lambda_k$ and $\tau \ll 1/\mu_k$).[2] Suppose the observation epoch is at the beginning of each slot. At the t-th observation epoch, the observed system state is $\chi_k t = \{h_k, \mathbf{Q}(t) = (q_1(t), ..., q_K(t))\}$ (assume the k-th user is about to be scheduled). Then at the $(t+1)$-th observation epoch, one of the following events may happen:

- *Packet arrival at the k-th buffer, with probability:*

$$p_{k,+1} \triangleq \Pr[q_k(t+1) = q+1 | q_k(t) = q] = \lambda_k \tau,$$

- *Packet departure from the k-th buffer, with conditional probability (conditioned on the QSI):*

$$p_{k,-1} \triangleq \Pr[q_k(t+1) = q-1 | q_k(t) = q] = \overline{\mu}_k(\mathbf{Q}(\mathbf{t}))\tau,$$

  where $\overline{\mu}_k(\mathbf{Q}(t)) = \mathbb{E}_{h_k}[\mu_k(\chi_k(t)) | \mathbf{Q}(t)]$.
- *No change in the k-th buffer, with probability:*

$$p_{k,0} \triangleq \Pr[q_k(t+1) = q | q_k(t) = q] = 1 - p_{k,+1} - p_{k,-1}.$$

Furthermore, since the slot duration τ is small, in one slot, the probability of multiple packet arrivals or departures in the same queue is negligible, namely $p_{k,\pm n} = 0$ for $n > 1$. Due to the same reason, the probability of packet arrival and departure in more than one buffers is also negligible, i.e.,

$$\Pr\left[q_i(t+1) \neq q_i(t) | q_k(t+1) \neq q_k(t),\, i \neq k\right] = 0.$$

From the above, we see that the state transition probability is a function of the current queue state $\mathbf{Q}(t)$ only. Hence, given a stationary policy $\mathcal{P}$, $\{\mathbf{Q}(t)\}$ is an *irreducible* Markov chain induced by $\mathcal{P}$. Since the CSIT $\{h_k\}$ is i.i.d. between slots and users, we may reformulate the $\mathcal{J}_1$-minimization problem into the following Markov Decision Problem (MDP) to minimize the average cost per stage:

$$\mathcal{J}_1(\mathcal{P}) = \limsup_T \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\mathbf{Q}} \left[\sum_k c_k\left(\mathbf{Q}(t), \overline{P}_k(\mathbf{Q}(t))\right) \right] \tag{12}$$

where $\sum_k c_k(\cdot)$ is the cost per stage and $c_k(\cdot)$ is

$$c_k\left(\mathbf{Q}(t), \overline{P}_k(\mathbf{Q}(t))\right) = \sum_k q_k(t)/\lambda_k + \gamma \overline{P}_k(\mathbf{Q}(t)) \tag{13}$$

B. Optimal Modulation Level Selection Policy

For the infinite horizon MDP formulated above, the optimizing policy can be obtained by solving the *K-dimension Bellman equation* [9] recursively w.r.t. the optimal result $\theta = \mathcal{J}(\mathcal{P}^*)$ and the potentials $\{V(\mathbf{Q}) \triangleq V(q_1, ..., q_K)\}$ for the reduced state $\mathbf{Q} = (q_1, ..., q_K)$ as below:

$$\begin{aligned}\theta + V(\mathbf{Q}) = \inf_{\mathcal{P}(\mathbf{Q})} \Big\{ & \sum_k c_k\left(q_k, \overline{P}_k(\mathbf{Q})\right) \\ & + \tau \sum_k \lambda_k V_k^+(\mathbf{Q}) + \tau \sum_k \overline{\mu}_k(\mathbf{Q}) V_k^-(\mathbf{Q}) \\ & + V(\mathbf{Q})(1 - \tau \sum_k \lambda_k - \tau \sum_k \overline{\mu}_k(\mathbf{Q})) \Big\}\end{aligned} \tag{14}$$

where $V_k^+(\mathbf{Q}) \triangleq V(q_1, ..., \min\{q_k + 1, L\}, ..., q_K)$ and $V_k^-(\mathbf{Q}) \triangleq V(q_1, ..., \max\{q_k - 1, 0\}, ..., q_K)$. By substituting

[1] It is shown [9] that for finite state MDP, stationary and history independent policy is optimal. Hence, there is no loss of generality to consider policy that is function of current system state only.

[2] In fact, this is a mild assumption. For example, in WiMAX, a frame duration is around 2ms while the target queueing delay for applications (such as video streaming) is around 200ms or more.

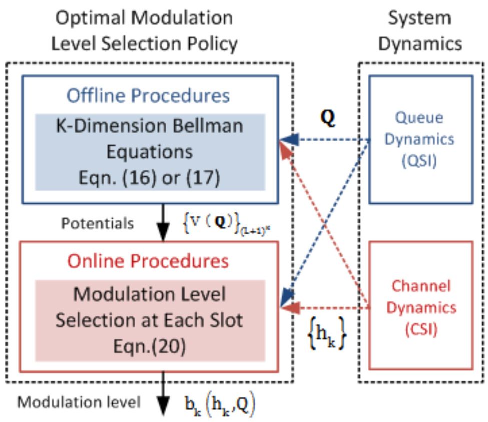

Fig. 3. Flowchart of the optimal modulation level selection policy.

(13) into (14) and further denoting $\psi_k^+(\mathbf{Q}) = V_k^+(\mathbf{Q}) - V_k(\mathbf{Q})$ and $\psi_k^-(\mathbf{Q}) = V_k^-(\mathbf{Q}) - V_k(\mathbf{Q})$, we have

$$\theta = \sum_k q_k/\lambda_k + \tau \sum_k \lambda_k \psi_k^+(\mathbf{Q}) + \mathcal{J}_2^* \tag{15}$$

where the sub-optimization problem is defined as follows

$$\mathcal{J}_2^* \triangleq \inf_{\mathcal{P}(\mathbf{Q})} \left\{ \gamma \sum_k \overline{P}_k(\mathbf{Q}) + \tau \sum_k \overline{\mu}_k(\mathbf{Q}) \psi_k^-(\mathbf{Q}) \right\}. \tag{16}$$

Note that $\overline{P}_k(\mathbf{Q}) = \mathbb{E}_{h_k}[P_k(\chi_k)|\mathbf{Q}]$ and $\overline{\mu}_k(\mathbf{Q}) = \mathbb{E}_{h_k}[\mu_k(\chi_k)|\mathbf{Q}]$, we may further change the form of the problem defined in (16) as below

$$\begin{aligned}\mathcal{J}_2^* &= \inf_{\mathcal{P}(\mathbf{Q})} \sum_k \mathbb{E}_{h_k}\left[(P_k(\chi_k) + \tau\mu_k(\chi_k)) | \mathbf{Q} \right] \\ &= \sum_k \mathbb{E}_{h_k}\left[\inf_{b_k(\chi_k)} \left\{ \gamma P_k(\chi_k) + \tau \mu_k(\chi_k) \psi_k^-(\mathbf{Q}) \right\} \right] \end{aligned} \tag{17}$$

Taking the expressions of $P_k(\chi_k)$ and $\mu_k(\chi_k)$ into (17), we derive the explicit relationship between $\mathcal{J}_2^*$ and the modulation level b_k selection scheme as follows

$$\mathcal{J}_2^* = \sum_k \mathbb{E}_{h_k}\left[\inf_{b_k} \left\{ \gamma P_c' + C_1 b_k + \gamma C_2 \left(\sqrt{2^{b_k}} - 1 \right)^2 \right\} \right] \tag{18}$$

where $P_c' = \frac{P_c + P_{BB} + P_{sta}}{\eta_{PS}(1-\eta_c)}$, and C_1 and C_2 are two constants defined as follows (denoting $\eta' = \frac{\eta_{PA}}{\eta_{PS}(1-\eta_c)}$)

$$C_1 = \frac{\tau B \psi_k^-(\mathbf{Q})}{K \overline{N}_k} \quad \text{and} \quad C_2 = \frac{2N_0 B}{\eta' |h_k|^2} \ln \frac{1 - 1/\sqrt{2^{b_{max}}}}{\epsilon_0},$$

The inner optimization problem of selecting the optimal modulation level b_k under given system state χ_k can be solved by searching in the whole set Φ of the modulation levels.

In summary, the original Bellman equation in (14) (or (15)) can be solved using the classic value iteration method in an offline manner. The details of the solutions are described in the appendix. The output of the offline procedure are the optimal average cost per stage θ and the $((L+1)^K)$ values of potentials $V(\mathbf{Q})$. These values are used as the input of the online modulation level selection procedure. In particular, when the k-th user is scheduled, the optimal modulation level

Notation	Value	Notation	Value
κ_0	0.03	P_c	10 W
α	3.76	P_{BB}	20 W
B	200 kHz	P_{sta}	10 W
η_{PA}	0.35	η_{PS}	0.9 W
η_c	0.05	n_a	1 W
N_0	-174 dBm/Hz	ϵ_0	0.01
τ	0.01 sec	$\overline{N}$	400 kbits
λ	1 packet/sec	L	10 packets
Φ	$\{2, ..., 8\}$	b_{max}	8

TABLE I
SIMULATION PARAMETERS.

under the given system state $\chi_k(t) = \{h_k(t), \mathbf{Q}(t)\}$ can be obtained by solving (18). Fig. 3 depicts the whole flow chart.

Finally, the Markov chain $\mathbf{Q}(t)$ under the optimal modulation level selection policy $\mathcal{P}^*$ is irreducible and ergodic, and we define $\omega(\mathbf{Q}) = \omega(q_1, ..., q_K)$ as the steady-state probability for state $\mathbf{Q} = (q_1, ..., q_K)$. As a result, the average delay for each user as well as the average power for each user are

$$\overline{D}_k = \frac{1}{\lambda_k} \sum_{q_1,...,q_K} q_k \cdot \omega(q_1, ..., q_K) \tag{19}$$

$$\overline{P}_k = \sum_{q_1,...,q_K} \mathbb{E}_{h_k}\left[P_k(h_k, \mathbf{Q}) | \mathbf{Q} \right] \cdot \omega(q_1, ..., q_K), \tag{20}$$

where $\omega(q_1, ..., q_K)$ can be calculated from the K-dimension detailed balance equations [10].

C. A Reduced Problem with Single Queue Dynamics

The Bellman equation in this case reduces from K-dimension to one-dimension, and can be expressed as

$$\begin{aligned}\theta + V(q_k) = \inf_{\mathcal{P}_{red}} \Big\{ & c_k\left(q_k, \overline{P}_k(q_k)\right) + \tau \lambda_k V_k^+(q_k) \\ & + \tau \overline{\mu}_k(q_k) V_k^-(q_k) + V(q_k)(1 - \tau\lambda_k - \tau\overline{\mu}_k(q_k)) \Big\} \end{aligned} \tag{21}$$

where $V_k^+(q_k) = V_k(min\{q_k+1, L\})$ and $V_k^-(q_k) = V_k(max\{q_k - 1, 0\})$, and the policy $\mathcal{P}_{red} : \{0, 1, ..., L\} \times \mathbb{C} \to \mathbb{R}$ is a mapping from the reduced system state $\chi_{k,red}$ to a modulation level b_k. Using similar derivations, the Bellman equation can be transformed into the following form

$$\theta = q_k/\lambda_k + \tau \lambda_k \psi_k^+(q_k) + \mathcal{J}_3^*, \tag{22}$$

$$\mathcal{J}_3^* = \mathbb{E}_{h_k}\left[\inf_{b_k} \left\{ \gamma P_c' + C_1 b_k + \gamma C_2 \left(\sqrt{2^{b_k}} - 1 \right)^2 \right\} \right] \tag{23}$$

The steady state distribution $\{\omega(q_k)\}$ of the queue lengths under the optimal policy can be obtained by solving the *one-dimensional* detailed balance equations: $(q_k = 0, 1, .., L-1)$

$$\lambda_k \omega(q_k) = \overline{\mu_k}(q_k + 1)\omega(q_k + 1) \tag{24}$$

with the boundary condition $\sum_{q_k=0}^{L} \omega(q_k) = 1$. Finally,

$$\omega(q_k) = \omega(L) \prod_{j=q_k+1}^{L} \frac{\overline{\mu_k}(j)}{\lambda_k}, \tag{25}$$

where $\omega(L) = \left[1 + ... + \frac{\overline{\mu_k}(L)...\overline{\mu_k}(1)}{\lambda_k^L} \right]^{-1}$.

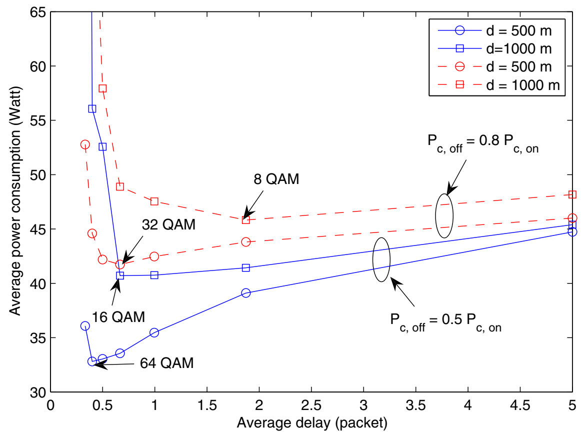

Fig. 4. Power-delay tradeoff curves under fixed modulation scheme with different residual power assumptions and different transmission distance.

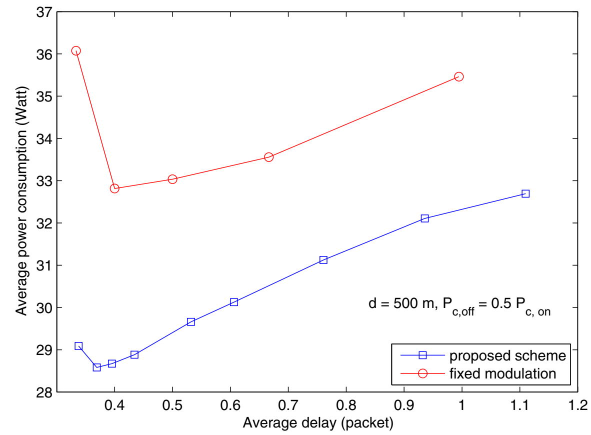

Fig. 5. Comparison of power-delay tradeoff curves under the proposed scheme and the fixed modulation scheme.

IV. Numerical Examples

In this section, we characterize the average power-delay tradeoff for the proposed modulation level selection scheme via numerical simulations. The system parameters for simulations are listed in Table III-B. In addition, we consider homogeneous users, i.e. $\lambda_k = \lambda, \overline{N}_k = \overline{N}, \forall k$. Given the TDMA scheduler and the i.i.d channel fadings, it can be proved that the steady state of each queue is the same. Therefore, in the following, we focus on the k-th user only and build the simulations for adaptive modulation based on the reduced problem, as stated in III-C.

Fig. 4 depicts the power-delay relation under fixed modulation levels. Each point on the curve corresponds to a specific modulation level. The purpose of this figure is to illustrate how the selection of optimal modulation level changes with the power consumption model in dynamic traffic and with transmission distance. Firstly, it is easy to see that the larger the transmission distance, the higher the power demand. Secondly, the lower the residual power consumption when no data is being transmitted, the lower the average power consumption. Finally, the joint impact of residual power level and transmission distance leads to different selection of the modulation levels in the sense of optimizing the power-delay tradeoff curve. Specifically, with small transmission distance and low residual power, high modulation level is suggested to be used (e.g., 64QAM for $d = 500$m, $P_{c,off} = 0.5P_{c,on}$ case); while with large transmission distance and comparatively high residual power, lower modulation level is instead preferred (e.g., 8QAM for $d = 1000$m, $P_{c,off} = 0.8P_{c,on}$ case).

Fig. 5 compares the power-delay tradeoff curve under the proposed adaptive modulation scheme and the fixed modulation scheme. Each point on the solid bold line corresponds to a operation point on the Pareto frontier defined by (11) with some specific multiplier γ. As we can observe from the figure that, under the fixed modulation level policy, changing from one modulation level to another, corresponds to the move of operation point along the given power-delay curve, i.e., trading delay for power or trading power for delay. On the contrary, with the optimal adaptive modulation scheme, both the power and delay performance is improved for the same simulation setting, i.e., lower power and shorter delay can be achieved simultaneously and the gain is around 20%.

V. Conclusions

In this paper, the tradeoff between the downlink system power consumption and average user delay is studied under practical concerns such as the practical power consumption model of a base station and the dynamic traffic of users. Simulation results shows that smaller delay and lower power can be achieved simultaneously given proper transmission scheme design, especially when under bad transmission conditions (e.g. long distance). Moreover, the selection of the optimal modulation level is closely related to the practical power consumption model, as well as the transmission distance.

References

[1] G. P. Fettweis and E. Zimmermann, "ICT energy consumption - trends and challenges," in *Proc. of 11th International Symposium on Wireless Personal Mulitimedia Communications*, Lapland, Finland, Sept 2008.

[2] E. Uysal-Biyikoglu, B. Prabhakar, and A. E. Gamal, "Energy-efficient packet transmission over a wireless link," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 487–499, Aug 2002.

[3] R. A. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135–1148, May 2002.

[4] M. J. Neely, "Optimal energy and delay tradeoffs for multi-user wireless downlinks," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3095–3113, Sept 2007.

[5] E. M. Yeh, "Multiaccess and fading in communication networks," Ph.D. dissertation, MIT, Sept 2001.

[6] V. Lau and Y. Chen, "Delay-optimal power and precoder adaptation for multi-stream mimo systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 3104–3111, June 2009.

[7] J. G. Proakis, *Digital Communication*, 4th ed. New York: McGraw Hill, 2000.

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

[9] D. P. Bertsekas, *Dynamic Programming - Deterministic and Stochastic Models*. Prentice Hall, New Jersey, NJ, USA, 1987.

[10] L. Kleinrock, *Queueing Systems Volume I: Theory*. John Wiley & Sons, 1975.