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

Hongseok Kim, *Student Member, IEEE*, Chan-Byoung Chae, *Member, IEEE*,
Gustavo de Veciana, *Fellow, IEEE*, and 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 included. 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.**

## 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 underutilized[1], 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 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

Manuscript received August 20, 2008; revised February 4, 2009; accepted May 8, 2009. The associate editor coordinating the review of this paper and approving it for publication was Y. J. (A.) Zhang.

H. Kim, G. de Veciana, and R. W. Heath, Jr. are with the Wireless Networking and Communications Group (WNCG), the University of Texas at Austin, TX 78712 USA (e-mail: hongseok@ieee.org, {gustavo, rheath}@ece.utexas.edu).

C.-B. Chae is with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA 02138 (e-mail: cbchae@seas.harvard.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.

Digital Object Identifier 10.1109/TWC.2009.081123

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

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]–[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]–[12], but the authors mainly focused on the physical layer. Power-efficient MIMO systems were studied in terms of transmit diversity [13] and input covariance [14]. 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 [3], [4], [15]–[17]. 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., [15], is limited to physical layer modulation techniques with a single sender and receiver pair for sensor networks. A link-level multiantenna approach was proposed in [17], which adapts the transmission mode packet-per-packet among space-division multiplexing, space-time coding, and single antenna transmission in a wireless local area networks. The work in [4], [18] addresses multiple users including the MAC layer, but only for a *fixed* number of users in a wireless local area network. Unlike previous work [3], [4], [15]–[18], our focus is on *dynamic* multi-users 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" [19]. 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. Based on the exponential correlation model [20], [21], a closed form expression for the crossover point is provided as a function of a correlation coefficient. 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. We also propose an energy-opportunistic scheduler exploiting both multi-user and multi-mode diversity to further enhance the energy efficiency.

**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% [22]. Similarly a substantial fraction of Wi-Fi hotspot capacity is unused [23]. 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 [24], [25]. Broadband cellular systems also have higher time variability in traffic loads and system capacity [25]. 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 channel correlation and $N_r$ on the crossover point mainly from static single user scenario. We address the dynamic multi-user scenario 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 fading[2] 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 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. Thus, the transmission mode decision, either MIMO or SIMO, is made at the BS and fed back to the MS, 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 consumption[4] 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. Nevertheless for the actual transmission, the MS requires another kinds of feedback, for example, to implement the adaptive modulation and coding, and to adjust the tranmist power level in accordance with the pathloss. 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 $\phi_1$ and $\phi_2$ 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]

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

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 $\min(N_t, N_r) = N_t$. Note that the channel matrix $\mathbf{H}$ is general. For example, if $\mathbf{H} = \mathbf{R}_r^{\frac{1}{2}} \mathbf{H}_w$ where $\mathbf{R}_r$ is a receive correlation matrix, then $\mathbf{H}$ captures the correlated channels at the receiver [20]. If $\mathbf{H} = \mathbf{H}_w$ where the elements of $\mathbf{H}_w$ are independent complex Gaussian random variables with zero mean, $\mathbf{H}$ models uncorrelated Rayleigh fading channels.

The MIMO transmission power model is derived from (1)

TABLE I
Notation Summary.

| | |
|---|---|
| $\mathbf{H}$ | $N_r \times N_t$ uplink channel matrix |
| $N_r$ | the number of receive antennas at the BS ($\geq 2$) |
| $N_t$ | the number of transmit antennas at the MT ($= 2$) |
| $\phi_k$ | 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) |
| $p_{\text{dc,m}}$ | circuit power consumption of MIMO |
| $p_{\text{dc,s}}$ | circuit power consumption of SIMO |
| $\eta$ | drain efficiency of power amplifier |
| $\mathbf{H}_w$ | complex Gaussian random matrix with mean 0 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 |
| $\lambda$ | file arrival rate |
| $\mu^{-1}$ | average file size |
| $\rho$ | $:=\frac{\lambda}{\mu}$ traffic load (bps) |
| $\mathbf{R}$ | receive correlation matrix |
| $\xi$ | correlation coefficient |
| $z$ | transmission mode index |
| $i$ | user index |
| $t$ | time frame index |
| $n(t)$ | number of flows (users) at time frame $t$ |
| $v_i(t)$ | the fraction of time allocated to user $i$ |
| $\nu$ | exponential average parameter |

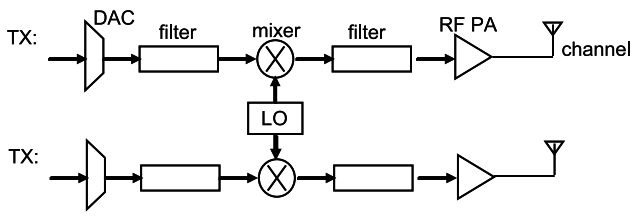


Fig. 1. Transmission chain for a MIMO system with two antennas.

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

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

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

as follows. With $N_t = 2$, we have

$$C = \sum_{k=1}^{2} \log_2\left(1 + \frac{P_o}{N_0 N_t}\phi_k\right) \quad (2)$$

$$= \log_2\left[\left(\frac{P_o}{2N_0}\right)^2 \phi_1\phi_2 + \frac{\phi_1+\phi_2}{2N_0}P_o + 1\right], \quad (3)$$

which leads to the following quadratic equation of $P_o$,

$$\phi_1\phi_2\left(\frac{P_o}{2N_0}\right)^2 + (\phi_1+\phi_2)\frac{P_o}{2N_0} + (1-2^C) = 0. \quad (4)$$

The feasible solution to the above quadratic equation is given by

$$P_o = \frac{2N_0}{\phi_1\phi_2}\left[\sqrt{\left[\frac{\phi_1+\phi_2}{2}\right]^2 + \phi_1\phi_2(2^C-1)} - \frac{\phi_1+\phi_2}{2}\right]. \quad (5)$$

Assuming that the power consumed by the power amplifier is linearly dependent on the output power [3], the power consumed in the power amplifier can be modeled as $\frac{P_o}{\eta}$ where $\eta$ is the drain efficiency of the power amplifier. Drain efficiency is the ratio of the output power, i.e., the electromagnetically radiated power into the air versus the power consumed in the power amplifier. The transmission power equation $f_m(r)$ for MIMO at transmission rate $r$ with spectral bandwidth $w$ and circuit power consumption $p_{\text{dc},m}$ is then given by

$$f_m(r) = \frac{2N_0}{\eta\phi_1\phi_2}\left(\sqrt{\left(\frac{\phi_1+\phi_2}{2}\right)^2 + \phi_1\phi_2(2^{r/w}-1)} - \frac{\phi_1+\phi_2}{2}\right) + p_{\text{dc},m} \quad (6)$$

where the subscript $m$ stands for MIMO. For simplicity, we assume that possible transmission rates are continuous.[5]

Note that (1) and (6) 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 [20], under an independent coding and detection assumption, (1) can be rewritten as,

$$C = \sum_{k=1}^{N_t} \log_2\left(1 + \frac{P_o}{N_0 N_t}\frac{1}{[(\mathbf{H}^*\mathbf{H})^{-1}]_{k,k}}\right) \quad (7)$$

where $[(\mathbf{H}^*\mathbf{H})^{-1}]_{k,k}$ denotes $k$th diagonal element of $(\mathbf{H}^*\mathbf{H})^{-1}$. Thus, if $\phi_k$ is replaced by $1/[(\mathbf{H}^*\mathbf{H})^{-1}]_{k,k}$ in (6), we obtain a transmission power model of 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], [24] as can be seen in Fig. 1. Then, the total circuit power consumption of MIMO is given by

$$p_{\text{dc},m} = N_t(p_{\text{dac}} + p_{\text{mix}} + p_{\text{filt}}) + p_{\text{syn}} \quad (8)$$

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

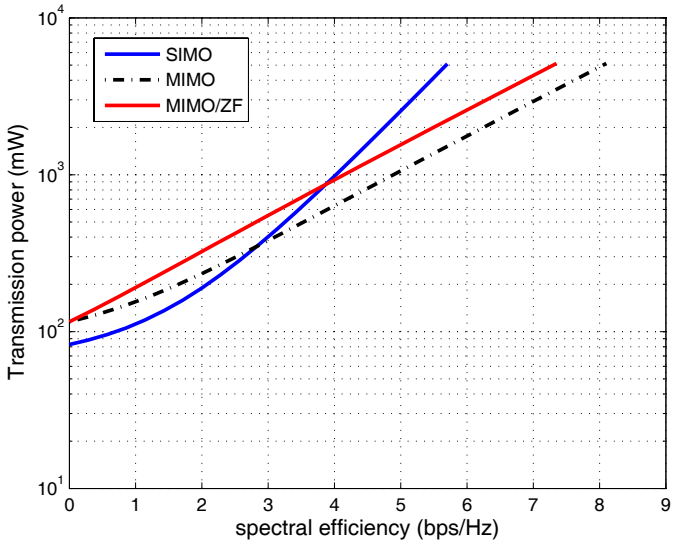


Fig. 2. Transmission power consumption of mobile terminals including circuit power.

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

*2) SIMO power model:* The SIMO power equation can be derived from (1) with $N_t = 1$ and $\phi_k = \mathbf{h}_k^*\mathbf{h}_k = \sum_{\ell=1}^{N_r}|h_{\ell,k}|^2$ where $\mathbf{h}_k$ is the $k$-th column vector of $\mathbf{H}$, and $k = 1, 2$. Then, in selecting the transmit antenna out of the two, $\mathbf{h}_k$ with higher $\phi_k$ is chosen for the transmission. This decision is made at the BS, which implies 1 bit feedback from the BS to the MS is required. Then, the capacity can be expressed as

$$C = \log_2\left[1 + \frac{P_o}{N_0}\sum_{\ell=1}^{N_r}|h_{\ell,\hat{k}}|^2\right] \quad (9)$$

where $\hat{k}$ denotes the index for the selected antenna. From (9), we have

$$P_o = \frac{2^C - 1}{\sum_{\ell=1}^{N_r}|h_{\ell,\hat{k}}|^2}N_0, \quad (10)$$

and substituting $C$ with $\frac{r}{w}$ and introducing the drain efficiency $\eta$ and the circuit power of SIMO, $p_{\text{dc},s}$, the transmission power $f_s(r)$ where the subscript $s$ stands for SIMO, is given by

$$f_s(r) = \frac{1}{\eta}\frac{2^{r/w}-1}{\sum_{\ell=1}^{N_r}|h_{\ell,\hat{k}}|^2}N_0 + p_{\text{dc},s}. \quad (11)$$

## 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 channel correlation on the crossover point*

Besides the circuit power, another important factor determining MIMO energy efficiency is spatial correlation among antennas. Since channel correlation degrades the capacity of MIMO systems [20], 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

$$\mathbf{H} = \mathbf{R}_r^{\frac{1}{2}} \mathbf{H}_w \tag{12}$$

where $\mathbf{R}_r$ is an $N_r \times N_r$ receive correlation matrix whose elements are $[\mathbf{R}_r]_{i,j} = \xi^{|i-j|}$, $0 \leq \xi < 1$. This model, called exponential correlation model, is extensively used in the literature [20], [21]. Since receive antennas at the BS are not surrounded by many 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 consider generally 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 1:* Assuming high received SNR, when a correlated channel matrix $\mathbf{H}$ is given by (12) with $N_t = N_r = 2$, the crossover point is explicitly given by

$$r^* \approx 2\log_2 \left\{ \frac{\varphi}{\sqrt{\phi_{w1}\phi_{w2}(1-\xi^2)}} + \sqrt{\frac{\varphi^2}{\phi_{w1}\phi_{w2}(1-\xi^2)} + \Delta p_{\mathrm{dc}} \frac{\eta \varphi g}{N_0}} \right\} w \tag{13}$$

where $\xi$ is the correlation coefficient, $\phi_{w1}$ and $\phi_{w2}$ 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 $2 \times 1$ vector whose elements are independent complex Gaussian random variables.

*Proof:* The achievable spectral efficiency of MIMO with $\mathbf{H}$ in (12) and equal power allocation is given by

$$C = \log_2 \left| \mathbf{I} + \frac{P_o g}{N_0 N_t} \mathbf{R}_r^{\frac{1}{2}} \mathbf{H}_w \mathbf{H}_w^* \mathbf{R}_r^{*\frac{1}{2}} \right|. \tag{14}$$

Since the correlation matrix $\mathbf{R}_r$ has full rank, (14) is approximated under the high SNR regime by [20]

$$C \simeq N_t \log_2 \frac{P_o g}{N_0 N_t} + \log_2 |\mathbf{H}_w \mathbf{H}_w^*| + \log_2 |\mathbf{R}_r|. \tag{15}$$

Note $\log_2 |\mathbf{R}_r| = (N_r - 1) \log_2 \left(1 - \xi^2\right)$. Thus, the more correlated (i.e., $\xi$ is close to 1), the more the capacity is degraded. Since $\phi_{w1}$ and $\phi_{w2}$ 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

$$f_m(r) = \frac{N_0 2}{\eta g} \frac{2^{\frac{r}{2w}}}{\sqrt{\phi_{w1}\phi_{w2}(1-\xi^2)}} + p_{\mathrm{dc},m}. \tag{16}$$

Similarly, the power equation for SIMO is given by

$$f_s(r) = \frac{N_0}{\eta g} \frac{2^{\frac{r}{w}}}{\varphi} + p_{\mathrm{dc},s}. \tag{17}$$

Then, the crossover point shown in (13) is obtained by equating $f_m(r)$ and $f_s(r)$. ■

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

### *C. The impact of the number of receive antennas on the crossover point*

The number of receiver antennas at the BS also has an impact on the crossover point. To obtain an analytically tractable result we assume uncorrelated Rayleigh fading channels with many receive antennas at the BS. We first provide the following Lemma.

*Lemma 2:* Suppose that $\mathbf{H}$ is an $N_r \times N_t$ matrix whose elements are independent complex Gaussian random variables with zero mean and variance $g$, i.e., $\mathbf{H}$ represents uncorrelated Rayleigh fading channels. Then, as $N_r$ goes to infinity, $\frac{1}{N_r g} \mathbf{H}^* \mathbf{H}$ converges to an identity matrix by the law of large numbers.

*Proof:* Let $\mathbf{A} = \frac{1}{N_r g} \mathbf{H}^* \mathbf{H}$. Let $[\mathbf{A}]_{i,j}$ denote the $(i,j)$ element of $\mathbf{A}$. Then, $[\mathbf{A}]_{i,j} = \sum_{\ell=1}^{N_r} \frac{1}{N_r g} [\mathbf{H}^*]_{i,\ell} [\mathbf{H}]_{\ell,j} = \sum_{\ell=1}^{N_r} \frac{1}{N_r g} \left([\mathbf{H}]_{\ell,i}\right)^* [\mathbf{H}]_{\ell,j}$. When $i = j$, $[\mathbf{A}]_{i,i} = \frac{1}{N_r} \sum_{\ell=1}^{N_r} |\frac{[\mathbf{H}]_{\ell,i}}{\sqrt{g}}|^2$, which is the mean of the square of $N_r$ independent normal complex Gaussian random variables. Then, as $N_r$ goes to infinity $[\mathbf{A}]_{i,i}$ converges to 1 by the law of large numbers. When $i \neq j$, $[\mathbf{A}]_{i,j}$ is the average of $N_r$ independent zero mean complex random variables. By the law of large numbers, $[\mathbf{A}]_{i,j}$ converges to zero as $N_r$ goes to infinity. Thus $\mathbf{A}$ converges to an identity matrix as $N_r$ goes to infinity. ■

Then, two eigenvalues of $\mathbf{H}^* \mathbf{H}$ are asymptotically given by $N_r g$, and we have the following proposition.

*Proposition 3:* 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 (6), the power equation is given by

$$f_m(r) = \frac{N_0}{\eta g} \frac{2}{N_r} \left(2^{\frac{r}{2w}} - 1\right) + p_{\mathrm{dc},m}. \tag{18}$$

Similarly, (11) can be rewritten as

$$f_s(r) = \frac{N_0}{\eta g} \frac{1}{N_r} \left(2^{\frac{r}{w}} - 1\right) + p_{\mathrm{dc},s}. \tag{19}$$

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

$$r^* = 2\log_2\left(1+\sqrt{\Delta p_{\rm dc}\frac{\eta N_r g}{N_0}}\right) w \tag{20}$$

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

$$r_{\max} = 2\log_2\left(1+\frac{P_o}{2}\frac{N_r g}{N_0}\right) w. \tag{21}$$

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

$$\frac{r^*}{r_{\max}} = \frac{\log_2\left(1+\sqrt{\Delta p_{\rm dc}\frac{\eta N_r g}{N_0}}\right)}{\log_2\left(1+\frac{P_o}{2}\frac{N_r g}{N_0}\right)}. \tag{22}$$

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

Proposition 3 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 (18) and (19), 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.

### *D. Asymptotic analysis for many receive antennas using flow-level dynamics*

So far we have focused on the link level performance, i.e., single user scenario. To better understand the impact of large $N_r$ on energy efficiency, we now proceed to a multi-user scenario and assume the number of ongoing users varies with time, i.e., a *dynamic* system. For large $N_r$, we have shown that 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* [19], 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. For analytical simplicity, we assume that users have the same target throughput $q$ and are served via temporally fair TDMA scheduling. Then, the system capacity[6] in state $n$ is given by

$$c(n) = \min(nq, r_{\max}). \tag{23}$$

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 $\lambda$ and have independent file sizes with mean $\mu^{-1}$. 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 the 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 $\pi$, if it exists, is given by

$$\pi(n) = \pi(0)\frac{\rho^n}{\Pi_{m=1}^n c(m)}, \tag{24}$$

where $\rho := \frac{\lambda}{\mu}$ is the traffic load (bits per second) and $\pi(0) = \left(1+\sum_{n=1}^{\infty}\frac{\rho^n}{\Pi_{m=1}^n c(m)}\right)^{-1}$. 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 $\pi$. 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 is given by

$$p(n) = \min\left[f_m(c(n)), f_s(c(n))\right]. \tag{25}$$

Note that, from the crossover point $r^*$ in (20), if $nq \leq r^*$ then SIMO is more energy-efficient, and vice versa. Thus,

$$p(n) = \begin{cases} f_s(c(n)) & \text{if } n \leq n^* \\ f_m(c(n)) & \text{if } n > n^* \end{cases}, \tag{26}$$

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 [26], which is akin to Little's result gives that

$$E[J] = \frac{1}{\lambda}E[P]. \tag{27}$$

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 (20), 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.

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

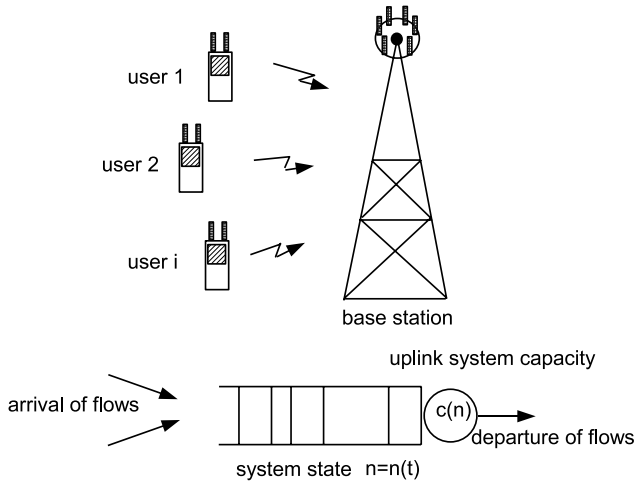


Fig. 3. Flow-level model for uplink transmission in a dynamic system. One user corresponds to one flow.

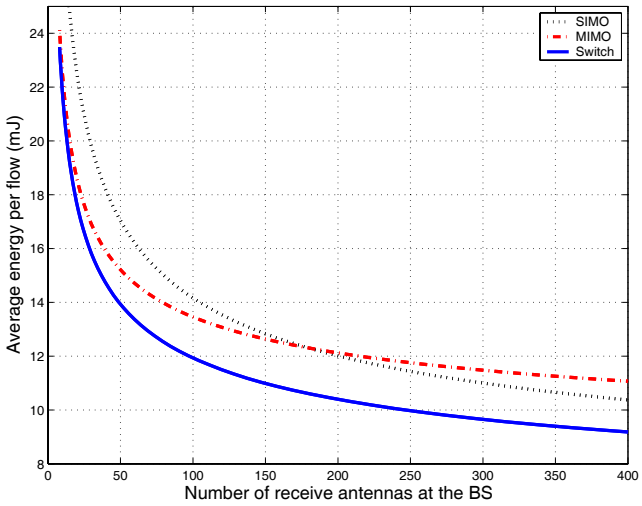


Fig. 4. Flow-level analysis result for average energy per file vs the number of receive antennas: offered load 30% and $q = 0.01 \times r_{\max}$, i.e., 100 users can share the system without congestion.

## IV. Energy-efficient adaptive MIMO in dynamic user populations

In this section we investigate the mode switching combined with rate selection considering *multi-user* scenario in *dynamic* systems and propose a practical solution to realize energy-efficient adaptive MIMO systems. Energy-opportunistic scheduling exploiting multi-mode and multi-user diversity is also proposed to further enhance the energy efficiency.

### 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

$$\hat{z}(r) = \arg\min_{z\in\{m,s\}} f_z(r). \tag{28}$$

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 (6) and (11), 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 current[7] [27], [28]. 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.

### 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 CUTE[8], 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 once.

***Rate selection:*** Let $t$ denote the frame index. Let $v_i(t)$ denote the fraction of time allocated to user $i$, and $\sum_i v_i(t) = 1$. Suppose that $n(t)$ users are sharing the uplink channel based on round-robin scheduling with weight $v_i(t)$. If temporally fair scheduling is used, $v_i(t)$ is simply $\frac{1}{n(t)}$. 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 each user is only allocated a fraction $v_i(t)$ of the time frame, the maximum achievable rate of user $i$ is $v_i(t)c_{i,z}(t)$. 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

[7]Idling 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 [27].

[8]This is an extended version of CUTE introduced in [26], which was for single-input single-output (SISO) system, but we use the same name here.

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 tradeoffs*. 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

$$r_{i,z}(t) = \min\left[\max\left[e_{i,z}(t), q_i(t)\right], v_i(t)c_{i,z}(t)\right], \quad (29)$$

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:*** To achieve $r_{i,z}(t)$ on average during one time frame, the instantaneous rate should be $r_{i,z}(t)/v_i(t)$ because user $i$ only uses a $v_i(t)$ fraction of a time frame, and the corresponding transmission power is $f_{i,z}\big(r_{i,z}(t)/v_i(t)\big)$. So, the energy per bit is given by $\frac{f_{i,z}\left(r_{i,z}(t)/v_i(t)\right)}{r_{i,z}(t)/v_i(t)}$, and the transmission mode of user $i$ is selected as that with the least energy per bit, i.e.,

$$\hat{z}_i = \arg\min_{z\in\{m,s\}} \frac{f_{i,z}\big(r_{i,z}(t)/v_i(t)\big)}{r_{i,z}(t)/v_i(t)}. \quad (30)$$

Note that $r_{i,z}(t)$ might be different for MIMO and SIMO because of different $e_{i,z}(t)$, $v_i(t)$ or $c_{i,z}(t)$. If $r_{i,m}(t) = r_{i,s}(t)$, then the rule in (30) is identical to simple mode switching in (28). After we determine the mode, the service rate of user $i$ is

$$r_i(t) = r_{i,\hat{z}_i}(t). \quad (31)$$

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

$$q_i(t) = \frac{q_i - \bar{r}_i(t-1)\nu}{1-\nu}, \quad (32)$$

which relaxes the time scale over which the performance target should be met. It is shown in [26] 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 $p_{\text{idle}}$, 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

$$e_{i,z}(t) = \arg\min_r \left[v_i(t) f_{i,z}\big(r/v_i(t)\big) + (1 - v_i(t))\, p_{\text{idle}}\right]\frac{1}{r}, \quad (33)$$

which means that user $i$ consumes $f_{i,z}\big(r/v_i(t)\big)$ power for $v_i(t)$ fraction of time and $p_{\text{idle}}$ for $(1-v_i(t))$ fraction of time. Note that $p_{\text{idle}}$ is independent of $z$. The proposed algorithm converges *exponentially fast* to an equilibrium rate given a fixed $v_i(t) = 1/n(t)$ (i.e., temporally fair scheduling) and channel gains – the proof is given in [26].

### D. Extension to energy-opportunistic scheduling

In order to further reduce transmit energy, CUTE algorithm can be easily extended to exploit multi-user diversity. In addition to the conventional opportunistic scheduling that selects the user who experiences the best channel condition [29], we need to consider the spatial transmission mode and the energy efficiency. The proposed *energy-opportunistic* scheduler selects the pair (the user and its transmission mode) that consumes the minimum energy per bit. To this end we slightly change the definition of the time frame. While in round-robin scheduling all users are scheduled a fraction $v_i(t)$ of each time frame, in energy-opportunistic scheduling only one user is selected and that user takes whole time frame. Hence, the length of time frame needs to shrink to a channel coherence time. Assuming users experience the same average channel gain (small-scale fading) and temporally fair scheduling, we use the following energy-opportunistic scheduling,

$$(i^*, \hat{z}_{i^*}) = \arg\min_{(i\in\mathcal{A}(t),\, z\in\{m,s\})} \frac{f_{i,z}\big(u_{i,z}(t)\big)}{u_{i,z}(t)} \quad (34)$$

where $\mathcal{A}(t)$ is a set of all active users and $u_{i,z}(t)$ is defined by

$$u_{i,z}(t) = \min\left[\max\left[e_{i,z}(t), q_i\right], c_{i,z}(t)\right]. \quad (35)$$

Note that $v_i(t)c_{i,z}(t)$ in (29) is replaced by $c_{i,z}(t)$ because the selected user takes the whole time frame. In addition, (32) is modified to

$$q_i(t) = \frac{n(t)q_i - \bar{r}_i(t-1)\nu}{1-\nu} \quad (36)$$

where $\bar{r}_i(t)$ is computed during the time frames where user $i$ has been served, and (33) is modified to

$$e_{i,z}(t) = \arg\min_r \frac{f_{i,z}\big(r\big) + (n(t)-1)\, p_{\text{idle}}}{r}. \quad (37)$$

Finally, the service rate of user $i$ is given by

$$r_{i,z}(t) = u_{i,z}(t)\mathbf{1}_{\{(i,z)=s_\theta(t)\}}, \quad (38)$$

where $s_\theta(t)$ denotes the pair of the scheduled user and the transmission mode; i.e., $(i^*, \hat{z}_{i^*})$.

## 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 [19]. 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 [19]. Users are assumed to experience $N_r \times 2$ spatially correlated Rayleigh fading channels. Our simulation parameters are $\eta = 0.2$, $\nu = 0.95$ for round-robin scheduling and $\nu = 0$ for energy-opportunistic scheduling, $\xi = 0.7$, $w = 1$ 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

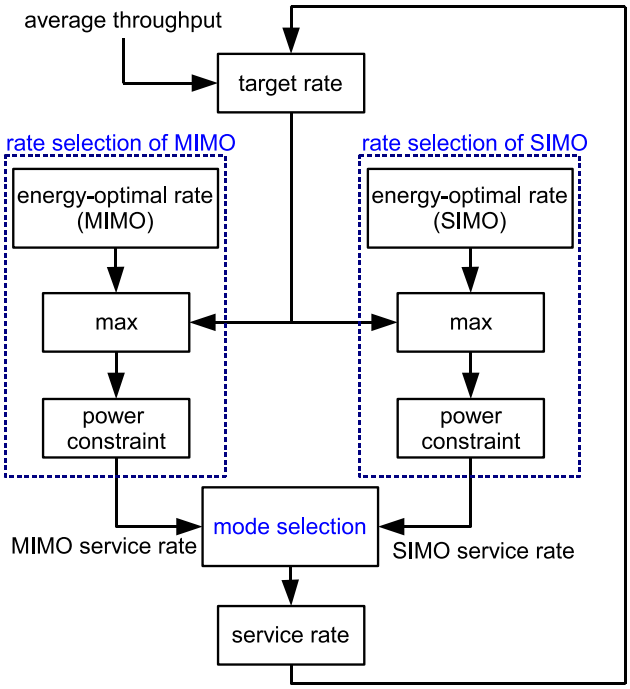


Fig. 5. Flow chart of the proposed algorithm.

output power of power amplifier is 27.5 dBm [3], [24], [27].[9] The duration of a time frame is 5 ms [24], and the number of time frames for the simulation is 1,000,000.

We choose $v_i(t) = \frac{1}{n(t)}$, i.e., temporally fair scheduling for our flow-level simulations. Interestingly in a dynamic system, temporally fair scheduling eventually gives more time resource to users with large files. This is because users with large files remain in the system for a long while users with small files quickly finish their uploads and leave the system. For example, suppose that user 1 with 1 Mbyte file and user 2 with 100 kbyte file share the uplink. As soon as user 2 finishes uploading, user 1 takes the whole time resource.

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 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 that 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.

Fig. 12 shows the results under the same environment of Fig. 11 except that energy-opportunistic scheduling is adopted instead of round-robin scheduling. Comparing two figures shows substantial performance improvement. For example, the delay/energy pair of (0.54 sec, 46.2 mJ) in Fig. 11 shifts to (0.37 sec, 35.3 mJ) in Fig. 12, i.e., shorter delay and smaller energy consumption. Furthermore, considering the lower bound of the energy consumption is determined by the circuit energy consumption 22.2 mJ in the case of SIMO, the energy-opportunistic scheduling achieves additional 45% reduction of transmit energy at least.

## 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

[9]Since the value of $p_{\text{filt}}$ in [15] is too low for cellular systems, we adjust it from 2.5 to 20 mW, but the simulation results are almost the same.

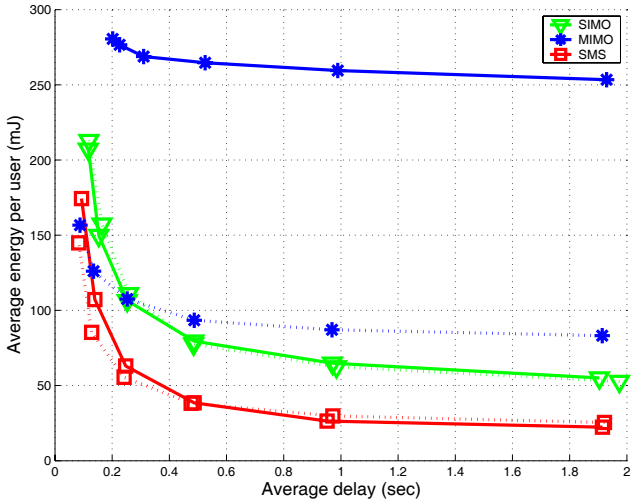


Fig. 6. Energy-delay tradeoff curves without circuit and idling power: zero forcing receiver for MIMO, $N_r = 2$, $N_t = 2$, traffic load $\rho$ = 2.51Mbps, $r_{\max}$ = 8.35Mbps, correlation coefficient $\xi = 0.7$ (solid line), $\xi = 0$ (dotted line).

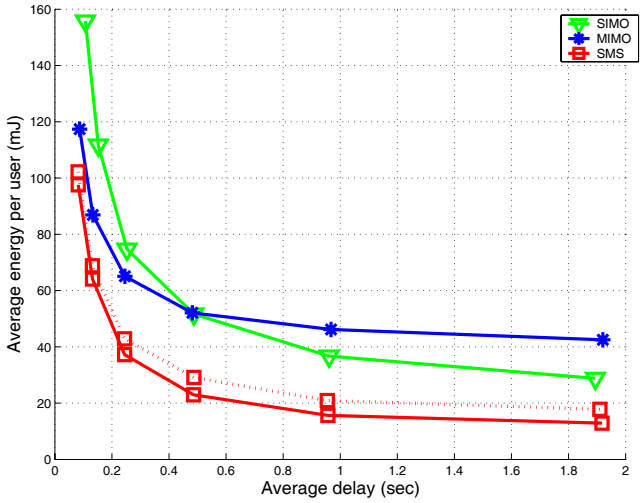


Fig. 7. Energy-delay tradeoff curves without circuit and idling power: zero forcing receiver for MIMO, $N_r = 4$, $N_t = 2$, traffic load $\rho$ = 3.70Mbps, $r_{\max}$ = 12.34Mbps. Dotted line shows SMS with random antenna selection, i.e, without using 1 bit antenna selection indicator for SIMO.

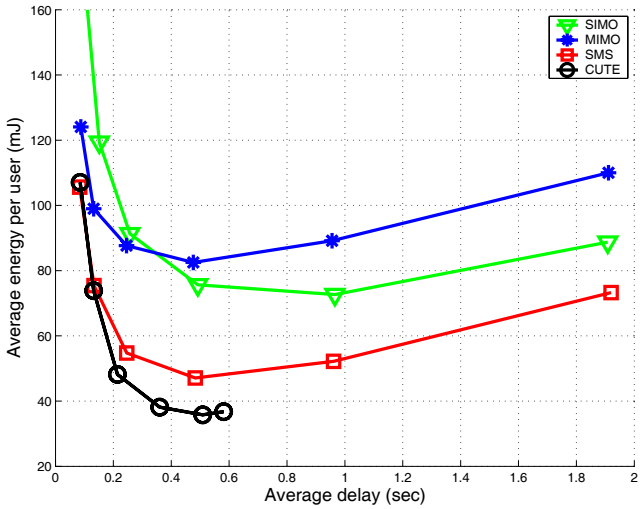


Fig. 8. Energy-delay tradeoff curves *with* circuit and idling power: zero forcing receiver for MIMO, $N_r = 4$, $N_t = 2$, traffic load $\rho$ = 3.70Mbps, $r_{\max}$ = 12.34Mbps.

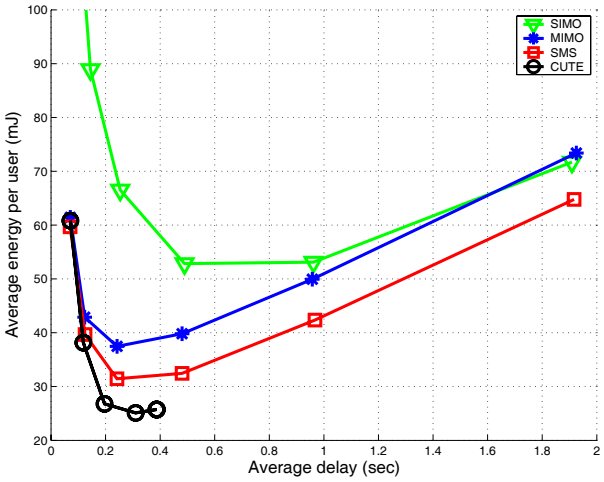


Fig. 9. Energy-delay tradeoff curves *with* circuit and idling power: zero forcing receiver for MIMO, $N_r = 8$, $N_t = 2$, traffic load $\rho$ = 4.51Mbps, $r_{\max}$ = 15.04Mbps.

rates, and MIMO consumes more circuit power than SIMO. 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. 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 $N_r$. 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. Finally, we proposed the energy-opportunistic scheduling to further enhance the energy efficiency. Investigating the optimality of energy-delay tradeoff in a dynamic system remains for future research.

## REFERENCES

[1] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. Infom. Theory*, vol. 50, pp. 125-144, 2004.

[2] B. Prabhakar, E. U. Biyikoglu, and A. E Gamal, "Energy-efficient transmission over a wireless link via lazy packet scheduling," in *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 J. Select. Areas Commun.*, 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 methology for

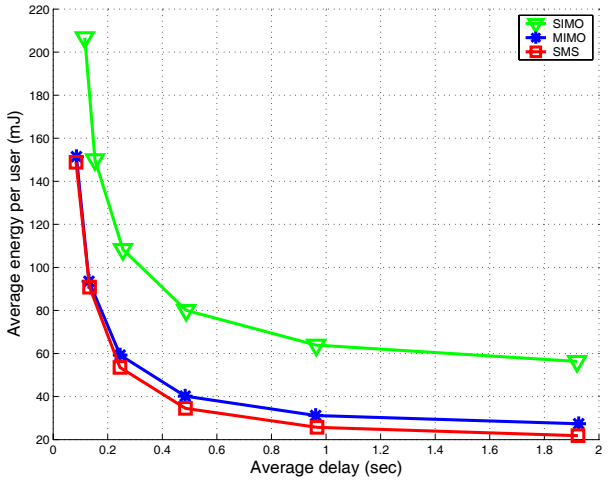


Fig. 10. Energy-delay tradeoff curves without circuit and idling power: ideal receiver for MIMO, $N_r = 2$, $N_t = 2$, traffic load $\rho$ = 2.51Mbps, $r_{\max}$ = 8.36Mbps.

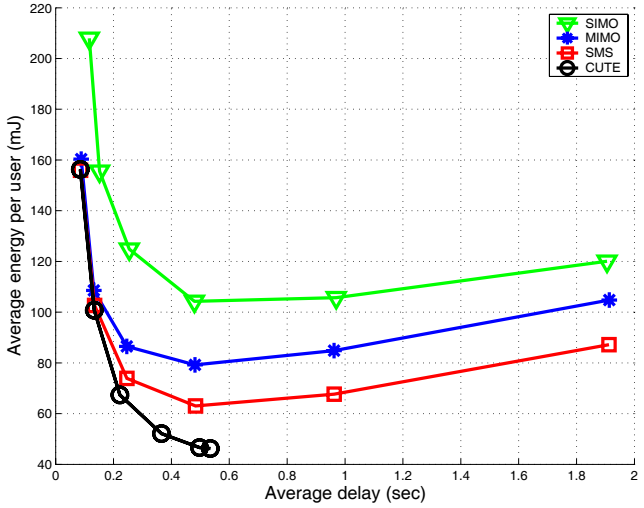


Fig. 11. Energy-delay tradeoff curves *with* circuit and idling power: ideal receiver for MIMO, $N_r = 2$, $N_t = 2$, traffic load $\rho$ = 2.51Mbps, $r_{\max}$ = 8.36Mbps.

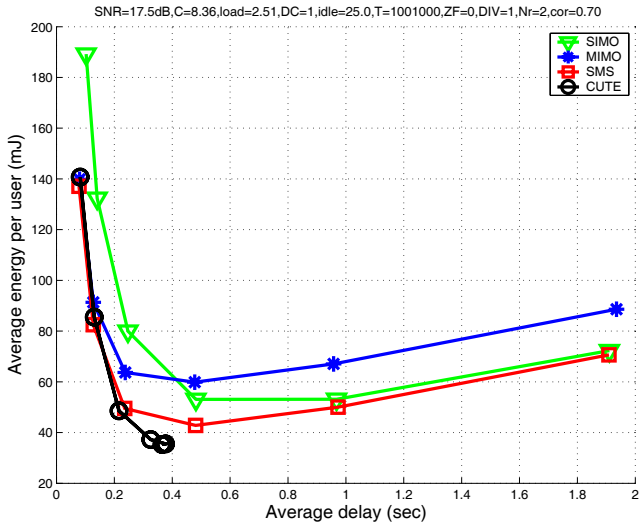


Fig. 12. Energy-delay tradeoff curves based on energy-opportunistic scheduling (circuit and idling power included): ideal receiver for MIMO, $N_r = 2$, $N_t = 2$, traffic load $\rho$ = 2.51Mbps, $r_{\max}$ = 8.36Mbps, $\nu = 0$.

energy efficient resource allocation in wireless networks," *IEEE Trans. Wireless Commun.*, 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 J. Select. Areas Commun.*, 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 Commun. 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. Commun.*, 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. 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," in *Proc. IEEE WCNC*, pp. 87-92, Mar. 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 Commun. Mag.*, Mar. 2007.

[13] D. S. Michalopoulos, A. S. Lioumpas, and G. K. Karagiannidis, "Increasing power efficiency in transmitter diversity systems under error performance constraints," *IEEE Trans. Commun.*, vol. 56, pp. 2025-2029, Dec. 2008.

[14] C.-K. Wen, K.-K. Wong, P. Ting, and C.-L. I, "Optimal power-saving input covariance for MIMO wireless systems exploiting only channel spatial correlations," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2005, pp. 2011-2015.

[15] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2349-32360, Sept. 2005.

[16] G. Miao, N. Himayat, Y. Li, and D. Bormann, "Energy efficient design in wireless OFDMA," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2008.

[17] B. Bougard, G. Lenoir, A. Dejonghe, L. V. der Perre, F. Catthoor, and W. Dehaene, "SmartMIMO: an energy-aware adaptive MIMO-OFDM radio link control for next-generation wireless local area networks," *EURASIP J. Wireless Commun. Networking*, vol. 2007, pp. 1-15, 2007.

[18] S. Cui, R. Madan, A. J. Goldsmith, 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.

[19] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," in *Proc. IEEE INFOCOM*, vol. 1, pp. 321-331, Apr. 2003.

[20] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[21] C. Martin and B. Ottersten, "Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1350-1359, July 2004.

[22] M. Kodialam, T. V. Lakshman, and S. Sengupta, "Traffic-oblivious routing for guaranteed bandwidth performance," *IEEE Commun. Mag.*, pp. 46-51, Apr. 2007.

[23] Goliath, "Financial assessment of citywide Wi-Fi/WiMAX deployment," [Online]. Available: http://goliath.ecnext.com/coms2/summary_0199-6709510_ITM, 2006.

[24] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX*. Prentice Hall, 2007.

[25] B. Rengarajan and G. de Veciana, "Architecture and abstraction for environment and traffic aware system-level coordination of wireless networks: the downlink case," in *Proc. IEEE INFOCOM*, pp. 1175-1183, 2008.

[26] 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.

[27] WiMAX power amplifier ADL5570 and 5571, Analog Devices, Sept. 2007. [Online]. Available: http://www.analog.com/uploadedfiless/data_sheets/adl5570.pdf, http://www.analog.com/uploadedfiless/data_sheets/adl5571.pdf

[28] H. Kim, C.-B. Chae, G. de Veciana, and R. W. Heath Jr., "Energy-efficient adaptive MIMO systems leveraging dynamic spare capacity," in *Proc. Conf. Inform. Sciences Systems (CISS)*, Mar. 2008.

[29] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 1995, pp. 331-335.

**Hongseok Kim** (S'06) received his B.S. and M.S. degree in Electrical Engineering from Seoul National University in 1998 and 2000, respectively. He is currently working towards the Ph.D. degree in Electrical and Computer Engineering at the University of Texas at Austin. From 2000 to 2005 he was a member of technical staff at Korea Telecom Network Laboratory. He participated in FSAN and ITU-T SG16 FS-VDSL standardization from 2000 to 2003. From 2008 June to December he worked at Intel Standard Division led by S. Alamouti. His research focuses on network resource allocation/optimization, energy-efficient wireless systems, flow-level network dynamics, and cross-layer design of MAC/PHY. He is the recipient of Korea Government Overseas Scholarship in 2005.

**Chan-Byoung Chae** (S'06 - M'09) is a Post-Doctoral Fellow/Lecturer in the School of Engineering and Applied Sciences (SEAS) at Harvard University, Cambridge, MA. He received the Ph. D. degree in the Electrical and Computer Engineering from the University of Texas (UT), Austin, TX in 2008. He was a member of the Wireless Networking and Communications Group (WNCG) at UT.

Prior to joining UT, he was a Research Engineer at the Telecommunications R&D Center, Samsung Electronics, Suwon, Korea, from 2001 to 2005. He was a visiting scholar at the WING Lab, Aalborg University, Denmark in 2004 and at University of Minnesota in August 2007. He participated in the IEEE 802.16e standardization, where he made several contributions and filed a number of related patents from 2004 to 2005. His current research interests include capacity analysis and interference management in wireless mobile networks and all aspects of MIMO communications.

Dr. Chae is the recipient of the IEEE Dan. E. Noble Fellowship in 2008, the Gold Prize in the 14th Humantech Paper Contest, and the KSEA-KUSCO scholarship in 2007. He also received the Korea Government Fellowship (KOSEF) during his Ph. D. studies.

**Gustavo de Veciana** (S'88-M'94-SM'01-F'09) received his B.S., M.S., and Ph.D. in Electrical Engineering from the University of California at Berkeley in 1987, 1990, and 1993 respectively. He is currently a Professor at the Department of Electrical and Computer Engineering at the University of Texas at Austin. He served as the Associate Director and then Director of the Wireless Networking and Communications Group (WNCG) 2004-2008. His research focuses on the design, analysis and control of telecommunication networks. Current interests include: measurement, modeling and performance evaluation; wireless and sensor networks; architectures and algorithms to design reliable computing and network systems. Dr. de Veciana has served as editor for the IEEE/ACM TRANSACTIONS ON NETWORKING, and as co-chair of ACM CoNEXT 2008. He is the recipient of General Motors Foundation Centennial Fellowship in Electrical Engineering, an NSF Foundation CAREER Award 1996, co-recipient of the IEEE William McCalla Best ICCAD Paper Award 2000, and co-recipient of the Best Paper in ACM Transactions on Design Automation of Electronic Systems, 2002-2004.

**Robert W. Heath, Jr.** (S'96 - M'01 - SM'06) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, VA, in 1996 and 1997 respectively, and the Ph.D. from Stanford University, Stanford, CA, in 2002, all in electrical engineering.

From 1998 to 2001, he was a Senior Member of the Technical Staff then a Senior Consultant at Iospan Wireless Inc, San Jose, CA where he worked on the design and implementation of the physical and link layers of the first commercial MIMO-OFDM communication system. In 2003 he founded MIMO Wireless Inc, a consulting company dedicated to the advancement of MIMO technology. Since January 2002, he has been with the Department of Electrical and Computer Engineering at The University of Texas at Austin where he is currently an Associate Professor and member of the Wireless Networking and Communications Group. His research interests include several aspects of MIMO communication: limited feedback techniques, multihop networking, multiuser MIMO, antenna design, and scheduling algorithms as well as 60GHz communication techniques and multi-media signal processing.

Dr. Heath has been an Editor for the IEEE TRANSACTIONS ON COMMUNICATION and an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is a member of the Signal Processing for Communications Technical Committee in the IEEE Signal Processing Society. He was a technical co-chair for the 2007 Fall Vehicular Technology Conference, is the general chair of the 2008 Communication Theory Workshop, and is a co-organizer of the 2009 Signal Processing for Wireless Communications Workshop. He is the recipient of the David and Doris Lybarger Endowed Faculty Fellowship in Engineering and is a registered Professional Engineer in Texas.