On the Energy Delay Tradeoff of HARQ-IR in Wireless Multiuser Systems

Jinho Choi, *Senior Member, IEEE,* Jeongseok Ha, *Member, IEEE,* and Hyoungsuk Jeon, *Member, IEEE*

***Abstract*—Energy delay tradeoff (EDT) is a fundamental tradeoff that plays a crucial role in understanding the energy efficiency of various transmission schemes. In particular, for hybrid automatic repeat request (HARQ) protocols, the EDT can be well defined. In this paper, in order to understand the EDT for a wireless multiuser system, we consider various downlink transmission schemes where the HARQ with incremental redundancy (HARQ-IR) protocol is employed for reliable transmissions to users. For a given total power, the impact of different transmission schemes in conjunction with power allocation on EDT is extensively studied. Among the transmission schemes, it is shown that the threshold-based transmission (TBT) scheme efficiently lowers the minimum energy per bit (EB) in spite of limited prior channel state information (CSI) feedback. The study is further extended to a more plausible scenario where users have different channel statistics. In doing so, we consider a proportional fairness scheduling based on relative channel gain that can be easily accommodated in the proposed TBT scheme, and study its EDT behaviors. Performance evaluations show that the schemes incorporating proportional fairness achieve good EDT performances while guaranteeing the fairness among users.**

***Index Terms*—Energy efficiency, hybrid automatic repeat request (HARQ), power allocation.**

I. INTRODUCTION

ENERGY efficiency will play a fundamental role in designing future wireless systems due to increasing energy cost and limiting energy consumption to reduce CO_2 emissions. There are various issues to be addressed for energy efficient wireless systems [1]. At a circuit-level, energy efficient transceivers can be designed [2], [3], [4], [5], [6]. At a link-level, there could be various energy efficient protocols. In a multiuser system, energy efficiency at a system-level would be the most important issue and decide overall energy efficiency.

At a link-level, energy efficient transmission schemes should take into account time-varying channels in wireless communications. Using the notion of power control [7], [8], energy saving for information transmission can be achieved. A simple approach is based on on/off transmission. A transmitter can transmit signals if the channel signal-to-noise ratio (SNR) is greater than a certain threshold, while no transmission is made if the channel SNR is lower than the threshold. Furthermore, the transmission rate can depend on channel conditions using adaptive modulation and coding (AMC). In [9], the energy efficiency is studied with delay constraints when a transmitter is equipped with AMC to mitigate fading channels. As shown in [10], [11], [12], schedulers can be designed to minimize transmission energy with delay constraints over wireless channels.

In addition, energy efficient transmission scheduling is studied in [10], [11], [13] and in [14] with energy harvesting. Relay-aided transmissions in wireless communications are also investigated as they can be energy (or power) efficient [15], [16], [17], [18]. In [19], energy efficient scheduling is studied with deadline for wireless relay networks. In a relay network of three nodes, with the notion of network coding, the energy efficiency of different relaying protocols is discussed in [20]. In [21], the energy-bandwidth tradeoff is studied in multihop wireless networks.

Hybrid automatic repeat request (HARQ) protocols are link layer protocols that are employed for reliable transmissions over wireless channels. Their energy efficiency has been studied in [22]. In this paper, we study HARQ protocols to exploit the tradeoff between energy consumption and transmission delay in multiuser systems. This tradeoff is referred to as the energy delay tradeoff (EDT) throughout this paper. In particular, the HARQ with incremental redundancy (HARQ-IR) protocol is considered as it can achieve the ergodic capacity over fading channels [23]. In multiuser systems, we will show that the power allocation with *limited* prior[1] CSI feedback plays an important role in improving the EDT of HARQ-IR. In order to take into account fairness issues when users have different channel characteristics, we proposed a transmission scheme with limited prior CSI feedback in conjuction with HARQ-IR. This scheme can be easily modified for fairness scheduling and provides a good EDT performance. It is noteworthy that, we only consider transmission delay that is the reciprocal of throughput in this paper (no queuing delay is not taken into account as it depends on characteristics of source). Thus, the EDT can also be considered as the tradeoff between throughput (or spectral efficiency) and energy efficiency.

Note that in [9], [24], [25], information-theoretic approaches

Manuscript received January 14, 2013; revised April 3, 2013. The editor coordinating the review of this paper and approving it for publication was L. K. Rasmussen.

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2012R1A1B3002684).

J. Choi is with the College of Engineering, Swansea University, Swansea, SA2 8PP, UK (e-mail: j.choi@swansea.ac.uk).

J. Ha and H. Jeon are with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 305-701, Korea (e-mail: jsha@kaist.edu, h.jeon@kaist.ac.kr).

Digital Object Identifier 10.1109/TCOMM.2013.053013.130031

[1]As opposed to the CSI feedback for HARQ-IR, which is called posterior CSI feedback (as the feedback signal is sent to a transmitter from a receiver after receiving signals), the CSI feedback prior to transmission is referred to prior CSI feedback in this paper. In practice, the transmitter can have such prior CSI when channels change much slower than the latency of CSI feedbacks from users.

are studied for the tradeoff between energy efficiency and rate, while, in this paper, the tradeoff between energy efficiency and delay is investigated by employing a specific link layer protocol (i.e., HARQ-IR). Thus, although the results would be less fundamental, they could be more practical.

In summary, the main contribution of the paper is two-fold. The first is the characterization of the EDT of HARQ-IR in a multiuser system, which allows us to understand asymptotic behaviors of EDT including the minimum energy per bit. In [9], it is shown that the minimum energy per bit of AMC can be lower than the Shannon limit, -1.59 dB, over fading channels, by exploiting prior CSI in the power control. With HARQ-IR in a multiuser system, we find that the role of prior CSI is similar to that in AMC. Furthermore, there is a multiuser diversity gain that lowers the minimum energy per bit in a multiuser system. The second contribution is to find a way to improve the EDT performance by accommodating limited prior CSI with HARQ-IR in a multiuser system, which can also be extended to take into account proportional fairness.

II. Background

In this section, we present a block-fading channel model and briefly explain the HARQ-IR protocol and its EDT that can show the energy efficiency in terms of effective delay or spectral efficiency.

A. HARQ Protocols for Reliable Transmissions

Suppose that a message sequence is transmitted over a fading channel. Provided that the channel SNR is not known prior to transmissions, we can use HARQ protocols for reliable communications. For HARQ-IR, we need to have a channel encoder that can generate a codeword of an arbitrary length. A codeword is subsequently divided into $\Lambda \in \mathbb{N}^+$ subblocks. A channel decoder should also be capable of decoding with a sequence of corrupted subblocks. A detailed procedure for HARQ-IR is given below.

1) For a given message block $\mathbf{m}$ of length b, a sequence of coded subblocks is generated as $\mathbf{c}_1, \ldots, \mathbf{c}_\Lambda$, where $\mathbf{c}_k = \mathrm{ENC}_k(\mathbf{m})$ for $k \in \{1, \ldots, \Lambda\}$. We will denote by $\mathrm{ENC}_k(\cdot)$ the kth encoding operation and by $\mathrm{DEC}_k(\cdot)$ the kth decoding operation.
2) A transmitter transmits $\mathbf{c}_k$ and a receiver attempts to recover $\mathbf{m}$ with $\{\mathbf{r}_1, \ldots, \mathbf{r}_k\}$ as $\hat{\mathbf{m}}_k = \mathrm{DEC}_k(\mathbf{r}_1, \ldots, \mathbf{r}_k)$, where $\mathbf{r}_k$ is the received signal corresponding to coded subblock $\mathbf{c}_k$ and $\hat{\mathbf{m}}_k$ denotes the decoded message block obtained with k transmissions.
3) If decoding is successful, the receiver sends positive acknowledgment (ACK). Otherwise, a retransmission request (RQ) or negative acknowledgment (NACK) is sent to the transmitter. Let $k \leftarrow k+1$ and move to Step 2.

For point-to-point communications, we can consider a block-fading channel with complex passband bandwidth of B where the received signal subblock is given by

$$\mathbf{r}_k = \sqrt{\alpha_k}\mathbf{x}_k + \mathbf{n}_k.$$

Here, $\mathbf{x}_k$ is a modulated signal subblock of length $L = \lfloor BT \rfloor$ corresponding to $\mathbf{c}_k$, T is a subblock transmission duration, α_k is the channel power gain over the kth subblock, and $\mathbf{n}_k \sim \mathcal{CN}(0, \mathbf{I})$. Let $Z_k = \log_2(1 + P\alpha_k)$, where P denotes the normalized transmitted signal power or SNR. Note that in this normalization, we assume that $\mathrm{SNR} = \frac{\tilde{P}}{N_0 B} = P$, where $\tilde{P}$ is the transmission power. According to [23], there exists a code $\mathcal{C} \subset \mathbb{C}^{L\Lambda}$ such that for a sufficiently large L and any $k \in \{1, \ldots, \Lambda\}$, an error-free transmission of the message block $\mathbf{m}$ is achievable if

$$\frac{1}{k}\sum_{i=1}^{k} Z_i > \frac{R}{k}, \tag{1}$$

where R/k is the effective code rate, and $R = b/L$ called the initial rate, which is a design parameter, [26] in bits per second per Hertz (bps/Hz). In HARQ-IR, if one transmission is sufficient for successful decoding (i.e., $Z_1 > R$), the transmission rate becomes R, which would be the maximum transmission rate. However, if k transmissions are required for successful decoding, the transmission rate becomes R/k. As $\Lambda \to \infty$, the probability that there exists k satisfying the inequality in (1) approaches 1, and the number of transmissions of HARQ-IR is given by [23]

$$K(R, P) = \min\left\{k \;\middle|\; \sum_{i=1}^{k} Z_i > R\right\}. \tag{2}$$

If a non-ideal code is employed, Z_k can be modified as $Z_k = g\log_2(1 + P\alpha_k)$, where $0 < g < 1$ represents the performance gap between an ideal capacity achieving code and a non-ideal code employed. The number of transmissions is a random variable that depends on R and P.

B. Energy Delay Tradeoff

Define the average number of transmissions of HARQ-IR as

$$\bar{K}(R, P) = \mathbb{E}[K(R, P)],$$

where the expectation, denoted by $\mathbb{E}[\cdot]$, is carried out over random channel power gains, $\{\alpha_i\}$. Throughout the paper, we assume that $\mathbb{E}[K(R, P)]$ is finite for a finite R and $P > 0$. In general, $\bar{K}(R, P)$ is a nondecreasing function of R and nonincreasing function of P. The *effective delay* is defined as the ratio of the average number of transmissions, $\bar{K}(R, P)$, to the initial rate as follows:

$$\mathsf{D}(R, P) \triangleq \frac{\bar{K}(R, P)}{R}. \tag{3}$$

The reciprocal of the effective delay is the spectral efficiency or throughput of the HARQ-IR protocol in bps/Hz. It can be seen as the normalized transmission time to deliver one bit. Note that in (3) we assume that there is no packet loss. In other words, we assume that $\Pr(K(R, P) > K_{\max}) = 0$, where $K_{\max}$ is the maximum number of transmissions. In (2), we assume the maximum number of transmissions is infinite, i.e., $K_{\max} = \infty$. Using Markov inequality, it can be shown that

$$\Pr(K(R, P) > K_{\max}) \leq \frac{\mathbb{E}[K(R, P)]}{K_{\max}}.$$

Provided that $\mathbb{E}[K(R, P)]$ is finite, we can show that there is no packet loss if $K_{\max} \to \infty$.

We can now define the energy per bit (EB) [24], which is the ratio of the SNR to the spectral efficiency or the product of the SNR and the effective delay, as

$$\mathsf{EB}(R,P) \triangleq \frac{P}{R/\bar{K}(R,P)} = P\mathsf{D}(R,P). \tag{4}$$

Note that the dimension of EB is identical to that of E_b/N_0 in [24], [9], where E_b represents the bit energy. From (3) and (4), we can see the relationship between EB and delay. Using the pair of $\{\mathsf{EB}(R,P), \mathsf{D}(R,P)\}$, which gives rise to the EDT curve, we can characterize the EDT of HARQ-IR for wireless communication systems.

In order to see the energy efficiency through EDT, we can consider an example for point-to-point communications. Suppose that the channel power gains are given by $\{\alpha_1, \ldots, \alpha_6\} = \{1, 3, 0, 1, 2, 1\}$, and two different power levels are used: $P_1 = 10$ and $P_2 = 3$. If $R = 10$, with P_1, the required number of transmissions is $K(R, P_1) = 4$ as $\sum_{i=1}^{3} \log_2(1 + P_1\alpha_i) = 8.41 < R < \sum_{i=1}^{4} \log_2(1 + P_1\alpha_i) = 11.87$. For the case of $R = 10$ and $P_2 = 3$, the required number of transmissions is $K(R, P_2) = 5$ as $\sum_{i=1}^{4} \log_2(1+P_2\alpha_i) = 7.32 < R < \sum_{i=1}^{5} \log_2(1+P_2\alpha_i) = 10.12$. Although P_2 results in a longer delay, its EB is $\mathsf{EB} = \frac{5P_2}{R} = 1.5$ that is lower than the EB with P_1 which is $\mathsf{EB} = \frac{4P_1}{R} = 4$.

Fig. 1 shows typical EDT curves for various values of R when the channel power gains, α_k, are independent and exponentially distributed (i.e., Rayleigh fading channels) and $\mathbb{E}[\alpha_k] = 1$. As R increases, the EDT of HARQ-IR can be improved. This improvement results from the increasing throughput of HARQ-IR when R increases [23]. In general, a large P or SNR requires a higher order modulation and a channel code of a high code rate, which may not be practical yet in most wireless communication systems. However, as shown in Fig. 1, as the effective delay increases (or the spectral efficiency decreases), the difference of EB's between large and small values of R vanishes. This implies that a system of a low spectral efficiency or a long effective delay can be energy efficient even with a low R. In addition, since R can be low without a significant loss of energy efficiency, its implementation would be affordable.

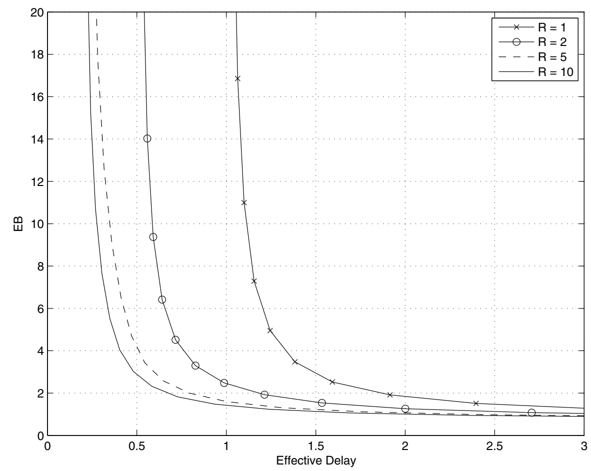


Fig. 1. EDT curves of HARQ-IR with various values of R.

III. Impact of Power Allocation on EDT in Multiuser HARQ-IR

In this section, we consider different power allocation approaches and their impact on the EDT performance of HARQ-IR in a multiuser system when R is fixed. As an example of a multiuser system, we consider downlink channels in a cellular system consisting of one base station (BS) and M users. Each user has an orthogonal channel, and the BS tries to allocate the total transmission power, denoted by P_T, to M orthogonal channels in an energy efficient way. Note that the bandwidth is an important resource and should also be optimally allocated to maximize the performance. However, in this paper, since we mainly focus on the energy efficiency, in particular, when the transmission power is low (i.e., the low power regime), we do not consider the bandwidth allocation and rather assume no constraint on the bandwidth.

We denote by $\alpha_{m,k}$ the channel power gain for the mth user over the kth subblock and consider the Rayleigh fading channels unless stated otherwise. In particular, we have the following assumption:

A) $\{\alpha_{m,k}\}$ are independent and identically distributed (iid) across users and subblocks and have the standard exponential probability density function (pdf) with $\mathbb{E}[\alpha_{m,k}] = 1, \forall m, k$. Note that for iid random channels, $\bar{K}(R,P)$ is finite for $R < \infty$ and $P > 0$ if $\mathbb{E}[\alpha_{m,k}]$ is finite.

A. Equal Power Allocation – No Prior CSI Feedback

A simple approach is to allocate an equal power to every user. For the kth subblock, the instantaneous capacity for user m, denoted by $Z_{m,k}$, is given by

$$Z_{m,k} = \log_2\left(1 + \alpha_{m,k}\frac{P_T}{M}\right).$$

The HARQ-IR protocol can be employed for each channel independently. The main advantage of the equal power allocation is that the channel gain is not required at the BS to allocate the powers. The average number of transmissions for user m is given by

$$\bar{K}_m(R, P_T/M) = \mathbb{E}[K_m(R, P_T/M)],$$

where $K_m(R, P_T/M) = \min\left\{k \mid \sum_{i=1}^{k} Z_{m,i} > R\right\}$. The average effective delay per user is given by

$$\mathsf{D}(R, P_T/M) = \frac{\sum_{m=1}^{M} \bar{K}_m(R, P_T/M)}{MR},$$

and EB per user is given by

$$\mathsf{EB}(R, P_T/M) = \frac{P_T}{M}\mathsf{D}(R, P_T/M).$$

It is expected that the EDT with equal power allocation is the same as that of the single user case as an equal power is allocated to all M orthogonal channels. Note that as $P_T \to \infty$, $\bar{K}_m(R, P_T/M)$ will approach 1 (see Appendix A) and we have

$$\lim_{P_T\to\infty} \mathsf{D}(R, P_T/M) = \frac{1}{R}.$$

As the effective delay is bounded for a finite R, the EB will approach ∞ as $P_T \to \infty$. This behavior is also valid for the case of single-user (or point-to-point communications) as shown in Fig. 1. In addition, when $P_T \to 0$, we can show that EB approaches $\ln 2 \approx 0.6931$ or -1.59dB. If P_T is sufficiently small, from the Taylor series expansion of logarithm function, we have

$$\sum_{i=1}^{k} Z_{m,k} = \frac{\sum_{i=1}^{k} \frac{\alpha_{m,k} P_T}{M}}{\ln 2} + O(P_T^2).$$

As $\mathbb{E}[\alpha_{m,k}] = 1$, the following approximation is valid as $P_T \to 0$:

$$\bar{K}_m(R, P_T/M) \to \frac{R}{P_T/M} \ln 2,$$

and it follows that

$$\lim_{P_T \to 0} \mathsf{EB}(R, P_T/M) = \ln 2 = -1.59\text{dB}. \tag{5}$$

As $P_T \to 0$ (in the low power regime), since the EB converges to $\ln 2$, the effective delay approaches infinity. This can be confirmed from Fig. 1 as the EDT curve with the equal power allocation is identical to that of single user.

For convenience, the HARQ-IR protocol with equal power allocation is referred to as HARQ-IR-EPA throughout the paper.

B. Multiuser Diversity – Limited Prior CSI Feedback

In the equal power allocation, no CSI feedback is required (note that the HARQ-IR protocol needs the CSI feedback after transmissions, which is referred to as the posterior CSI feedback). In the power allocation, CSI can be exploited for better performance. In order to achieve the multiuser diversity (MD) through the power allocation, the BS should know all users' CSI prior to transmissions to select a user (who will be assigned the total power P_T). This type of CSI feedback is referred to as the prior CSI feedback in order to differentiate the CSI feedback after transmissions (or posterior CSI feedback) for HARQ-IR. Although the BS needs to know the full CSI, $\{\alpha_{m,k}\}$, to choose the best user, there are efficient schemes that can reduce the amount of feedback to send CSI to the BS for MD [27]. For example, for a given scheduling metric, we can consider the ordered feedback scheduling that the higher metric the user has, the earlier the user sends the feedback message to the BS. Here, the metric can be a function of several parameters such as the channel power gain, $\mathsf{EB}(R, P)$ and/or $\mathsf{D}(R, P)$. Then, the BS allocates the total power to the user who first sends the feedback message. As a result, the BS can exploit the MD gain only with *limited* prior CSI feedback. This scheme is also attractive as we do not need to have M orthogonal channels, but only one channel could be sufficient.

In this subsection, we now study the performance gain in terms of EDT by the power allocation with limited prior CSI feedback. Suppose that a user is selected at a time for downlink transmissions with the total transmission power P_T. The scheduling scheme is designed in a way that only one user who has the minimum instantaneous EB is selected from the BS. The index of the selected user is given by

$$m^* = \arg\min_m \frac{P_T}{\log_2(1 + \alpha_{m,k} P_T)}. \tag{6}$$

The resulting power allocation is referred to as the multiuser diversity (MD)-based power allocation (MDPA). As the MD in [28], the user whose channel gain is the maximum can have the maximum transmission power to transmit signals effectively and this approach maximizes the energy efficiency.

Property 1. *We have the following asymptotic results for the MDPA:*

$$\lim_{P_T \to 0} \mathsf{EB} = \frac{\ln 2}{\sum_{m=1}^{M} m^{-1}};$$
$$\lim_{P_T \to \infty} \mathsf{D} = \frac{M}{R}, \tag{7}$$

where $\sum_{m=1}^{M} m^{-1}$ can be considered as the MD gain.

Proof: Define the the largest channel gain as $\alpha_{(M),k} = \max_m \alpha_{m,k}$. Since the probability that an user has the best channel gain is $1/M$, the average number of transmissions for each user becomes $M\mathbb{E}[K_{(M)}(R, P_T)]$, where

$$K_{(M)}(R, P_T) = \min \left\{ k \mid \sum_{i=1}^{k} \log_2(1 + \alpha_{(M),i} P_T) > R \right\}. \tag{8}$$

Since $K_{(M)}(R, P_T)$ is a stopping time, according to Wald's equality, we have

$$\mathbb{E}\left[\sum_{i=1}^{K_{(M)}(R,P_T)} \log_2(1 + \alpha_{(M),i} P_T)\right] = \mathbb{E}\left[K_{(M)}(R, P_T)\right] \mathbb{E}\left[\log_2(1 + \alpha_{(M),i} P_T)\right]$$

or

$$\mathbb{E}\left[K_{(M)}(R, P_T)\right] = \frac{\mathbb{E}\left[\sum_{i=1}^{K_{(M)}(R,P_T)} \log_2(1 + \alpha_{(M),i} P_T)\right]}{\mathbb{E}\left[\log_2(1 + \alpha_{(M),i} P_T)\right]}.$$

Due to the stopping criterion in (8), we have

$$\sum_{i=1}^{K_{(M)}(R,P_T)} \log_2(1 + \alpha_{(M),i} P_T) > R.$$

From this, the following lower bound can be obtained:

$$\mathbb{E}\left[K_{(M)}(R, P_T)\right] > \frac{R}{\mathbb{E}\left[\log_2(1 + \alpha_{(M),i} P_T)\right]}. \tag{9}$$

For an upper-bound, consider the following stopping time:

$$K'_{(M)}(R, P_T) = \min \left\{ k - 1 \mid \sum_{i=1}^{k} \log_2(1 + \alpha_{(M),i} P_T) > R \right\} = K_{(M)}(R, P_T) - 1. \tag{10}$$

Again, from Wald's equality and the stopping criterion in (10), we have

$$\mathbb{E}\left[\sum_{i=1}^{K'_{(M)}(R,P_T)} \log_2(1 + \alpha_{(M),i} P_T)\right] = \mathbb{E}\left[K_{(M)}(R, P_T) - 1\right] \mathbb{E}\left[\log_2(1 + \alpha_{(M),i} P_T)\right] < R$$

or

$$\mathbb{E}\left[K_{(M)}(R, P_T)\right] < \frac{R}{\mathbb{E}\left[\log_2(1+\alpha_{(M),i}P_T)\right]} + 1. \quad (11)$$

From (9) and (11), since the average transmssion power per user is P_T/M (as each user has an equal probability to be selected due to iid channel gains), we have

$$\frac{P_T}{\mathbb{E}\left[\log_2(1+\alpha_{(M),i}P_T)\right]} < \mathsf{EB} < \frac{P_T}{\mathbb{E}\left[\log_2(1+\alpha_{(M),i}P_T)\right]} + \frac{P_T}{R}. \quad (12)$$

If $P_T \to 0$, using the Taylor series expansion, we have

$$\log_2(1+\alpha_{(M),i}P_T) = \frac{P_T}{\ln 2}\alpha_{(M),i} + O(P_T^2).$$

As $P_T \to 0$, both the upper- and lower-bounds in (12) meet and converge to the following limit:

$$\lim_{P_T \to 0} \mathsf{EB} = \frac{\ln 2}{\mathbb{E}[\alpha_{(M),i}]}. \quad (13)$$

Using order statistics [29], we have

$$\mathbb{E}[\alpha_{(M),i}] = \mathbb{E}\left[\max_{m \in \{1,\ldots,M\}} \alpha_{(m),i}\right] = \sum_{m=1}^{M} \frac{1}{m},$$

which implies that the additional MD gain from $(M-1)$ to M users is $\frac{1}{M}$, which decreases with M.

Meanwhile, as $P_T \to \infty$, the BS can send a message to the selected user over one transmission duration. Thus, $M\mathbb{E}[K_{(M)}(R, P_T)] = M$ (see Appendix A), which gives

$$\lim_{P_T \to \infty} \mathsf{D} = \frac{M}{R}.$$

■

Note that since

$$\sum_{m=1}^{M} \frac{1}{m} \geq \int_1^{M+1} \frac{1}{x} dx = \ln(M+1),$$

which is the Mth harmonic number[2], we have an upper-bound on the minimum EB as follows:

$$\lim_{P_T \to 0} \mathsf{EB} \leq \frac{1}{\log_2(M+1)}. \quad (14)$$

This shows that the minimum EB decreases with M. With the reciprocal of EB, which is the bits per Joule (BJ), we can show that

$$\lim_{P_T \to 0} \mathsf{BJ} \geq \log_2(1+M). \quad (15)$$

It is interesting to note that this expression is similar to the Shannon capacity formula. In (15), as the number of users, M, is equivalent to the SNR in the Shannon capacity formula, we can see that the bits per Joule (resp., bits per second) grows logarithmically with M (resp., SNR).

A general result on the minimum EB for arbitrary fading channels can be found below.

[2]A tight approximate for the harmonic number, which is given by $\sum_{m=1}^{M} \frac{1}{m} \approx \ln M + \gamma + \epsilon_M$, where $\epsilon_M \to 0$ as $M \to \infty$ and γ is the Euler constant, can be used to provide tighter approximates for (14) and (15). In this case, we can also observe a logarithmic growth in terms of M, in particular in (15).

Theorem 1. *For any fading channels (not necessarily Rayleigh fading channels), if $\{\alpha_{m,k}\}$ are iid and $\mathbb{E}[\alpha_{m,k}] = \mu\ (< \infty)$ and $\mathbb{E}[\alpha_{m,k}^2] = \sigma^2\ (< \infty)$, we have*

$$\frac{\ln 2}{\mu + \frac{(M-1)\sigma}{\sqrt{2M-1}}} \leq \lim_{P_T \to 0} \mathsf{EB} \leq \frac{\ln 2}{\mu + \frac{c_M \sigma M}{2}}, \quad (16)$$

where

$$c_M = \sqrt{\frac{2\left(1 - \frac{1}{\binom{2M-2}{M-1}}\right)}{2M-1}}.$$

Proof: The bounds on the mean value of $\alpha_{(M),k}$ in [29, pp.61–63] are applied to (13) to obtain (16). ■

The implication of (16) is interesting. With any distribution for $\alpha_{m,k}$ as long as μ and σ^2 are finite, the minimum EB that is achieved by $P_T \to 0$ can approach 0 as $M \to \infty$. This is different from the case of equal power allocation in (5), where the minimum EB is -1.59 dB. That is, if μ and σ^2 are finite, we have

$$\lim_{M \to \infty} \lim_{P_T \to 0} \mathsf{EB} = 0 \text{ or } -\infty \text{ dB}.$$

This result implies that the minimum EB can approach 0 if $P_T \to 0$ when $M \to \infty$. If there are a number of users, some of them have very high channel gains over fading channels. Thus, the required bit energy can be very small if the user of the highest channel gain is chosen provided that CSI is avaialble in advance. Consequently, we can see that the availability of the CSI feedback prior to transmissions or prior CSI feedback can decrease the minimum EB and allows a very small EB (approaching 0 when $M \to \infty$). A similar result can also be found in [9], where the minimum EB approaches 0 using the power control (with prior CSI). We will discuss this issue further in Subsection IV-A.

For convenience, the HARQ-IR protocol with MDPA is referred to as HARQ-IR-MDPA.

C. Water-filling Theorem – Full Prior CSI Feedback

As shown earlier, the MD gain can be achieved through the MDPA that requires limited prior CSI feedback. Since the HARQ-IR protocol relies on limited posterior CSI feedback, the HARQ-IR-MDPA protocol requires both limited prior and posterior CSI feedback. On the other hand, since the HARQ-IR-EPA protocol only requires limited posterior CSI feedback (just for HARQ-IR), it cannot achieve the MD gain. In this subsection, we are interested in the MD gain in the case where full prior CSI feedback is available for the power allocation.

Provided that full prior CSI feedback is available, we can consider an optimal power allocation in terms of energy efficiency with the following instantaneous EB as follows:

$$\mathsf{EB}_k = \frac{\sum_{m=1}^{M} P_{m,k}}{\sum_{m=1}^{M} \log_2(1+\alpha_{m,k}P_{m,k})},$$

where $P_{m,k}$ is the transmission power allocated to the mth user over the kth subblock. If the total transmission power over subblock k is fixed as $\sum_{m=1}^{M} P_{m,k} = P_T$, the power allocation that minimizes the instantaneous EB becomes

$$\{P_{m,k}^*\} = \arg\max_{\{P_{m,k}\}} \sum_{m=1}^{M} \log_2(1+\alpha_{m,k}P_{m,k})$$

$$\text{subject to } \sum_{m=1}^{M} P_{m,k} = P_T. \quad (17)$$

The solution can be found by using the water-filling theorem [30]. Since the total power is fixed, the optimal power allocation that maximizes the spectral efficiency can also be the optimal solution maximizing the energy efficiency. However, the main drawback of this approach is that the BS should know the CSI, $\{\alpha_{m,k}\}$, perfectly. Users should send their channel gains to the BS through feedback channels. This feedback overhead could be exceedingly high when the channel variation is fast. Furthermore, the requirement of full prior CSI offsets the advantage of HARQ-IR over AMC, which is no requirement of CSI prior to (re)transmissions. Therefore, the EDT with the optimal power allocation using the water-filling theorem would be considered as a bound.

Property 2. *We have the following asymptotic results with the water-filling power allocation:*

$$\lim_{P_T \to 0} \mathsf{EB} = \frac{\ln 2}{\sum_{m=1}^{M} m^{-1}};$$
$$\lim_{P_T \to \infty} \mathsf{D} = \frac{1}{R}. \quad (18)$$

Proof: As $P_T \to 0$, based on the water-filling theorem, the power to the user who has the largest channel gain approaches P_T, while the powers to the other users become 0. Thus, the asymptotic EB is identical to that of the MDPA.

If $P_T \to \infty$, then $P_{m,k}$ also becomes ∞ from the water-filling theorem for $\alpha_{m,k} > 0$. Thus, the average number of transmissions, $\mathbb{E}[K_m(R, P_{m,k})]$, becomes 1 for all m since $\Pr(\alpha_{m,k} > 0) = 1$ almost surely. This implies that the effective delay is $1/R$ for all m, which completes the proof. ∎

Note that the asymptotic EBs in (5), (13), and (18) are also the minimum EBs as the EB decreases with P_T.

As $P_T \to \infty$, the equal power allocation is optimal and thus CSI at the transmitter is not necessary. For convenience, the HARQ-IR protocol with water-filling power allocation (WPA) is referred to as HARQ-IR-WPA. According to Property 2, we can see that the HARQ-IR-WPA protocol performs better than both the HARQ-IR-EPA and HARQ-IR-MDPA protocols in the low and high power regimes, respectively. This performance improvement results from the availability of full prior CSI feedback.

As mentioned earlier, however, AMC is a better scheme than HARQ protocols if full prior CSI feedback is available. Thus, the performance of the HARQ-IR-WPA protocol should only be considered as bounds. From this point of view, it is interesting to observe that the minimum EB that can be achieved with full prior CSI feedback can also be obtained with limited prior CSI feedback by comparing (18) and (7). Furthermore, with limited prior CSI feedback, in order to approach the performance with full prior CSI feedback, we could use one of HARQ-IR-EPA and HARQ-IR-MDPA selectively such as HARQ-IR-EPA in the high power regime and HARQ-IR-MDPA in the low power regime. This can be confirmed by Fig. 2 in Section V.

IV. Modification and Scheduling

In this section, we propose a simple but effective and practical power allocation approach with *limited* prior CSI feedback for HARQ-IR in a wireless multiuser system. In particular, we focus on the binary prior CSI feedback case, where each user sends binary feedback, either a transmission request (TQ) or no transmission request (NTQ), to the BS. Thus, in practice, this scheme requires one bit per user for the feedback. Then, the BS may allocate the transmit power to the only one user or multiple users according to the power allocation and user selection schemes. We will also address fairness issues, when users' channel gains are statistically independent but not necessarily to be identical.

A. A Threshold-based Multiuser Diversity Scheme

We first consider a threshold-based scheme, which combines the MDPA scheme with EPA. Let $\Delta \geq 0$ denote a threshold value for the channel gain. If the mth user has the channel gain that is greater than or equal to Δ for the kth subblock (i.e., $\alpha_{m,k} \geq \Delta$), the user sends a TQ signal to the BS. Otherwise, the user sends a NTQ signal to the BS. Then, the BS sends data packets based on HARQ-IR to the users transmitting TQs using EPA. The resulting approach is referred to as the threshold-based transmission (TBT) scheme with EPA (in short, the TBT-EPA scheme).

In the HARQ-IR protocol with TBT-EPA, we can assume that there are M independent HARQ-IR protocols that can be activated depending on channel gains. Let P_k denote the realization of the power allocation at the kth subblock to an activated user. With EPA, we have $P_k = P_T/M_k$, where M_k denotes the number of users who send TQs prior to the kth subblock as their channel gains are greater than or equal to Δ. Note that P_k is a random variable as M_k is. Define

$$\alpha^+_{m,k} = \begin{cases} \alpha_{m,k}, & \text{if } \alpha_{m,k} \geq \Delta; \\ 0, & \text{if } \alpha_{m,k} < \Delta. \end{cases} \quad (19)$$

Then, the number of transmissions for the mth user becomes

$$K_m(R, P_T) = \left\{ k \mid \sum_{i=1}^{k} W_{m,i} > R \right\},$$

where

$$W_{m,i} = \log_2(1 + P_i \alpha^+_{m,i}).$$

The average transmission power per user is given by

$$\mathbb{E}[P_k] = \frac{P_T}{M}\left(1 - \Pr(M_k = 0)\right). \quad (20)$$

Here, we assume that $P_k = 0$ if there is no user sent TQ back to the BS (i.e., $M_k = 0$) and a proof of (20) is shown in Appendix B. For iid channel gains, the probability that q out of M users sent TQ back to the BS is given by

$$P_{q,M} = \binom{M}{q} p_\Delta^q (1 - p_\Delta)^{M-q},$$

where $p_\Delta = \Pr(\alpha_{m,k} \geq \Delta)$. Since $p_\Delta = e^{-\Delta}$ for $\mathbb{E}[\alpha_{m,k}] = 1$, we have

$$P_{0,M} = (1 - e^{-\Delta})^M.$$

Property 3. *We have the following asymptotic results with TBT-EPA:*

$$\lim_{P_T \to 0} \mathsf{EB} = \frac{\ln 2}{1+\Delta}; \tag{21}$$

$$\lim_{P_T \to \infty} \mathsf{D} = \frac{e^{\Delta}}{R}.$$

Proof: Using Wald's identity, we have

$$\frac{R\ln 2}{\mathbb{E}[P_k \alpha^{+}_{m,k}]} < \mathbb{E}[K_m(R, P_T)] \leq \frac{R\ln 2}{\mathbb{E}[P_k \alpha^{+}_{m,k}]} + 1.$$

From this, the upper- and lower-bounds on the EB can be found as

$$\frac{\mathbb{E}[P_k]}{R}\frac{R\ln 2}{\mathbb{E}[P_k \alpha^{+}_{m,k}]} < \mathsf{EB} = \frac{\mathbb{E}[P_k]}{R}\mathbb{E}[K_m(R,P_T)]$$
$$\leq \frac{\mathbb{E}[P_k]}{R}\left(\frac{R\ln 2}{\mathbb{E}[P_k \alpha^{+}_{m,k}]} + 1\right).$$

In Appendix B, we can show that

$$\mathbb{E}[P_k \alpha^{+}_{m,k}] = \frac{P_T(1+\Delta)}{M}\left(1 - \Pr(M_k = 0)\right), \forall m. \tag{22}$$

From (20) and (22), the upper- and lower-bounds on the EB become

$$\frac{\ln 2}{1+\Delta} < \mathsf{EB} \leq \frac{\ln 2}{1+\Delta} + \frac{P_T\left(1 - \Pr(M_k = 0)\right)}{MR}. \tag{23}$$

As $P_T \to 0$, the upper-bound approaches the lower-bound in (23), which implies (21).

Meanwhile, as $P_T \to \infty$, the BS can send messages to the selected users over only one transmission duration. Since the probability of $\alpha_{m,k} \geq \Delta$ is p_Δ, the average number of transmissions per user is given by

$$\lim_{P_T \to \infty} \mathbb{E}[K_m(R,P_T)] \to \sum_{q=1}^{\infty} q p_\Delta \left(1 - p_\Delta\right)^{q-1} \overset{(a)}{=} e^{\Delta}.$$

where (a) follows from the arithmetic-geometric series $\sum_{q=1}^{\infty} q r^q = r/(1-r)^2$ for $|r| < 1$. Since $\{\alpha_{m,k}\}$ are iid, each user has the same average number of transmissions. Thus, the asymptotic effective delay becomes

$$\lim_{P_T \to \infty} \mathsf{D} = \frac{e^{\Delta}}{R}.$$

■

If $M = 1$, the minimum EB becomes

$$\lim_{P_T \to 0} \mathsf{EB} = \frac{\ln 2}{1+\Delta}.$$

If $\Delta = 0$, the HARQ-IR protocol with TBT-EPA becomes identical to the HARQ-IR-EPA protocol. In this case, the resulting minimum EB becomes

$$\lim_{P_T \to 0} \mathsf{EB} = \ln 2,$$

which is identical to that in (5).

Since Δ is a design parameter, we can decide Δ for certain desirable properties. Let

$$\Delta = \ln(cM),$$

where c is a positive constant. Then, we have the following scaling law for the minimum EB:

$$\lim_{P_T \to 0} \mathsf{EB} = \frac{\ln 2}{1 + \ln(cM)} \approx \frac{1}{\log_2(cM)}, \text{ for a large } M.$$

The minimum EB decreases logarithmically with M, which is the same as that of the HARQ-IR-MDPA protocol. In addition, since

$$\mathbb{E}[P_i] = \frac{P_T}{M}\left(1 - \left(1 - \frac{1}{cM}\right)^M\right)$$
$$\approx \frac{P_T}{M}\left(1 - e^{-\frac{1}{c}}\right), \text{ for a large } M,$$

the total transmission power becomes invariant with respect to M as $M \to \infty$, i.e.,

$$\lim_{M \to \infty} M\mathbb{E}[P_i] = P_T\left(1 - e^{-\frac{1}{c}}\right).$$

Thus, we can satisfy the total power constraint in terms of the average total power in the HARQ-IR protocol with TBT-EPA.

B. Proportional Fairness

Although the MDPA protocol is simple and effective, there could be a fairness issue if statistical properties of channel gains are different. In this subsection, we assume that $\{\sqrt{\alpha_{m,k}}\}$ have independent Rayleigh distributions with different mean values varying according to the users's channel conditions, i.e., the $\mathbb{E}[\alpha_{m,k}]$'s are different. In this case, users with small channel gains rarely access channels as the BS performs the MDPA scheme with the user having the highest channel gain. To mitigate this fairness problem, each user's delay should be considered to make sure that the users who experience longer delay can access channels [31]. Alternatively, users can have fair opportunities to access to the BS if the scheduler selects the user m with the highest *relative channel gain* [32]

$$\frac{\alpha_{m,k}}{\mathbb{E}[\alpha_{m,k}]}. \tag{24}$$

Since this metric guarantees that each user statistically has an equal probability to gain an access opportunity to the BS, the users with small channel gains are expected to reduce the average number of transmissions. Note that we should consider both the delay and energy efficiency for a better EDT curve, and the metric in (24) does not fully take into account energy efficiency. To address this issue, we can combine the TBT protocol with this proportional fairness scheduling. As the TBT scheme in the previous subsection, the user sends the TQ signal to the BS for the kth subblock if the user has

$$\frac{\alpha_{m,k}}{\mathbb{E}[\alpha_{m,k}]} \geq \Delta, \text{ for } m = 1, \cdots, M. \tag{25}$$

Then the BS sends a data packet to the user with the highest relative channel gain among the users to feed TQ signals back. This fairness scheduling is referred to as the proportional fairness TBT-MDPA scheme (in short, the PF-TBT-MDPA scheme). Likewise, if the BS sends data packets to all the selected users satisfying the condition (25) with EPA, i.e., $\frac{P_T}{M_k}$, we can combine the TBT-EPA scheme with the proportional fairness scheduling. We call this scheme the proportional fairness TBT-EPA scheme (in short, the PF-TBT-EPA scheme). Note that the user selection by comparing with the relative channel gains in (25) guarantees the fairness among the users, and the threshold Δ provides the efficient use of energy in a way to select the users with high relative channel gains. Thus it is expected to achieve a better EDT curve by adjusting the threshold Δ.

With full prior CSI, a better power allocation scheme can improve the energy efficiency. Again, this is not practical, but could provide a performance bound. We can modify the power allocation based on the water-filling theorem in (17) for the proportional fairness as follows:

$$\{P^*_{m,k}\} = \arg\max_{\{P_{m,k}\}} \sum_{m=1}^{M} \log_2(1 + \alpha^+_{m,k} P_{m,k}) \quad \text{subject to } \sum_{m=1}^{M} P_{m,k} \le P_T,$$

where

$$\alpha^+_{m,k} = \begin{cases} \alpha_{m,k}, & \text{if } \frac{\alpha_{m,k}}{\mathbb{E}[\alpha_{m,k}]} \ge \Delta; \\ 0, & \text{otherwise.} \end{cases}$$

For convenience, we call this scheme the proportional fairness TBT based on the water-filling power allocation (in short, the PF-TBT-WPA scheme). Note that as mentioned earlier, due to full prior CSI feedback from the selected users to the BS, the improved delay and energy efficiency by the PF-TBT-WPA scheme can be considered as performance bounds.

We finally note that the PF-TBT-EPA and PF-TBT-WPA schemes with $\Delta = 0$ become the conventional TBT-EPA scheme and optimal one in (17), respectively since they activate all the users. Thus the fairness issue is ignored for $\Delta = 0$. On the other hand, the PF-TBT-MDPA scheme can still take into account the fairness among users even though we set Δ to zero.

V. Simulation Results

In this section, we will show the EDT curves under various scenarios through simulations. We assume independent Rayleigh fading channels for simulations. For given K and R, we first observe the asymptotic behaviors of $\mathsf{EB}(R, P_T)$ and $\mathsf{D}(R, P_T)$ for different power allocation schemes by changing the total transmission power P_T. The EDT curve is then obtained from multiple pairs of $\{\mathsf{EB}(R, P_T), \mathsf{D}(R, P_T)\}$.

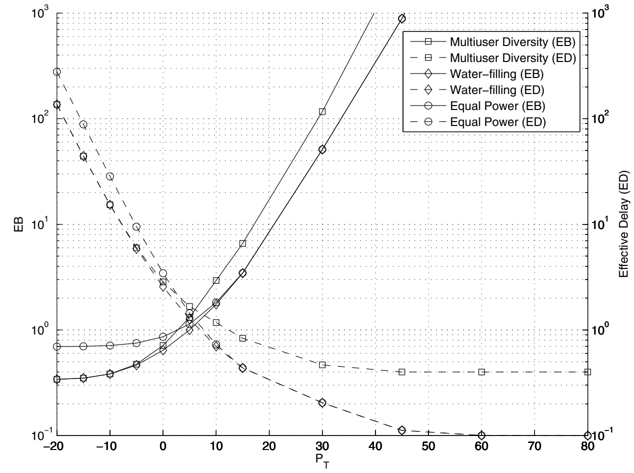


Fig. 2. $\mathsf{EB}(R, P)$ and $\mathsf{D}(R, P)$ curves of HARQ-IR for different power allocation schemes when $K = 4$ and $R = 10$.

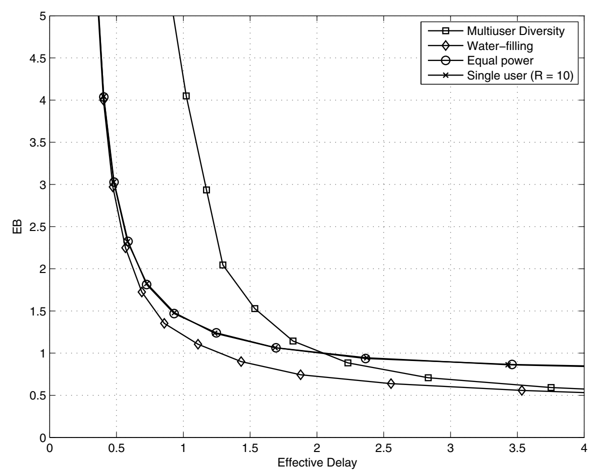


Fig. 3. EDT curves of HARQ-IR for different power allocation schemes when $K = 4$ and $R = 10$.

Fig. 2 shows $\mathsf{EB}(R, P_T)$ and $\mathsf{D}(R, P_T)$ curves of HARQ-IR with respect to different values of P_T when $K = 4$ and $R = 10$. We consider the HARQ-IR with EPA, optimal power allocation (i.e., WPA), and MDPA when $\mathbb{E}[\alpha_{m,k}] = 1$ for all m (thus, it is not necessary to consider the fairness issue). As expected, we can observe the tradeoff between the EB and effective delay. As P_T increases, the effective delays (the right y-axis in Fig. 2) with EPA and WPA decrease and approach $1/R = 0.1$, while that for MDPA converges to $M/R = 0.4$ as we derived. On the other hand, as P_T decreases, EB (the left y-axis in Fig. 2) with EPA converges to $\ln 2 \approx 0.693$, while those with MDPA and WPA converge to $\ln 2/(\sum_{m=1}^{4} m^{-1}) \approx 0.333$. We can also see that with limited prior CSI, the performance HARQ-IR can approach that with full prior CSI (i.e., the HARQ-IR with WPA) by selectively using HARQ-IR with EPA in the high power regime ($P_T > 5$ dB) and HARQ-IR with MDPA in the low power regime ($P_T < 5$ dB).

For each value of P_T, we can obtain a pair of $\{\mathsf{EB}(R, P), \mathsf{D}(R, P)\}$, from which the EDT curve is shown in Fig. 3. For comparison, we also plot the EDT of the single-user system (point-to-point communications), which is equivalent to the EDT of HARQ-IR-EPA.

Fig. 3 shows that the EDTs for multiuser systems with the optimal power allocation can be better than that for a single-user system. We can also see that the EDT curve of HARQ-IR-EPA is close to that of HARQ-IR-WPA for low spectral efficiency or short effective delays (i.e., $\mathsf{D} < 0.5$ or less than 5 transmissions) or high total transmission powers (i.e., $P_T > 14$ dB). This shows that the equal power allocation could be effective in the high power regime. However, in the lower power regime, the EDT curve of HARQ-IR-MDPA approaches that of HARQ-IR-WPA. In this case, since a low code rate, R, does not degrade EDT significantly, a lower order modulation can be used, which is more practical in wireless applications.

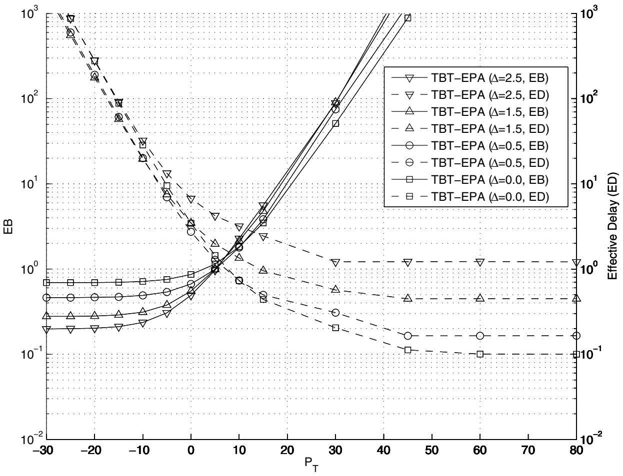


Fig. 4. $\mathsf{EB}(R, P_T)$ and $\mathsf{D}(R, P_T)$ curves of HARQ-IR with the TBT-EPA scheme for different thresholds in a wide range of P_T when $K = 4$ and $R = 10$.

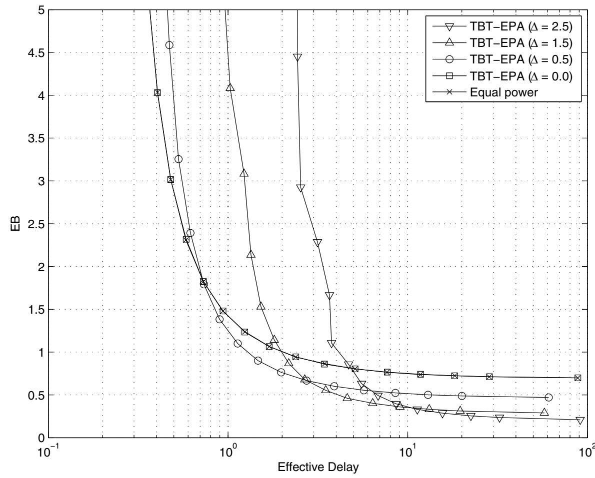


Fig. 5. EDT curves of HARQ-IR with the TBT-EPA scheme for different thresholds when $K = 4$ and $R = 10$.

TABLE I
THE AVERAGE NUMBER OF TRANSMISSIONS FOR DIFFERENT POWER ALLOCATIONS (SNR = 20 dB AND $R = 10$).

	Near users	Far users
MDPA	4.11	138.28
PF-TBT-MDPA ($\Delta = 1.0$)	8.19	9.49
PF-TBT-EPA ($\Delta = 1.0$)	3.31	5.70
PF-TBT-WPA ($\Delta = 0.0$)	2.50	5.62
PF-TBT-WPA ($\Delta = 1.0$)	3.31	5.80

Fig. 4 shows $\mathsf{EB}(R, P_T)$ and $\mathsf{D}(R, P_T)$ curves of HARQ-IR with the TBT-EPA scheme (for convenience, we will omit HARQ-IR hereafter, e.g., HARQ-IR with the TBT-EPA scheme will be referred to as the TBT-EPA scheme) for different values of Δ when $M = 4$ and $R = 10$. It is also assumed that $\mathbb{E}[\alpha_{m,k}] = 1$. We can confirm from Fig. 4 that, for low and high P_T regimes, EBs and effective delays are approaching the limiting values derived in Property 3. Fig. 5 shows the EDT curves of the TBT-EPA scheme, which is obtained from the multiple pairs of $\{\mathsf{EB}(R, P), \mathsf{D}(R, P)\}$ in Fig. 4. As expected, the TBT-EPA scheme with $\Delta = 0.0$ and the EPA scheme have the same EDT performance. With growing Δ, we can achieve a better EDT curve in the low power regime. However, comparing the results of the TBT-EPA scheme with $\Delta = 1.5$ and that with $\Delta = 2.5$, the EB gain improvement resulting from increasing Δ can be observed for large effective delays ($\mathsf{D} > 10$). Thus, we may allow a very large effective delay to achieve a very low EB. This is because the probability that a user is selected for downlink transmission exponentially decreases with increasing Δ. In turn, the EDT curve of TBT-EPA when $\Delta = 0.5$ shows that even for a small Δ, the TBT-EPA scheme could be more energy efficient than EPA in a wide range of effective delays.

Suppose that there are four users. Two of them are close to the BS (they are called near users) and the other two users are far away from the BS (they are called far users). We have $\mathbb{E}[\alpha_{m,k}] = 1.8$ for the near users, while $\mathbb{E}[\alpha_{m,k}] = 0.2$ for the far users. Without taking into account the fairness, the average number of transmissions, $\bar{K}$, becomes 4.11 and 138.28 for the near and far users, respectively, when the MDPA scheme is employed at a total SNR (P_T) of 20 dB and with $R = 10$. If the PF-TBT-MDPA scheme is carried out with $\Delta = 1.0$, the average number of transmissions, $\bar{K}$, becomes 8.19 and 9.49 for the near and far users, respectively. The average numbers of transmissions for the other schemes are also found in Table I. Note that all the schemes except the MDPA scheme provide the same access probabilities to the near and far users, but users' different average channel gains make a difference in the average numbers of transmissions. In turn, it is shown that the PF-TBT-WPA scheme with $\Delta = 0.0$ can also improve the fairness. This is because the water-filling theorem results in a reasonably fair power allocation to far users without taking into account proportional fairness.

Fig. 6 shows the EDT curves of HARQ-IR with all the schemes in Table I when Δ is set to 0.0 and 2.5. The same parameters considered in Table I are also used for simulations except the SNR, which is varying to obtain the EDT curves. For the MDPA scheme, we can see that the far users suffer from the unfairness, resulting in a long effective delay. For the PF-TBT-MDPA scheme with $\Delta = 0.0$, it is shown that the user selection in (24) effectively reduces the effective delay for the far users since it gives them more opportunities to get access to the BS. However, the effective delay for the near users in the PF-TBT-MDPA schem becomes longer in return. We can also see interesting results for the low power regime (or for long effective delays) that the far users for the PF-TBT-MDPA scheme have higher EBs than those for the MDPA scheme while the opposite result is observed for the near users. In the case of the PF-TBT-MDPA scheme, the far users selected by the BS are generally lower absolute channel gains than those in the MDPA scheme. Thus, although the BS allocates the same power to the selected far users in both schemes, the lower absolute channel gains in the PF-TBT-MDPA scheme result in the energy inefficient power allocation in the figure. We expect that this loss can be compensated by increasing Δ. For $\Delta = 0.0$, we can also observe that the EDT curves of MDPA and PF-TBT-EPA converge to that of PF-

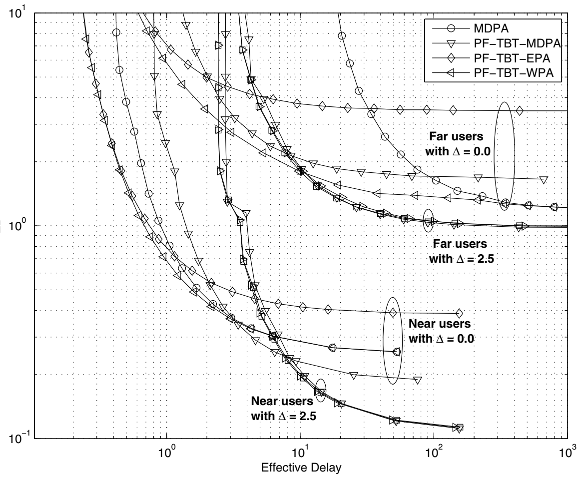


Fig. 6. EDT curves of HARQ-IR with various power allocation and proportional fairness schemes when there are two different types of users (near and far users).

TBT-WPA for long and short effective delays, respectively. This is because the PF-TBT-EPA and PF-TBT-WPA schemes become the equal and optimal power allocation schemes in Section III, respectively, and we can see the same results shown in Fig. 3 in this figure as well. For $\Delta = 2.5$, as expected, the energy efficiency of both near and far users in the proportional fairness protocols is improved for the low power regime. Note that the probability that the only one user is selected by the BS becomes higher with growing Δ in all the proportional fairness protocols. In this case, since they have the same power allocation scheme that the transmit power is used only for one user, they have identical EDT curves regardless of their power allocation schemes.

VI. Concluding Remarks

The EDT of HARQ-IR has been studied for multiuser systems in this paper. In particular, we focused on the power allocation schemes to allocate powers to multiple users in downlink channels. It was shown that the EPA can provide a near optimal EDT in the high power regime. On the other hand, the MDPA can have a near optimal EDT in the low power regime. Since the MDPA can also be spectral efficient (only one channel is used at a time), this power allocation scheme is attractive in the lower power regime. Furthermore, it was shown that the MDPA can be easily modified to take into account the proportional fairness. The user selection based on their relative channel gains and the TBT scheme were applied to the MDPA, EPA, and WPA schemes in HARQ-IR, and we showed that this modification can provide better EDT performances. Thus, by adjusting the threshold, we can have a wide choice for the flexible design of the energy efficient HARQ-IR system.

An interesting observation in this paper was that the minimum EB can be lower than -1.59 dB, which is the minimum EB for a point-to-point communication, in a multiuser system using prior CSI feedback. Another obervation was that various schemes with limited CSI feedback such as the TBT scheme in conjuction with HARQ-IR can be derived to improve EDT performance.

In this paper, we did not consider bandwidth allocation schemes (i.e., it was assumed that all users have orthogonal channels) in order to focus on the energy efficiency with power allocation only. However, since the system bandwidth can be limited, we need to consider joint power and bandwidth allocation, which will be further studied in terms of EDT in the future.

Appendix A
The number of transmissions when $P_T \to \infty$

First, we consider the case of equal power allocation where we have $Z_{m,k} = \log_2\left(1 + \alpha_{m,k}\frac{P_T}{M}\right)$. It can be shown that

$$\begin{aligned} & K_m(R, P_T/M) \\ & = \min\left\{ k \mid \sum_{i=1}^{k} Z_{m,k} > R \right\} \\ & \le \min\left\{ k \mid \sum_{i=1}^{k} \log_2\left(\alpha_{m,k}\frac{P_T}{M}\right) > R \right\} \\ & = \min\left\{ k \mid \frac{1}{k}\sum_{i=1}^{k} \log_2 \alpha_{m,k} > \frac{R}{k} + \log_2 M - \log_2 P_T \right\}. \end{aligned}$$

Thus, we have $K_m(R, P_T/M) = 1$, if

$$\alpha_{m,1} > 2^{R+\log_2 M} 2^{-\log_2 P_T} = \frac{2^{R+\log_2 M}}{P_T}. \tag{26}$$

Under **A)**, The probability of the condition in (26) becomes

$$\Pr\left(\alpha_{m,1} > \frac{2^{R+\log_2 M}}{P_T}\right) = \exp\left(-\frac{2^{R+\log_2 M}}{P_T}\right).$$

For finite R and M, this probability approaches 1 as $P_T \to \infty$. This means that as $P_T \to \infty$, $K_m(R, P_T/M) = 1$ with probability 1. Thus, $\bar{K}_m(R, P_T/M) \to 1$ as $P_T \to \infty$.

Now, we consider the case of MD. From (8), a sufficient condition for $K_{(M)}(R, P_T) = 1$ is

$$\log_2(\alpha_{(M),1} P_T) > R$$

or

$$\alpha_{(M),1} > \frac{2^R}{P_T}. \tag{27}$$

Under **A)**, the probability of the condition in (27) is

$$\begin{aligned} \Pr\left(\alpha_{(M),1} > \frac{2^R}{P_T}\right) &= 1 - \prod_{m=1}^{M} \Pr\left(\alpha_{m,k} < \frac{2^R}{P_T}\right) \\ &= 1 - \left\{1 - \exp\left(-\frac{2^R}{P_T}\right)\right\}^M. \end{aligned}$$

For a finite R, as $P_T \to \infty$, this probability approaches 1. This implies that $K_{(M)}(R, P_T) = 1$ with probability 1 as $P_T \to \infty$. Thus, $\bar{K}_{(M)}(R, P_T/M) \to 1$ as $P_T \to \infty$.

APPENDIX B
PROOFS OF EQ. (20) AND EQ. (22)

We first derive (20). Using the conditional expectation, we have

$$\begin{aligned}\mathbb{E}[P_k] &= \sum_{q=1}^{M} \mathbb{E}[P_k \mid M_k = q, \text{active}] \Pr(M_k = q, \text{active}) \\ &\quad + \mathbb{E}[P_k \mid M_k = q, \text{inactive}] \Pr(M_k = q, \text{inactive}) \\ &\overset{(a)}{=} \sum_{q=1}^{M} \frac{P_T}{q} \Pr(M_k = q, \text{active}) \\ &= \sum_{q=1}^{M} \frac{P_T}{q} \Pr(\text{active} \mid M_k = q) \Pr(M_k = q) \\ &\overset{(b)}{=} \sum_{q=1}^{M} \frac{P_T}{q} \frac{q}{M} \Pr(M_k = q) \\ &= \frac{P_T}{M} \sum_{q=1}^{M} \Pr(M_k = q), \end{aligned} \tag{28}$$

where (a) is due to the fact that the transmission power for an inactive user is 0, and (b) results from that the probability that a user is active is q/M for given $M_k = q$. Then, from (28), we can easily obtain (20).

In order to show (22), using the conditional expectation, we can show that

$$\begin{aligned}\mathbb{E}[P_k \alpha_{m,k}^{+}] &= \sum_{q=0}^{M} \mathbb{E}[P_k \alpha_{m,k}^{+} | M_k = q] \Pr(M_k = q) \\ &= \sum_{q=1}^{M} \frac{P_T}{q} \mathbb{E}\left[\alpha_{m,k}^{+} | M_k = q\right] \Pr(M_k = q). \end{aligned} \tag{29}$$

Note from (19) that

$$\begin{aligned}\mathbb{E}[\alpha_{m,k}^{+} | M_k = q] &= \mathbb{E}[\alpha_{m,k}^{+} = \alpha_{m,k} | M_k = q, \alpha_{m,k} \geq \Delta] \\ &\quad \times \Pr(\alpha_{m,k} \geq \Delta | M_k = q) \\ &\quad + \underbrace{\mathbb{E}[\alpha_{m,k}^{+} = 0 | M_k = q, \alpha_{m,k} < \Delta]}_{=0} \\ &\quad \times \Pr(\alpha_{m,k} < \Delta | M_k = q). \end{aligned} \tag{30}$$

When the number of users sending TQ back is q, the probability that the mth user also sent TQ back becomes q/M since we assume iid channel gains across users. Therefore, $\Pr(\alpha_{m,k} \geq \Delta | M_k = q)$ in (30) is q/M. Note in (30) that

$$\begin{aligned}&\mathbb{E}[\alpha_{m,k}^{+} = \alpha_{m,k} | M_k = q, \alpha_{m,k} \geq \Delta] \\ &= \frac{\mathbb{E}[\alpha_{m,k}^{+} = \alpha_{m,k}, \alpha_{m,k} \geq \Delta]}{p_\Delta}, \end{aligned}$$

where $\mathbb{E}[\alpha_{m,k}^{+} = \alpha_{m,k}, \alpha_{m,k} \geq \Delta] = \int_{\Delta}^{\infty} \alpha e^{-\alpha} d\alpha = (1 + \Delta)\, e^{-\Delta}$. Substituting (30) into (29), we finally have

$$\mathbb{E}[P_k \alpha_{m,k}^{+}] = \frac{P_T (1 + \Delta)}{M} \sum_{q=1}^{M} \Pr(M_k = q).$$

REFERENCES

[1] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, pp. 30–37, June 2011.

[2] M. Horowitz, T. Indermaur, and R. Gonzalez, "Low-power digital design," in *Proc. 1994 IEEE Symp. Low Power Electron.*, pp. 8–11.

[3] R. Gonzalez, B. Gordon, and M. Horowitz, "Supply and threshold voltage scaling for low power CMOS," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1210–1216, Aug. 1997.

[4] P. I. Pénzes and A. J. Martin, "Energy-delay efficiency of VLSI computations," in *Proc. 2002 ACM Great Lakes Symp. VLSI*, pp. 104–111.

[5] S.-C. Lin, N. Srivastava, and K. Banerjee, "A thermally-aware methodology for design-specific optimization of supply and threshold voltages in nanometer scale ICs," in *Proc. 2005 IEEE Int. Conf. Comput. Design*, pp. 411–416.

[6] M. Alioto, E. Consoli, and G. Palumbo, "General strategies to design nanometer flip-flops in the energy-delay space," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, pp. 1583–1596, July 2010.

[7] F. Meshkati, H. V. Poor, S. C. Schwartz, and N. B. Mandayam, "An energy-efficient approach to power control and receiver design in wireless networks," *IEEE Trans. Commun.*, vol. 53, pp. 1885–1894, Nov. 2005.

[8] M. Neely, "Energy optimal control for time-varying wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2915–2934, July 2006.

[9] M. C. Gursoy, D. Qiao, and S. Velipasalar, "Analysis of energy efficiency in fading channels under QoS constraints," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 4252–4263, Aug. 2009.

[10] E. Uysal-Biyikoglu, B. Prabhakar, and A. El Gamal, "On adaptive transmission for energy efficiency in wireless data networks," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3081–3094, Dec. 2004.

[11] E. Uysal-Biyikoglu and A. El Gamal, "Energy-efficient packet transmission over a wireless link," *IEEE/ACM Trans. Netw.*, vol. 10, pp. 487–499, Aug. 2002.

[12] X. Zhong and C.-Z. Xu, "Energy-efficient wireless packet scheduling with quality of service control," *IEEE Trans. Mobile Comput.*, vol. 6, pp. 1158–1170, Oct. 2007.

[13] M. Zafer and E. Modiano, "Minimum energy transmission over a wireless channel with deadline and power constrains," *IEEE Trans. Autom. Control*, vol. 54, pp. 2841–2852, Dec. 2009.

[14] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 1732–1743, Sept. 2011.

[15] D. Rajan, "Power efficient delay allocation in multihop wireless networks," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 1813–1825, July 2007.

[16] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. H. Kim, "On the energy efficiency of hybrid-ARQ protocols in fading channels.," in *Proc. 2007 IEEE Int. Conf. Commun.*, pp. 3173–3177.

[17] M. Chen, S. Serbetli, and A. Yener, "Distributed power allocation strategies for parallel relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 552–561, Feb. 2008.

[18] Z. Zhou, S. Zhou, J.-H. Cui, and S. Cui, "Energy-efficient cooperative communication based on power control and selective single-relay in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3066–3078, Aug. 2008.

[19] C. K. Ho, P. H. Tan, and S. Sun, "Relaying for energy-efficient scheduling with deadline," in *Proc. 2010 IEEE Int. Conf. Commun.*, pp. 1–6.

[20] J. Choi and D. To, "Energy efficiency of HARQ-IR for two-way relay systems with network coding (invited paper)," in *Proc. 2012 European Wireless*, pp. 1–5.

[21] C. Bae and W. Stark, "End-to-end energybandwidth tradeoff in multihop wireless networks," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4051–4066, Sept. 2009.

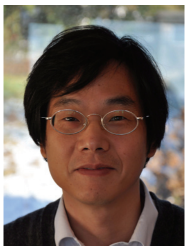
[22] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. H. Kim, "Energy efficiency of non-collaborative and collaborative hybrid-ARQ protocols," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 326–335, Jan. 2009.

[23] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1971–1988, July 2001.

[24] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1319–1343, June 2002.

[25] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Minimum energy to send k bits with and without feedback," in *Proc. 2010 IEEE Int. Symp. Inf. Theory*, pp. 221–225.

[26] P. Wu and N. Jindal, "Performance of hybrid-ARQ in block-fading channels: a fixed outage probability," *IEEE Trans. Commun.*, vol. 58, pp. 1129–1141, Apr. 2010.
[27] J. So and J. M. Cioffi, "Feedback reduction scheme for downlink multiuser diversity," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 668–672, Feb. 2009.
[28] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. 1995 IEEE Int. Conf. Commun.*, pp. 331–335.
[29] H. A. David, *Order Statistics*. John Wiley & Sons, 1980.
[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley, 2006.
[31] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1277–1294, June 2002.
[32] L. Yang, M. Kang, and M.-S. Alouini, "On the capacity-fairness tradeoff in multiuser diversity systems," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 1901–1907, July 2007.



Jinho Choi (Senior Member of IEEE) was born in Seoul, Korea.He received B.E. (magna cum laude) degree in electronics engineering in 1989 from Sogang University, Seoul, and the M.S.E. and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, in 1991 and 1994, respectively. He is now with the College of Engineering, Swansea University, the United Kingdom, as a Professor/Chair in Wireless. His research interests include wireless communications and array/statistical signal processing. He authored two books published by Cambridge University Press in 2006 and 2010. Prof. Choi received the 1999 Best Paper Award for Signal Processing from EURASIP, 2009 Best Paper Award from WPMC (Conference), and is Senior Member of IEEE. Currently, he is an Associate Editor of IEEE COMMUNICATIONS LETTERS and an Editor of the *Journal of Communications and Networks* (JCN) since 2005 and served as an Associate Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2005 to 2007 and ETRI journal.

Jeongseok Ha (M'06) received the B.E. degree in electronics from Kyungpook National University, Daegu, Korea in 1992, the M.S. degree in electronic and electrical engineering from Pohang University of Science and Technology, Pohang, Korea, in 1994, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, in 2003. He is now with the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, as an Associate Professor. His research interests include theories and applications of error-control codes and physical layer security.

Hyoungsuk Jeon received the B.S. degree in electrical engineering from Dongguk University, Seoul, Korea, in 2004, the M.S. degree in the school of engineering from Information and Communications University, Daejeon, Korea, in 2005, and the Ph.D. degree in information and communications engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2010. From 2010, he has been with KAIST as a post doctoral researcher. He is now with the School of Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta, as a visiting faculty. His research interests are in the fields of communications, signal processing, and information theory, especially design and analysis of distributed detection and physical layer security.