

Energy-Efficient Multi-User Resource Management with IR-HARQ

Ye Wu and Shugong Xu

GREAT Research Team

Huawei Technologies Co. Ltd.

Email: {matthew.wu, shugong}@huawei.com

Abstract—A general energy-efficient multi-user management framework for IR-HARQ is proposed. We first develop a basic tradeoff between energy efficiency, power, bandwidth and delay, which reveals that frequency domain, rather than power domain, offers a larger space for exploiting energy efficiency with delay constraints. Based on the basic tradeoff, the proposed energy-efficient multi-user management framework, consisting of user scheduling and power/bandwidth allocation, optimizes energy efficiency (EE) in both power and frequency domain. It is shown that while achieving EE, power-domain optimization is more effective and provide higher throughput for large-bandwidth scenarios, and frequency-domain optimization produces shorter delay in small bandwidth scenarios.

I. INTRODUCTION

In recent years, wireless communications has been experiencing an in-depth transformation, aiming to significantly improve energy efficiency (EE). This transformation has motivated diverse breakthroughs in different energy-efficiency research scopes. Therein, multi-user scheduling is directly related to transmission power at air interface and requires thorough research to minimize negative effect on system capacity while achieving EE. Meanwhile, hybrid automatic retransmission request (HARQ), especially incremental redundancy (IR) HARQ, has been widely employed to exploit EE due to its capability of achieving channel capacity [1]–[3] and its adoption in industrial standardization.

Multi-user scheduling can be basically classified into two categories: MaxSNR [4] and proportional fair (PF) [5][6]. The former maximizes the system capacity by sacrificing throughput of users with low SNRs, and the latter emphasizes resource allocation fairness for all users but lose some system capacity compared to the former. [7] proposed a cross-layer multi-user scheduling approach, which shows the feasibility of ensuring fairness without diminishing the throughput. However, there still requires more research efforts to demonstrate the energy-saving effect and impact on system capacity while employing energy-efficient multi-user scheduling.

IR-HARQ has been shown to have substantial potential to achieve EE. Meanwhile, it is observed that a great challenge of energy-efficient IR-HARQ is how to keep transmission delay under delay constraint [8]. In order to address this issue, the EE-delay tradeoff for IR-HARQ has been investigated in [9]. In particular, a special multi-user case of IR-HARQ, which schedules users based on transmission delay fairness, has been studied in [9]. However, in previous literature on EE of IR-HARQ, whether focusing on basic tradeoff or scheduling approaches, bandwidth resource for each user is fixed, and transmission delay and EE are only influenced by

power for a given data rate. This contradicts a general resource allocation principle, in which bandwidth resource is variable for all users in practice. Therefore, a basic tradeoff for IR-HARQ incorporating four elements, i.e. EE, power, transmission bandwidth and delay, is required. Based on the four-element tradeoff, a general energy-efficient multi-user resource management framework, and the impact of power/bandwidth allocation on EE and system capacity in this framework require more research as well.

In this paper, we first develop the basic tradeoff between EE, power, transmission bandwidth and delay for IR-HARQ. Based on this tradeoff, we further propose a general framework for energy-efficient multi-user resource management with IR-HARQ, comprising user scheduling and power/bandwidth allocation. The basic tradeoff reveals that frequency region has more potential to improve EE with delay constraints compared to power region. The multi-user resource management framework for IR-HARQ is considered to optimize IR-HARQ performance in terms of EE. In particular, we exploit two dimensions of multi-user resource management by flexible power (FP) optimizing EE in power region and flexible bandwidth (FB) optimizing EE in frequency region. By employing a long term evolution (LTE)/LTE advanced (LTE-A) like configuration, we demonstrate that FP outperforms FB in most scenarios, especially in large bandwidth region. However, the effectiveness of FB still can be found in small bandwidth region by providing shorter delay compared to FP.

In Section II, we present the basic tradeoff between EE, power, bandwidth and delay for IR-HARQ. The general framework for efficient multi-user resource management with IR-HARQ is proposed in Section III. Section IV shows simulation results, and the conclusion is drawn in Section V.

II. EE-POWER-BANDWIDTH-DELAY TRADEOFF FOR IR-HARQ

The most appealing feature of HARQ is that the transmission based on HARQ protocol can provide reliable communications even without *a priori* knowledge of signal-to-noise ratio (SNR) [9]. Among all sorts of HARQ protocols, IR-HARQ, capable of providing channel capacity, is the focus of this paper, and its basic procedure can be captured as follows [10].

1) Start with $k = 1$. For a given message block \mathbf{m} , a sequence of coded subblocks is generated as $\mathbf{c}_1, \mathbf{c}_2, \dots$, where $\mathbf{c}_k = \text{ENC}_k(\mathbf{m})$. $\text{ENC}_k(\cdot)$ denotes the k th encoding operation and $\text{DEC}_k(\cdot)$ the k th decoding operation.

2) A transmitter transmits \mathbf{c}_k and a receiver attempts to recover \mathbf{m} with $\{\mathbf{r}_1, \dots, \mathbf{r}_k\}$ as $\hat{\mathbf{m}}_k = \text{DEC}_k(\mathbf{r}_1, \dots, \mathbf{r}_k)$, where \mathbf{r}_k is the received signal subblock and $\hat{\mathbf{m}}_k$ denotes the decoded

message block obtained with k (re)transmissions.

3) If decoding is successful, the receiver sends positive acknowledgment (ACK) and go to step 1. On receiving an ACK, a successful HARQ (re)transmission process is finished. Otherwise, a negative acknowledgment (NACK) is sent to the transmitter. Let $k = k + 1$ and go to step 2.

For point-to-point transmission, a block fading channel varying from subblock to subblock is assumed and the k th received signal subblock is given by

$$\mathbf{r}_k = \alpha_k \mathbf{x}_k + \mathbf{n}_k \quad (1)$$

where \mathbf{x}_k denotes the modulated baseband signal of \mathbf{c}_k , α_k is the Rayleigh fading channel gain over the k th subblock, and \mathbf{n}_k is the additive white Gaussian noise (AWGN) with normalized power spectral density (PSD) of $N_0 = 1$. For each single (re)transmission, IR-HARQ can achieve the following channel capacity

$$Z_k = \log_2(1 + P|\alpha_k|^2) \quad (2)$$

where the power of modulated signal is P and a capacity achieving code of rate R [9] (i.e., instantaneous spectral efficiency) is used on bandwidth B . It is shown in [9] that the total number of (re)transmissions is given by

$$K(R, P) = \min \left\{ k \mid \sum_{i=1}^k Z_i \geq R \right\} \quad (3)$$

Whether $\sum_{i=1}^k Z_i \geq R$ decides feedback type, ACK or NACK.

We can evaluate effective spectral efficiency for each subblock as

$$SE = \frac{R}{K(R, P)} \quad (4)$$

Assuming successful delivery of a message block \mathbf{m} requires $K(R, P)$ subblock (re)transmissions and each subblock occupies time resource of T and bandwidth resource of B , the transmission delay of each bit becomes

$$D(R, P, B) = \frac{K(R, P)T}{BTR} = \frac{K(R, P)}{BR} \quad (5)$$

where $D(R, P, B)$ is measured based on unit of sec/bit. EE in terms of bit per joule is given by

$$EE(R, P, B) = \frac{BR}{K(R, P)P} = \frac{1}{D(R, P, B)P} \quad (6)$$

It requires $PK(R, P)T$ joules to successfully deliver a message block of BTR bits. In [9], bandwidth resource of B is fixed in (5) and (6), and variables D and EE are therefore functions of variables of R and P . However, B varies resulting from multi-user scheduling in practice, it is more reasonable to incorporate bandwidth when examining delay and EE.

It has been numerically shown that given the same R , lower power results in better EE with extra delay for IR-HARQ [9]. Suppose channel gains are given by $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} = \{1, 3, 2, 3\}$, two power levels of $P_1 = 5$ and $P_2 = 2$, two bandwidth levels of $B_1 = 10$ and $B_2 = 20$ can be applied. If $R = 10$, the required number of (re)transmissions is $K(R, P_1) = 3$ and $K(R, P_2) = 4$. $B_2 = 20$ and $K(R, P_1) = 3$ result in $D(R, P_1, B_2) = 0.015$ and

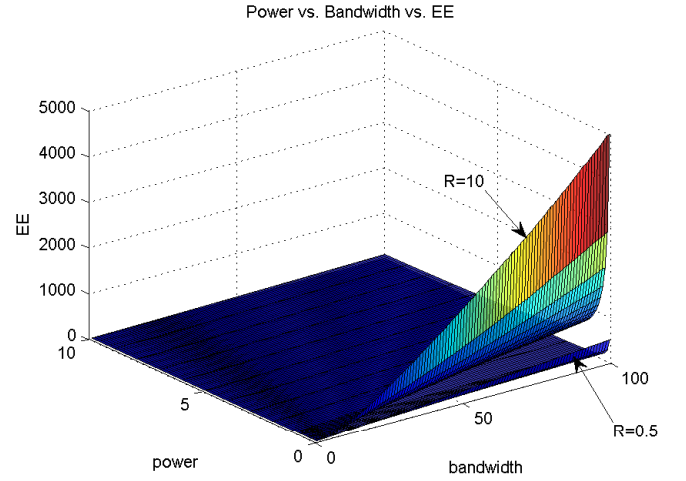


Figure 1. Relationship between power, bandwidth and EE

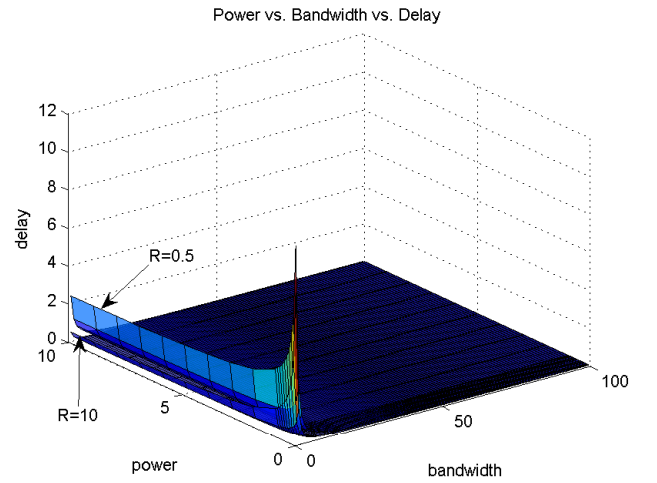


Figure 2. Relationship between power, bandwidth and delay

$EE(R, P_1, B_2) = 13.33$. $B_1 = 10$ and $K(R, P_2) = 4$ result in $D(R, P_2, B_1) = 0.04$ and $EE(R, P_2, B_1) = 12.5$. Although $P_1 > P_2$, $EE(R, P_1, B_2) > EE(R, P_2, B_1)$, which implies that not only power, but also bandwidth can significantly influence EE. In order to fully demonstrate the tradeoff between EE, power, bandwidth and delay, we employ Monte-Carlo simulation to obtain Fig. 1 and Fig. 2, where $E(|\alpha_k|^2) = 1$. Using the triplets $\{P, B, EE\}$ and $\{P, B, D\}$, we can depict the EE-power-bandwidth-delay tradeoff for IR-HARQ.

It is observed in Fig. 1 that in low power region the decrease of power and increase of bandwidth contribute to improve EE, and EE keeps flat and low regardless of variation of power and bandwidth in high power region. Fig. 2 shows that the increase of both power and bandwidth can lower delay in small bandwidth region, and delay remains flat and low in large bandwidth region. However, the increase of power can only help reduce delay when power is low enough. Fig. 1 and Fig. 2 reveal that with delay constraints frequency domain has more potential to improve EE compared to power domain.

TABLE I: Tradeoff between EE, delay and bandwidth (Given $P = 5$ and $R = 10$)

B	1	20	40	60	80
D	6.2e-1	8.8e-3	3.5e-3	2.1e-3	1.5e-3
EE	3.2e-1	22.8	56.5	93.9	133.0

It is difficult to directly show the EE-power-bandwidth-delay tradeoff in a single four-dimension figure. In order to clearly show the tradeoff between EE, power, bandwidth and delay, we fix power P or bandwidth B to find the triplet $\{EE, B, D\}$ or $\{EE, P, D\}$. For example, given $P = 5$ and $R = 10$, the tradeoff between EE, delay and bandwidth is shown in Table I.

III. FRAMEWORK FOR ENERGY-EFFICIENT MULTI-USER RESOURCE MANAGEMENT WITH IR-HARQ

A. Energy-Efficient Power and Bandwidth Allocation

For point-to-point communications, power and bandwidth are available to a single user, which requires power and bandwidth as small as possible. Therefore, given R , P and B that maximize EE are given by

$$\{P^*, B^*\} = \arg \max_{\{P, B\}} EE(R, P, B) \quad (7)$$

subject to $D(R, P, B) \leq D_{th}$, $B \leq B_{th}$ and $P \leq P_{th}$

where $EE(R, P, B)$ is an increasing function of D [9], D increases while B or P decreases as shown in (5), D_{th} , B_{th} and P_{th} denote delay threshold, bandwidth constraint and power constraint respectively. Then, the solution to (7) becomes

$$\{P^*, B^*\} = \{P, B | D(R, P, B) = D_{th} \text{ and } B \leq B_{th} \text{ and } P \leq P_{th}\} \quad (8)$$

It is clear that there exists no solution to (8) if $D_{th} < D(R, P_{th}, B_{th})$ and there exists at least one solution if $D_{th} \geq D(R, P_{th}, B_{th})$.

Throughput or capacity for point-to-point IR-HARQ is much easier to evaluate, since the instantaneous throughput for IR-HARQ is defined as the reciprocal of $D(R, P, B)$ as in (5). When delay threshold is satisfied, throughput requirement must be equivalently fulfilled.

Different from point-to-point communications, multi-user systems concentrate more on system throughput, while achieving energy efficiency. For the sake of simplicity, the total fixed system power P_{max} , the total fixed system bandwidth B_{max} , and the maximum number of scheduled users for each scheduling instant N are assumed. In multi-user IR-HARQ systems, the system throughput for J scheduling instants is evaluated as

$$C = \frac{1}{J} \sum_{j=1}^J \sum_{l=1}^{\tilde{N}_j} \frac{K_{l,j}}{D_{l,j}} = \frac{1}{J} \sum_{j=1}^J \sum_{l=1}^{\tilde{N}_j} \frac{R_{l,j}}{K_{l,j}} \sum_{k_l} B_{k_l,j} \quad (9)$$

where $D_{l,j}$, $R_{l,j}$ and $K_{l,j}$ represent transmission delay, instantaneous spectral efficiency, and total number of (re)transmissions for the l th user of \tilde{N}_j ($\tilde{N}_j \leq N$) scheduled users finishing IR-HARQ retransmission processes at the j th scheduling instant. It is worth noting that for each scheduling instant each scheduled IR-HARQ user only (re)transmits a

subblock, and $D_{l,j}$, $R_{l,j}$, $B_{k_l,j}$ and $K_{l,j}$ are given according to the IR-HARQ (re)transmission process that the l th user has experienced.

Since power and bandwidth with respect to each user may vary during the multi-user resource allocation process, (9) justifies the evaluation method of transmission delay $D_{l,j}$ as

$$D_{l,j} = \frac{K_{l,j}^2}{R_{l,j} \sum_{k_l} B_{k_l,j}} \quad (10)$$

The optimum power and bandwidth allocation in terms of EE as well as spectral efficiency for multi-user IR-HARQ systems is given by

$$\{P_{i,j}^*, B_{i,j}^*\} = \arg \max_{\{P_{i,j}, B_{i,j}\}} \frac{\sum_{i=1}^N B_{i,j} \log_2 \left(1 + P_{i,j} |\alpha_{i,j}|^2 \right)}{P_{max}} \quad (11)$$

subject to $P_{i,j} > 0 \forall i$, $\sum_{i=1}^N P_{i,j} = P_{max}$ and $\sum_{i=1}^N B_{i,j} = B_{max}$

(11) is intended to provide the optimal solution with both bandwidth and power variable, which, however, requires a prohibitive complexity. To simplify the problem in (11), we approximate the problem by assuming power is equally allocated in (12)

$$\{B_{i,j}^*\} = \arg \max_{\{B_{i,j}\}} \sum_{i=1}^N B_{i,j} \log_2 \left(1 + \frac{P_{max} |\alpha_{i,j}|^2}{N} \right) \quad (12)$$

subject to $B_{i,j} > 0 \forall i$ and $\sum_{i=1}^N B_{i,j} = B_{max}$

or assuming bandwidth is equally allocated in (13)

$$\{P_{i,j}^*\} = \arg \max_{\{P_{i,j}\}} \sum_{i=1}^N \frac{B_{max}}{N} \log_2 \left(1 + P_{i,j} |\alpha_{i,j}|^2 \right) \quad (13)$$

subject to $P_{i,j} > 0 \forall i$ and $\sum_{i=1}^N P_{i,j} = P_{max}$

Using Lagrange multipliers and numerical method, there always exists only one solution to (12) or (13). Due to the introduction of bandwidth allocation in (12), the IR-HARQ system has an extra dimension to improve energy efficiency by achieving the most multi-user diversity.

B. Energy-Efficient Multi-User Scheduling

Although optimum power and bandwidth allocation in (12) and (13) are capable of offering high energy efficiency, there is still a scheduling priority issue to be addressed in resource allocation for multi-user IR-HARQ. System throughput and scheduling fairness are well balanced by traditional PF scheduling [5][6], which, however, lacks necessary metrics to incorporate bandwidth and power for evaluating energy efficiency priority.

During IR-HARQ scheduling, each scheduled user feedbacks ACK or NACK, and each unscheduled user has no feedback. Let $\tilde{D}_{k,j}$ denote the statistical delay for the k th user (of all users waiting for scheduling) until the j th scheduling instant, which can be given by

$$\tilde{D}_{k,j} = \begin{cases} \frac{1}{\frac{(F-1)}{F\tilde{D}_{k,j-1}} + \frac{1}{FD_{k,j}}} & \text{if ACK} \\ \tilde{D}_{k,j-1} & \text{if NACK or no feedback} \end{cases} \quad (14)$$

where the delays over previous F successful IR-HARQ (re)transmission processes until the j th scheduling instant are filtered while ACK is received.

The metric of deciding scheduling priority of user k is given by

$$\text{PRI}_k = \frac{\tilde{B}_{k,j} \log 2 \left(1 + \tilde{P}_{k,j} |\alpha_{k,j}|^2 \right) \tilde{D}_{k,j}}{\tilde{P}_{k,j}} \quad (15)$$

where bandwidth $\tilde{B}_{k,j}$ and power $\tilde{P}_{k,j}$ with respect to the k th user equal bandwidth and power actually used most recently, respectively.

Since $\tilde{D}_{k,j}$ is the reciprocal of statistical throughput for a single user, the priority metric in (15) can be interpreted as a traditional PF scheduling plus additional bandwidth and power optimization in terms of energy efficiency for multi-user IR-HARQ.

C. Framework for Energy-Efficient Multi-User Resource Management

Multi-user resource management framework generally consists of a scheduling process which decides scheduling priorities and a user-specific resource allocation process which allocate power and bandwidth for each user to be scheduled. We use pseudo codes to show how the framework handles multi-user resource management with IR-HARQ.

While(1)

j th scheduling instant starts

If user k has not been scheduled before

$$\tilde{P}_{k,j} = P_{\max}, \tilde{B}_{k,j} = B_{\max} \text{ and } \tilde{D}_{k,j} = 0XFFFF$$

Else

If user k feedbacks ACK

$$\text{Update } \tilde{D}_{k,j} = \frac{1}{\frac{(F-1)}{F\tilde{D}_{k,j-1}} + \frac{1}{FD_{k,j}}}$$

Elseif user k has no feedback

$$\tilde{D}_{k,j} = \tilde{D}_{k,j-1}$$

End

Update $\tilde{P}_{k,j} = P_{i,r}^*$, $\tilde{B}_{k,j} = B_{i,r}^*$ where user index k of all users waiting for scheduling and user index i of N scheduled users at the most recent r th scheduling instant denotes the same user

End

All users with NACK feedback are scheduled

Compute PRI_k according to (14) for users without NACK feedback

Choose users with the highest PRI_k until there are N users to be scheduled

Allocate $\{P_{i,j}^*, B_{i,j}^*\}$ for all scheduled N users according to (11) or (12)

if $P_{i,j}^* = 0$ or $B_{i,j}^* = 0$

user i is not scheduled

End

Each scheduled user feedbacks ACK or NACK according to (3)

j th scheduling instant ends

$j = j + 1$

End

It is worth noting that channel gain is known and feedback of channel gain by user is required in practice. The user with NACK feedback is always scheduled, except that power or bandwidth allocated equals zero. This ensures that each message block for a user is transmitted in a continuous manner with best effort.

IV. SIMULATION RESULTS

Simulations have been carried out to determine the performance of the proposed multi-user resource management framework with IR-HARQ. Different from single-user scenarios, system power and bandwidth are always invariant in order to achieve maximum system throughput. We employ LTE/LTE-A like system bandwidths ranging from 1.4MHz to 100MHz, and system power varying from 1W to 40W. There are 4 users simulated, where $N = 2$ users are scheduled at each scheduling instant at most. Let $N_0 = -174\text{dBm/Hz}$ for AWGN

PSD and $E(|\alpha_{i,j}|^2) = 1e^{-20}$ for block fading channel. Each user has the same coding rate (i.e., instantaneous spectral efficiency) $R_{i,j} = 2$. Two power and bandwidth allocation criteria for the energy-efficient multi-user resource management framework are considered. One is flexible bandwidth (FB) in (12), and the other is flexible power (FP) in (13).

Fig. 3 illustrates how EE is influenced by system power and bandwidth constraints with FB and FP. FP always provides higher EE than FB. Particularly, the EE performances differ evidently between FB and FP in small-power-and-large-bandwidth region, where FP significantly outperforms FB. It can be explained that in the region FP has larger optimization freedom than FB due to that $\log 2(\cdot)$ of power in (13) can help achieve a much higher value than linear combination of bandwidth in (12). While FP is adopted in LTE/LTE-A like systems, the larger bandwidth is, the more substantial potential systems have to improve energy efficiency by lowering power. Meanwhile, throughput must be degraded to achieve high EE, which implies FP is most suitable in low-traffic scenarios. When more system throughput is required, both FP and FB can provide similar EE.

Throughput has to be taken care of while achieving EE. Although throughput is generally the reciprocal of EE, varying power significantly influence the absolute value of throughput. As in Fig.3, FP is more robust, offering consistently higher throughput than FB. In high power region ($>10\text{W}$), system

throughput for both FP and FB, especially for the former, are stable enough. It indicates improving power is not capable of improving EE as well as throughput in LTE/LTE-A like state-of-the-art systems, which usually have system power larger than 10W. In contrast, increasing bandwidth can effectively improve throughput regardless of power region.

Transmission delay is another key performance metric systems are concerned about. It is shown in Fig. 5 that delay can be significantly reduced for both FB and FP with larger bandwidth. For a typical small LTE/LTE-A bandwidth (i.e., 5MHz), system power must be above a threshold (i.e., 15W) to maintain a low-enough delay. In such high-power-and-small-bandwidth region, FB can achieve almost half the delay of FP. It implies that FB is more suitable for small-bandwidth systems with stringent delay requirement, and FP is regarded as more effective in balancing delay, EE and throughput.

V. CONCLUSION

We have developed the basic tradeoff between EE, power, transmission bandwidth and delay for IR-HARQ, based on which an energy-efficient multi-user resource management framework with IR-HARQ has been proposed. With diverse combinations of system bandwidth, system power, which are close to in-practice configurations, it is observed that FP can efficiently handle large bandwidth scenarios in optimizing EE, while FB can be applied in small bandwidth scenarios due to its shorter delay. Both FB and FP demonstrate that EE superiority hardly compromise the loss of throughput and delay in low power region.

REFERENCES

- [1] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1971–1988, July 2001.
- [2] M. Luby, "LT codes," in *Proc. 43rd Symp. Foundations of Computer Science*, (Washington, DC, USA), pp. 271–290, 2002.
- [3] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1988.
- [4] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. on Communications (ICC)*, June 1995.
- [5] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, pp. 1277 – 1294, June 2002.
- [6] H. J. Zhu and R. H. Hafez, "Novel scheduling algorithms for multimedia service in OFDM broadband wireless systems," in *Proc. IEEE Int. Conf. on Communications (ICC)*, vol. 2, June 2006, pp. 772 – 777.
- [7] C. Gueguen and S. Baey, "An efficient and fair scheduling scheme for multiuser OFDM wireless networks," in *Proc IEEE Int. Wireless Communications and Networking Conference (WCNC)*, 2008.
- [8] A. K. Karmakar, D. V. Djonin, and V. K. Bhargava, "Delay-aware power adaptive for incremental redundancy hybrid ARQ over fading channels with memory," in *Proc. IEEE International Conf. Commun.*, June 2006, vol. 9, pp. 4315–4320.
- [9] J. Choi, "On the energy efficient power allocation for HARQ-IR in multiuser systems," submitted to *IEEE Int. Conf. on Communications (ICC)*, 2012.
- [10] D. J. Costello, Jr., J. Hagenauer, H. Imai, and S. B. Wicker, "Applications of Error-Control Coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2531–2560, Oct. 1998.

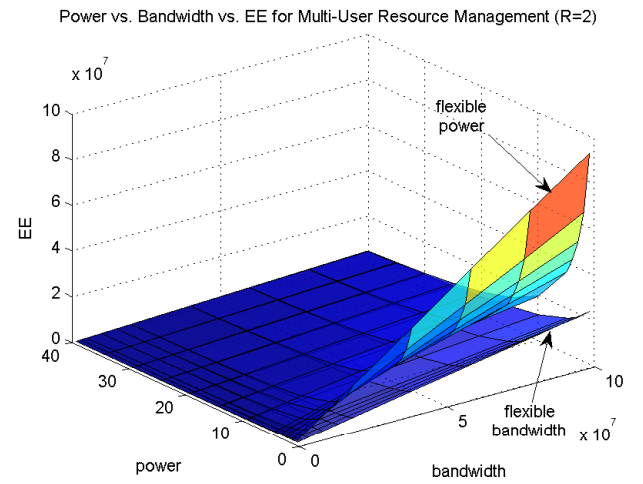


Figure 3. Power vs. bandwidth vs. EE for multi-user resource management ($R = 2$)

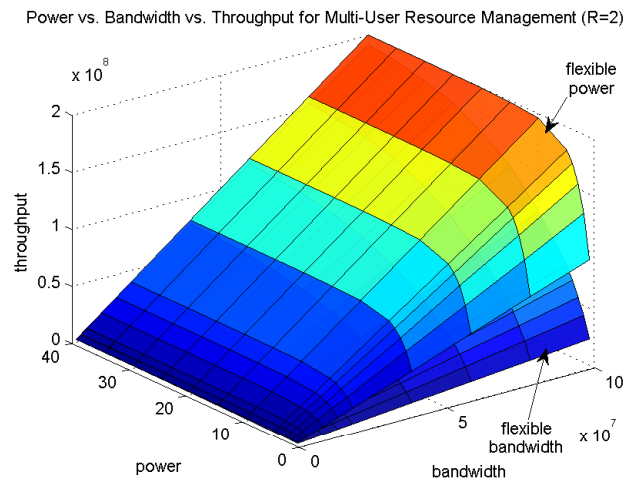


Figure 4. Power vs. bandwidth vs. throughput for multi-user resource management ($R = 2$)

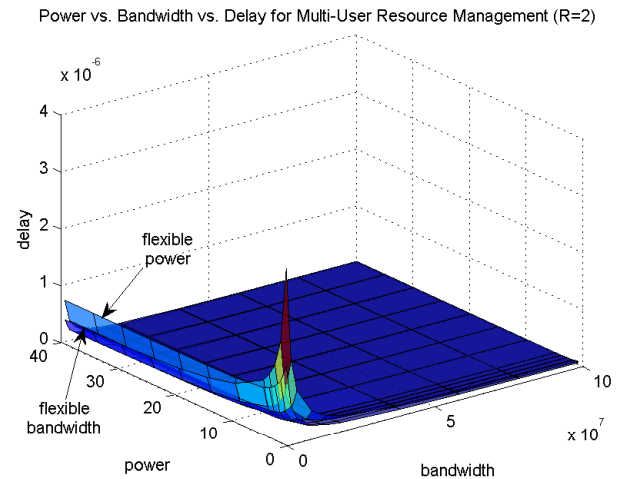


Figure 5. Power vs. bandwidth vs. delay for multi-user resource management ($R = 2$)