

Energy-Efficient Communications in MIMO Systems Based on Adaptive Packets and Congestion Control With Delay Constraints

Kunlun Wang and Wen Chen, *Senior Member, IEEE*

Abstract—In this paper, we propose adaptive techniques for multiple input and multiple output (MIMO) systems, to solve the problem of energy efficient communications with delay constraint, where the energy efficiency is defined as the number of bits per second correctly received per power consumed. We first investigate the optimal multiple quadrature amplitude modulation (MQAM) constellation size for each transmission stream and the optimal packet size. By exploiting the intrinsic relationship among the constellation size, the packet size, the symbol error rate (SER) and delay, we propose an adaptive transmission mode for different delay demands. For the case of user's buffer overflow, we use the congestion control algorithm to schedule the average queue length, and maintain the optimal delay performance for energy efficiency. Simulations show that to maximize the energy efficiency and offer different Quality of Service (QoS) of delay simultaneously, the transmitter should adaptively choose the constellation size and the packet size as well as the transmission mode. In this framework, the tradeoff between energy efficiency and delay demand are well demonstrated.

Index Terms—MIMO, energy efficiency, packet size, MQAM constellation size, delay, congestion control.

I. INTRODUCTION

RECENTLY, energy efficient communications in wireless networks have attracted much research attention. In communication theory, the throughput and the power are respectively the common measures of the benefit and the cost of a communication system, while the energy efficiency, expressed as the throughput per power, is to use the power as efficiently as possible.

The packet size, the constellation size and the delay constraint are some key factors to influence the energy efficiency of a communication system [1], [2]. A large packet is suscep-

tible to error which may cause retransmission, so as to affect the throughput [1] and the energy efficiency. On the other hand, when the packets generated from upper layer, they will be transmitted by the physical layer. Large constellation size of physical layer requires more power for transmission with particular symbol error probability, while small constellation size can reduce the transmit power. Thus, a specific modulation has impact on the energy efficiency. When the packet size, the constellation size and the symbol error rate (SER) are given, the packet delay can be derived. Then, under different delay constraints, different throughput and transmission power are required [2], which hence influences the energy efficiency. Motivated by these issues of energy efficiency, we try to find the optimal packet size and the optimal constellation size for energy efficient communications, and investigate the optimal energy efficiency for given delay constraints, which is a cross-layer optimization model. Cross-layer design based on energy and delay optimization in small-scale sensor networks are considered in [3], [4], although multiple input and multiple output (MIMO) energy efficiency is not the main objective of these works. Similar cross-layer methodology can be applied.

There exists several studies on the packet size optimization for wireless networks [5], [6], which consider the maximization of energy efficiency subject to transmission power or data rates. In [7], [8], energy efficient packet size for sensor network has been studied. However, these works did not consider the inherent relation between the packet size and the delay aware energy efficiency. On the other hand, in [9], the authors use the convex-optimization method to study the optimal modulation that can minimize the energy consumption per bit under the delay constraint in multi-hop time division multiple access (TDMA) networks, but not considering different modulation sizes for each stream. In [10], the influence of the constellation size to the energy efficiency of a MIMO system has been shown, and it has been shown that the optimal constellation size can dramatically increase the energy efficiency. However, the influence of the constellation size to the power consumption of the MIMO streams has not been considered there, although the optimal power allocation for MIMO streams has been studied in [11]. On the other hand, MIMO systems are not always superior to the single input and multiple output (SIMO) systems due to different circuit power consumption [2]. Since multiple transmission antennas require multiple circuits (mixers, synthesizers, digital-to-analog converters, filters, etc). Hence the circuit power of MIMO transmission is higher than that of SIMO [10].

Manuscript received January 20, 2014; revised June 22, 2014 and October 11, 2014; accepted December 8, 2014. Date of publication December 18, 2014; date of current version April 7, 2015. This work is supported by the National 973 Project #2012CB316106, by NSF China #61161130529 and #61328101, by STCSM Science and Technology Innovation Program #13510711200, and by SEU National Key Lab on Mobile Communications #2013D11. Part of this paper was presented at the International Conference on Wireless Communications and Signal Processing, Hefei, China, October 23–25, 2014. The associate editor coordinating the review of this paper and approving it for publication was S. Valaee.

The authors are with the Department of Electronics Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the School of Electronic Engineering and Automation, Guilin University of Electronics Technology, Guilin 541004, China (e-mail: wenchen@sjtu.edu.cn; kunlun1228@sjtu.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/TWC.2014.2382117

Therefore, a better transmission mode from MIMO and SIMO can be chosen to improve the energy efficiency [10], [12].

The delay demand can influence the energy efficiency [3], [13]–[15]. The existing works majorally focus on the trade-off between the average delay and the average transmission power, not considering the MIMO/SIMO mode switching. In [12], the delay aware MIMO/SIMO switching strategy is proposed, however, which is based on the flow delay not including the optimization of the constellation size. On the other hand, long delay may cause buffer overflow, which has not been considered in the area of energy efficiency in most of the recent research works. In fact, buffer overflow can be solved by congestion control algorithm, which is the main function of the transmission control protocol (TCP). Note that the congestion problem has been considered in [16], and the congestion information is transmitted back to the source by putting the congestion price in the acknowledgement (ACK) packets. Therefore, one can utilize the congestion control algorithm to solve the buffer overflow in energy efficiency by taking the queue length in the buffer of the user as the congestion price.

In this paper, we optimize the constellation size for each stream to achieve the energy efficient transmission in MIMO systems. Meanwhile, we simultaneously derive the optimal packet size to maximize the energy efficiency. Since SIMO systems may be more energy efficient than the MIMO systems in some packet delay circumstances, we propose an adaptive MIMO/SIMO transmission strategy to improve the energy efficiency, and select the optimal antenna for SIMO mode. To our best knowledge, considering the packet size, the modulation size and the antenna selection in the delay aware energy efficiency, has not been considered so far, and the prior works in this area did not explicitly take into account the effect of the packet delay constraint to antenna selection in MIMO systems.

Due to the burst service and the long delay, there will be congestion causing buffer overflow in the user, which can influence the energy efficiency. To deal with this problem, we use the congestion control algorithm and the virtual queue algorithm to offset its influence, and then we can see the superiority and the convergence property of the algorithms. In all, we consider the cross-layer model in the user station, from which the optimal energy efficiency solution can be asymptotically obtained. The energy-delay tradeoffs in wireless networks have been studied in various works [12]–[14], [17]. But the energy-delay tradeoff investigated in this paper is based on energy efficient transmission in MIMO systems considering the circuit power.

The rest of the paper is organized as follows. Section II describes the system model, including the transmission model and the queuing model. In Section III, we analyze the optimal constellation size for the MIMO wireless transmission. Section IV shows the delay performance. In Section V, we show the simulation results, and the conclusions are made in Section VI.

Notations: $E(\cdot)$, $\|\cdot\|_F$, $\|\cdot\|_2$ and $(\cdot)^H$, denote the expectation, the Frobenius norm, the Euclidean norm and the conjugate transpose, respectively. $\text{Tr}\{\mathbf{X}\}$ is the trace of matrix \mathbf{X} . $\text{diag}(\cdot)$ is a diagonal matrix. $\lfloor x \rfloor$ is the largest integer number that is not larger than x , and $\lceil x \rceil$ is the smallest number that is not smaller than x .

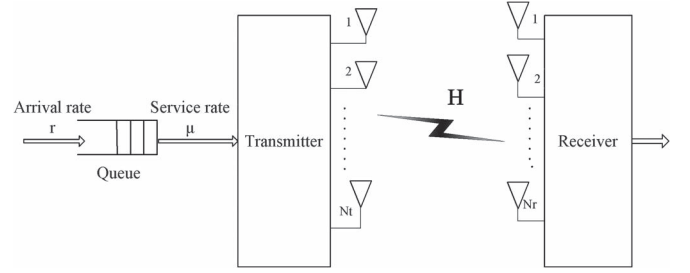


Fig. 1. System model.

II. SYSTEM MODEL

A. Physical Layer Channel Model

Consider a MIMO system with the bandwidth of B , shown in Fig. 1, where the transmitter is equipped with N_t antennas and the receiver is equipped with N_r antennas. Without loss of generality, we assume $N_t \leq N_r$, so that N_t independent data streams can be transmitted simultaneously through the MIMO channel. Let $\mathbf{s} = [s_1, s_2, \dots, s_{N_t}]^T$ denote the information symbol vector to be transmitted at each time instant. Each element s_i can come from a 2^{b_i} -QAM modulation and is subject to a unit power constraint $E\{|s_i|^2\} = 1$. We assume that Nyquist pulses are used and hence the MQAM symbol rate is approximately equal to the transmission bandwidth B . The total number of information bits that can be transmitted at each time is given by

$$b = \sum_{i=1}^{N_t} b_i.$$

Let $\mathbf{P} = \text{diag}\{\sqrt{p_{t1}}, \sqrt{p_{t2}}, \dots, \sqrt{p_{tN_t}}\}$ be the power allocation matrix of the information symbols and \mathbf{Q} be the precoding matrix. Then the received signal vector at the receiver, denoted as $\mathbf{y} = [y_1, y_2, \dots, y_{N_r}]^T$, can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{Q}\mathbf{P}\mathbf{s} + \mathbf{n}, \quad (1)$$

where $\mathbf{H} = N_r \times N_t$ is the channel matrix and $\mathbf{n} = [n_1, n_2, \dots, n_{N_r}]^T$ is a Gaussian distribution noise with zero mean and the covariance matrix $(N_0 B) \mathbf{I}_{N_r}$, and \mathbf{I}_{N_r} is the identity matrix. The knowledge of the channel matrix \mathbf{H} is assumed to be known at both the transmitter and the receiver.

B. Link Layer Queuing Model

The transmitted bits at the physical layer come from the link layer in a packet basis. Each packet has a size of L bits, among which L_h bits are the header and

$$L_p = L - L_h$$

are the payload bits. When the receiver correctly receives a packet, it will send back an ACK packet to the transmitter, which is the signal passed between transmitter and receiver to signify acknowledgement or receipt of response. If the receiver can not correctly receive the packet, the transmitter will repeat transmitting the packet until it is received correctly.

The link layer packets arrive at the transmitter into a first-in-first-out (FIFO) queue, and the buffer is finite with size Q_0 . Consider that the user's link layer constructs packet streams with the packet size of L bits. With regard to the delay performance of the packet, assume that the user's queuing model is a single server M/G/1 queue [16], shown in Fig. 1, where r is the mean packet generation rate from the data link layer, and μ is the mean service rate at the physical layer. Clearly, the service rate μ depends on b , the total number of transmitted bits through the channel at each time as determined by the channel model in (1).

III. POWER MINIMIZED CONSTELLATION SIZE

In this section, we determine the optimal constellation size for power consumption minimization in the MIMO system subject to a given target rate, and the power consumption includes transmission power and the circuit power.

A. Power Calculation

Since we assume that \mathbf{H} is known at both the transmitter and the receiver, we can apply singular value decomposition (SVD) based transmission [18]. In specific, let $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, where \mathbf{U} and \mathbf{V} are respectively $N_r \times N_r$ and $N_t \times N_t$ unitary matrices, and

$$\mathbf{\Lambda} = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_{N_t}], \quad (2)$$

where $\lambda_i \geq 0$ represents the singular value of \mathbf{H} , and we assume that \mathbf{H} has a rank of N_t . We set the precoding matrix $\mathbf{Q} = \mathbf{V}$ and the receiving matrix $\mathbf{B} = \mathbf{U}^H$. Correspondingly, the decision vector for the transmitted symbols is

$$\hat{\mathbf{s}} = \mathbf{\Lambda}\mathbf{P}\mathbf{s} + \mathbf{U}^H\mathbf{n} \triangleq \mathbf{\Lambda}\mathbf{P}\mathbf{s} + \hat{\mathbf{n}}, \quad (3)$$

where $\hat{\mathbf{n}} = \mathbf{U}^H\mathbf{n}$ is the effective noise.

From (3), the receiver can detect each symbol independently, and derive the achieved SNR of all the received symbols as

$$\Gamma \triangleq \left[\frac{p_{t1}\lambda_1^2}{N_0B}, \frac{p_{t2}\lambda_2^2}{N_0B}, \dots, \frac{p_{tN_t}\lambda_{N_t}^2}{N_0B} \right]. \quad (4)$$

We know that the SNR per symbol is

$$\gamma_{s_i} \triangleq \frac{p_{ti}\lambda_i^2}{N_0B} \frac{B}{R_s} = \frac{p_{ti}\lambda_i^2}{N_0R_s}, \quad (5)$$

for $i = 1, \dots, N_t$, where R_s is the symbol rate.

It is well known that [19], the SER of MQAM modulation with size 2^{b_i} , is given by

$$p_e(b_i, \gamma_{s_i}) = 2(1 - 2^{-b_i/2})Q\left(\sqrt{\frac{3}{2^{b_i}-1}}\gamma_{s_i}\right), \quad (6)$$

where $Q(\cdot)$ is the complementary cumulative distribution function of the standard Gaussian random variable and b_i is the constellation size for symbol s_i . Using Chernoff upper bound, we have

$$p_e(b_i, \gamma_{s_i}) \leq 2(1 - 2^{-b_i/2})e^{-\frac{3}{2^{b_i}-1}\gamma_{s_i}}. \quad (7)$$

For the purpose of simplicity, we assume that the target SER for each stream is the same and given by p_e . Therefore, by (4), (5) and (7), the transmission power for symbol s_i is given by

$$p_{ti}(b_i) \approx \frac{N_0R_s}{\lambda_i^2} \frac{2(2^{b_i}-1)}{3} \ln \frac{2(1-2^{-b_i/2})}{p_e}. \quad (8)$$

Next we introduce the power consumption model. The total power consumption of the transmitter consists of the over-the-air transmission power and the circuit power.

From (1) and (8), the total transmission power $P_m(\{b_i\})$ of the MIMO system can be calculated as

$$P_m(\{b_i\}) \approx \sum_{i=1}^{N_t} \frac{N_0R_s}{\lambda_i^2} \frac{2(2^{b_i}-1)}{3} \left(\ln \frac{2}{p_e} + \ln(1-2^{-b_i/2}) \right). \quad (9)$$

The circuit power consumption P_c of the MIMO system includes two parts. The first part circuit power consumption at each transmit antenna is a constant power P_0 , and second part circuit power consumption at each antenna linearly scales with the transmission rate b_i , that can be denoted as αb_i for some constant α [20]. Thus

$$P_c = \sum_{i=1}^{N_t} (\alpha b_i + P_0). \quad (10)$$

In all, we can get the total power consumption of the transmitter as

$$\hat{P}_m = P_m(\{b_i\}) + \sum_{i=1}^{N_t} (\alpha b_i + P_0). \quad (11)$$

B. Constellation Size Optimization

Now, we consider that the transmission bit rate from the upper-layer is a constant. Then $b = \sum_{i=1}^{N_t} b_i$ is a constant for given symbol rate R_s . Consider $b_i \geq b_{\min}$ for some lower bound b_{\min} . Our objective is to optimize b_i of the symbol s_i , to get the minimal power consumption.

Since the design variable b_i is an integer number, exhaustive search is a feasible way to solve this problem. However, for the purpose of reducing complexity, we relax b_i to be a real number, and prove that this is a convex optimization. Since (11) is complicated, we can give a tight upper bound of (11) and minimize the upper bound instead. Since $1 - 2^{-b_i/2} \leq 1$, we have

$$P_m(\{b_i\}) \leq \sum_{i=1}^{N_t} \frac{N_0R_s}{\lambda_i^2} \frac{2(2^{b_i}-1)}{3} \ln \frac{2}{p_e}. \quad (12)$$

From (11) and (12), we can get

$$\hat{P}_m \leq \sum_{i=1}^{N_t} \frac{N_0R_s}{\lambda_i^2} \frac{2(2^{b_i}-1)}{3} \ln \frac{2}{p_e} + \alpha \sum_{i=1}^{N_t} b_i + N_t P_0. \quad (13)$$

Thus, From (13), we can formulate the optimization problem to derive the optimal constellation size as

$$\begin{aligned} \min \quad & \sum_{i=1}^{N_t} \frac{N_0R_s}{\lambda_i^2} \frac{2(2^{b_i}-1)}{3} \ln \frac{2}{p_e} + \alpha \sum_{i=1}^{N_t} b_i, \\ \text{s.t.} \quad & b_i \geq b_{\min}, \\ & \sum_{i=1}^{N_t} b_i = b. \end{aligned} \quad (14)$$

Note that the objective function in problem (14) is convex. We refer to Lagrangian multiplier to solve the problem. Let κ and ν denote the Lagrange multipliers associated with the constraints in the optimization problem (14). The Lagrangian function is then given by

$$\mathcal{L}(\mathbf{b}, \kappa, \nu) = \sum_{i=1}^{N_t} \left(\frac{N_0 R_s}{\lambda_i^2} \frac{2(2^{b_i} - 1)}{3} \ln \frac{2}{P_e} + \alpha b_i \right) - \sum_{i=1}^{N_t} \kappa_i (b_i - b_{\min}) + \nu \left(\sum_{i=1}^{N_t} b_i - b \right). \quad (15)$$

So the necessary and sufficient conditions for optimality are given by the Karush-Kuhn-Tucker (KKT) conditions [21],

$$\begin{cases} b_i^* \geq b_{\min}, \\ \sum_{i=1}^{N_t} b_i^* = b, \\ \kappa_i^* \geq 0, \\ \kappa_i^* (2 - b_i^*) = 0, \quad i = 1, \dots, N_t, \\ \beta_i 2^{b_i^*} + \alpha - \kappa_i^* + \nu^* = 0, \quad i = 1, \dots, N_t, \end{cases} \quad (16)$$

where

$$\beta_i \triangleq \frac{2 \ln 2}{3} \ln \frac{2}{P_e} \frac{N_0 R_s}{\lambda_i^2},$$

κ_i^* and ν^* denote the optimal multipliers, b_i^* is the optimal b_i . We can directly solve the equations in (16) to find b_i^* , κ_i^* and ν^* . Thus we have

$$b_i^* = \begin{cases} \log_2 \left(\frac{-\alpha - \nu^*}{\beta_i} \right), & \nu^* \leq -\alpha - \beta_i 2^{b_{\min}}, \\ b_{\min}, & \nu^* > -\alpha - \beta_i 2^{b_{\min}}. \end{cases} \quad (17)$$

Substituting (17) into the condition $\sum_{i=1}^{N_t} b_i^* = b$, we can obtain the optimal ν^* and b_i^* . Since b_i is an integer number, we choose the optimal constellation size b_i^{opt} as

$$b_i^{opt} \triangleq \arg \min_{b_i \in \{\lfloor b_i^* \rfloor, \lceil b_i^* \rceil\}} |b_i - b_i^*|.$$

From (9) and $b_i \geq b_{\min}$, we can get the lower bound of the transmission power as

$$P_m(\{b_i\}) \geq \sum_{i=1}^{N_t} \frac{N_0 R_s}{\lambda_i^2} \frac{2(2^{b_i^{opt}} - 1)}{3} \ln \frac{2(1 - 2^{-b_{\min}/2})}{p_e}. \quad (18)$$

Thus, we can use (18) to lower bound the total power consumption of the transmitter.

Denote \hat{P}_m^{opt} and \hat{P}_m^* as the total powers corresponding to the constellation size b_i^{opt} and b_i^* respectively. Define the relative error ε caused by approximating b_i^{opt} using b_i^* as

$$\varepsilon = \frac{\hat{P}_m^{opt} - \hat{P}_m^*}{\hat{P}_m^*}. \quad (19)$$

If the relative error ε exceeds a predefined tolerance level, e.g., 10^{-2} , we can further refine the solution by applying the exhaustive search method. As a consequence, the user determines the optimal power on each stream. So we can get the optimal transmission strategy for required p_e . The optimal b_i is determined by the singular value λ_i of the channel matrix \mathbf{H} and the SER p_e . It is clear that the packet delay is directly related to p_e ,

so as to affect the power consumption. In the next section, we will investigate the delay-aware energy efficiency for the data link layer packet.

IV. DELAY PERFORMANCE

In this section, we investigate how delay affects the optimal energy efficiency in details. The energy efficient constellation size obtained in the previous section will be used to derive the delay-aware optimal energy efficiency.

A. Throughput and Energy-Efficiency

Assume each packet contains L bits. Define the transmission time per packet as

$$t_L = \frac{L}{R_s \sum_{i=1}^{N_t} b_i}. \quad (20)$$

Since not all the transmitted data in the packet are information bits, we define the effective throughput T as the payload information that can be correctly received per second as [22]–[24]:

$$T = \frac{L_p p_s}{t_L} = \frac{L - L_h}{L} \sum_{i=1}^{N_t} b_i R_s p_s. \quad (21)$$

To facilitate the analysis of packet throughput, the probability of successful packet transmission at the link layer, denoted as p_s , needs to be derived. Based on the relationship of the packet and the symbol, p_s can be expressed as a function of the SER p_e for each data stream.

The packet can be divided into N_t streams to be transmitted by the physical layer and the number of information bits for each stream linearly scales with the data rate. Then we can get the number of information bits from each stream in a packet as:

$$L_i = \frac{b_i L}{\sum_{i=1}^{N_t} b_i}.$$

So the number of symbols for each stream in a packet is L_i/b_i . With the assumption of the same SER, we can get

$$p_s = \prod_{i=1}^{N_t} (1 - p_e)^{\frac{L_i}{b_i}} = (1 - p_e)^{N_t L / \sum_{i=1}^{N_t} b_i}. \quad (22)$$

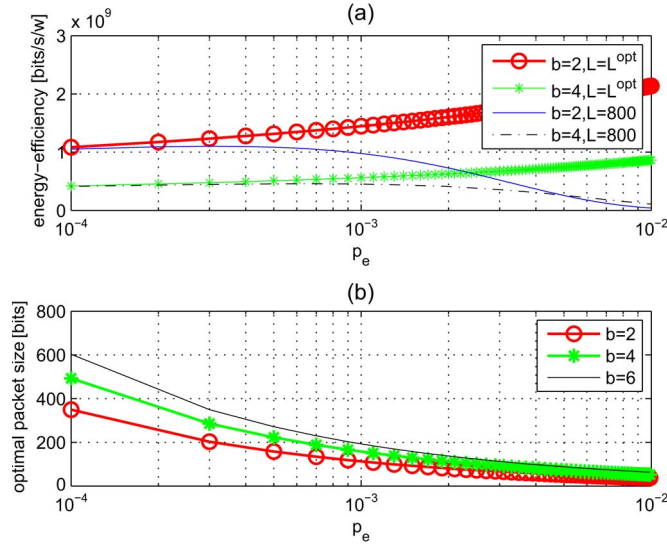
Now we come to define the energy-efficiency as

$$f_{ee} \triangleq \frac{T}{P_t + P_c}, \quad (23)$$

where P_t is the total transmission power as in (9) and P_c is the circuit power at the transmitter as in (10).

Denote P_m^{opt} as the total transmission power corresponding to the constellation size b_i^{opt} . We have the data rate $\sum_{i=1}^{N_t} b_i = b$, and assume that L can take continuous value. So differentiating (23) with respect to L leads to:

$$\begin{aligned} \frac{\partial f_{ee}}{\partial L} = f(b_i^{opt}, p_e) & \left[\frac{L_h}{L^2} (1 - p_e)^{\frac{L}{b}} \right. \\ & \left. + \frac{1}{b} \left(1 - \frac{L_h}{L} \right) (1 - p_e)^{\frac{L}{b}} \ln(1 - p_e) \right], \end{aligned} \quad (24)$$

Fig. 2. The header information $L_h = 32$ bits.

where $f(b_i^{opt}, p_e) = [P_m^{opt} + \sum_{i=1}^{N_t} (\alpha b_i^{opt} + P_0)]^{-1}$. By setting (24) to zero, we find that there is only one real solution. Thus we obtain the closed form solution of the optimal packet size expression as:

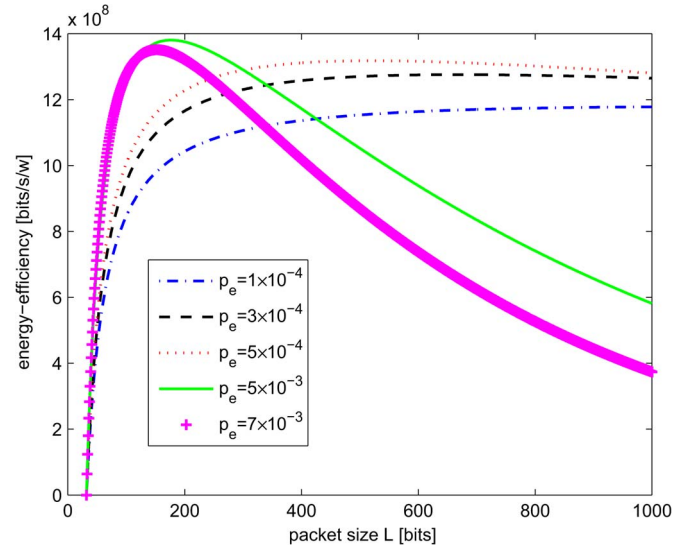
$$L^* = \frac{L_h + \sqrt{L_h^2 - \frac{4bL_h}{\ln(1-p_e)}}}{2}. \quad (25)$$

This stationary point is a maximum follows from the second order derivative. Since the packet size is an integer number, we can take the optimal packet size as

$$L^{opt} \triangleq \arg \min_{L \in \{[L^*], [L^*]\}} |L - L^*|.$$

It can be seen that when the required SER p_e and the data rate b are determined, the optimal packet size L^{opt} will be determined. Fig. 2 shows the energy efficiency and the optimal packet size versus the SER in the scenario of single transmitter and single receiver, which demonstrates that the optimal packet size can improve the energy efficiency greatly compared with the constant packet size. However, this doesn't mean that the packet size should always be variable. Since additional overhead and resource management costs are involved for the variable packet size, we should therefore determine the optimal fixed packet size based on parameter available at the time of design.

In Fig. 3, we can see that in the low SER regime, i.e., $p_e \leq 5 \times 10^{-4}$, there is no great difference in energy efficiency for different SER when using a fixed optimal packet size instead of the optimal packet size. Meanwhile, in the high SER regime, i.e., $p_e \geq 5 \times 10^{-3}$, there is again no big difference in energy efficiency for different SER when using another fixed optimal packet size instead of the optimal packet size. With the fixed optimal packet size, we can still achieve the near optimal energy efficiency. Thus, we assume that while the receiver knows the SER during one time block, it will control the transmitter to set the optimal packet size fixed over a long time horizon, which is practical and convenient to derive the delay in the next subsection.

Fig. 3. Energy-efficiency versus the packet size L , with $L_h = 32$ bits and $b = 4$.

B. Delay Analysis

In M/G/1 queue model, the packet service time S_T has the following probability mass function:

$$P\{S_T = n\tau\} = p_s(1 - p_s)^{n-1}, \quad n = 1, 2, \dots, \infty, \quad (26)$$

where τ represents the packet transmission time when the queue is serving one packet in one time slot, which is given by

$$\tau = \frac{L}{bR_s}. \quad (27)$$

From (26), we can get the mean service time:

$$\begin{aligned} E\{S_T\} &= \sum_{n=1}^{\infty} n\tau p_s(1 - p_s)^{n-1} \\ &= \frac{\tau}{p_s}. \end{aligned} \quad (28)$$

From (27) and (28), the service rate μ is given by:

$$\mu = \frac{1}{E\{S_T\}} = \frac{p_s}{\tau} = \frac{bR_s p_s}{L}. \quad (29)$$

Note that the service rate μ is a constant regardless of the packet number Q in the buffer. Let r_Q and μ_Q be respectively the generation rate from source and the service rate when there are Q packets in the buffer. Thus, the M/G/1 queue is a birth-death process with $r_Q = r(Q \geq 0)$ and $\mu_Q = \mu(Q \geq 0)$. By [25], using the Pollaczek-Khintchine formula, we can get the mean queue length as

$$\bar{Q}_q = \frac{r^2 E\{S_T^2\}}{2(1 - \delta)}, \quad (30)$$

where $\delta = r/\mu$ is the traffic intensity or utilization, and $E\{S_T^2\}$ is the expectation of the second moment of the service distribution. Using (26), we can get

$$E\{S_T^2\} = \frac{2\tau^2}{p_s^2} - \frac{\tau^2}{p_s}. \quad (31)$$

It is known that for an M/G/1 queue the average waiting time of a packet consists of queuing time and service time, and the queuing delay is

$$\bar{D}_q = \frac{\bar{Q}_q}{r}, \quad (32)$$

which is akin to Little's formula [25]. In summary, with (28), (30) and (32), the whole delay for a packet is given by

$$\bar{D} = \bar{D}_q + E\{S_T\} = \frac{\bar{Q}_q}{r} + \frac{1}{\mu} = \frac{2bR_sL - rL^2}{2b^2R_s^2p_s - 2rbR_sL}. \quad (33)$$

C. Adaptive MIMO/SIMO Transmission Mode

So far, we have derived the optimal constellation size for energy efficiency in a MIMO system and the packet delay for given p_e . Since p_e is related to the numbers of the transmission for a packet, which is proportional to the delay, different delay requirements will result in different values of p_e , which will create different transmission power consumption P_t by (9).

Since the total circuit power P_c for SIMO and MIMO are different, the transmission power P_t and the circuit power P_c can alternately become the dominant part of the total power consumption at particular delay performance for SIMO and MIMO. In the following, we will study the switching between the SIMO and MIMO systems by the energy efficiency criteria.

We consider the SIMO systems by performing antenna selection on the user's antennas in the MIMO systems. Let h_{ij} be the channel fading coefficient from the j -th transmit antenna to the i -th receive antenna. Then the best channel gain is chosen as

$$g_{SIMO} = \max_{j \in \{1, \dots, N_t\}} \mathbf{h}_j^H \mathbf{h}_j = \max_{j \in \{1, \dots, N_t\}} \sum_{i=1}^{N_r} |h_{ij}|^2, \quad (34)$$

where \mathbf{h}_j is the j -th column vector of \mathbf{H} .

By [26], we can get the SNR of the SIMO systems as

$$SNR_s = \frac{P_s^{opt} g_{SIMO}}{N_0 B},$$

where P_s^{opt} is the optimal transmission power in the SIMO systems. From (5) and (7), the transmission power for the user is

$$P_s^{opt} = \frac{N_0 R_s}{g_{SIMO}} \times \frac{2(2^b - 1)}{3} \ln \frac{2(1 - 2^{-b/2})}{p_e}, \quad (35)$$

and the total power consumption for SIMO systems is

$$\hat{P}_s^{opt} = P_s^{opt} + \alpha b + P_0. \quad (36)$$

To select the transmission mode with the maximum energy efficiency, we only need to select the transmission mode which consumes less power at the same throughput, that can be denoted as

$$t^* = \arg \min_{t \in \{m, s\}} \hat{P}_t^{opt}, \quad (37)$$

where m and s stand for MIMO and SIMO modes respectively. Therefore, we can adaptively change the transmission mode to

meet the user's QoS of delay, and allocate the transmission power to ensure the optimal energy efficiency of the user at the same time. In the simulations, we will show the energy efficiency performance of MIMO and SIMO for different QoS of delay.

Proposition 1: There exists a delay threshold

$$\bar{D}^* = \frac{2bR_sL - rL^2}{2b^2R_s^2p_s^* - 2rbR_sL}$$

between the optimal MIMO transmission mode and optimal SIMO transmission mode, making $f_{ee}^s = f_{ee}^m$, which is a crossover point, where f_{ee}^s and f_{ee}^m are the energy efficiency for the SIMO and MIMO respectively. When the user's service is delay sensitive sessions, we will choose MIMO transmission mode, otherwise, we choose SIMO transmission mode, from which we can get the optimal energy efficiency.

Proof: From (9), when the SER p_e is small, we know that the transmission power P_t will be large and dominates the total power consumption, since $P_t \rightarrow \infty$ as $p_e \rightarrow 0$. Hence the circuit power is negligible compared to the transmission power and $\hat{P}_m \approx P_m(\{b_i\})$ and $\hat{P}_s^{opt} \approx P_s^{opt}$, and the transmission SNR is high. By (5) and (7), for a particular p_e and the same symbol transmission, we assume that the MIMO systems transmit the same copy of the symbol s_s per antenna. Then we have

$$\gamma_s^s = \gamma_s^m = \text{SNR} \frac{B}{R_s}.$$

Thus, the received SNR of the SIMO, SNR_s , and the received SNR of MIMO, SNR_m have the relation of $SNR_s = SNR_m$. From (3), we can get

$$\hat{\mathbf{s}} = \sqrt{P_m} \Lambda \tilde{\mathbf{s}} + \mathbf{U}^H \mathbf{n},$$

where $\tilde{\mathbf{s}} = \underbrace{[s_s, s_s, \dots, s_s]^T}_{N_t}$ and P_m is the transmission power for

the transmission symbol s_s with constellation size $b = \sum_{i=1}^{N_t} b_i$. Therefore

$$SNR_m = \frac{P_m \text{Tr}\{\Lambda \Lambda^H\}}{N_0 B} = \frac{P_m \|\mathbf{H}\|_F^2}{N_0 B}.$$

This is equivalent to the SNR of the symbol transmission with the space-time block coding [27].

For the channel \mathbf{H} , while we select the transmit antenna with best channel gain to the receive antennas from (34), we can get

$$SNR_s = \frac{P_s^{opt} g_{SIMO}}{N_0 B}.$$

Therefore, we have $P_m = \frac{SNR_m N_0 B}{\|\mathbf{H}\|_F^2}$, and $P_s^{opt} = \frac{SNR_s N_0 B}{g_{SIMO}}$. Note that

$$g_{SIMO} = \max_{j \in \{1, \dots, N_t\}} \sum_{i=1}^{N_r} |h_{ij}|^2,$$

and

$$\|\mathbf{H}\|_F^2 = \sum_{i=1}^{N_r} \lambda_i^2 = \sum_{i=1}^{N_r} \sum_{j=1}^{N_t} |h_{ij}|^2,$$

we have $\|\mathbf{H}\|_F^2 > g_{SIMO}$, which results in $P_m < P_s^{opt}$ with $SNR_s = SNR_m$. And in the high SNR regime, in comparison with multiplexing and space time block coding, we can get $P_m(\{b_i^{opt}\}) \leq P_m$ at the same transmission rate [27], then we have $P_m(\{b_i^{opt}\}) < P_s^{opt}$. Therefore

$$f_{ee}^s \approx T/P_s^{opt} < T/P_m(\{b_i^{opt}\}) \approx f_{ee}^m.$$

This shows that MIMO mode outperforms the SIMO mode in terms of energy efficiency for fixed rate b , that is the optimal transmission mode $t^* = m$, where m stands for the MIMO mode.

On the other hand, when p_e is close to 1, the circuit power P_c will dominate the total power consumption, that is $\hat{P}_m^{opt} \approx P_c^m$ and $\hat{P}_s^{opt} \approx P_c^s$. Since

$$P_c^m = \sum_{i=1}^{N_t} \alpha b_i + N_t P_0 > \sum_{i=1}^{N_t} \alpha b_i + P_0 = P_c^s,$$

we have

$$f_{ee}^s \approx T/P_c^s > T/P_c^m \approx f_{ee}^m.$$

This shows that the SIMO transmission mode can be selected to improve the energy efficiency, resulting in the optimal transmission mode $t^* = s$, where s stands for SIMO mode.

By (33), the average delay \bar{D} is a function of the successful transmission probability p_s . Since $p_s = (1 - p_e)^{N_t L/b}$, therefore, \bar{D} is a function of p_e . Since $f_{ee}^s < f_{ee}^m$ for small p_e and $f_{ee}^s > f_{ee}^m$ for large p_e , there must exist a p_e^* such that $f_{ee}^s = f_{ee}^m$. This means that

$$\bar{D}^* = \frac{2bR_s L - rL^2}{2b^2 R_s^2 p_s^* - 2rbR_s L}$$

is the delay threshold value for selecting the transmission mode, where $p_s^* = (1 - p_e^*)^{N_t L/b}$. ■

Proposition 1 suggests that if the user has the delay tolerant service, the optimal SIMO transmission mode can be superior to the optimal MIMO transmission mode, otherwise the optimal MIMO mode is better. As a result, the tradeoff between the delay and the optimal energy efficiency is shown. Note that Proposition 1 is derived under the condition $r < \mu$, which ensures a finite average queue length.

Another important issue is that the optimization of energy efficiency may be influenced by user's congestion, which can make the optimal solution not satisfy the delay demand. In the next subsection, the virtual queue with congestion control algorithm is proposed to reduce the queuing delay. Meanwhile, the service rate is preserved, so as to keep the power consumption at low level and make the optimal transmission efficient and robust simultaneously.

D. Buffer Overflow

In the previous subsection, we have defined the delay of the packet transmission, and the user can choose the energy efficient transmission mode based on the delay sensitive and delay tolerant services. This is particular useful for applications where delay is an important measure of the quality of service. In the case of serving the delay tolerant applications, however,

the SER makes the service rate low. Furthermore, in the practical systems, the low service rate and the time varying data generation rate from upper layer may generate congestion, that is, buffer overflow, i.e., $r_Q \geq \mu_Q$. Thus, we should optimize the delay performance to reduce the congestion influence.

The user's queue model and congestion control algorithm will affect the transmission time of the packet [16], and make the queue length stable. We consider that the queue length of the user at the beginning of the time slot t is $Q(t)$, the arrival process $\{r(t)\}$ is assumed to be independent and identically distributed from one time slot to another. Therefore, We can get the evolution of the queue length in the buffer as

$$Q(t+1) = (Q(t) + r(t) - \mu(t))^+, \quad (38)$$

where $(x)^+$ is equal to x when $x > 0$, and zero otherwise.

1) *Congestion Control Algorithm*: For the user, given its current queue length $Q(t)$ at time slot t , the data generation rate as $r(t)$, and the threshold of the queue length for the congestion control signal as Q_0 , the congestion control algorithm is shown in Algorithm 1.

Algorithm 1 Congestion Control Algorithm

- 1: initialize the maximize transmission rate r_s and the arrival rate bound M ;
 - 2: **while** $Q(t) \geq Q_0$ **do**
 - 3: select the positive constant K ;
 - 4: **if** $\frac{K}{Q(t)^m} \leq M$ **then**
 - 5: $r(t) = \frac{K}{Q(t)^m}$;
 - 6: **else**
 - 7: $r(t) = M$;
 - 8: **end if**
 - 9: at the same time, compute the maximal service rate $bR_s p_s / L$ from (29);
 - 10: **if** $Q(t)r_s \leq \frac{bR_s p_s}{L}$ **then**
 - 11: $\mu(t) = Q(t)r_s$;
 - 12: **else**
 - 13: $\mu(t) = \frac{bR_s p_s}{L}$;
 - 14: **end if**
 - 15: **end while**
-

In the congestion control algorithm, m and K are positive constants assuming that the utility function of user is $K \frac{r^{1-m}}{1-m}$ when it transmits at rate r , and r_s is the maximum transmission rate for the user at particular channel state [16]. Here we consider that the rate r_s is constant. With the congestion control algorithm, the mean queue length can eventually reach stable, and the packet arrival rate with the service rate are scheduled by the controller, and we will see the convergence property of them in the simulation part. To accomplish optimal delay at the same μ , we can design an algorithm to reduce the mean queue length. We will see the superiority of energy efficiency in the next part.

2) *Virtual Queue*: By now, we have solved the problem of buffer overflow. In many cases, we want to make the delay below a certain value while keeping the energy efficiency. We can solve this problem by implementing the user's scheduler

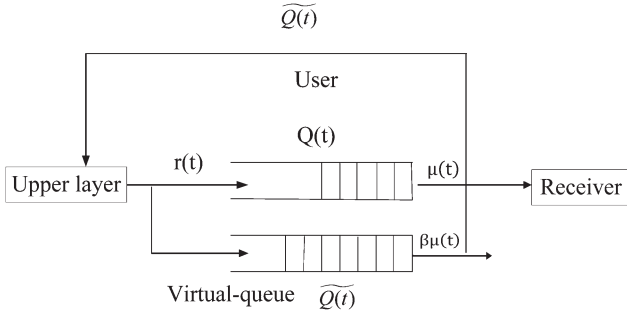


Fig. 4. Virtual queue model at the user.

using virtual queue [28]–[30], which is shown in Fig. 4, and can be described into three steps:

- Step 1: For each flow i , the user maintains a counter called the virtual queue. The virtual queue of flow i will count the length of the virtual queue, and the arrival rate $r(t)$ of the virtual queue is as that of the flow i .
- Step 2: The service rate $\mu(t)$ is always a fixed fraction $\beta < 1$ of the actual service rate $\mu_i(t)$ of flow i , that is $\mu(t) = \beta\mu_i(t)$. However, the control signal which feedbacks to data generation source is the virtual queue length $\tilde{Q}_i(t)$.
- Step 3: When the virtual queue is evolved to the case of $\tilde{Q}_i(t) \geq Q_0$, the congestion control algorithm is scheduled.

So we can get the result that the virtual queue length $\tilde{Q}_i(t)$ for flow i will always be larger than the actual queue length $Q_i(t)$ for the reason of virtual service rate. The upper layer of the user will reduce its arrival rate r_i before its actual queue being busy. So the congestion is virtual, making the stable queue length shorter. Choosing β close to 1, we can sacrifice throughput minimally, however, the queue-length can be dramatically reduced.

3) *Delay Analysis*: Note that the total packet delay includes queueing delay and transmission delay. By the congestion control algorithm, we can reduce the average queue length, but the mean service rate for the user does not decrease significantly. Define the mean queueing delay of the virtual queue with congestion control as \bar{D}_q^v . We can get the stable queue length of the virtual queue with congestion control as $\bar{Q}_q^v = r\bar{D}_q^v$. Therefore, the total packet delay \bar{D}_v of the virtual queue with congestion control is given by:

$$\bar{D}_v = E\{S_T\} + \bar{D}_q^v = \frac{1}{\mu} + \frac{\bar{Q}_q^v}{r}. \quad (39)$$

The congestion control and virtual queue algorithm can keep the queue length finite and low. Let \bar{Q}_q be the average queue length of the actual queue with congestion control. Then, we can get $\bar{Q}_q^v < \bar{Q}_q$. Let \bar{D}_q be the queueing delay of the actual queue with congestion control. Then $\bar{D}_q^v < \bar{D}_q$, which makes a shorter delay with the same successfully received packet probability. Under the condition of the same delay demand, from (22) and (9), we can get the relationship $P_{t^*}^{opt} > P_v$, where P_v is the transmission power of the optimal delay performance of virtual queue, which makes f_{ee}^* perform best at particular delay D . Then, we can get the optimal transmission mode $t^* = \arg \max_{t \in \{m, s\}} f_{ee}^t$ from the delay optimization algorithm, which makes \bar{D}_v short and stable.

TABLE I
PARAMETERS

Number of transmit antennas N_t	2
Number of receive antennas N_r	4
Circuit power P_0	60mW
Power spectral density of noise	-90dBm/Hz
Total transmission rate b	8bits/s/Hz
Minimum constellation size b_{min}	2bits/s/Hz
Packet size L	320bits
α	25mw/bit/s/Hz

Under the congestion condition, if we do not consider the congestion control algorithm, it will cause the increasing of energy consumption, and reducing the energy efficiency. However, the congestion control can improve the energy efficiency for a specific delay, making the optimization robust. In summary, the proposed algorithm can make the optimization practical for the buffer overflow and energy efficient for delay tolerant sessions, which will be demonstrated by simulations in the next section. The delay that we analyze in this paper is the average value. If we want to analyze the instantaneous delay alternatively, we can use discrete time queueing analysis in [31], which can derive the probability generating function for the packet delay. In general, obtaining a closed form expression for the instantaneous delay probability distribution is still difficult. Therefore, it is difficult to analyze the optimal energy efficiency in such case.

V. NUMERICAL RESULTS

This section presents simulation results to evaluate the adaptive transmission strategy and the proposed congestion control algorithm in energy efficient communications. We consider the uplink system with a total bandwidth $B = 5$ MHz, and the symbol rate $R_s = B$. The user's upper layer packets arrival follows a Poisson process. The other parameters are listed in Table I.

A. Performance of the Optimization

Fig. 5 shows the energy efficiency as a function of the delay performance for different transmission modes. For the purpose of simplicity, we set $r/\mu = 499/500$. We can see that the optimal MIMO transmission can always offer better energy efficiency than that by allocating random rate b_i with different delay demands in MIMO transmission mode, and the optimal MIMO transmission is approaching the MIMO upper bound derived from (18). Furthermore, there exists a crossover point between the optimal MIMO and the optimal SIMO transmission, which coincides with the analytic results. When the delay requirement is lower than that corresponding to the crossover point, the optimal MIMO is superior in energy efficiency. Otherwise, the SIMO offers better energy efficiency. These results further indicate that the energy efficiency can be improved for non-real time sessions by turning off the antennas with low gain. In general, the service can be divided into two classes: delay tolerant and delay sensitive. Specifically, when the delay requirement is higher than 20 ms, we call the service be delay tolerant, otherwise, the service is delay sensitive. Therefore, we

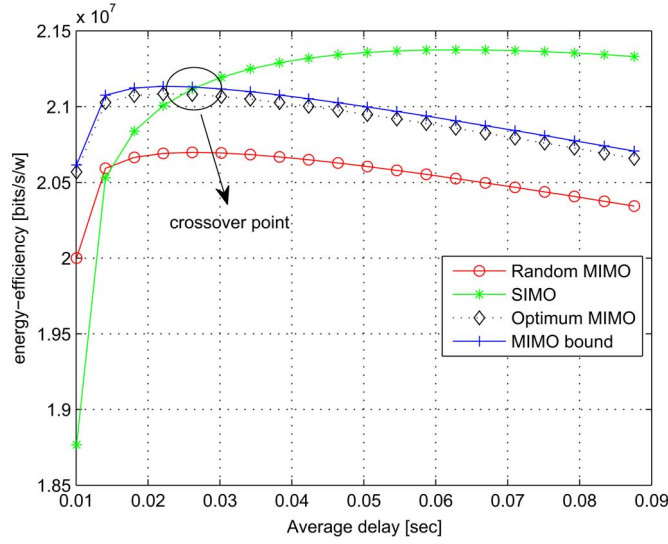


Fig. 5. Energy efficient transmission for MIMO and SIMO. $b = 8$ for the user, $\alpha = 25$ mw/bit/s/Hz, and $P_0 = 60$ mW.

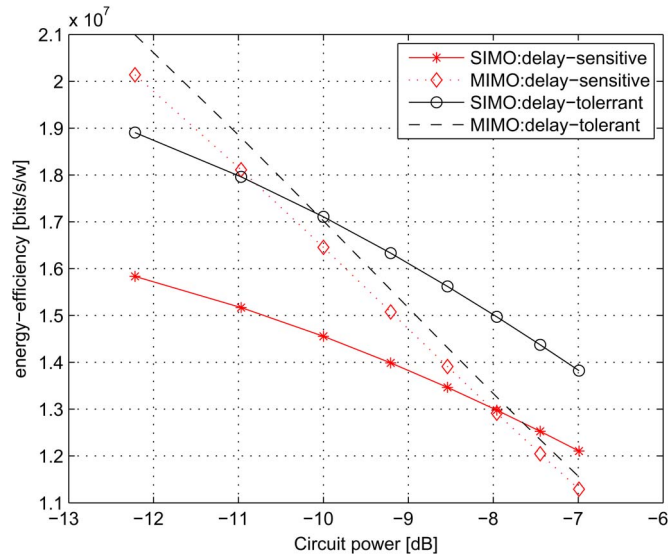


Fig. 6. Energy efficiency versus P_0 for optimal MIMO and SIMO. $b = 8$.

can choose the most energy efficient transmission mode for the upper layer service according to delay tolerant or sensitive.

Fig. 6 shows the energy efficiency versus P_0 . We observe that the energy efficiency decreases with P_0 increasing, due to more power consumption. We can see that the circuit power of the crossover point $f_{ee}^m = f_{ee}^s$ for the delay sensitive service is larger than that for the delay tolerant service. This can be explained by the fact that the transmission power is dominant in the delay sensitive service. Thus, we should increase P_0 to make $P_c = P_t$. Here, we set average delay $\bar{D} = 20$ ms for delay-sensitive service and $\bar{D} = 30$ ms for delay tolerant service. As expected, the delay tolerant service can be superior in energy efficiency.

B. Congestion in the Buffer

Specifically, to measure the fluctuation of the congestion control algorithm, we examine the variations of the service

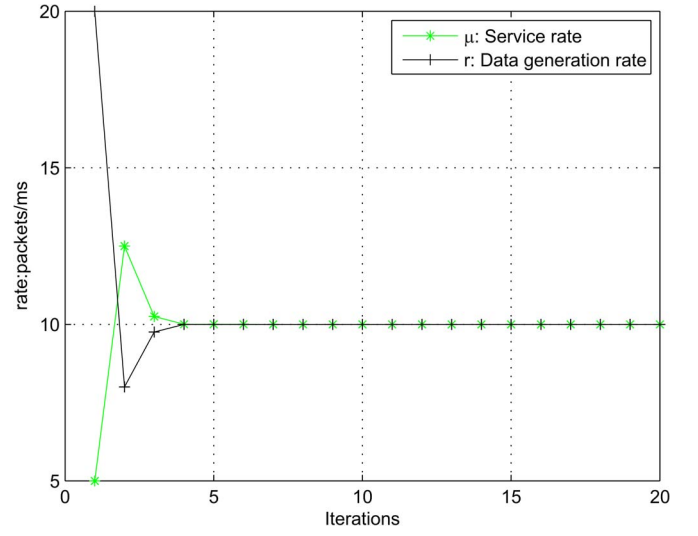


Fig. 7. Convergence properties of the service rate and data generation rate for the user with specific p_e .

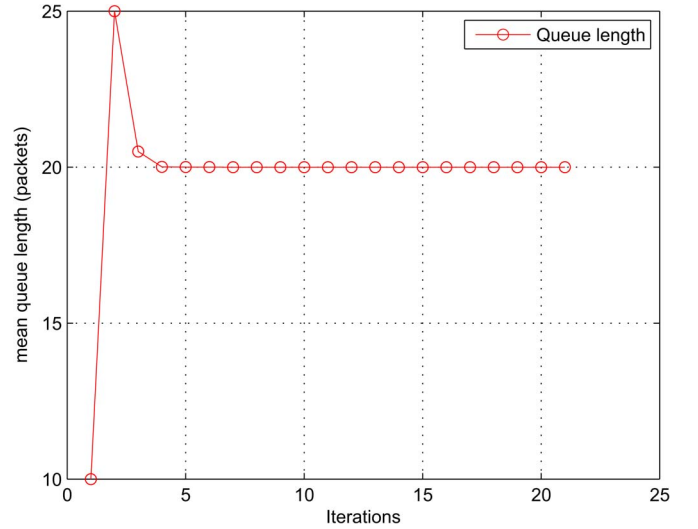


Fig. 8. Convergence property of queue length for the user with specific p_e .

rate and packet generation rate for specific p_e . With the buffer overflow, Fig. 7 shows the convergence properties of μ and r , both of which can be kept stable. Fig. 8 shows the average queue length, which converges to a stable value. In addition, to optimize the queuing delay, we can choose the virtual queue parameter β close to 1, from which the queue length can be decreased, but the service rate do not decrease. Simultaneously, the mean queuing delay can be computed from the mean queue length and the data generation rate from upper layer. Consequently, the throughput T does not change obviously with shorter delay. Hence, we can get the energy efficient gain with the same delay condition.

Fig. 9 shows the superiority of the energy efficiency with the delay optimization. Under the condition of congestion control, the adaptive SIMO/MIMO is the transmission strategy choosing the MIMO transmission mode or the SIMO transmission mode. We choose the optimal transmission mode t^* for different delay demands. Finally, we can see that with the same delay, the

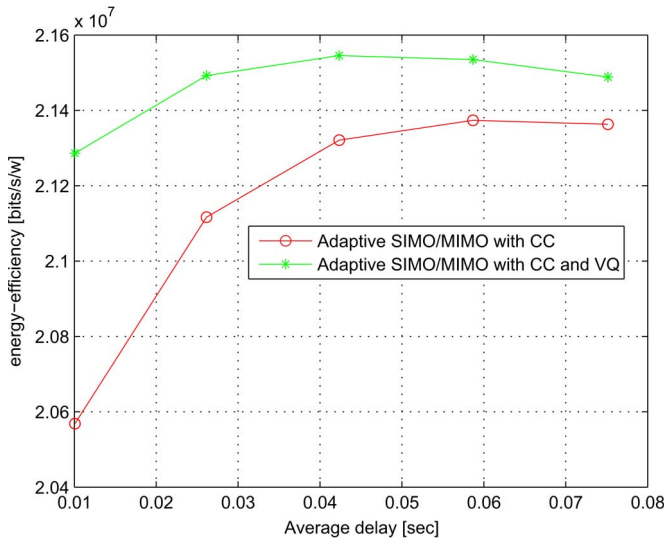


Fig. 9. Comparison of energy-efficiency with and without virtual queue model, when $b = 8$, where CC and VQ represent the congestion control algorithm and virtual queue algorithm respectively.

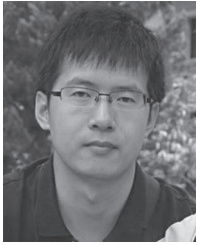
proposed algorithm can increase the energy efficiency greatly in different delay conditions compared with that without virtual queue. At the same time, we can see that the energy efficiency gap between the two curves is big when the delay is short, due to more power consumption when μ is increased.

VI. CONCLUSION

In this paper, we propose a cross-layer design model for the delay-aware energy efficiency. We first investigate the energy efficient communications with the effect of the constellation size in wireless physical layer. To achieve energy efficient communications, we optimize the constellation size for each stream, based on which, we study the delay aware energy efficiency for the upper layer service. We design the optimal packet size with the specific SER and derive the delay performance. By considering the circuit power consumption, we find the crossover point to select the energy efficient transmission mode between the optimal MIMO and the optimal SIMO for different delay-aware applications. In dealing with the user's buffer overflow from the burst service, we propose a congestion control with virtual queue algorithm to offset the influence, which can further enhance the energy efficiency in certain delay condition in comparison with that without virtual queue.

REFERENCES

- [1] T. Yoo, R. J. Lavery, A. Goldsmith, and D. J. Goodman, "Throughput optimization using adaptive techniques." [Online]. Available: <http://eeweb.poly.edu/dgoodman/icc04modulate.pdf>
- [2] G. Y. Li et al., "Energy-efficient wireless communications: Tutorial, survey, and open issues," *IEEE Wireless Commun. Mag.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.
- [3] S. Cui, A. J. Goldsmith, R. Madan, and S. Lall, "Cross-layer energy and delay optimization in small-scale sensor networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3688–3699, Oct. 2007.
- [4] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 74–85, Jan. 2004.
- [5] P. Letteri and M. B. Srivastava, "Adaptive frame length control for improving wireless link through, range, and energy efficiency," in *Proc. IEEE INFOCOM*, 1998, pp. 564–571.
- [6] C. K. Siew and D. J. Goodman, "Packet data transmission over mobile radio channels," *IEEE Trans. Veh. Technol.*, vol. 38, no. 2, pp. 95–101, May 1989.
- [7] Y. Sankarasubramanian, I. F. Akyildiz, and S. W. McLaughlin, "Energy efficiency based packet size optimization in wireless sensor networks," in *Proc. IEEE Workshop Sensor Netw. Protocols Appl.*, Anchorage, AK, USA, 2003, pp. 1–8.
- [8] M. C. Oto and O. B. Akan, "Energy-efficient packet size optimization for cognitive radio sensor networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1544–1553, Apr. 2012.
- [9] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 2349–2360, Sep. 2005.
- [10] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of mimo and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004.
- [11] G. Miao, "Energy-efficient uplink multi-user MIMO," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2302–2313, May 2013.
- [12] H. Kim, C. B. Chae, G. de Veciana, and R. W. Heath, "A cross-layer approach to energy efficiency for adaptive mimo systems exploiting spare capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4264–4275, Aug. 2009.
- [13] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.
- [14] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 125–144, Jan. 2004.
- [15] L. Musavian and T. Le-Ngoc, "Energy-efficient power allocation for delay-constrained systems," in *Proc. IEEE GLOBECOM*, Dec. 2012, pp. 3554–3559.
- [16] A. Eryilmaz and R. Srikant, "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.
- [17] R. Berry, "Power and delay trade-offs in fading channels," Ph.D. dissertation, MIT, Cambridge, MA, USA, Jun. 2000.
- [18] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [19] A. J. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [20] A. Mezghani and J. A. Nossek, "Power efficiency in communication systems from a circuit perspective," in *Proc. IEEE ISCAS*, May 2011, pp. 1896–1899.
- [21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [22] M. B. Purslet and J. M. Shea, "Adaptive nonuniform phase-shift-key modulation for multimedia traffic in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 8, pp. 1394–1407, Aug. 2000.
- [23] R. J. Lavery, "Throughput optimization for wireless data transmission," M.S. thesis, Telecommun. Netw., Polytechnic University, Brooklyn, NY, USA, Jun. 2001. [Online]. Available: <http://eeweb.poly.edu/dgoodman/lavery.pdf>
- [24] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Data throughputs using Multiple-Input Multiple-Output (MIMO) techniques in a noise-limited cellular environment," *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 226–235, Apr. 2002.
- [25] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of Queueing Theory*, 4th ed. Hoboken, NJ, USA: Wiley, 2008.
- [26] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [27] J. Huang and S. Signell, "On spectral efficiency of low complexity adaptive mimo systems in rayleigh fading channel," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4369–4374, Sep. 2009.
- [28] R. J. Gibbens and F. P. Kelly, "Resource pricing and the evolution of congestion control of congestion control," *Automatica*, vol. 35, no. 12, pp. 1969–1985, Dec. 1999.
- [29] S. Kunniyur and R. Srikant, "Analysis and design of an adaptive virtual queue algorithm for active queue management," in *Proc. ACM SIGCOMM*, Aug. 2001, pp. 123–134.
- [30] A. Lakshmikantha, C. Beck, and R. Srikant, "Robustness of real and virtual queue based active queue management schemes," *IEEE/ACM Trans. Netw.*, vol. 13, no. 1, pp. 81–93, Feb. 2005.
- [31] P. Gao, S. Wittevrongel, and H. Bruneel, "Discrete-time multiserver queues with geometric service times," *Comput. Oper. Res.*, vol. 31, no. 1, pp. 81–99, Jan. 2004.



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.



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 at 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, where he is also the Director of the Institute for Signal Processing and Systems. He is also serving the dean of school of Electronic Engineering and Automation, Guilin University of Electronic Technology. His interests cover network coding, cooperative communications, cognitive radio, and MIMO-OFDM systems.