2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP)

Energy-Efficient Cross-Layer Design of Delay-Aware MIMO Systems

Kunlun Wang, Wen Chen, and Meixia Tao
Department of Electronic Engineering, Shanghai Jiao Tong University, China
Emails: {kunlun1228; wenchen; mxtao}@sjtu.edu.cn

***Abstract*—In this paper, we propose a cross-layer design model for multiple input and multiple output (MIMO) cellular systems, to solve the problem of energy efficient communications with delay demand. We first investigate the energy efficient multiple quadrature amplitude modulation (MQAM) constellation size for each transmission stream. With the demand of the packet delay, then we propose an adaptive MIMO/SIMO transmission mode by exploiting the intrinsic relationship between the upper layer packet delay and the constellation size, the symbol error rate (SER) from the physical layer. Simulations show that in order to maximize the energy efficiency and offer different Quality of Service (QoS) of delay simultaneously, a user should adaptively choose the constellation 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; Cross-layer Design; MQAM Constellation Size; Delay.**

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 physical layer constellation size and the different delay demands from upper layer service can be combined to yield cross-layer design for energy efficiency.

When the packets generate 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 rate (SER), and small constellation size can reduce the transmit power. When the constellation size and the SER are given, the packet delay can be derived. Then, under different delay demand, different throughput and transmission power are required [1], which hence influences the energy efficiency. Motivated by these issues, we try to find the energy efficient communications for given delay demand.

Cross-layer design based on energy and delay optimization in sensor networks are considered in [2], [3], although multiple input and multiple output (MIMO) energy efficiency is not the main objective of these works. Similar cross-layer methodology can be applied. In [4], the influence of the constellation size to the energy efficiency of a MIMO system has been shown. However, the influences of the constellation size to the power consumption of the MIMO systems and the delay of the upper layer packet have not been considered there, although the offset QAM is often employed in [5] and the optimal power allocation has been studied in [6]. The delay demand can influence the energy efficiency [2], [7]. The existing works mainly focus on the tradeoff between the average delay and the transmission power, not considering the MIMO/SIMO mode switching, since MIMO systems are not always superior to the single input and multiple output (SIMO) systems due to different circuit power consumption [1]. Therefore, a better transmission mode from MIMO and SIMO can be chosen to improve the energy efficiency [4], [8]. In [8], 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.

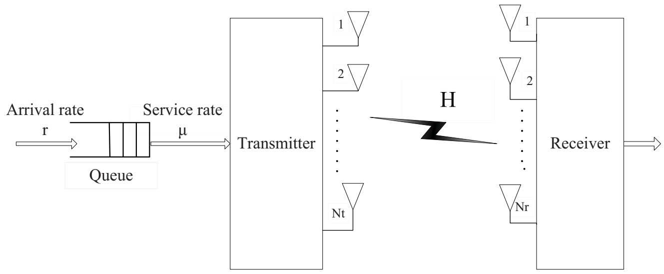

Fig. 1. System model.

In this paper, we formulate and solve an energy efficient cross-layer problem. In order to achieve the energy efficient transmission for the link layer packet, we optimize the constellation size for each symbol. Since SIMO systems may be more energy efficient than the optimal MIMO systems in some packet delay demand, we propose an adaptive MIMO/SIMO transmission strategy, and select the optimal antenna for SIMO mode. To our best knowledge, considering the cross-layer model 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.

II. System Model

A. Physical Layer Channel Model

Consider a MIMO system with the bandwidth of B Hz, 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 $\mathbf{s} = [s_1, s_2, \cdots, s_{N_t}]^T$ 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 $\mathsf{E}\left\{|s_i|^2\right\} = 1$. 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 $\mathbf{P} = \text{diag}\left\{\sqrt{p_{t1}}, \sqrt{p_{t2}}, \cdots, \sqrt{p_{tN_t}}\right\}$ 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 $\mathbf{y} = [y_1, y_2, \cdots, y_{N_r}]^T$, can be written as

$$\mathbf{y} = \mathbf{HQPs} + \mathbf{n}, \tag{1}$$

where $\mathbf{H} = N_r \times N_t$ is the channel matrix and $\mathbf{n} = [n_1, n_2, \cdots, n_{N_r}]^T$ 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. When the receiver correctly receives a packet, it will send back an ACK packet to the transmitter. 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 arrival at the transmitter in a first-in-first-out (FIFO) queue. 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 [9], 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 Minimization 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 [10]. In specific, let $\mathbf{H} = \mathbf{U\Lambda V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are respectively $N_r \times N_r$ and $N_t \times N_t$ unitary matrices, and

$$\mathbf{\Lambda} = \text{diag}\left[\lambda_1, \lambda_2, \cdots, \lambda_{N_t}\right], \tag{2}$$

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

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

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

$$\mathbf{\Gamma} \triangleq \left[\frac{p_{t1}\lambda_1^2}{N_0 B}, \frac{p_{t2}\lambda_2^2}{N_0 B}, \cdots, \frac{p_{tN_t}\lambda_{N_t}^2}{N_0 B}\right]. \tag{4}$$

We know that the SNR per symbol is

$$\gamma_{s_i} \triangleq \frac{p_{ti}\lambda_i^2}{N_0 B}\frac{B}{R_s} = \frac{p_{ti}\lambda_i^2}{N_0 R_s}, \tag{5}$$

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

It is well known that [11], the SER of MQAM modulation with size 2^{b_i}, is given by

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

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

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

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

$$p_{ti}(b_i) \approx \frac{N_0 R_s}{\lambda_i^2}\frac{2(2^{b_i} - 1)}{3}\ln\frac{2(1 - 2^{-b_i/2})}{p_e}, \tag{8}$$

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

$$P_m(\{b_i\}) \approx \sum_{i=1}^{N_t} \frac{N_0 R_s}{\lambda_i^2}\frac{2(2^{b_i} - 1)}{3}\left(\ln\frac{2}{p_e} + \ln(1 - 2^{-\frac{b_i}{2}})\right). \tag{9}$$

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 [12]. Thus

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

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

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

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 can get

$$\hat{P}_m \leq \sum_{i=1}^{N_t} \frac{N_0 R_s}{\lambda_i^2} \frac{2(2^{b_i}-1)}{3} \ln\frac{2}{p_e} + \alpha \sum_{i=1}^{N_t} b_i + N_t P_0. \quad (12)$$

Thus, From (12), we can formulate the optimization problem to derive the optimal constellation size as

$$\begin{array}{ll} \min & \sum_{i=1}^{N_t} \frac{N_0 R_s}{\lambda_i^2} \frac{2(2^{b_i}-1)}{3} \ln\frac{2}{p_e} + \alpha \sum_{i=1}^{N_t} b_i, \\ s.t. & b_i \geqslant b_{min}, \\ & \sum_{i=1}^{N_t} b_i = b. \end{array} \quad (13)$$

Note that the objective function in problem (13) 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 (13). So the necessary and sufficient conditions for optimality are given by the Karush-Kuhn-Tucker (KKT) conditions [13],

$$\begin{cases} b_i^* \geq b_{min}, \\ \sum_{i=1}^{N_t} b_i^* = b, \\ \kappa_i^* \geq 0, \\ \kappa_i^*(2 - b_i^*) = 0, \quad i = 1, \cdots, N_t, \\ \beta_i 2^{b_i^*} + \alpha - \kappa_i^* + \nu^* = 0, \quad i = 1, \cdots, N_t, \end{cases} \quad (14)$$

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 (14) to find b_i^*, κ_i^* and ν^*. Thus we have

$$b_i^* = \begin{cases} \log_2\left(\frac{-\alpha-\nu^*}{\beta_i}\right), & \nu^* \leq -\alpha - \beta_i 2^{b_{min}}, \\ b_{min}, & \nu^* > -\alpha - \beta_i 2^{b_{min}}. \end{cases} \quad (15)$$

Substituting (15) 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

$$P_m(\{b_i\}) \geq \sum_{i=1}^{N_t} \frac{N_0 R_s}{\lambda_i^2} \frac{2(2^{b_i^{opt}}-1)}{3} \ln\frac{2(1-2^{-b_{min}/2})}{p_e}. \quad (16)$$

Thus we can use (16) 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

$$\varepsilon = \frac{\hat{P}_m^{opt} - \hat{P}_m^*}{\hat{P}_m^*}. \quad (17)$$

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

We define the transmission throughput T as the information that can be correctly received per second as [14]:

$$T = R_s p_s \sum_{i=1}^{N_t} b_i, \quad (18)$$

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. Assume each packet contains L bits.

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

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

then we can get

$$p_s = \prod_{i=1}^{N_t} (1-p_e)^{\frac{L_i}{b_i}} = (1-p_e)^{N_t L / \sum_{i=1}^{N_t} b_i}. \quad (19)$$

Now we come to define the energy-efficiency as

$$f_{ee} \triangleq \frac{T}{P_t + P_c}, \quad (20)$$

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

B. Delay Analysis

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

$$P\{S_T = n\tau\} = p_s(1-p_s)^{n-1}, \quad n = 1, 2, \cdots, \infty, \tag{21}$$

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

$$\tau = \frac{L}{bR_s}. \tag{22}$$

From (21), we can get the mean service time:

$$\begin{aligned} \mathsf{E}\{S_T\} &= \sum_{n=1}^{\infty} n\tau p_s(1-p_s)^{n-1} \\ &= \frac{\tau}{p_s}. \end{aligned} \tag{23}$$

From (22) and (23), the service rate μ is given by:

$$\mu = \frac{1}{\mathsf{E}\{S_T\}} = \frac{p_s}{\tau} = \frac{bR_s p_s}{L}. \tag{24}$$

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 [15], using the Pollaczek-Khintchine formula, we can get the mean queue length as

$$\bar{Q}_q = \frac{r^2 \mathsf{E}\{S_T^2\}}{2(1-\delta)}, \tag{25}$$

where $\delta = r/\mu$ is the traffic intensity or utilization, and $\mathsf{E}\{S_T^2\}$ is the second moment of the service distribution. Using (21), we can get

$$\mathsf{E}\{S_T^2\} = \frac{2\tau^2}{p_s^2} - \frac{\tau^2}{p_s}.$$

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 $\bar{D}_q = \frac{\bar{Q}_q}{r}$, which is akin to Little's formula [15]. In summary, the whole delay for a packet is given by

$$\bar{D} = \frac{\bar{Q}_q}{r} + \mathsf{E}\{S_T\} = \frac{2bR_sL - rL^2}{2b^2R_s^2p_s - 2rbR_sL}. \tag{26}$$

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

$$g_{SIMO} = \max_{j\in\{1,\cdots,N_t\}} \mathbf{h}_j^H \mathbf{h}_j = \max_{j\in\{1,\cdots,N_t\}} \sum_{i=1}^{N_r} |h_{ij}|^2, \tag{27}$$

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

By [10], 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

$$P_s^{opt} = \frac{N_0 R_s}{g_{SIMO}} \times \frac{2(2^b-1)}{3} \ln \frac{2(1-2^{-b/2})}{p_e}, \tag{28}$$

and the total power consumption for SIMO systems is

$$\hat{P}_s^{opt} = P_s^{opt} + \alpha b + P_0. \tag{29}$$

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

$$t^* = \arg \min_{t\in\{m,s\}} \hat{P}_t^{opt}, \tag{30}$$

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 at the same time. In the simulations, we will show the energy efficiency performance of MIMO and SIMO for different delay demands.

Proposition 1. *There exists a delay threshold*

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

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 $P_t \to \infty$ as $p_e \to 0$. Hence the circuit power is negligible compared to the transmission power and $\hat{P}_m \approx P_m(\{b_i\})$ and $\hat{P}_s^{opt} \approx P_s^{opt}$. 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

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

where $\tilde{\mathbf{s}} = \underbrace{[s_s, s_s, \cdots, s_s]^T}_{N_t}$ 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

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

This is equivalent to the SNR of the symbol transmission with the space-time block coding [16].

For the channel $\mathbf{H}$, while we select the transmit antenna with best channel gain to the receive antennas from (27), we can get

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

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

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

and

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

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 [16], 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 (26), 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 $p_e^\star$ such that $f_{ee}^s = f_{ee}^m$. This means that

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

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

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.

V. Numerical Results

This section presents simulation results to evaluate the delay-aware energy efficient communications with cross-layer design. We consider the uplink system with a total bandwidth B=5MHz, 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.

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

Fig. 2 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 (16). Furthermore, there exits a crossover point between the optimal MIMO and the optimal SIMO transmission, which is coincide 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 20ms, we call the service be delay tolerant, otherwise, the service is delay sensitive. Therefore we can choose the most energy efficient transmission mode for the upper layer service according to delay tolerant or sensitive.

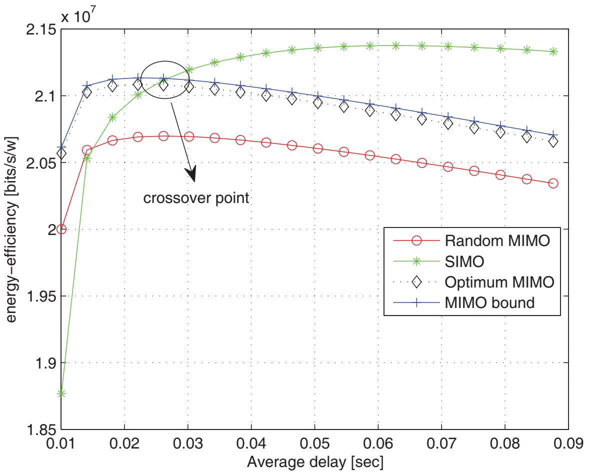

Fig. 2. Energy efficient transmission for MIMO and SIMO. b=8 for the transmitter, $\alpha = 25mw/bit/s/Hz$, and $P_0 = 60mW$.

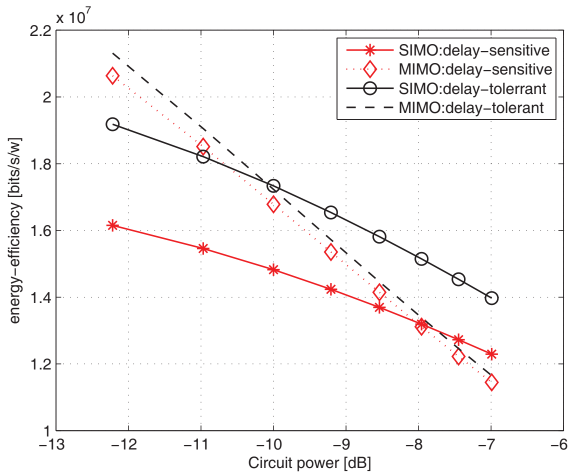

Fig. 3. Energy efficiency versus P_0 for optimal MIMO and SIMO. b=8.

Fig. 3 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 $f^m_{ee} = f^s_{ee}$ 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.

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. Secondly, we study the delay aware energy efficiency for the upper layer service. 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 at last.

Acknowledgment

This work is supported by the National 973 Project #2012CB316106, and by NSF China #61161130529.

References

[1] G. Y. Li, Z. Xu, C. Xiong, C. yang, S. Zhang, Y. Chen, and S. Xu, “Energy-efficient wireless communications: tutorial, survey, and open issues,” *IEEE Wireless Commun. Mag.*, vol. 18, no.6, pp. 28–35, Dec 2011.

[2] S. Cui, A. J. G. R. Madan, and S. Lall, “Cross-layer energy and delay optimization in small-scale sensor networks,” *IEEE Trans. Wireless Commun.*, vol. 6, pp. 3688–3699, Oct 2007.

[3] T. ElBatt and A. Ephremides, “Joint scheduling and power control for wireless ad hoc networks,” *IEEE Trans. Wireless Commun*, vol. 1, pp. 74–85, 2004.

[4] S. Cui, A. J. Goldsmith, and A. Bahai, “Energy-efficiency of mimo and cooperative mimo techniques in sensor networks,” *IEEE Jour. Select. Aeras in Commun*, vol. 22, pp. 1089–1098, Aug 2004.

[5] D. Darsena, G. Gelli, and F. Verde, “Widely-linear precoders and decoders for mimo channels,” in *Proc. of the Tenth International Symposium on Wireless Communication System(ISWCS), Ilmenau, Germany*, pp. 1–5, Aug. 2013.

[6] G. Miao, “Energy-efficient uplink multi-user mimo,” *IEEE Trans. Wireless Commun.*, vol. 12, no.5, pp. 2302–2313, May 2013.

[7] L. Musavian and T. Le-Ngoc, “Energy-efficient power allocation for delay-constrained systems,” in *Proc. of the IEEE GLOBECOM, Anaheim, CA*, pp. 3554–3559, Dec 2012.

[8] 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/ACM Trans. Networking*, vol. 8, pp. 4264–4275, August 2009.

[9] R. S. Atilla Eryilmaz, “Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control,” *IEEE/ACM Trans. Networking*, vol. 15, no.6, pp. 1333–1344, Dec 2007.

[10] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[11] A. J. Goldsmith, *Wireless Communications*. New York, NY: Cambridge University Press, 2005.

[12] A. Mezghani and J. A. Nossek, “Power efficiency in communication systems from a circuit perspective,” in *Proc. of the IEEE International Symposium on Circuit and Systems (ISCAS), Rio de Janeiro*, pp. 1896–1899, May 2011.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[14] 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, pp. 226–235, Apr 2002.

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

[16] 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.