

A Cross-Layer Approach to Energy Efficiency for Adaptive MIMO Systems Exploiting Spare Capacity

Hongseok Kim, *Student Member, IEEE*, Chan-Byoung Chae, *Student Member, IEEE*, Gustavo de Veciana, *Senior Member, IEEE*, Robert W. Heath, Jr., *Senior Member, IEEE*

Abstract

In this paper, we propose a mechanism to switch between multiple-input multiple-output (MIMO) with two transmit antennas and single-input multiple-output (SIMO) to conserve mobile terminals' energy. We focus on saving uplink RF transmission energy of mobile terminals in cellular systems supporting best effort traffic. The key idea is to judiciously slow down transmission rates when a base station is underutilized. We show that there exists a *crossover point* on the transmission rate below which SIMO consumes less power than MIMO when circuit power is factored. The crossover point is an increasing function of the circuit power, the number of receive antennas and channel correlation, all of which increase the potential energy savings resulting from mode switching. We propose an adaptive mode switching algorithm combined with rate selection to maintain a user's target throughput while achieving energy efficiency. Extensive flow-level simulations under dynamic loads confirm that the proposed technique can reduce the transmission energy by more than 50% and enables an effective tradeoff between file transfer delay and energy conservation.

Index Terms

MIMO, energy conservation, spare capacity, adaptive switching, cross-layer design

The authors are with the Wireless Networking and Communications Group (WNCG), the University of Texas at Austin, TX 78712 USA. Emails are {hkim4, cbchae, gustavo, rheath}@ece.utexas.edu. This work was presented in part at CISS 2008, Princeton, NJ. This work was supported in part by a gift from the Intel Research Council, NSF CNS-0721532, Samsung Electronics and NSF CCF-514194, CNS-626797.

I. INTRODUCTION

Energy efficiency is critical for mobile terminals supporting broadband connections such as WiMAX or 3GPP-LTE because high transmission rates are usually achieved at the expense of higher energy consumption. Out of the many key energy consuming components in mobile terminals such as the display or CPU, increasing attention is being paid to power consumption due to RF transmission – this is one of the main contributors to battery consumption, e.g., about 60% in time division multiple access (TDMA) phones [1]. In this paper we focus on reducing the uplink RF transmission energy of mobile terminals to extend the battery’s lifetime.

Fortunately, unlike voice service which requires a sustained constant bit rate, data services (e.g., uploading of files, pictures or emails) allow mobile terminals some latitude in exploiting delay-tolerance to save energy. Specifically, when the base station (BS) is underutilized¹, which is likely due to changing user populations and traffic loads, a simple way to save energy at mobile terminals (MTs) is to exploit spare capacity – i.e., slow down file transfers as long as the user-perceived performance is acceptable. Indeed, even though file transfer delays (or simply *delay* hereafter) would be prolonged, the transmission power drops sharply (e.g., Shannon’s result suggests an exponential drop with rate decreases) and thus the transmission energy – the product of power and delay – is reduced [2]. We refer to this as the *energy-delay tradeoff*.

The underlying character of this tradeoff changes when circuit power is taken into account [3], [4]. By circuit power we refer to power consumption in the RF transmission chain which can be assumed to be constant irrespective of the transmission rate. Thus, energy expenditures resulting from circuit power are proportional to transmission time. The impact of circuit power on energy efficiency is more critical when the MT has multiple transmit antennas because of the multiplicity of associated circuits such as mixers, synthesizers, digital-to-analog converters, filters, etc. The circuit power for a multiple-input multiple-output (MIMO) system is thus higher than that of a single-input multiple-output (SIMO) system by approximately N_t times where N_t is the number of transmit antennas. It is generally accepted that MIMO achieves better energy-efficiency than SIMO thanks to spatial multiplexing gain [5]. Altogether when circuit power is considered, however, MIMO may consume more power than SIMO at low spectral efficiency, thus circuit power hinders the use of MIMO on the uplink. This is one of the reasons why

¹Utilization is defined as the average fraction of time when the system is busy.

several emerging standards do not use MIMO for the uplink [6]. Mitigating the adverse impact of circuit power remains crucial to enabling the use of MIMO for the uplink .

Towards addressing the circuit power problem of MIMO systems, we identify a *crossover point* on the transmission rate (or spectral efficiency) below which SIMO is more energy efficient than MIMO. We will focus on the case where $N_t = 2$ at the MT, which is perhaps the most practical assumption at this point given the antenna configurations of the IEEE802.16m standard [7]. We propose an *adaptive* switching mechanism between MIMO and SIMO. The key idea is simple. When the system is underutilized, the MT operates with SIMO at low spectral efficiency to save energy, but when congested, the MT operates with MIMO at high spectral efficiency to increase throughput. This is done in an adaptive way considering two aspects – dynamic network traffic and channel variations. In determining the crossover point, the circuit power is the main factor, but we will see that two other factors, the number of receive antennas and channel correlation also increase the crossover point and make mode switching more beneficial.

Prior work on adaptive MIMO techniques [8], [9], [10] has not specifically addressed energy conservation. The authors in [9] proposed mode selection criteria to improve link level bit error rate (BER) performance for a fixed rate. To increase throughput, several adaptive MIMO and link adaptation techniques have been proposed [10], [11], [12], but the authors mainly focused on the physical layer. By contrast, our work is a *cross-layer* energy saving approach considering the role of *circuit power* at the circuit level, *multiple antennas* at the physical layer, and *dynamic user load* at the medium access control (MAC) layer and above.

One of the challenges in saving energy lies in the tradeoff between transmit energy and circuit energy; slowing down the transmission rate reduces transmit energy [2] but in turn increases circuit energy [4], [3], [13], [14]. Thus, the total energy consumption becomes a convex function of the transmission rate, and an energy-optimal transmission rate exists. Previous work towards achieving energy-optimal transmission, e.g., [13], is limited to physical layer modulation techniques with a single sender and receiver pair for sensor networks. The work in [15], [4] addresses multiple users including the MAC layer but only for a *fixed* number of users in a wireless local area network. Unlike previous work, our focus is on dynamic user populations in a *cellular* system where new file transfers are initiated at random and leave the system after being served – we refer to this as a system with “flow-level dynamics” [16]. Such dynamic models capture the characteristics of a practical system supporting data traffic, but in general are

hard to analyze and have not been studied as extensively as their static counterparts, i.e., with a fixed set of backlogged users.

Our contributions. This paper makes three main contributions. First, we propose a mechanism for adaptively switching between MIMO and SIMO to conserve mobile terminals' energy. In a practical MIMO system with two transmit antennas at the MT and many receive antennas at the BS, we demonstrate that adaptive switching can save uplink transmission energy by more than 50% as compared to MIMO without substantially changing user-perceived performance. In addition our asymptotic analysis shows that the crossover point scales as $O(\log_2 N_r)$ where N_r is the number of receive antennas at the BS, and thus increasing N_r may improve mode switching benefits.

Our second contribution is to show that mode switching benefits are more significant when channels are correlated. A closed form expression for the crossover point is provided as a function of channel correlation. If MIMO uses a zero forcing receiver, the benefit of mode switching further increases because SIMO is more robust in ill-conditioned channels.

Our third contribution lies in that we are the first to consider exploiting *dynamic spare capacity* to realize energy savings in MIMO systems. Dynamic spare capacity is available when the system is underutilized, occasionally, due to changes in user population and/or bursty traffic loads. Energy is saved by slowing down transmission rates when the system is underutilized. Circuit and idling power, however, deteriorate the energy saving benefit, and the total energy consumption may increase if the user's target throughput is too low. The proposed algorithm effectively avoids this problem by exploiting an energy-optimal transmission rate.

Discussion of the assumptions. Let us briefly discuss two of the underlying assumptions of this work.

1) *Are wireless BSs really underutilized?* One might argue that given the scarcity of spectrum wireless networks should not be underutilized. However as a result of time varying, non-stationary loads, or unpredictable bursty loads these networks are often overdesigned to be able to support a peak load condition, and so are often underutilized. For example Internet service providers' networks see a long term utilization as low as 20% [17]. Similarly a substantial fraction of Wi-Fi hotspot capacity is unused [18]. More generally, due to the high variations in capacity that a wireless access system can deliver to various locations in its coverage area, e.g., up to three

orders of magnitude difference, one can also expect high variability in the system load [19], [20]. Broadband cellular systems also have higher time variability in traffic loads and system capacity[20]. Thus, bursty, uncertain traffic loads and fluctuating capacity necessitate conservative design, so BSs are likely to continue to be underutilized, even with evolving pricing strategies.

2) *Are slow downs acceptable for energy savings?* One might argue that users may prefer file transfers be realized as quickly as possible rather than conserve energy. This may be true if saving energy compromises user-perceived performance. Specifically, for the downlink, fast transmission is important for users to be satisfied with say web browsing or file download applications. For the uplink, which is the main concern in this paper, however, uploading of files, e.g., pictures or emails, may be quite delay-tolerant and could be carried out as a *background* process after users click the ‘send button’.

The paper is organized as follows. In Section II we describe our system model and assumptions. Section III analyzes the impact of N_r and channel correlation on the crossover point. We address the dynamic user populations and develop a practical energy-efficient adaptive MIMO algorithm in Section IV. Section V provides simulation results followed by conclusion in Section VI.

II. SYSTEM MODEL

A. Assumptions

We consider a centralized wireless communication system with one BS serving multiple MTs. Target systems could be, but are not limited to, WiMAX or 3GPP-LTE. We assume that the system is based on MIMO and shared via time division multiple access (TDMA). Since energy savings are more important at the MTs than at the BS, we focus on uplink transmissions. Our work is, however, also applicable to saving downlink energy at the BS. We assume that the channels experience flat fading² and the dimension of channel matrix \mathbf{H} is $N_r \times N_t$ where N_r is the number of receive antennas at the BS and N_t is the number of transmit antennas at the MT. We focus on the case where $N_r \geq 2$ (at the BS) and $N_t = 2$ (at the MTs). The assumption of two transmit antennas at the MTs is in accordance with the antenna configurations of the IEEE802.16m standard [7].³ One might question that current practical systems only use single

²Flat fading can be obtained in practice using multiple input multiple output orthogonal frequency division multiplexing (MIMO-OFDM).

³Even though the 3GPP LTE considers two antennas at the MT but only one antenna is used for uplink.

transmit antenna on the uplink because of the implementation issues such as antenna spacing and circuit power [6]. In that case, our study becomes more meaningful because adaptive mode switching resolves the adverse impact of circuit power and justifies the use of two transmit antennas on the uplink.

We consider MIMO systems where the transmitter does not have channel state information (CSI), i.e., no instantaneous channel feedback. Nevertheless, we assume that the BS informs each MT of the appropriate transmission mode, either MIMO or SIMO, which requires 1 bit of feedback. In addition, in the case of SIMO, the BS informs the MT of the index for the antenna with the highest channel gain, which requires an additional 1 bit of feedback. Our focus is on delivering delay-tolerant (best effort) traffic.

B. Problem definition

The key questions addressed in our paper are 1) how to change transmission mode between MIMO and SIMO to save energy in a system supporting dynamic user populations (*mode switching*), and 2) how to determine the appropriate transmission rate considering circuit and idling power consumption⁴ as well as the average target throughput of each user (*rate selection*).

C. Transmission power models

Fig. 1 (redrawn from [3]) illustrates the transmission chain for MIMO. A key element of our work is to have a reasonably accurate transmission power model, which we discuss next.

1) *MIMO power model*: Assuming the MTs do not have access to CSI, we consider equal power allocation to each antenna, and thus do not consider precoding in this paper. Our work is also applicable to closed loop MIMO, but it is harder to derive a closed form expression for the transmission power as a function of the rate. Let ϕ_1 and ϕ_2 be the eigenvalues of $\mathbf{H}^*\mathbf{H}$ where \mathbf{H}^* is a complex conjugate of the channel matrix \mathbf{H} . Then, the achievable spectral efficiency of MIMO using equal power allocation into each antenna given \mathbf{H} is expressed as [5]

$$C = \sum_{k=1}^{N_t} \log_2 \left(1 + \frac{P_o}{N_0 N_t} \phi_k \right) \quad (1)$$

⁴The definition and the impact of idling power, which plays a crucial role in dynamic systems, will be addressed in Section IV in detail.

where $N_t = 2$ and P_o is the output power (from the power amplifier) that is dissipated into the air, and N_0 is the noise power. We assume that \mathbf{H} has a rank of $\min(N_t, N_r) = N_t$. Based on (1) and assuming that the power consumed by the power amplifiers is linearly dependent on the output power [3], the transmission power equation $f_m(r)$ for MIMO at transmission rate r with spectral bandwidth w and circuit power consumption $p_{dc,m}$ is given by

$$f_m(r) = \frac{1}{\eta} \frac{2}{\phi_1 \phi_2} N_0 \times \left(\sqrt{\left(\frac{\phi_1 + \phi_2}{2} \right)^2 + \phi_1 \phi_2 (2^{r/w} - 1)} - \frac{\phi_1 + \phi_2}{2} \right) + p_{dc,m} \quad (2)$$

where the subscript m stands for MIMO and η is the drain efficiency, which is defined as the ratio of the output power and the power consumed in the power amplifier. For simplicity, we assume that possible transmission rates are continuous.⁵

Note that (1) and (2) are based on an ideal MIMO receiver. As an example of a practical linear receiver, we here consider a zero forcing receiver which gives us analytical tractability. Then as shown in [21], under an independent coding assumption, (1) can be rewritten as,

$$C = \sum_{k=1}^{N_t} \log_2 \left(1 + \frac{P_o}{N_0 N_t} \frac{1}{[(\mathbf{H}^* \mathbf{H})^{-1}]_{k,k}} \right) \quad (3)$$

where $[(\mathbf{H}^* \mathbf{H})^{-1}]_{k,k}$ denotes k th diagonal element of $(\mathbf{H}^* \mathbf{H})^{-1}$. Thus, if ϕ_k is replaced by $1/[(\mathbf{H}^* \mathbf{H})^{-1}]_{k,k}$ in (2), we obtain a transmission power model of a MIMO with a zero forcing receiver. In the case of MIMO with a minimum mean square error (MMSE) receiver, we can obtain a transmission power model in a similar way, so for simplicity we only consider a zero forcing receiver.

In computing the circuit power, we assume that MIMO requires N_t transmission blocks, but that the frequency synthesizer, i.e., local oscillator (LO), is shared by multiple antennas [3], [19] as can be seen in Fig. 1. Then, the total circuit power consumption of MIMO is given by

$$p_{dc,m} = N_t (p_{dac} + p_{mix} + p_{filt}) + p_{syn} \quad (4)$$

where p_{dac} , p_{mix} , p_{filt} , p_{syn} stand for the power consumption from a digital-to-analog converter, a mixer, a filter, and a frequency synthesizer, respectively.

⁵For the discrete transmission rate, i.e., finite modulation order with BER constraint, see [3].

2) *SIMO power model*: Similarly, the transmission power $f_s(r)$ where the subscript s stands for SIMO, is given by

$$f_s(r) = \frac{1}{\eta} \frac{2^{r/w} - 1}{\sum_{k=1}^{N_r} |h_k|^2} N_0 + p_{\text{dc},s} \quad (5)$$

where h_1, \dots, h_{N_r} are channel coefficients and the circuit power for SIMO is given by $p_{\text{dc},s} = p_{\text{dac}} + p_{\text{mix}} + p_{\text{filt}} + p_{\text{syn}}$. The antenna with higher gain is selected using 1 bit antenna selection feedback in the case of SIMO.

III. ANALYSIS OF THE CROSSOVER POINT

A. Motivation for mode switching

Fig. 2 exhibits the transmission powers for both MIMO and SIMO in the case of Rayleigh fading channels. Note the crossover around $r/w = 3$ bps/Hz below which SIMO is more energy-efficient than MIMO. (This figure is an example of one realization for an uncorrelated Rayleigh fading channel – different realizations will give different results.) In addition, as an example of MIMO with a linear receiver, we plot the transmission power of MIMO with a zero forcing receiver. It is noticeable that the crossover point is higher than that for the ideal receiver, i.e., $r/w = 3.9$ bps/Hz. The crossover points exhibit the need for a smart switching policy between MIMO and SIMO considering the transmission rate, user-perceived throughput and energy efficiency.

B. The impact of the number of receive antennas on the crossover point

In Rayleigh fading channels with many receive antennas at the BS, the crossover point can be analytically computed. Let $\mathbf{H} = \mathbf{H}_w$ where elements of \mathbf{H}_w are independent complex Gaussian random variables with zero mean and variance g . When N_r goes to the infinity, $\frac{1}{N_r g} \mathbf{H}^* \mathbf{H}$ converges to an identity matrix by the law of large numbers, and thus two eigenvalues of $\mathbf{H}^* \mathbf{H}$ can be asymptotically given by $N_r g$. In this case, we have the following proposition.

Proposition 1: Assuming uncorrelated Rayleigh fading channels, when N_r is large enough (and N_t is 2), the energy efficiency crossover point between SIMO and MIMO scales with $O(\log_2 N_r)$ as N_r grows. When N_r goes to infinity, the ratio between the crossover point and the maximum achievable rate of MIMO converges to $\frac{1}{2}$.

Proof: Assuming large N_r and thus substituting $\phi_1 = \phi_2 = N_r g$ in (2), the power equation is given by

$$f_m(r) = \frac{N_0}{\eta g} \frac{2}{N_r} \left(2^{\frac{r}{2w}} - 1 \right) + p_{dc,m}. \quad (6)$$

Similarly, (5) can be rewritten as

$$f_s(r) = \frac{N_0}{\eta g} \frac{1}{N_r} \left(2^{\frac{r}{w}} - 1 \right) + p_{dc,s}. \quad (7)$$

The crossover point r^* satisfies $f_s(r^*) = f_m(r^*)$. Solving this equation gives

$$r^* = 2 \log_2 \left(1 + \sqrt{\Delta p_{dc} \frac{\eta N_r g}{N_0}} \right) w \quad (8)$$

where $\Delta p_{dc} = p_{dc,m} - p_{dc,s}$. Thus, r^* scales with $O(\log_2 N_r)$. The maximum transmission rate for MIMO is given by

$$r_{\max} = 2 \log_2 \left(1 + \frac{P_o}{2} \frac{N_r g}{N_0} \right) w. \quad (9)$$

Thus, the ratio of r^* and r_{\max} is

$$\frac{r^*}{r_{\max}} = \frac{\log_2 \left(1 + \sqrt{\Delta p_{dc} \frac{\eta N_r g}{N_0}} \right)}{\log_2 \left(1 + \frac{P_o}{2} \frac{N_r g}{N_0} \right)}. \quad (10)$$

Since $\Delta p_{dc} > 0$, as N_r goes to infinity, $\frac{r^*}{r_{\max}}$ converges to $\frac{1}{2}$. ■

Proposition 1 implies that as the number of receive antennas at the BS grows, the rate regime where SIMO is more energy-efficient than MIMO expands. This is because, as can be seen in (6) and (7), increasing N_r makes $f_m(r)$ and $f_s(r)$ grow more slowly in r , and thus the impact of circuit power becomes dominant, which makes SIMO more energy-efficient. Finally, if N_r is sufficiently large, the system operates at SIMO for the lower half of the feasible rates and then switched to MIMO for the higher half of the feasible rates.

1) Asymptotic analysis using flow-level dynamics: To better understand the impact of large N_r on energy efficiency, we consider a *dynamic* system where the number of ongoing users varies with time. For large N_r , the eigenvalues of $\mathbf{H}^* \mathbf{H}$ are approximately $N_r g$. Then, we perform stationary analysis as follows.

Users randomly arrive to the system according to a Poisson process and leave the system after finishing the file transfer. We are interested in the average energy consumption per file. To capture this, we use a *flow-level queuing model* [16], see Fig. 3. Flow-level analysis tracks the

arrival and departure process of users. We will assume that each user arrives with exactly one file and thus corresponds to a single flow. We refer to the number of flows in the system n as the system's state in the sequel.

Our objective is to minimize the average energy per file by switching between MIMO and SIMO transmission modes. We assume for simplicity that users have the same target throughput q . Then, the system capacity⁶ in state n is given by

$$c(n) = \min(nq, r_{\max}). \quad (11)$$

The system capacity increases linearly to satisfy the individual targets until the system is overloaded, i.e., $c(n) = r_{\max}$. Assuming a processor sharing scheduling discipline, if the system is not overloaded each user should see his target throughput q . This policy represents a simple approach towards exploiting *dynamic spare capacity* to conserve energy; when the system is congested, it operates at the maximum rate r_{\max} , however, when underutilized, the overall transmit power and the system capacity are reduced with n .

Given the above simple model for system capacity, we now obtain a Markov chain model for the number of ongoing flows in the system. We assume that the arrivals of file transfer requests follow an independent Poisson process with arrival rate λ and have independent file sizes with mean μ^{-1} . Let $\mathbf{N} = (N(u), u \geq 0)$ denote a random process representing the number of ongoing file transfers at time u . Then, if file sizes are exponentially distributed, \mathbf{N} is a Markov process with state space \mathbb{Z}^+ and rate matrix Q is given by

$$\begin{aligned} q(n, n+1) &= \lambda \\ q(n+1, n) &= \mu c(n+1) \quad \text{for } n \geq 0. \end{aligned}$$

The stationary distribution π , if it exists, is given by

$$\pi(n) = \pi(0) \frac{\rho^n}{\prod_{m=1}^n c(m)}, \quad (12)$$

where $\rho := \frac{\lambda}{\mu}$ is the traffic load (bits per second) and $\pi(0) = \left(1 + \sum_{n=1}^{\infty} \frac{\rho^n}{\prod_{m=1}^n c(m)}\right)^{-1}$. Note that the insensitivity property for processor sharing queue ensures this distribution also holds for general file size distributions. In the sequel we let N be a random variable with distribution π .

⁶System capacity $c(n)$ is not the same as the information theoretic capacity but implies the system throughput.

Let P be random variable denoting the stationary system power consumption. In steady state, the average system power consumption is given by $E[P] = \sum_{n=0}^{\infty} p(n)\pi(n)$ where $p(n)$ is a function which captures the overall system power expenditure in state n and given by

$$p(n) = \min [f_m(c(n)), f_s(c(n))]. \quad (13)$$

Note that, from the crossover point r^* in (8), if $nq \leq r^*$ then SIMO is more energy-efficient, and vice versa. Thus,

$$p(n) = \begin{cases} f_s(c(n)) & \text{if } n \leq n^* \\ f_m(c(n)) & \text{if } n > n^* \end{cases}, \quad (14)$$

where $n^* = \lfloor \frac{r^*}{q} \rfloor$. Thus if the number of users is small, e.g., less than or equal to n^* , then SIMO is selected, otherwise, MIMO.

Let J be a random variable denoting the energy consumed to serve a typical user's flow. Then, *energy-power equivalence* in a stationary system [22], which is akin to Little's result gives that

$$E[J] = \frac{1}{\lambda} E[P]. \quad (15)$$

Fig. 4 shows that $E[J]$ for a system with mode switching decreases faster than that of MIMO as N_r grows. This result can be also anticipated from (8), i.e., r^* scales with $O(\log_2 N_r)$. Although practical systems would not be able to employ a large number of antennas at the BS, we expect that our results provide an insight on the impact of receive antennas at the BS in designing practical systems, i.e., increasing N_r improves the energy-saving benefit of mode switching.

C. The impact of channel correlation on crossover point

Another important factor determining MIMO energy efficiency is spatial correlation among antennas. Since channel correlation degrades the capacity of MIMO systems [21], it further motivates mode switching. The rate regime where SIMO is more energy-efficient than MIMO expands further and thus mode switching becomes more plausible. To capture correlated channels, consider a channel model

$$\mathbf{H} = \mathbf{R}^{\frac{1}{2}} \mathbf{H}_w \quad (16)$$

where \mathbf{R} is an $N_r \times N_r$ receive correlation matrix whose elements are $[\mathbf{R}]_{i,j} = \xi^{|i-j|}$, $0 \leq \xi < 1$. This model is extensively used in the literature [21]. Since receive antennas at the BS do not

have rich scatters, it is reasonable to assume spatial correlation at receive antennas. By contrast, we assume that transmit antennas at the MT are not correlated for simplicity; otherwise, MIMO capacity is further degraded, and we have more chance to switch into SIMO. Even though in this paper we generally consider the case where $N_t = 2$ and $N_r \geq 2$, here we shall focus on the case where $N_t = N_r = 2$ to allow simple analysis. In this case, we can explicitly compute the crossover point in a high SNR regime. This result is given in the following proposition.

Proposition 2: Assuming high received SNR, when a correlated channel matrix \mathbf{H} is given by (16) with $N_t = N_r = 2$, the crossover point is explicitly given by

$$r^* \approx 2 \log_2 \left(\frac{\varphi}{\sqrt{\phi_{w1}\phi_{w2}(1-\xi^2)}} + \sqrt{\frac{\varphi^2}{\phi_{w1}\phi_{w2}(1-\xi^2)} + \Delta p_{dc} \frac{\eta\varphi g}{N_0}} \right) w \quad (17)$$

where ξ is the correlation coefficient, ϕ_{w1} and ϕ_{w2} are the eigenvalues of $\mathbf{H}_w \mathbf{H}_w^*$, $\varphi := \mathbf{h}^* \mathbf{h}$, $\mathbf{h} = \mathbf{R}^{\frac{1}{2}} \mathbf{h}_w$ and \mathbf{h}_w is 2×1 vector whose elements are independent complex Gaussian random variables.

Proof: The achievable spectral efficiency of MIMO with \mathbf{H} in (16) and equal power allocation is given by

$$C = \log_2 \left| \mathbf{I} + \frac{P_o g}{N_0 N_t} \mathbf{R}^{\frac{1}{2}} \mathbf{H}_w \mathbf{H}_w^* \mathbf{R}^{*\frac{1}{2}} \right|. \quad (18)$$

Since the correlation matrix \mathbf{R} has full rank, (18) is approximated under the high SNR regime by [21]

$$C \simeq N_t \log_2 \frac{P_o g}{N_0 N_t} + \log_2 |\mathbf{H}_w \mathbf{H}_w^*| + \log_2 |\mathbf{R}|. \quad (19)$$

Note $\log_2 |\mathbf{R}| = (N_r - 1) \log_2 (1 - \xi^2)$. Thus, the more correlated (i.e., ξ is close to 1), the more the capacity is degraded. Since ϕ_{w1} and ϕ_{w2} are the eigenvalues of $\mathbf{H}_w \mathbf{H}_w^*$, $|\mathbf{H}_w \mathbf{H}_w^*|$ is given by $\phi_{w1}\phi_{w2}$. Then, in the case of $N_t = N_r = 2$, the power equation for MIMO is given by

$$f_m(r) = \frac{N_0 2}{\eta g} \frac{2^{\frac{r}{2w}}}{\sqrt{\phi_{w1}\phi_{w2}(1-\xi^2)}} + p_{dc,m}. \quad (20)$$

Similarly, the power equation for SIMO is given by

$$f_s(r) = \frac{N_0 2^{\frac{r}{w}}}{\eta g \varphi} + p_{dc,s}. \quad (21)$$

Then, the crossover point shown in (17) is obtained by equating $f_m(r)$ and $f_s(r)$. ■

Thus, r^* is an increasing function of ξ , and channel correlation makes the crossover point higher. Note that even if the circuit power is not factored, and thus $\Delta p_{dc} = 0$, the crossover point

still exists such as $r^* = 2 \log_2 \left(\frac{2\varphi}{\sqrt{\phi_{w1}\phi_{w2}(1-\xi^2)}} \right) w$, which further emphasizes the importance of mode switching in correlated channels.

IV. ENERGY-EFFICIENT ADAPTIVE MIMO IN DYNAMIC USER POPULATIONS

A. Simple mode switching

If SIMO and MIMO use the same transmission rate r , it is straightforward to choose the best transmission mode; we pick the transmission mode that consumes least power at rate r , and the selected mode \hat{z} at rate r is

$$\hat{z}(r) = \arg \min_{z \in \{m, s\}} f_z(r). \quad (22)$$

Let us call this *simple mode switching*, where m and s denote MIMO and SIMO respectively.

B. Challenges in mode switching and rate selection

In each mode z , however, we need to be careful to choose the transmission rate r considering the tradeoff between transmit and circuit power consumption. As can be seen in (2) and (5), MIMO and SIMO have different transmit and circuit powers, and thus different energy-optimal transmission rates.

A dynamic user population makes realizing such energy-delay tradeoffs more challenging. To better understand the challenges involved, consider a TDMA system supporting a stationary dynamic load of file transfer requests. If one slows down the uplink transmission rate to save energy then the number of users in the system may grow, resulting in excess power consumption associated with users that *idle* while awaiting transmission. Indeed although ideally idling users turn off their transmission chains, in practice they still consume power due to leakage current⁷ [23], [24]. Hence, in a dynamic system, if the transmission rates are excessively reduced, the number of users that are idling may accumulate resulting in excessive overall *idling power* consumption. Consequently, we need to judiciously select the transmission rate to avoid excessive idling power consumption. This makes tradeoffs between energy conservation and delay somewhat complex.

⁷Idling power consumption depends on the specific power amplifier design. For example, power amplifier for WiMAX from Analog Devices consumes 2.5 to 25 mW during idling period [23].

C. Proposed algorithm: CUTE

Next, we describe our proposed rate selection and mode switching algorithm for multiple users with time varying MIMO channels. This algorithm is named CUTE⁸, which stands for *Conserving User Terminals' Energy*. The CUTE algorithm resolves two objectives: saving energy and achieving (or exceeding) a target user-perceived throughput. The underlying principle is to switch between SIMO and MIMO adaptively in accordance to the number of users, throughput history and channel fluctuations. In a TDMA system, we assume that time is divided into equal-sized *frames*. A frame is defined as the time period during which all users are scheduled an equal fraction of time, i.e., temporally fair scheduling. Round robin scheduling falls into this category. Let t denote the frame index.

Rate selection: Suppose that $n(t)$ users share the uplink channel for an equal fraction of time. Let $r_{i,z}(t)$ be the transmission rate of user i using transmission mode $z \in \{m, s\}$. We specify the maximum possible transmission rate as $c_{i,z}(t)$, which is determined by the time varying MIMO channel matrix \mathbf{H} and the maximum output power. Since we assume that users share the channel in a temporally fair way, each user is only allocated a fraction $1/n(t)$ of the time frame, and the maximum achievable rate of user i should be $c_{i,z}(t)/n(t)$. Let $q_i(t)$ denote the target rate of user i . Since file transfers are delay-tolerant, users can specify their own target rate considering their preferences between energy savings and fast transmission. For example, a user with sufficient residual battery may prefer fast transmission, but another user with scarce battery may prefer slow transmission to benefit from the *energy-delay tradeoff*. Note that the target rate should be independent of z , so we do not have a subscript z in $q_i(t)$. Finally, we define an energy-optimal transmission rate as $e_{i,z}(t)$, which captures the circuit and idling power consumption. Then, the transmission rate $r_{i,z}(t)$ is given by

$$r_{i,z}(t) = \min \left[\max [e_{i,z}(t), q_i(t)], \frac{c_{i,z}(t)}{n(t)} \right], \quad (23)$$

which means that we pick up the maximum of the energy-optimal rate and the target rate (if feasible). The specification of $e_{i,z}(t)$ and $q_i(t)$ are given later.

Mode switching: Suppose that $r_{i,z}(t)$ is given. Since each user uses $1/n(t)$ fraction of time frame, the instantaneous rate should be $n(t)r_{i,z}(t)$ to achieve $r_{i,z}(t)$, and the corresponding trans-

⁸This is an extended version of CUTE introduced in [22], which was for single-input single-output (SISO) system, but we use the same name here.

mission power is $f_{i,z}(n(t)r_{i,z}(t))$. So, the energy per bit is given by $f_{i,z}(n(t)r_{i,z}(t))/n(t)r_{i,z}(t)$, and the transmission mode of user i is selected as that with least energy per bit, i.e.,

$$\hat{z}_i = \arg \min_{z \in \{m,s\}} \frac{f_{i,z}(n(t)r_{i,z}(t))}{r_{i,z}(t)}. \quad (24)$$

Note that $r_{i,z}(t)$ might be different for MIMO and SIMO because of different $e_{i,z}(t)$ or $c_{i,z}(t)$. If $r_{i,m}(t) = r_{i,s}(t)$, then the above rule is identical to simple mode switching in (22). After we determine the mode, the service rate of user i is

$$r_i(t) = r_{i,\hat{z}_i}(t). \quad (25)$$

Fig. 5 shows the overall operation of the proposed algorithm.

Target rate $q_i(t)$: Suppose that user i wants to achieve a throughput q_i . Since we focus on best effort traffic, which is assumed to be tolerant to transmission rate variation, we do not need to achieve q_i instantaneously. Instead, we consider achieving q_i on average. Based on an exponential averaging of $r_i(t)$, let us define the *average* rate $\bar{r}_i(t)$ seen by user i up to time frame t as $\bar{r}_i(t) = \bar{r}_i(t-1)\nu + r_i(t)(1-\nu)$ where $0 < \nu < 1$ corresponds to averaging weight on the past. We define a *relaxed target rate* $q_i(t)$ to satisfy $q_i = \bar{r}_i(t-1)\nu + q_i(t)(1-\nu)$ so $q_i(t)$ is given by

$$q_i(t) = \frac{q_i - \bar{r}_i(t-1)\nu}{1-\nu}, \quad (26)$$

which relaxes the time scale over which the performance target should be met. It is shown in [22] that the use of a relaxed target rate enables additional energy savings.

Energy-optimal rate $e_{i,z}(t)$: Given $f_z(r)$ and idling power consumption p_{idle} , we define the energy-optimal transmission rate $e_{i,z}(t)$ as that which minimizes the energy per bit during a time frame such as

$$e_{i,z}(t) = \arg \min_r \left(\frac{1}{n(t)} f_{i,z}(n(t)r) + \frac{n(t)-1}{n(t)} p_{\text{idle}} \right) \frac{1}{r}, \quad (27)$$

which means that user i consumes $f_{i,z}(n(t)r)$ power for $\frac{1}{n(t)}$ fraction of time and p_{idle} for $\frac{n(t)-1}{n(t)}$ fraction of time. Note that p_{idle} is independent of z . The proposed algorithm converges *exponentially fast* to an equilibrium rate given a fixed $n(t)$ and channel gains – the proof is given in [22].

V. SIMULATION RESULTS

To validate the proposed algorithm, we estimate the average energy consumption per file transfer versus the average delay using flow-level event-driven simulations [16]. On each time frame, new user requests arrive according to a Poisson process. Each user requests exactly one file that is log normally distributed with mean 60 kbyte [16]. Users are assumed to experience $N_r \times 2$ spatially correlated Rayleigh fading channels. Our simulation parameters are $\eta = 0.2$, $\nu = 0.95$, $\xi = 0.7$, $w = 1$ MHz, $N_0 = -114$ dBm, $g = -124$ dB, $p_{\text{mix}} = 30.3$ mW, $p_{\text{syn}} = 50.0$ mW, $p_{\text{filt}} = 20.0$ mW, $p_{\text{idle}} = 25$ mW, $p_{\text{dac}} = 15.6$ mW, and the maximum output power of power amplifier is 27.5 dBm [23], [3], [19].⁹ The duration of a time frame is 5 ms [19], and the number of time frames for the simulation is 1,000,000. We plot the energy-delay tradeoff curves for $q_i = (1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32})$ of maximum achievable rate to show how user's preference on energy savings versus fast transmission impacts energy-delay tradeoff. The offered load is 30% of the maximum system capacity.

Fig. 6 to Fig. 9 show simulation results for MIMO with zero forcing receivers. Fig. 6 plots the pair of average delay and average energy per file transfer when circuit or idling power are not present. Three curves correspond to SIMO with antenna selection, MIMO with zero forcing receivers, and simple mode switching (SMS). Interestingly, we see significant energy savings with SMS even though circuit or idling power are not factored. This is because spatial correlation at the receive antennas makes the channel ill-conditioned and degrades the energy efficiency of MIMO. To compare the impact of spatial correlation we also plot the uncorrelated case (dotted lines). We see that energy efficiency of MIMO is greatly affected by the channel correlation while SIMO and SMS are not. Fig. 7 shows the result when N_r is changed from 2 to 4. Comparing Fig. 7 with Fig. 6 exhibits that increasing N_r alleviates the impact of channel correlation on MIMO. We still see that SMS reduces the energy per file significantly against MIMO, e.g., more than 60% when the delay is 0.5 sec or larger. A dotted line represents SMS with random antenna selection for SIMO, which demonstrates additional energy savings by 1 bit antenna selection indicator.

Fig. 8 shows the energy-delay curves when circuit and idling power are factored. As can be

⁹Since the value of p_{filt} in [13] is too low for cellular systems, we adjust it from 2.5 to 20 mW, but the simulation results are almost the same.

seen, SMS saves energy significantly against MIMO. An interesting observation is that three energy curves of SIMO, MIMO and SMS go up again as the delays grow. This is because the effect of idling energy emerges when the file transfer delay is long. Hence, we cannot fully exploit energy-delay tradeoff. This problem is effectively solved by the proposed algorithm CUTE; CUTE removes the undesirable points, (i.e., long delay and large energy consumption) by incorporating the energy-optimal transmission rate; even if the user specifies an excessively low target throughput (and large delay), the proposed algorithm automatically sends *faster* than the user's requirement to save energy. We see that energy savings of CUTE versus MIMO are significant, e.g., more than 50% at 0.5 second delay. Fig. 9 shows the energy-delay curves when N_r is changed from 4 to 8. We see that MIMO performance is improved because increasing N_r alleviates the impact of channel correlation. However, still the SMS and CUTE algorithm substantially improve the energy-delay performance.

Fig. 10 illustrates the results for MIMO with ideal receivers when $N_r = 2$. Comparing Fig. 6 and Fig. 10 shows that the energy-delay performance of MIMO is better than that of MIMO with zero forcing receivers. Nevertheless, the performance of SMS and CUTE are almost the same as before, which implies SIMO plays a major role in energy saving. In Fig. 10, we see that SMS performs better than MIMO even without circuit or idling power. This gain comes from SIMO antenna selection and MIMO channel correlation. In Fig. 11, we also see that CUTE removes the undesirable delay and energy pairs and further improves the energy efficiency.

VI. CONCLUSION

In this paper, we showed that significant energy-saving is achieved by transmission mode switching between MIMO and SIMO under dynamic loads. Even though MIMO is more energy-efficient than SIMO thanks to multiplexing gains, this may not be true when circuit power is factored. This is because circuit power can be dominant at low transmission rates, and MIMO consumes more circuit power than SIMO. For large N_r we showed that crossover point scales as $O(\log_2 N_r)$ and thus the benefit of mode switching increases with N_r . Mode switching saves more energy for the case of MIMO with a zero forcing receiver, which occasionally suffers from ill-conditioned channels. In addition, spatial correlation among receive antennas further requires the mode switching because the energy efficiency of MIMO is degraded due to channel correlation. To capture the dynamic user population, we performed flow-level simulations under Rayleigh

fading channels. In doing this, we considered the effect of idling power consumption, which led us to investigate the energy-optimal transmission rates, and solved the mode switching problem combined with rate selection. The proposed algorithm CUTE not only exhibited significant energy savings but also eliminated the undesirable operating points with excessive delay and/or energy consumption.

REFERENCES

- [1] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. Information Theory*, vol. 50, pp. 125–144, 2004.
- [2] B. Prabhakar, E.U. Bıyıkoglu, and A. E. Gamal, "Energy-efficient transmission over a wireless link via lazy packet scheduling," *Proc. IEEE INFOCOM*, vol. 1, pp. 386–393, 2001.
- [3] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE Jour. Select. Areas in Comm.*, vol. 22, pp. 1089–1098, Aug. 2004.
- [4] S. Pollin, R. Mangharam, B. Bougard, L. V. der Perre, I. Moerman, R. Rajkumar, and F. Catthoor, "MEERA: Cross-layer methodology for energy efficient resource allocation in wireless networks," *IEEE Trans. Wireless Comm.*, vol. 6, pp. 617–628, Feb. 2007.
- [5] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE Jour. Select. Areas in Comm.*, vol. 21, pp. 281–302, Apr. 2003.
- [6] 3GPP Long Term Evolution, "Physical layer aspects of UTRA high speed downlink packet access," *Technical Report TR25.814*, 2006.
- [7] "IEEE P802.16m-2007 draft standards for local and metropolitan area networks part 16: Air interface for fixed broadcast wireless access systems," *IEEE Standard 802.16m*, 2007.
- [8] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath, Jr., "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Comm. Mag.*, vol. 2, pp. 108–115, June 2002.
- [9] R. W. Heath, Jr. and A. Paulraj, "Switching between spatial multiplexing and transmit diversity based on constellation distance," *IEEE Trans. Communications*, vol. 53, pp. 962–968, June 2005.
- [10] A. Forenza, M. R. McKay, A. Pandharipande, R. W. Heath, Jr., and I. B. Collings, "Adaptive MIMO transmission for exploiting the capacity of spatially correlated channels," *IEEE Trans. on Veh. Technol.*, vol. 56, pp. 619–630, Mar. 2007.
- [11] C.-B. Chae, M. Katz, C. Suh, and H. Jeong, "Adaptive spatial modulation for MIMO-OFDM," *Proc. IEEE WCNC*, pp. 87–92, March 2004.
- [12] C.-B. Chae, A. Forenza, R. W. Heath, Jr., M. R. McKay, and I. B. Collings, "Adaptive MIMO transmission techniques for broadband wireless communication systems," *submitted to IEEE Comm. Mag.*, March 2007.
- [13] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Comm.*, vol. 4, pp. 2349–32360, Sep. 2005.
- [14] G. Miao, N. Himayat, Y. Li, and D. Bormann, "Energy efficient design in wireless OFDMA," in *Proc. IEEE Int. Conf. on Comm. (ICC)*, May 2008.
- [15] S. Cui, R. Madan, A. J. Goldsmith, and S. Lall, "Cross-layer energy and delay optimization in small-scale sensor networks," *IEEE Trans. Wireless Comm.*, vol. 6, pp. 3688–3699, Oct. 2007.

- [16] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *Proc. IEEE INFOCOM*, vol. 1, pp. 321–331, Apr. 2003.
- [17] M. Kodialam, T. V. Lakshman, and S. Sengupta, "Traffic-oblivious routing for guaranteed bandwidth performance," *IEEE Comm. Mag.*, pp. 46–51, Apr. 2007.
- [18] Goliath, "Financial assessment of citywide Wi-Fi/WiMAX deployment.," http://goliath.ecnext.com/coms2/summary_0199-6709510_ITM, 2006.
- [19] J. G. Andrews, A. Ghosh, and R. Muhamed, "Fundamentals of WiMAX," *Prentice Hall*, 2007.
- [20] B. Rengarajan and G. de Veciana, "Architecture and abstraction for environment and traffic aware system-level coordination of wireless networks: the downlink case," *Proc. IEEE INFOCOM*, pp. 1175–1183, 2008.
- [21] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.
- [22] H. Kim and G. de Veciana, "Leveraging dynamic spare capacity in wireless systems to conserve mobile terminals' energy," *submitted to IEEE/ACM Trans. Networking*, May. 2008.
- [23] WiMAX power amplifier ADL5570 and 5571, "http://www.analog.com/uploadedfiles/data_sheets/adl5570.pdf, http://www.analog.com/uploadedfiles/data_sheets/adl5571.pdf," *Analog Device*, Sep. 2007.
- [24] H. Kim, C-B. Chae, G. de Veciana, and R. W. Heath Jr., "Energy-efficient adaptive MIMO systems leveraging dynamic spare capacity," *Proc. Conference on Information Sciences and Systems (CISS)*, Mar. 2008.

TABLE I
NOTATION SUMMARY

\mathbf{H}	$N_r \times N_t$ uplink channel matrix
N_r	the number of receive antennas at the BS (≥ 2)
N_t	the number of transmit antennas at the MT ($= 2$)
ϕ_k	the eigenvalue of $\mathbf{H}^* \mathbf{H}$, $k = 1, 2$
N_0	noise power
P_o	output power dissipated into the air from power amplifier
w	spectral bandwidth (Hz)
$p_{dc,m}$	circuit power consumption of MIMO
$p_{dc,s}$	circuit power consumption of SIMO
η	drain efficiency of power amplifier
\mathbf{H}_w	complex Gaussian random matrix with zero mean and variance g
g	path loss
r^*	crossover point below which SIMO is more energy-efficient than MIMO
q	target throughput per user
r_{\max}	maximum system throughput
λ	file arrival rate
μ^{-1}	average file size
ρ	$:= \frac{\lambda}{\mu}$ traffic load (bps)
\mathbf{R}	receive correlation matrix
ξ	correlation coefficient
z	transmission mode index
i	user index
t	time frame index
$n(t)$	number of flows (users) at time frame t
ν	exponential average parameter

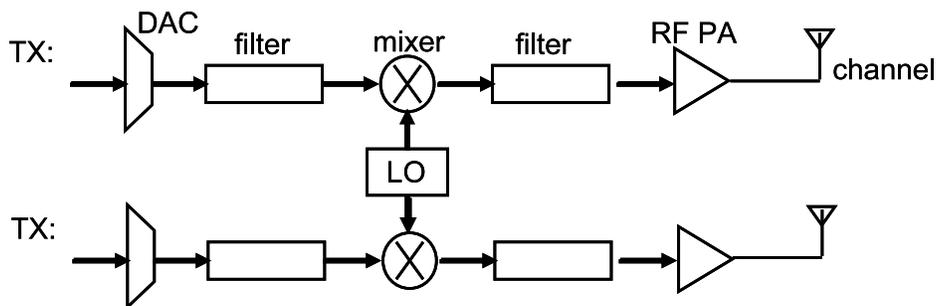


Fig. 1. Transmission chain for a MIMO system with two antennas.

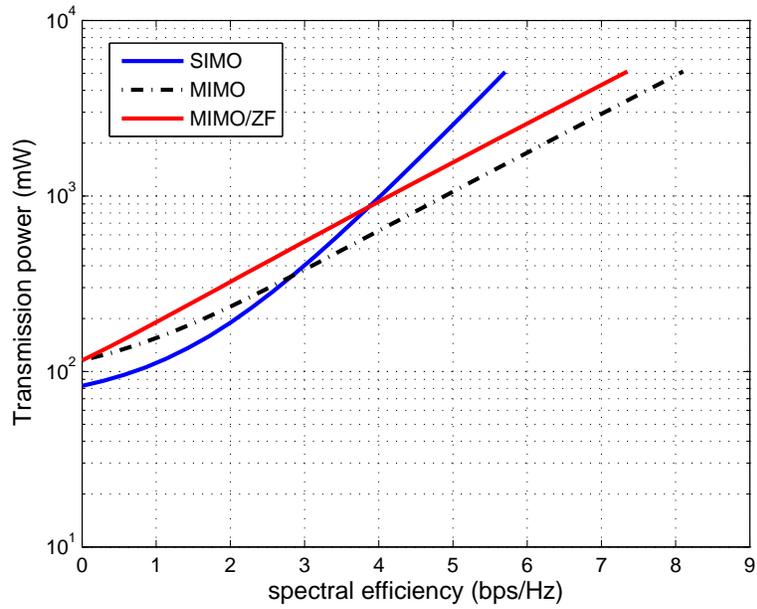


Fig. 2. Transmission power consumption of mobile terminals including circuit power.

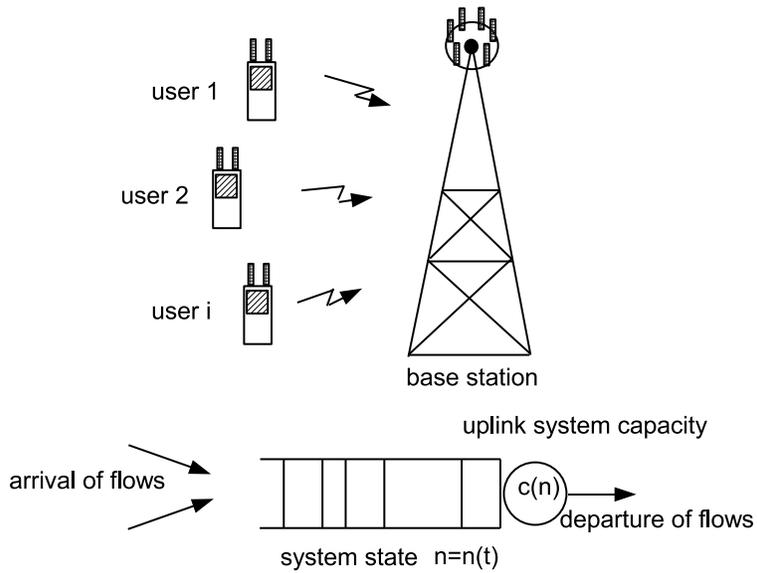


Fig. 3. Flow-level model for uplink transmission in a dynamic system. One user corresponds to one flow.

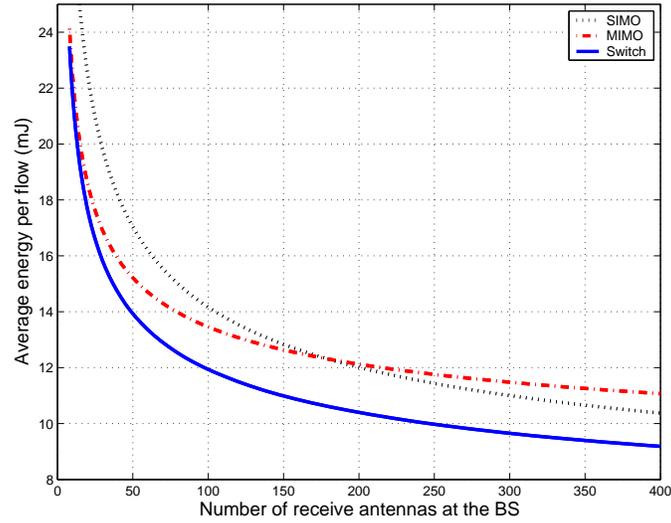


Fig. 4. Flow-level analysis result for average energy per file vs the number of receive antennas: offered load 30% and $q = 0.01 \times r_{\max}$, i.e., 100 users can share the system without congestion.

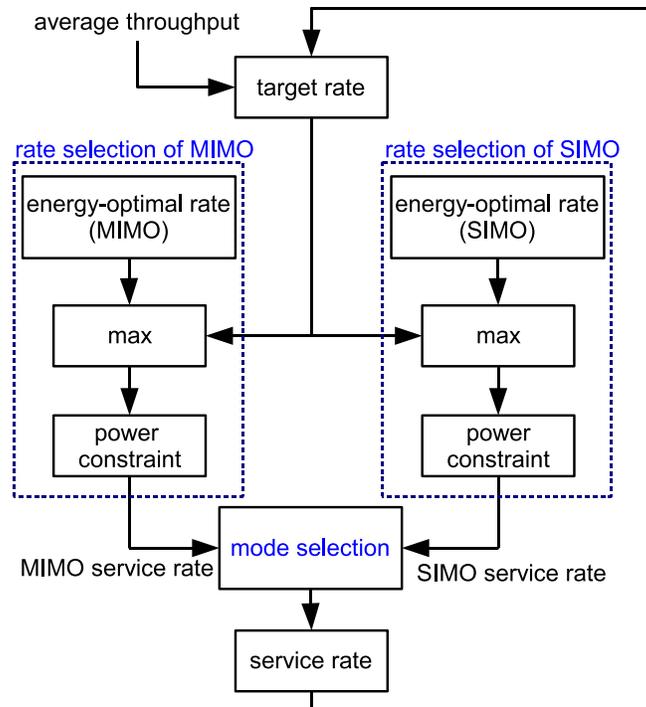


Fig. 5. Flow chart of the proposed algorithm.

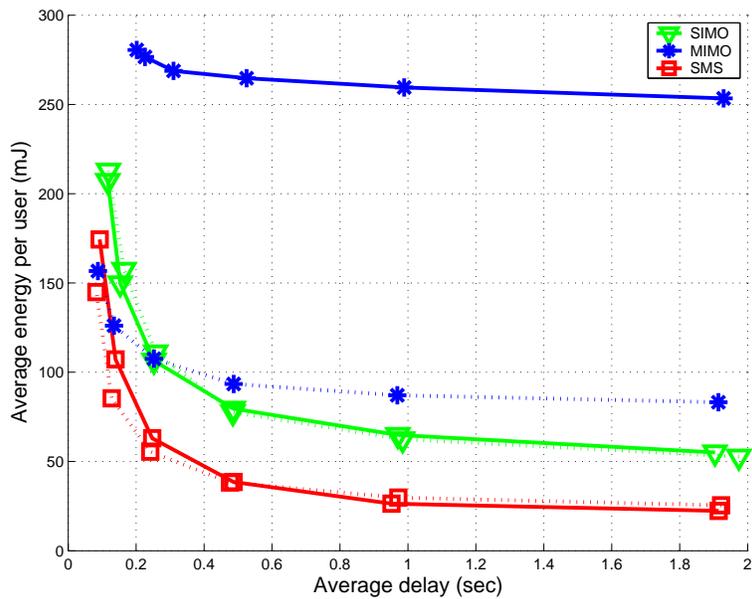


Fig. 6. Energy-delay tradeoff curves without circuit and idling power: zero forcing receiver for MIMO, $N_r = 2$, $N_t = 2$, traffic load $\rho = 2.51\text{Mbps}$, $r_{\max} = 8.35\text{Mbps}$, correlation coefficient $\xi = 0.7$ (solid line), $\xi = 0$ (dotted line)

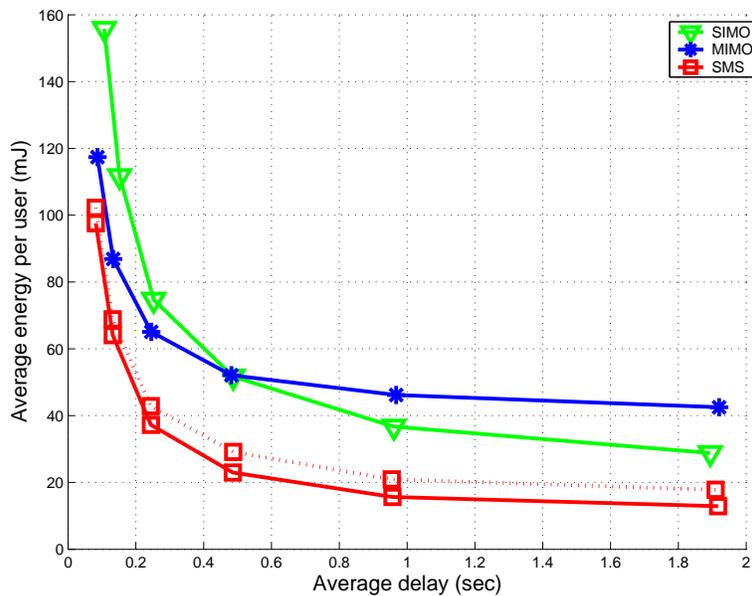


Fig. 7. Energy-delay tradeoff curves without circuit and idling power: zero forcing receiver for MIMO, $N_r = 4$, $N_t = 2$, traffic load $\rho = 3.70\text{Mbps}$, $r_{\max} = 12.34\text{Mbps}$. Dotted line shows SMS with random antenna selection, i.e, without using 1 bit antenna selection indicator for SIMO.

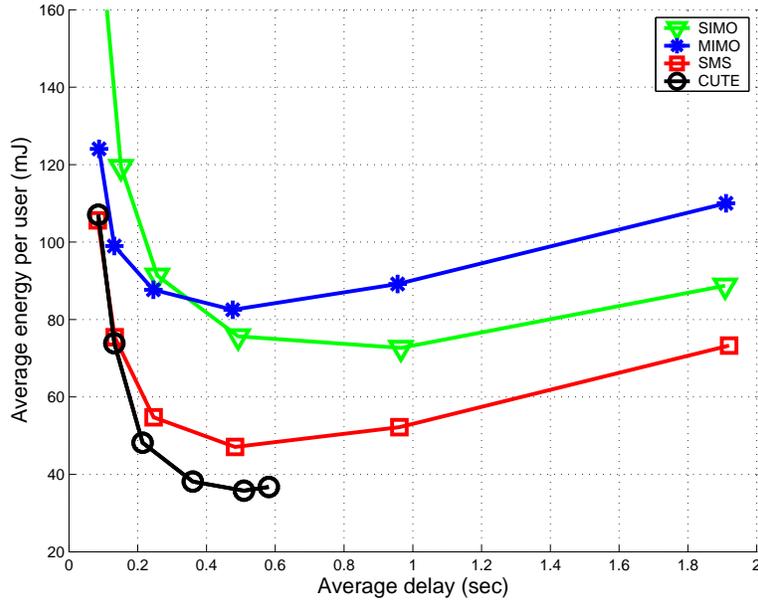


Fig. 8. Energy-delay tradeoff curves *with* circuit and idling power: zero forcing receiver for MIMO, $N_r = 4$, $N_t = 2$, traffic load $\rho = 3.70\text{Mbps}$, $r_{\max} = 12.34\text{Mbps}$.

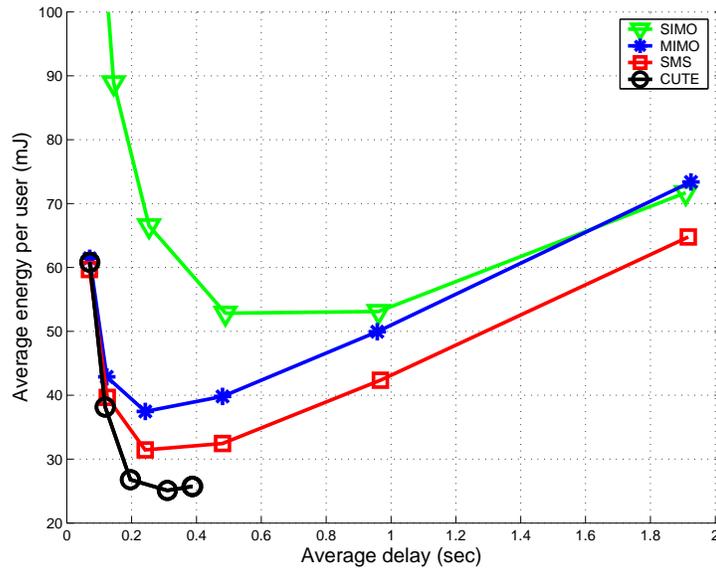


Fig. 9. Energy-delay tradeoff curves *with* circuit and idling power: zero forcing receiver for MIMO, $N_r = 8$, $N_t = 2$, traffic load $\rho = 4.51\text{Mbps}$, $r_{\max} = 15.04\text{Mbps}$.

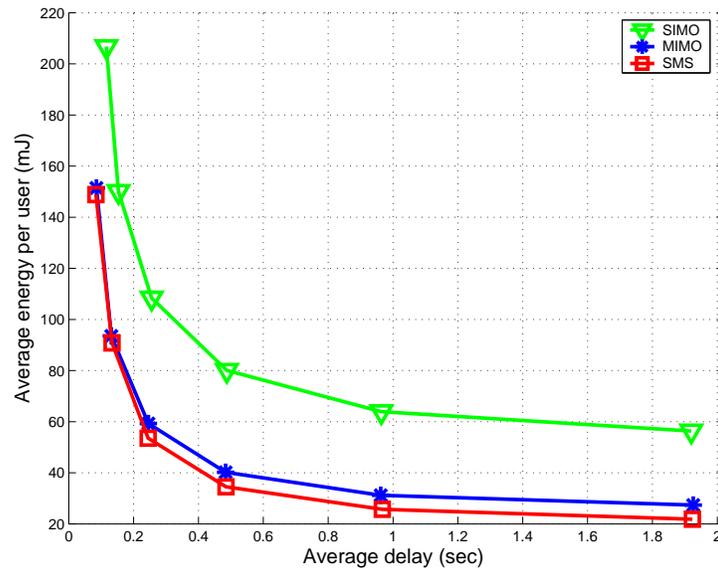


Fig. 10. Energy-delay tradeoff curves without circuit and idling power: ideal receiver for MIMO, $N_r = 2$, $N_t = 2$, traffic load $\rho = 2.51\text{Mbps}$, $r_{\max} = 8.36\text{Mbps}$.

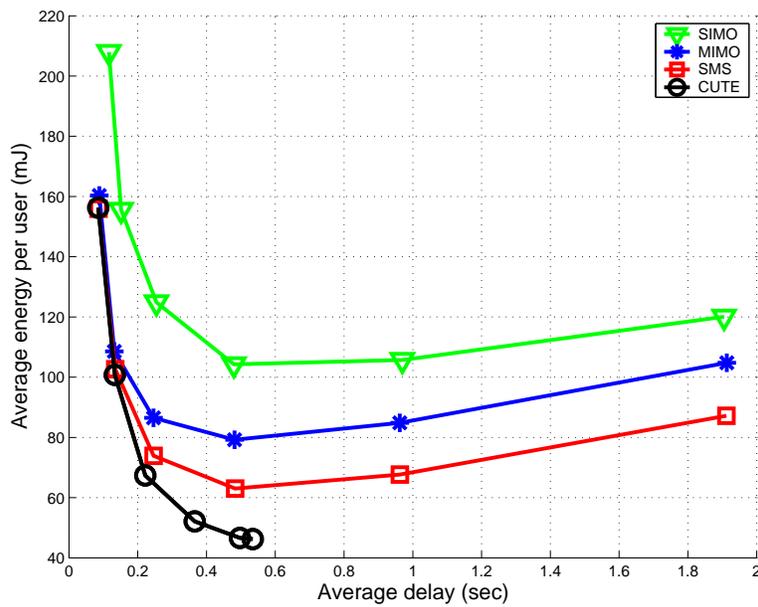


Fig. 11. Energy-delay tradeoff curves *with* circuit and idling power: ideal receiver for MIMO, $N_r = 2$, $N_t = 2$, traffic load $\rho = 2.51\text{Mbps}$, $r_{\max} = 8.36\text{Mbps}$.