

RF Chain and User Selection for Multiuser MIMO Systems under Random Data Arrival

Xu Zhang, Sheng Zhou, Zhisheng Niu, Xiaokang Lin

Abstract—Multiuser Multiple input multiple output (MIMO) systems are now having more and more radio frequency (RF) chains, with larger capacity and at the same time higher energy consumption. With random data arrival, it is desired to turn off RF chains to save energy according to the traffic variations. In this paper a low-complexity traffic-aware scheme is proposed, whereby RF chains and users are selected at each frame based on the channel quality and the data queue-length. Particularly, the number of active RF chains is determined by comparing the current queue-length to the predefined thresholds, the values of which are able to control the tradeoff between the energy saving and quality of service, i.e., delay. Simulation results show that the proposed scheme saves more energy compared with conventional schemes which is designed regardless the traffic variations, and the saving gain increases when the average traffic load decreases.

Index Terms—Energy efficiency, multiuser MIMO, RF chain sleeping, user scheduling.

I. INTRODUCTION

Multiuser MIMO systems are attractive for mobile networks, since it achieves a promising capacity gain without an increase in the transmitting power, such as Massive MIMO systems [1]. However, the energy consumption increases due to the increased number of radio frequency (RF) circuits. It is predicted that the energy consumption of RF chains will dominate the energy consumption of transmissions, as its number is extremely large [2]. On the other hand, the traffic of a cell is dynamic, and there is a substantial amount of time during which the system is underutilized [3] [4]. If all RF chains are turned on regardless of the traffic variations, there will be a significant energy waste when traffic load is low. Therefore, it is desired to adjust the number of RF chains according to the traffic variations without sacrificing the Quality of Service (QoS).

In previous works, the circuit energy consumption is often neglected for energy-efficient designs in MIMO systems until Cui et. al. firstly investigated the energy efficiency of MIMO systems with the static circuit power consumption [5]. They show that MIMO systems with Alamouti code are not always energy-efficient compared with single input single output (SISO) systems. Then many works focused on energy-efficient designs through link adaptation [2] [6] [7] [8]. In [2], a RF

chain selection scheme is proposed to maximize the energy-efficiency of a massive MIMO system by turning off the unselected RF chains. By further considering the data rate requirements, the number of RF chains is optimized in [6] for MIMO-OFDMA systems, and the RF chain selection schemes are proposed in [7] and [8] for multiuser MIMO systems and cooperated BS transmissions respectively.

Previous RF chain adaptation schemes are commonly designed according to the channel variations, and generally ignore the traffic variations. Those schemes work well when the data arrives constantly or the traffic load is very high. However, if the adaptive scheme neglects the traffic variations, a substantial amount of energy will be wasted when the traffic load is low. Moreover, Active Antenna Array Systems are proposed in 3GPP [9] recently, which provides better support for the dynamic control of RF chains. Therefore, it is worthwhile to investigate traffic-aware adaptive schemes for better energy efficiency.

Notice that user selection should be jointly optimized with RF chain selection. Because that some users may not have buffered packets temporarily or can wait for longer time due to the low data arrival rate. Therefore, in this paper, a low-complexity traffic-aware RF chain and user selection scheme is proposed. At the beginning of each frame, the sets of active RF chains and users are selected based on both channel quality and the queue-length. The variations of queue-length are actually due to those control actions and the randomly arrived traffic. When the queue-length is stabilized within a certain range, it indicates that the control actions match with the traffic [10]. Therefore, the traffic pattern can be obtained by only observing the variations of the queue-length.

In the proposed scheme, the number of active RF chains is selected based on the comparison between the current queue-length and predefined threshold values. Then, the set of active RF chains is selected through maximizing the data rate of the user who has the longest queue. At last, the users are selected based on a greedy algorithm which aims to maximize the rate-weighted-queue. The complexity of the proposed heuristic policy is also analyzed. Moreover, simulation results indicates that the performance of the proposed scheme saves much more energy compared with previous traffic-unaware schemes, especially for the cases with low average traffic load or large traffic variations.

The rest of this paper is organized as follows. In section II, the system model is described. In section III, a low-complexity RF chain and user selection scheme is proposed. The complexity of the scheme is also analyzed in this section. Simulation results are shown in section IV. At last, section V

The authors are with Tsinghua National Laboratory for Information and Technology, Dept. of Electronic Engineering, Tsinghua University, Beijing 100084, P.R. China. Email: zhang-xu10@mails.tsinghua.edu.cn, sheng.zhou, niuzhs@tsinghua.edu.cn, linxk@sz.tsinghua.edu.cn.

This work is sponsored in part by the National Basic Research Program of China (973 Program: 2012CB316001), the National Science Foundation of China (NSFC) under grant No. 61201191, and No. 61321061, Hitachi R&D Headquarter.

concludes this paper.

II. SYSTEM MODEL

A. Multiuser MIMO System

A downlink multiuser MISO system with one BS and K single-antenna users is considered as shown in Fig. 1. There are M RF chains at the BS. The set of all RF chains and users are denoted by \mathcal{M} and \mathcal{S} respectively, where $|\mathcal{M}| = M$, $|\mathcal{S}| = K$, and $|\cdot|$ denotes the cardinality of a set. We assume that each RF chain corresponds to one transmit antenna and therefore we do not distinguish the transmit antenna from RF chain in the sequel. The channel is assumed to be a Rayleigh block-fading channel. The duration of one transmission frame is assumed to be the duration of one block. This assumption is also reasonable from the perspective of practical systems. For example, a 100 km/hr mobile operating at 2GHz has a fading interval of about 5 ms. On the other hand, the minimum transmission slot of current LTE (Long-Term Evolution) networks [11] is 1 ms which is smaller than the fading interval. A transmission frame can be constructed by multiple transmission slots. The channel gain matrix of one frame is denoted by $\mathbf{G} = [\mathbf{g}_1^T, \mathbf{g}_2^T, \dots, \mathbf{g}_K^T]^T \in \mathbb{C}^{M \times K}$, where $[\cdot]^T$ denotes the matrix transpose, and $\mathbf{g}_k \in \mathbb{C}^{1 \times M}$ denotes the channel gain vector from all RF chains to user k which is the product of the channel coefficient vector and the square root of the path loss of user k . The perfect channel state information (CSI) is assumed to be known at the BS.

Denote the set of *active* RF chains as $\mathcal{M}_A \subset \mathcal{M}$ and $|\mathcal{M}_A| = m$. The received signal of the k -th user is

$$y_k = \sqrt{P_k} \tilde{\mathbf{g}}_k \mathbf{w}_k s_k + \sum_{j=1, j \neq k}^K \sqrt{P_j} \tilde{\mathbf{g}}_k \mathbf{w}_j s_j + n_k, \quad (1)$$

where $\mathbf{w}_k \in \mathbb{C}^{m \times 1}$ is the beamforming weight vector, s_k is the transmitted data symbol of user k on the downlink, P_k is the allocated power to user k , $n_k \in \mathbb{C}$ denotes the additive white Gaussian noise with the zero mean and variance one, and $\tilde{\mathbf{g}}_k \in \mathbb{C}^{1 \times m}$ denotes the channel gain vector with the active RF chains.

The zero-forcing beamforming (ZFBF) [12] is assumed. The precoding vectors satisfy the zero-interference condition $\tilde{\mathbf{g}}_k \mathbf{w}_j = 0$, for $j \neq k$. Let $\tilde{\mathbf{G}}(\mathcal{S}_A, \mathcal{M}_A) = [\tilde{\mathbf{g}}_1^T \dots \tilde{\mathbf{g}}_{|\mathcal{S}_A|}^T]^T$, where $\mathcal{S}_A \subset \mathcal{S}$ is the *selected* user set. The sum data rate is given by

$$R(\mathcal{S}_A, \mathcal{M}_A) = \sum_{k \in \mathcal{S}_A} \log_2[1 + P_k \gamma_k], \quad (2)$$

where $\sum_{k \in \mathcal{S}_A} P_k \leq P$, P is the total transmitting power, and $\gamma_k = \frac{1}{\|\mathbf{w}_k\|^2} = \frac{1}{[(\tilde{\mathbf{G}}(\mathcal{S}_A, \mathcal{M}_A) \tilde{\mathbf{G}}(\mathcal{S}_A, \mathcal{M}_A)^*)^{-1}]_{k,k}}$ is the effective channel gain.

B. Power Consumption Model

According to [13] [14] [15], the energy consumption model of a BS can be divided into three parts: the pre-transceiver part, the transceiver part and the power amplifier part. In the pre-transceiver part, the energy consumption is a constant value which includes the energy consumption of cooling systems and cable losses at the BS. This part of energy consumption

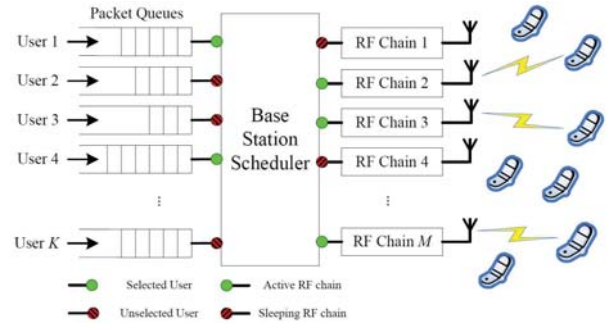


Fig. 1. A Downlink multiuser MIMO system with RF chain sleeping and user scheduling.

does not change when the number of RF chains is changed. The transceiver part includes the circuit-energy consumption of the transceiver except for the amplifier, such as the digital-to-analog converter, the mixer, the transmission filter and the synthesizer. The energy consumption of this part is highly related to the operation mode. It is assumed that there are three operation modes: *Active mode*, *Non-active mode*, and *Transient mode*.

Active mode: All components of the transceiver circuit are active.

Non-active mode: Only the synthesizer is working.

Transient mode: It occurs when the transmitter switches between different modes such as sleep-to-active and active-to-sleep.¹

For the last part, it includes the energy consumption of the power amplifier and is determined by the transmitting power.

Therefore, the total energy consumption at frame t is given by

$$\begin{aligned} E_{\text{total}}(t) &= \tau \left[\frac{P}{\zeta} + m(t)P_{\text{RF}} + P_{\text{non-Act}} + P_{\text{FIX}} \right] \\ &= \tau \left[\frac{P}{\zeta} + m(t)P_{\text{RF}} + P_c \right], \end{aligned} \quad (3)$$

where τ denotes the frame length, P is the total transmitting power, ζ denotes the efficiency of power amplifier, $m(t) \in \{0, 1, \dots, M\}$ denotes the number of active RF chains at frame t , P_{RF} denotes the power consumption of the transceiver circuit except the energy consumed by the non-active mode and the amplifier, $P_{\text{non-Act}}$ denotes the power consumption of the transceiver circuit when it is in the non-active mode, P_{FIX} is the power consumption of cooling systems and cable losses of the base station, and $P_c = P_{\text{non-Act}} + P_{\text{FIX}}$.

The energy consumption of the pre-transceiver part is τP_{FIX} ; the energy consumption of the transceiver part is $\tau [m(t)P_{\text{RF}} + P_{\text{non-Act}}]$; the energy consumption of the power amplifier part is $\tau \frac{P}{\zeta}$. When all RF chains are sleeping, the total energy consumption equals to τP_c .

C. Queuing Model

As shown in Fig. 1, the BS keeps a data queue with an infinite buffer for each user, and thus there are totally K

¹Since the duration of the transient mode is very small, thus it is neglected [15].

queues at the BS. User's packets arrive randomly according to a stationary and ergodic processes with mean $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_K\}$. The arrived data during frame t is denoted by $\mathbf{a}(t) = \{a_1(t), a_2(t), \dots, a_K(t)\}$, where $a_k(t)$ denotes the newly arrived data of user k during the t -th frame. Let $Q_k(t)$ represents the number of bits buffered at the beginning of t -th frame for user k . It is a non-negative value. At the beginning of each frame, if there are no data for transmission, the BS becomes idle and does not perform any transmissions. The queue length of each user evolves as

$$Q_k(t+1) = [Q_k(t) - r_k(t)]^+ + a_k(t), \forall k \in \{1, 2, \dots, K\}, \quad (4)$$

where $r_k(t)$ denotes the transmitted data per frame of user k at frame t and $[x]^+$ is the operation of selecting the largest value between x and 0.

III. LOW-COMPLEXITY RF CHAIN AND USER SELECTION

In order to exploit the ERG, a heuristic traffic-aware RF chain and user selection scheme is proposed which decides the selection of RF chains and users at the beginning of each frame. In the proposed scheme, the selection is accomplished by three steps: 1. Decide the number of active RF chains; 2. Decide the set of active RF chains; 3. Decide the set of scheduled users.

A. Decide the number of active RF chains

To decide the number of active RF chains, a threshold-based policy is proposed. The basic idea is to select the number of active RF chains based on the sum queue-length of the system. When the sum of the queue-length increases, the selected number of active RF chains should increase. Otherwise, smaller number of active RF chains is selected.

The candidate set for selection is given by $\{0, 1, \dots, M\}$, which includes all possible number of active RF chains. Accordingly, there exists M threshold values denoted as $\{\theta_1, \theta_2, \dots, \theta_M\}$, where $\theta_1 = 0$ and $\theta_1 < \theta_2 < \dots < \theta_M$. The selection policy is defined as follows:

$$m(t) = \begin{cases} j - 1, & \text{if } \theta_j < \sum_k Q_k(t) \leq \theta_{j+1}, \\ \forall j \in \{1, 2, \dots, M - 1\} \\ M, & \text{else} \end{cases}, \quad (5)$$

where $m(t)$ is the *selected* number of active RF chains at frame t .

The threshold values $\{\theta_1, \theta_2, \dots, \theta_M\}$ are determined by the QoS requirements of users. In this proposed policy, a constant interval value is assumed $\nu = \theta_{j+1} - \theta_j, \forall j \in \{1, 2, \dots, M - 1\}$. The smaller value of ν , the more active RF chains is eagerly used even when the sum queue-length is same. It indicates that a smaller value of ν leads to a higher QoS, but also a more energy consumption. In the following sections, simulation results show that, when a larger value of ν is used, the system saves more energy, but it results in a longer average queue-length.

B. Select the RF chain set and User set

For the selection of RF chains and users, the well-known maximum weighted queue (MWQ) scheduling policy is used [10]. This policy stabilizes the system whenever the data arrival rate vector locates within the capacity region, i.e., the throughput optimal policy. Then, at the beginning of each frame, based on the queue-length and channel state information, the sets of RF chains and users are selected by solving the following optimization problem.

$$\max_{\substack{\mathcal{S}_A(t), \mathcal{M}_A(t) \in \mathcal{A} \\ \sum_{k \in \mathcal{S}_A(t)} P_k \leq P}} \left\{ \sum_{k \in \mathcal{S}_A(t)} Q_k(t) r_k(\mathcal{S}_A(t), \mathcal{M}_A(t)) \right\}, \quad (6)$$

where the expression of the data rate is shown in (2), and the inner maximization is the power allocation which maximizes the queue-weighted-rate which can be solved by iterative water-filling algorithm [16]. Since the decisions is per-frame based, the frame index t is omitted in this section.

Then, define the set function $U(\mathcal{S}_A, \mathcal{M}_A)$ as the following:

$$U(\mathcal{S}_A, \mathcal{M}_A) = \begin{cases} \max_{\sum_{k \in \mathcal{S}_A} P_k \leq P} \sum_{k \in \mathcal{S}_A} Q_k \log_2(1 + P_k \gamma_k), & (\mathcal{S}_A, \mathcal{M}_A) \neq \emptyset \\ 0, & (\mathcal{S}_A, \mathcal{M}_A) = \emptyset, \end{cases} \quad (7)$$

where $(\mathcal{S}_A, \mathcal{M}_A) \in \mathcal{A}$. The selected RF chains and users should maximize this set function $U(\mathcal{S}_A, \mathcal{M}_A)$ according to MWQ scheduling policy.

1) *RF Chain Selection*: The selection of active RF chains determines the spatial signal space which affects the transmission data rate of each user. Suppose all RF chains are active, the channel gain vector $\mathbf{g}_k = \{g_k^1, g_k^2, \dots, g_k^M\}$, $\mathbf{g}_k \in \mathbb{C}^{1 \times M}$ can be seen as the coordinate of user k where he/she locates in the spatial signal space. With the linear precoding, the orthogonality of those *selected* users determines the data rates. When some of RF chains are deactive by the selection scheme, the spatial signal space is changed and the orthogonality relationships between those selected users are changed. In order to maximize the sum data rate, there are a lot of works which aim to obtain the optimal set of active RF chains with low complexity algorithms such as [17]. However, those previous works do not work in our case. It is because that the selection scheme of this paper is not only require to maximize the sum data rate, but also the queue-length is required to be jointly considered, as shown in (7).

Since the queue-length of each user can be seen as the urgency of its transmissions, the users are actually weighted by its own urgency. Thus, the users with larger weighted value should have higher priority. This insight motivates the heuristic scheme design. For the antenna selection, the antenna set should maximize the data rate of the user with the longest queue-length. This heuristic algorithm is shown as follows:

(a) The largest queue-length user is selected by $k^* = \arg \max_k Q_k$;

(b) The set of active RF chains is selected by maximizing the data rate of user k^* with the desired number of active RF chains m which has been decided by (5).

2) *User Selection*: Upon deciding the set of active RF chains, the objective function (7) is only determined by the set of users. A greedy algorithm is proposed to select the set of users, where one user is selected at a time among all of the unselected users by maximizing the incremental contribution of $U(\mathcal{S}_A, \mathcal{M}_A)$ until the number reaches m or the remaining users can not provide any increments.

Suppose $\mathcal{S}_A \subset \mathcal{S}$ and $\mathcal{S}'_A \subset \mathcal{S}$, where $\mathcal{S}'_A = \mathcal{S}_A \cup \{j\}$, $\{j\} \in \mathcal{S}$ and $\{j\} \cap \mathcal{S}_A = \emptyset$. The user selection scheme is as follows:

(a) At each iteration, the user among all of the unselected users is selected by maximizing $U(\mathcal{S}'_A, \mathcal{M}_A) - U(\mathcal{S}_A, \mathcal{M}_A)$;

(b) If the number of selected users equals to the desired number of active RF chains m which has been decided by (5), or there are no remaining users who satisfy $U(\mathcal{S}'_A, \mathcal{M}_A) - U(\mathcal{S}_A, \mathcal{M}_A) > 0$, the selection algorithm stops.

C. Complexity Analysis

In this subsection, the complexity of the proposed scheme is analyzed. As the performance benchmark, the exhaustive search algorithm is considered to obtain the optimal sets of RF chains and users based on a stochastic optimization [10]. Since it requires to search all possible sets of active RF chains and users, the total number of possible combinations is given by

$$\sum_{m=1}^M \binom{M}{m} \left[\sum_{k=1}^m \binom{K}{k} \right] = \sum_{m=1}^M \sum_{k=1}^m \left(\frac{M^m + \dots}{m!} \right) \left(\frac{K^k + \dots}{k!} \right), \quad (8)$$

where $\binom{M}{m}$ is the total number of possible sets of active RF chains when the number of active RF chains is m and $\binom{K}{k}$ is the total number of possible sets of users when the number of active RF chains is k . For each set of users \mathcal{S}_A and active RF chains \mathcal{M}_A , we need to calculate $[\mathbf{G}(\mathcal{S}_A, \mathcal{M}_A)\mathbf{G}(\mathcal{S}_A, \mathcal{M}_A)^*]^{-1}$ in the sum data rate as shown in (2) based on the stochastic optimization. It includes a matrix multiplication and a matrix inversion. Since the dimension of $\mathbf{G}(\mathcal{S}_A, \mathcal{M}_A)$ is $|\mathcal{S}_A| \times |\mathcal{M}_A|$, the complexity is on the order of $\mathcal{O}(|\mathcal{S}_A|^2 |\mathcal{M}_A|)$ for the matrix multiplication and $\mathcal{O}(|\mathcal{S}_A|^3)$ for the matrix inversion [18]. Therefore, the total complexity is

$$\begin{aligned} & \mathcal{O} \left[\sum_{m=1}^M \sum_{k=1}^m \left(\frac{M^m + \dots}{m!} \right) \left(\frac{K^k + \dots}{k!} \right) (k^2 m + k^3) \right] \\ & \approx \mathcal{O} \left[\sum_{m=1}^M \sum_{k=1}^m \left(\frac{M^m K^k}{m! k!} (k^2 m + k^3) \right) \right]. \end{aligned} \quad (9)$$

On the other hand, there are three steps to select the sets of active RF chains and users in the proposed scheme:

In the first step, the optimal number of active RF chains is obtained based on the total queue-length. With the binary search, the complexity is on the order of $\mathcal{O}(\log M)$.

In the second step, the RF chains are selected based on the channel gain of the user who has the longest queue-length. Firstly, the search of the longest queue-length user is achieved by the binary search with the complexity $\mathcal{O}(\log K)$. Secondly, suppose the number of active RF chains is m , $m \in \{0, 1, \dots, M\}$, the set of active RF chains is obtained

by finding m largest norms in the channel gain vector $\mathbf{g}_{k^*} = \{g_{k^*}^1, g_{k^*}^2, \dots, g_{k^*}^M\}$, $\mathbf{g}_{k^*} \in \mathbb{C}^{1 \times M}$, where k^* represents the *selected* user. Then, the set of active RF chains is found by sorting algorithms with the complexity $\mathcal{O}(M \log M)$ [19].

In the third step, the greedy user selection scheme is employed. Considering the worst case, the optimal number of RF chains is M . Then, the total number of iterations for the user selection is $\sum_{k=1}^M (K - k)$. For each iteration, we need to calculate $[\mathbf{G}(\mathcal{S}'_A, \mathcal{M}_A)\mathbf{G}(\mathcal{S}'_A, \mathcal{M}_A)^*]^{-1}$. Therefore, the complexity of the user selection is

$$\mathcal{O} \left\{ \sum_{k=1}^M (K - k) (k^3 + k^2 M) \right\} \approx \mathcal{O}(KM^3). \quad (10)$$

Consequently, the total computation complexity of the proposed scheme is given by

$$\mathcal{O}(KM^3 + M \log M + \log K + \log M) \approx \mathcal{O}(KM^3), \quad (11)$$

which is much smaller than the exhaustive search algorithm. For example, when $M = 8$ and $K = 10$, the complexity reduces at least 10^3 times by using the heuristic scheme. Moreover, the reduction of complexity increases with the scale of the system, e.g., when $M = 16$ and $K = 10$, the complexity reduces at least 10^5 times.

D. Discussion for Realistic Implementation

In the proposed scheme, the perfect CSI is assumed at the transmitter, which is difficult to be realized in practical systems. Training is commonly used to obtain the channel knowledge. It can be roughly classified into two groups: (1) channel feedback [20] [21], in which the CSI (either in a quantized version or an analog version) is obtained through a feedback channel, (2) channel approximation [22], in which the uplink pilots is used to estimate the downlink channel coefficients by exploiting the channel reciprocity. Moreover, the *invisible CSIT problem* [7] will occur when the RF chain and user selections are performing. Notice that it is impossible to do channel estimation for those inactive antennas, since the circuits of those antennas are turned off. Therefore, one may add one dedicated training period to turn all the RF chains on or directly use the statistic CSI to perform the RF chain and user selection [7]. In addition, motivated by the channel reciprocity, the channel estimation could be achieved by using the uplink pilots. In that case, the receiving circuits of the RF chain are not turned off.

Notice that, in order to combat the invisibility and inaccuracy of the CSI, more training overhead will be induced [23]. Those overhead will eventually degrade the energy-saving performance of the proposed scheme. The tradeoff between the training overhead and the energy-saving performance is left to our future work.

IV. SIMULATION RESULTS

In order to give a general insight, we focus on a very simple downlink multiuser MISO system in which all K users are homogeneous. There are total $K = 10$ users in this system. The parameters of the power consumption model are $P =$

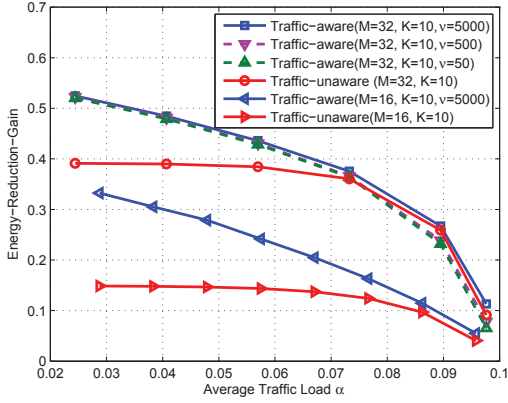


Fig. 2. The Energy-Reduction-Gain over the average traffic load, where M and K represent the total number of RF chains and users respectively, and ν represents the constant interval between the threshold values as shown in the step 1 of the proposed scheme.

2dB, $\zeta = 0.38$, $P_{\text{RF}} = 1\text{W}$, $P_{\text{non-active}} = 2\text{W}$, $P_{\text{FIX}} = 18\text{W}$ and $P_c = 20\text{W}$ according to a typical LTE (long-term-evolution) BS [13] [24] [25]. The data arrival is assumed as poisson processes.

For comparison, the *traffic-unaware scheme* is proposed according to previous works [6] [7] [8]. It is assumed that the average data arrival rate and channel state information are known. The average data arrival rate can be seen as the data rate requirement in those previous schemes. The channel state information is assumed to be known at the BS. At the beginning of each frame, the minimum number of RF chains is selected by maximizing the energy-efficiency. The antenna set selection is based on the algorithm proposed in [7].

Before giving the results, several technical terms should be defined. To characterize the energy-saving, we use the **Energy-Reduction-Gain** (ERG) [26] as the metric, which is the ratio of the saved energy to the maximum energy consumption which is given by

$$\eta = 1 - \frac{\bar{E}_f}{E_f^{\max}}, \quad (12)$$

where \bar{E}_f denotes the frame average of the energy consumption of a scheme, and $E_f^{\max} = \tau[\frac{D}{\zeta} + MP_{\text{RF}} + P_c]$ represents the maximum energy consumption without adaptive control.

Moreover, we further define the **Average Traffic Load** α as the ratio of the sum average data arrival rate to the system capacity, which is motivated by the definition of server utilization in queueing theory [27],

$$\alpha = \frac{\sum_{k=1}^K \lambda_k}{C_M}, \quad (13)$$

where C_M represents the system capacity when all RF chains are active.

In Fig. 2, the ERG performances of the proposed scheme and the traffic-unaware scheme are plotted over the average traffic load. Both of the curves decrease with the average traffic load. It is because that there are little opportunities to turn off the RF chains when the average data arrival rate approaches to the system capacity. It is also shown that the

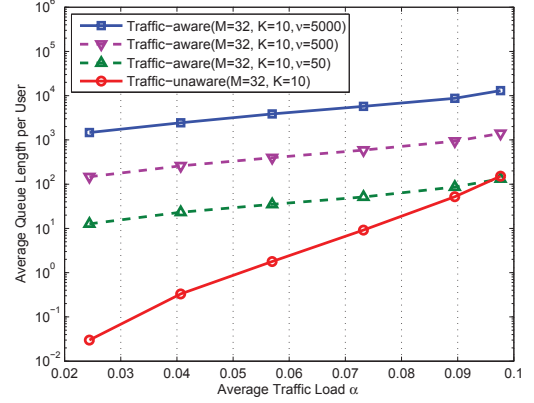


Fig. 3. Average queue-length per user over the average traffic load, where M and K represent the total number of RF chains and users respectively, and ν represents the constant interval between the threshold values as shown in the step 1 of the proposed scheme.

ERG of the proposed scheme is larger than the traffic-unaware scheme, especially when the average traffic load is low. In this case, the buffers of some users are relative low and even empty. The proposed scheme tracks the traffic and turns off some redundant RF chains to further save energy. Since the traffic-unaware scheme presumes that all users always have packets, it provides extra services even when the average traffic load is low. When the average traffic load increases, buffers of all users will be always full of packets. Then, both of schemes achieve much similar ERG performance. Moreover, the threshold interval ν of the proposed scheme affects the ERG performances. When using smaller value of ν , the ERG decreases. With small value of ν , the control actions will become more sensitive to the queue-length, and even some small amount of arrived packets will let the BS turn on more RF chains.

In Fig. 3, the average queue length performances of both the proposed scheme and the traffic-unaware scheme are plotted over the average traffic load. It is shown that the traffic-unaware scheme has much smaller average queue-length. It indicates that the proposed scheme turns off some RF chains to save energy while sacrificing some of QoS. Moreover, it is observed that the average queue-length of the proposed scheme decreases with the reduction of the value ν . Therefore, the value ν is able to control the tradeoff between the saved energy and the average queue-length.

V. CONCLUSIONS

In this paper, a low-complexity traffic-aware RF chain and user selection scheme is proposed. Unlike previous traffic-unaware schemes which neglect traffic variations, the proposed scheme controls the radio resources according to both of the traffic and channel variations. It is shown that the proposed scheme saves a great amount of energy compared with those traffic-unaware schemes under random traffic arrivals where the traffic pattern varies severely since some users frequently have empty buffers.

The threshold interval introduced by the scheme is used to control the tradeoff between the saved energy and the increased average queue-length. When the threshold interval increases, more energy can be saved but it results in a large average queue-length and thus average delay is increased. By reducing the threshold interval, the number of active RF chains will overmatch the average traffic load, which leads to better QoS, but more energy consumption.

REFERENCES

- [1] E.G. Larsson, O. Edfors, F. Tufvesson, T.L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186-195, Feb. 2014.
- [2] H. Li, L. Song, M. Debbah, "Energy efficiency of large-scale multiple antenna systems with transmit antenna selection," *IEEE Trans. Communications*, vol. 62, no. 2, pp. 638-647, Feb. 2014.
- [3] L.M. Correia, D. Zeller, Y. Jading, I. Godor, G. Auer, and L.V.D. Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 66-72, Nov. 2010.
- [4] Z. Niu, "TANGO: Traffic-aware network planning and green operation," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 25-29, Oct. 2011.
- [5] S. Cui, A. J. Goldsmith and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1089-1098, Aug. 2004.
- [6] Z. Xu, C. Yang, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient configuration of spatial and frequency resources in MIMO-OFDM systems," *IEEE Trans. Communications*, vol. 61, no. 2, pp. 564-575, Feb. 2013.
- [7] J. Xu and L. Qiu, "Energy efficiency optimization for MIMO broadcasting channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 690-701, Feb. 2013.
- [8] Q. Zhang, C. Yang, H. Haas, and J.S. Thompson, "Energy efficient downlink cooperative transmission with BS and Antenna switching off," *IEEE Trans. Wireless Communications*, vol. 13, no. 9, pp. 5183-5195, Sept. 2014.
- [9] 3GPP TR 37.840, "Study of Radio Frequency (RF) and Electromagnetic Compatibility (EMC) requirements for Active Antenna Array System (AAS) base station," Version 12.1.0, 2014.
- [10] M. J. Neely, "Stochastic network optimization with application to communication and queuing systems," Morgan & Claypool, 2010.
- [11] 3GPP TS 36.300 (2014-06), Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Overall description, Stage 2 (Release 12).
- [12] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal of Selected Areas in Communications*, vol. 24, no.3, pp. 528-541, March 2006.
- [13] R. Kumar and J. Gurugubelli, How Green the LTE Technology Can be? in *Intern. Conf. on Wireless Commun., Veh. Techn., Inform. Theory and Aerosp. Electron. Syst. Techn.*, Mar. 2011.
- [14] H. Yang, T.L. Marzetta, "Total energy efficiency of cellular large scale antenna system multiple access mobile networks," *IEEE Online Conf. Green Commun.*, 2013.
- [15] Y. Li, B. Bakkaloglu, and C. Chakrabarti, "A system level energy model and energy-quality evaluation for integrated transceiver front-ends," *IEEE Trans. Very Large Scale Integration Systems*, vol. 15, no. 1, pp. 90-103, Jan. 2007.
- [16] M. Kobayashi and G. Caire, "An iterative water-filling algorithm for maximum weighted sum-rate of gaussian MIMO-BC," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1640-1646, Aug. 2006.
- [17] P. Lin and S-H. Tsai, "Performance analysis and algorithm designs for transmit antenna selection in linearly precoded multiuser MIMO systems," *IEEE Trans. Vehicular Tech.*, vol. 61, no. 4, pp. 1698-1708, May 2012.
- [18] G.H. Golub and C.F. Van Loan, *Matrix Computation*, 3rd ed. Baltimore, MA: The Johns Hopkins Univ. Press, 1996.
- [19] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, *Introduction to Algorithms*, 3rd ed. Cambridge, MA: The MIT Press, 2009.
- [20] N. Jindal, MIMO Broadcast Channels with Finite Rate Feedback, *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045-5060, November 2006.
- [21] M. Kobayashi, On the use of multiple antennas in a downlink of wireless systems, Ph.D dissertation, ENST, June 2005.
- [22] T. L. Marzetta and B. M. Hochwald, Fast Transfer of Channel State Information in Wireless Systems, *IEEE Trans. Signal Processing*, vol. 54, no. 4, pp. 1268-1278, June 2004.
- [23] G. Caire, N. Jindal, M. Kobayashi, N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Information Theory*, vol. 56, no. 6, pp. 2845-2866, June 2010.
- [24] E. Bjornson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy-efficiency: when is massive MIMO the answer?," In *Proc. IEEE Wireless Communications and Networking Conference (WCNC) 2014*, Istanbul, Turkey, Apr. 2014.
- [25] ETSI TS 136 104 V9.4.0, Base Station (BS) radio transmission and reception (3GPP TS 36.104 version 9.4.0 Release 9), July 2010.
- [26] W. Guo and T.O. Farrell, "Dynamic cell expansion with self-organizing cooperation," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 5, pp. 851-860, May 2013.
- [27] N. Gautam, "Analysis of queues: Methods and applications," First Edition, Boca Raton, FL, USA, CRC Press, 2012.