Efficiency Based Feedback Reduction

Jeongho Jeon, Kyuho Son, Hyang-Won Lee, and Song Chong
Dept. of Electrical Engineering, KAIST
{jjeon, skio, mslhw}@netsys.kaist.ac.kr, song@ee.kaist.ac.kr

***Abstract*—Most of up-to-date wireless data systems are based on the OFDM technology and employ the opportunistic scheduling to exploit channel itself fully. Thus, feedback reduction in multiuser multi-carrier systems becomes more of an issue due to the large amount of feedback required to pass to the base station. In this paper, we propose a novel feedback reduction scheme preserving the essential of multi-user diversity. In our proposed scheme, active users determine their feedback amount based on the feedback efficiency factor in a distributed manner. The objective is to reduce the feedback load remarkably while achieving almost the same performance to the full feedback condition. Our proposed scheme offers several advantages over existing ones. First, it does not distort the property of the scheduling policies. Second, total feedback load can be explicitly controlled to a target level regardless of the number of users in a system by adjusting each user's feedback load adaptively.**

I. Introduction

There has been ever increasing demand for higher data rates in wireless systems to support various data services. Thus, most of the current and promising wireless data systems such as IEEE 802.11/16/20 have adopted OFDM (orthogonal frequency division multiplexing) technology. By employing OFDM and other predominant technologies such as MIMO (multiple input multiple output), we could earn much higher spectral efficiency than before. Meanwhile, there were other approaches to improve the system capacity from a different perspective. In [1], opportunistic scheduling and multiuser diversity concept was introduced as a means of exploiting independently fading wireless channels between users. However, in order to employ the opportunistic scheduling, CQI (channel quality indicator) of all users at each time slot is required. Thus, when we consider multi-carrier systems such as OFDM, the total feedback amount will be substantially large. Hence a proper feedback reduction scheme must be incorporated with a system to prevent extravagant channel feedback. Unless, it will incur inefficient use of uplink resources and unnecessary consumption of the MS (mobile station) power.

There was some previous work on the subject of this paper. In [2], best-M feedback scheme was introduced. Under this scheme, each user sends CQI of its best M subbands. In consequence, each user's feedback amount is identical. From that standpoint, we can expect that this scheme will well cooperate with PF scheduler due to the *equal time sharing* property of it. However, we can't guarantee that the best-M

This work was supported by the center for Broadband OFDM Mobile Access (BrOMA) at POSTECH through the ITRC program of the Korean MIC, supervised by IITA. (IITA-2006-C1090-0603-0037)

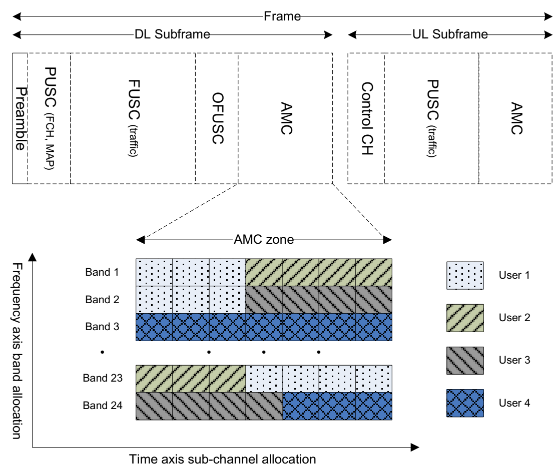


Fig. 1. Frame structure for IEEE 802.16e band-AMC mode

scheme will be suitable for other scheduling policies such as Max C/I or Max-Min Fair scheduler. We will provide more detail in the subsequent section.

In [3], *selective multiuser diversity* concept was introduced, in which each user sends its CQI based on the absolute SNR thresholding scheme. Under this scheme, the kth user sends feedback if and only if $\gamma_k(t) \geq \gamma_{th}$ where $\gamma_k(t)$ denote the instantaneous received SNR of the user k. However, it can cause serious unfairness problem, because the "weak users" have no chance of being scheduled at all due to the lack of feedback information. Hence [4] presented a modified SNR thresholding scheme, referred to as the normalized SNR thresholding scheme. In this scheme, the kth user sends feedback when the normalized value of the SNR, $\gamma_k(t)/\overline{\gamma_k}$, exceeds a certain threshold, A. However, in the average sense, the behavioral characteristic of the normalized SNR thresholding scheme is very similar with that of best-M scheme under PF scheduler.

On the other hand, the above works reduced total feedback load quite much, yet it still increases with the number of users. Thus, a random access based feedback protocol that uses SNR thresholding scheme, named as opportunistic feedback, was proposed in [5]. However, it inherently possesses the same problem of the absolute SNR thresholding scheme and their frame structure is not compatible with the current wireless data systems. Moreover, much longer feedback delay will be aroused if we use this protocol.

Here, we briefly outline the feedback mechanism of IEEE 802.16e, the main target system of this paper. Fig. 1 shows a frame structure of IEEE 802.16e band-AMC (adaptive modulation and coding) operation. This band-AMC operation is possible only if the coherence time of the channel is much longer than the lag between the time at which the channel is measured at the mobile and the time at which the packet is actually transmitted [6]. In case of 1024-FFT AMC operation, the number of bands is 24. In order to trigger band-AMC operation, mobile sends the REP-RSP (channel measurement report response) message which includes the CQI of five best bands and the indicating bitmaps. After the base station acknowledges the trigger, the mobile starts reporting the differential of channel state for the five selected bands on its allocated fast-feedback channel (CQICH) with a step of 1 dB [7]. In short, the feedback mechanism of the IEEE 802.16e system is comprised of the best-M feedback and 1-bit differential feedback.

In this paper, we propose a novel feedback reduction scheme preserving the essential of the multi-user diversity. At first, we introduce a new parameter, feedback efficiency factor. Under our proposed scheme, active users adjust their feedback amount in order to maintain the target efficiency in a distributed manner. The objective is to reduce the feedback load remarkably while achieving almost the same performance to the full feedback condition. Our proposed scheme offers several advantages over existing ones. First, it does not distort the property of the scheduler under various scheduling policies. To show this, we will adopt the α-proportional fair scheduler which is the generalized and unified form of the schedulers. Second, the total feedback load can be explicitly controlled to a target level regardless of the number of users in a system by adjusting each user's feedback load adaptively. Thus, we can predict the required portion of the uplink feedback channel. All these above advantages are achieved without additional information or control overhead.

II. α-PROPORTIONAL FAIR SCHEDULING

In a wireless network, many scheduling algorithms have been proposed. When designing a wireless scheduler, both throughput and fairness issues are very important. In order to maximize the total system throughput, Max C/I scheduler can be used to utilize the multiuser diversity fully. When we take the fairness into the first consideration, Max-Min Fair scheduler can be used. By definition, a feasible flow rate R is referred to as *max-min fair* if any rate R_i of ith user cannot be increased without decreasing R_j of arbitrary jth user which is smaller than or equal to R_i. To compromising these two challenging issues, the PF scheduler has been proposed [8].

α-proportional fair scheduler is a generalized and unified form of the schedulers also including above three ones. By allocating a certain time slot to the user k^* satisfying

$$k^* = \arg\ \max \left\{ \frac{R_k(t)}{(T_k(t))^\alpha} \right\}, \tag{1}$$

T_k, the average throughput of each user, is the solution of the following utility maximization problem [9]. Here, $R_k(t)$ is the achievable data rate that the kth user's channel can currently support.

$$\mathbf{P}: \quad \max\ U_\alpha(T_k)$$

where

$$U_\alpha(T) = \begin{cases} \sum_k log(T_k)\ , & if\ \alpha = 1, \\ \sum_k (1-\alpha)^{-1} {T_k}^{1-\alpha}\ , & if\ \alpha \neq 1. \end{cases} \tag{2}$$

When $\alpha = 1$, it simply reduces to that of PF scheduler. When $\alpha = 0$, the achieved data rate T=(T_k, $k \in K$) maximizes the total throughput $\sum_k T_k$, i.e. Max C/I scheduling. On the other hand, the long-term average throughput T satisfies max-min fairness as $\alpha \to \infty$.

Because our work is on the subject of multi-carrier systems, we devised multi-channel α-proportional fair scheduling algorithm which is a simple extended version of (1). The procedure is as following. At each time slot, we construct the $K \times N$ scheduling criterion matrix for totally K users and N channels. In the matrix, each element is obtained by computing $R_{k,n}(t)/(T_k(t))^\alpha$ where $R_{k,n}(t)$ is the achievable data rate that the kth user's nth channel can currently support. Next, we find the highest priority from the matrix, update the selected user's average rate, and remove the corresponding channel from the matrix until all the channel will be allocated.

III. EFFICIENCY BASED FEEDBACK REDUCTION

A. Proposed Algorithm

In this section, we will introduce our proposed algorithm, the efficiency based feedback reduction. In our proposed scheme, we will use each user's average number of allocated subbands, $\overline{s}_k(t)$, as the metric of deciding each user's feedback amount, rather than the SNR which was used for the previous works. The intuition behind is, if some users are more likely to be scheduled often than others, then it is natural for such users to send feedback more. Here, we introduce a new parameter, the feedback efficiency factor e. The efficiency factor is defined by the ratio of the average number of allocated subbands, $\overline{s}_k(t)$, to the number of feedback, $f_k(t)$. From now on, we will simply denote the number of subbands whose CQI is fed back to the base station as $f_k(t)$. This feedback efficiency factor has the value between 0 and 1. Conceptually, if we increase the target feedback efficiency factor toward 1, then each user will reduce their feedback in order to increase feedback efficiency.

A proper target efficiency factor will be predetermined from the system aspect, and all mobile users are expected to maintain the same efficiency by adjusting their current number of feedback. Now, the kth user's number of feedback at time slot t can be computed as following for the given target feedback efficiency factor e.

$$f_k(t) = nint\left(\frac{\overline{s}_k(t)}{e}\right), \tag{3}$$

where $nint(\cdot)$ is the nearest integer function. After then, each user sends feedback up to best $f_k(t)$th subbands. The

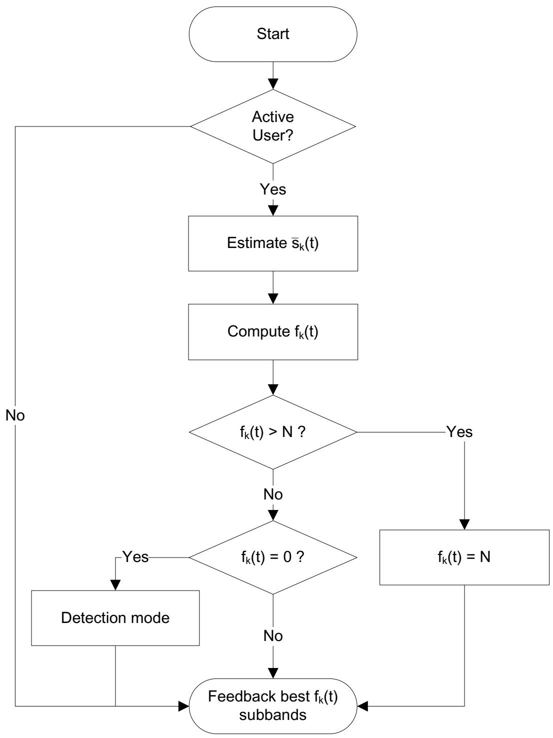


Fig. 2. Operation of MS according to the efficiency factor

average number of allocated subband can be updated using the exponential weighted low-pass filter

$$\overline{s}_k(t+1) = \left(1 - \frac{1}{t_e}\right)\overline{s}_k(t) + \frac{1}{t_e}s_k(t), \tag{4}$$

where t_e is the length of the window used for updating $\overline{s}_k(t)$. In our proposed algorithm, $f_k(t)$ was designed to be proportional to the average number of allocated subbands in order to obey the tendency of the incorporating scheduling policy.

Fig. 2 shows the proposed operation of the mobile station. To provide a safeguard against the extremes, we restrict the number of feedback to the maximum number of subbands or equivalently the total number of channels, N. If the number of feedback becomes zero, then the user will not be scheduled hereafter. This is because such a user can not participate the scheduling competition due to the lack of the feedback information. Thus, we devised the detection mode and users in that mode will try to send CQI of the subband having the best channel quality in order to check whether the channel statistics were changed for the better or not.

In this paper, we assume that the mobile can estimate the average number of allocated subband by itself, because the information about the resource allocation is contained in the downlink broadcasting message. For example in the IEEE 802.16e system, such a information is in the MAP message. We also assumed full queue condition meaning that there is always backlogged data for that user in the base station.

B. Total Feedback Load

In our proposed algorithm, total feedback load can be maintained at a target level regardless of the number of users in a system, because active users adjust their feedback load adaptively by themselves. Conceptually, if the number of users in a cell increases, then each user's average number of allocated subbands, $\overline{s}_k(t)$, will be decreased. Then, each user's feedback load, $f_k(t)$, will also be reduced in proportion to $\overline{s}_k(t)$ by (3) in order to maintain the target efficiency. Now, we can estimate the total feedback load at each time slot t as following,

$$E\left[F(t)\right] = E\left[\sum_{k=0}^{K} f_k(t)\right] = E\left[\sum_{k=0}^{K} nint\left(\frac{\overline{s}_k(t)}{e}\right)\right], \tag{5}$$

where $F(t)$ is the aggregate feedback load of all users in a cell at time slot t, and $E[\cdot]$ is the expectation function. We can divide $f_k(t)$ by deterministic term and the error term arisen by the nearest integer function.

$$f_k(t) = \frac{\overline{s}_k(t)}{e} + err_k(t). \tag{6}$$

Because the error term is arouse by rounding some floating point number to the nearest integer, it can be bounded as

$$-\frac{1}{2} < err_k(t) \leq \frac{1}{2}, \tag{7}$$

and we can assume without loss of generality that the $err_k(t)$ will have uniform probability density over the specified error range in (7). Now we can rewrite (5) as following,

$$\begin{aligned} E\left[F(t)\right] &= E\left[\sum_{k=0}^{K}\left(\frac{\overline{s}_k(t)}{e} + err_k(t)\right)\right] \\ &= \frac{N}{e} + E\left[\sum_{k=0}^{K}\left(err_k(t)\right)\right] \\ &\simeq \frac{N}{e}. \end{aligned} \tag{8}$$

In (8), sum of all the allocated subbands to each user is equal to N, because the maximum number of available subbands is totally N.

As we can observe in (8), the expectation value of the total feedback load at each time slot is not a function of the number of users but a function of the feedback efficiency factor. However, total feedback load of best-M, absolute SNR thresholding and normalized SNR thresholding will be proportional to the number of users. Thus, we can surmise that the previous schemes will reveal both under-utilization of the resources and the excessive feedback of the mobile users according to the varying number of users. In our proposed scheme, the total feedback load can be explicitly controlled to an appropriate value by adjusting each user's feedback load adaptively.

C. Impact on the Scheduler

In this section, we will inspect whether our proposed algorithm distorts the property of scheduling policies or not. To do this, it is worthwhile to know how each user shares the resource under α-proportional fair scheduling. Here, we obtained the time sharing property of single channel system, but we believe that the analysis also gives significant insight to the multi-channel case. We will start from the definition of α-proportional fairness [9].

Definition 1: A vector of rates x^* is α-proportional fair if it is feasible and for any other feasible vector x

$$\sum_i \frac{x_i - x_i^*}{x_i^{*\alpha}} \leq 0. \tag{9}$$

Now, we will regard the distribution of the rate vector as symmetric in the sense that the relative fluctuations in the feasible rates for the various users around the respective time-average values are statistically identical. This is roughly valid when the users for example have Rayleigh fading channels [10].

Lemma 1: If we consider single channel system with two users case, each user will share the unit resource as following manner in the average sense with varying α:

$$u_1^* = \frac{1}{1 + \left(\frac{R_1}{R_2}\right)^{\frac{(\alpha-1)}{\alpha}}}, \qquad u_2^* = 1 - u_1^*. \tag{10}$$

Proof: If ith user's average achievable rate is R_i, and if the fraction of the time slot allocated to the ith user is u_i^* by α-proportional fair scheduling, then the user's average rate can be expressed as $x_i^* = u_i^* R_i$, where $0 \leq u_i^* \leq 1$, and $\sum_i u_i^* = 1$. Also denote arbitrary feasible rate x_i as $u_i R_i$, where $0 \leq u_i \leq 1$, and $\sum_i u_i = 1$. Then, in order for x^* to be α- proportional fair, (9) must be satisfied for any feasible vector x. Hence, we can rewrite (9) for two user case as follows,

$$\frac{u_1 R_1 - u_1^* R_1}{(u_1^* R_1)^\alpha} + \frac{(1-u_1)R_2 - (1-u_1^*)R_2}{((1-u_1^*)R_2)^\alpha} \leq 0, \forall u_i \in [0,1] \tag{11}$$

After simple manipulation, (11) can be reduced as,

$$u_1 A - u_1^* A \leq 0, \qquad \forall u_i \in [0,1], \tag{12}$$

where $A = (1-u_1^*)^\alpha R_2^{\alpha-1} - u_1^{*\alpha} R_1^{\alpha-1}$. By inspection of (12), we can identify that it is always satisfied for $\forall u_i \in [0,1]$ when $A = 0$. ∎

Fig. 3 plots the time sharing ratio of the two users when $R_1/R_2 = 2$ using (12). When $\alpha = 0$, the "strong user" monopolizes all the resources. Therefore, we can deduce that the "weak user" does not need to send feedback at all. In this sense, the absolute SNR thresholding scheme will be harmonious with Max C/I scheduling. However, best-M and normalized SNR thresholding scheme will cause redundant feedback in the case of Max C/I scheduling. On the other hand, when $\alpha = 1$, each user will share equal fraction of time slot. Generally, if there are K users in a system, each user receives a fraction $1/K$ of the time slots [10]. We simply call it as the *equal time sharing* property of the PF scheduling. Thus, without a doubt, we can conclude that the best-M feedback scheme is suitable for the PF scheduling, because each user's feedback amount is identical under best-M feedback scheme.

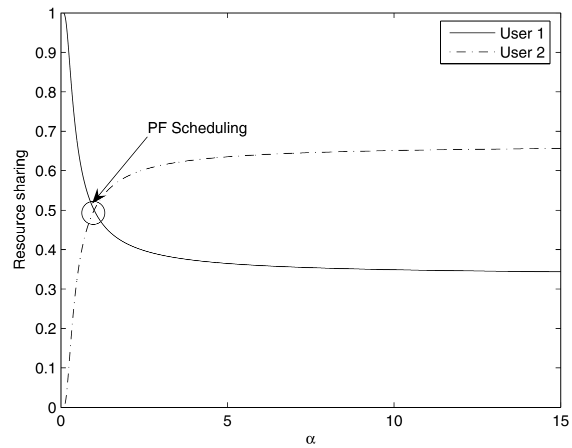


Fig. 3. Resource sharing in two user scenario, $R_1/R_2 = 2$

Normalized SNR thresholding scheme is also the same with best-M feedback in the average sense. However, absolute SNR thresholding scheme will exhibit serious unfairness problem with PF scheduling, because the "weak users" are excluded from the scheduling set. The case of Max-Min Fair scheduling can also be explained analogously.

In our proposed scheme, each user's feedback amount is given by $f_k(t) = nint\left(\overline{s}_k(t)/e\right)$, where $\overline{s}_k(t)$ is the fraction of resource sharing in itself. Thus, we can state that our algorithm will well cooperate with various scheduling policies, unlike the previous schemes. For example, when $\alpha = 0$, the efficiency based feedback reduction scheme acts like the absolute SNR thresholding scheme. When $\alpha = 1$, it acts like the best-M scheme or normalized SNR thresholding scheme. This property will provide more freedom to the system designer when deciding or changing the scheduling policy.

D. Selection of the Feedback Efficiency Factor

The feedback efficiency factor should be chosen carefully to meet a certain required performance ratio with respect to the full feedback condition. Cell throughput, fairness, or utility can be such a performance index. As mentioned before, if the efficiency factor increases, total feedback load will be decreased inverse proportionally to the factor. At the same time, the performance will be deteriorated due to the insufficient feedback information. Also, it can be decided to adjust the target feedback load to a maximum allowable total feedback load given by the system aspect using the property explained in III-B.

Fig. 4 shows the performance ratio in terms of total system throughput obtained by empirical method. Here, PF scheduling was performed, and all the other simulation environment is identical with that of IV. When $1/e$ is relatively small, we can observe certain degree of performance degradation, but the performance ratio approach to 1 as the value of $1/e$ increases. At $1/e = 6$, the performance of the efficiency based feedback is always higher than 99% of full feedback performance

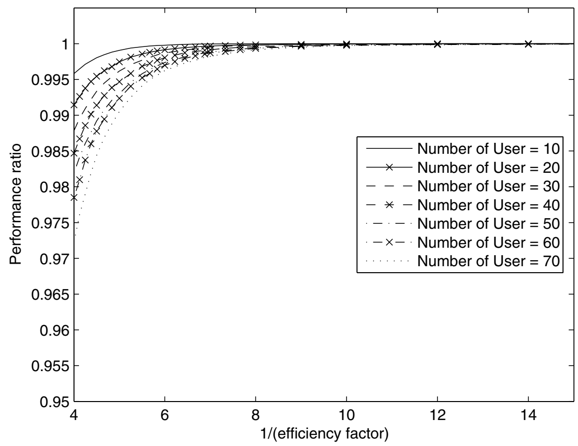


Fig. 4. Performance ratio in terms of throughput with PF scheduling

TABLE I
MEAN PERFORMANCE RATIO OVER EFFICIENCY FACTOR

1/(efficiency factor)	4	6	8
mean performance ratio	0.98167	0.99724	0.99936
1/(efficiency factor)	10	12	14
mean performance ratio	0.99976	0.99990	0.99994

for the various number of user cases. However, there is no more remarkable improvement above near $1/e = 10$. It is important to select proper value of the efficiency factor to provide sufficient margin to cope with the instantaneous channel variation.

IV. SIMULATION RESULTS

In order to validate the performance gain by applying our proposed feedback reduction scheme, IEEE 802.16e 1024-FFT OFDMA AMC mode [7] has been adopted as the target system. As mentioned, band-AMC mode is possible when the channel coherence time is much longer than the feedback delay. Thus, stationary or low mobility users can be served in this mode. Due to the reason, we consider ITU pedestrian B model [11] as the user mobility model. In IEEE 802.16e, high mobility users are served in the diversity mode, and such users send only one representative feedback information which is the average CQI value of the overall subband.

In the simulation, we further make the following assumptions. The cell diameter is 1km, and the distance, d_k, between kth user and the base station is a 2-D uniformly distributed random variable. The path loss model is $PL(d_k) = 16.62 + 37.6log_{10}(d_k)$[dB], and the shadowing is generated by normal distribution with zero mean and standard deviation σ=8dB. For frequency-selective fast fading, we employ standard delay-spread models [12]. Here, we assume that the scheduling decision is made by the base station at every 5 msec time slot for the multiple subbands.

In Fig. 5, we compared the performance of the previous and our proposed scheme with respect to that of full feedback condition. For fair comparison, we make the threshold of each scheme to exhibit the same feedback load at "number of user=30", as shown in Fig. 5(c). The parameters are $M = 5$ (best-M feedback), $\gamma_{th} = 11.86$ in dB scale (absolute SNR thresholding feedback), $A = 1.6$ in linear scale (normalized SNR thresholding feedback), and $e = 1/6.25$ (efficiency based feedback).

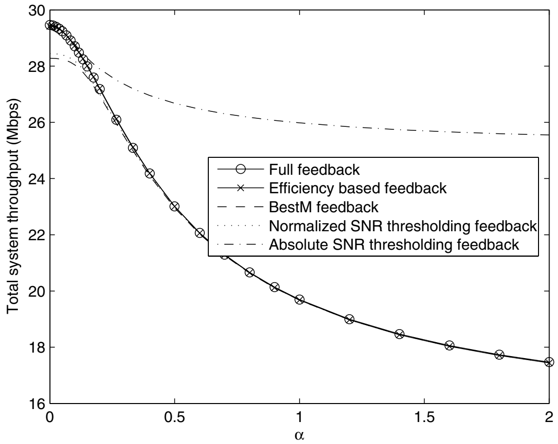
From Fig. 5(a)-(b), we can observe that the total system throughput and Jain's fairness index [13] of efficiency based feedback always keep track of the full feedback case under α-proportional fair scheduling. For the case of best-M and normalized SNR thresholding feedback, we can observe throughput degradation near $\alpha = 0$ (Max C/I scheduling), because each user sends the same amount of feedback regardless of their relative channel strengths to others. However, as α approaches to 1 (PF scheduling), the throughput and fairness converge to the full feedback performances. On the other hand, for the case of absolute SNR thresholding feedback, we can see that the throughput and fairness performances diverge from the full feedback case as α goes to infinity (Max-Min Fair scheduler). This is because the "weak users" have no chance of being scheduled at all due to the absolute thresholding.

Fig. 5(c)-(d) plots the total feedback load and total system throughput of each scheme under PF scheduling. From Fig. 5(c), we can observe the total number of feedback which is the sum of all individual user's feedback load increases with the number of users under full feedback and previous feedback schemes. However, in our proposed scheme, the total feedback load is controlled to a target level N/e, where $N = 24$ and $e = 1/6.25$. In Fig. 5(d), our proposed scheme exhibits the performance closely to that of full feedback for all the number of user cases. This is because each user adjust their feedback load according to the total number of users in a cell. However, there is considerable throughput degradation in the case of best-M and normalized SNR thresholding feedback when the number of user is small. Owing to the adaptive control mechanism of efficiency based feedback scheme, we can prevent both under-utilization of the downlink resources due to the insufficient feedback and extravagant use of uplink resources due to the redundant feedback.

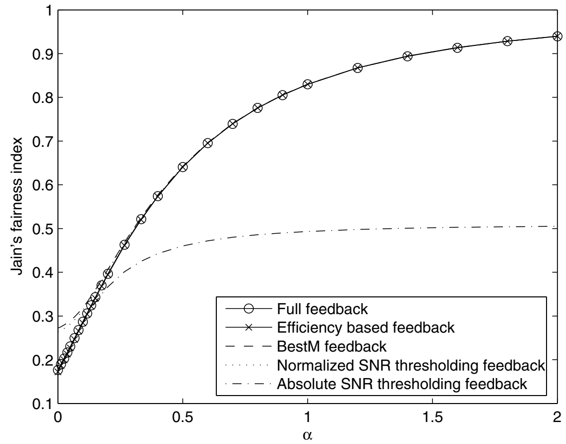
V. CONCLUDING REMARKS

In this paper, we developed a novel feedback reduction scheme preserving the essential of the multi-user diversity. We have showed that the total feedback load can be reduced remarkably while maintaining system performance the same as the full feedback condition. Moreover, it does not distort the property of the scheduling policies for any α-proportional fair scheduling. Also, the total feedback load can be maintained constantly regardless of the number of users by adjusting each user's feedback load adaptively. All these advantages are achieved without additional control overhead.

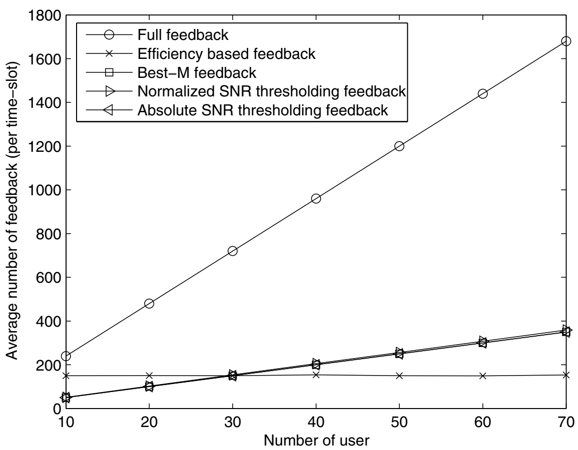
On the other hand, determining an optimal feedback efficiency factor is quite important. To find the optimal factor analytically, it is necessary to formulate the throughput achieved with our feedback reduction scheme which is daunting task. Even formulating the throughput of the multicarrier system



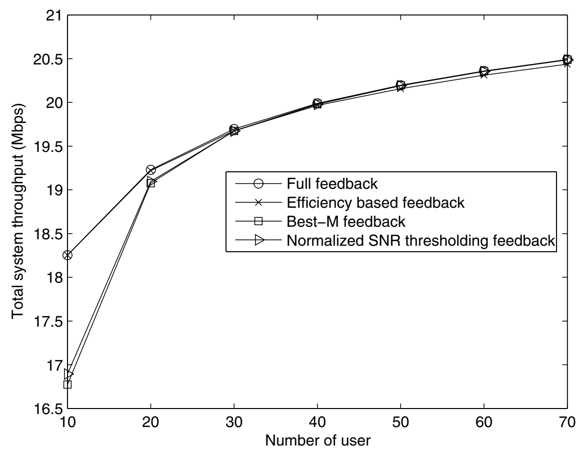
(a) System throughput ($e = 1/6.25$, number of user=30)



(b) Fairness index ($e = 1/6.25$, number of user=30)



(c) Total number of feedback ($e = 1/6.25$, $\alpha = 1$)



(d) System throughput ($e = 1/6.25$, $\alpha = 1$)

Fig. 5. Comparison of conventional and proposed feedback scheme

under specific scheduling policies is not easy alone. For this reason, we choose the efficiency factor based on the empirical method. Providing the mathematical way of finding the optimal feedback efficiency factor could be an interesting issue of further study.

REFERENCES

[1] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE ICC'95*, Seattle, Jun. 1995.

[2] "CQI report and scheduling procedure," *3GPP Tdoc R1-051045*, 2005.

[3] D. Gesbert and M. -S. Alouini, "How much feedback is multi-user diversity really worth?," in *Proc. IEEE ICC'04*, Paris, Jun. 2004.

[4] L. Yang, M. -S. Alouini and D. Gesbert, "Further Results on Selective Multiuser Diversity," in *Proc. ACM MSWiM'04*, Itary, Oct. 2004.

[5] T. Tang and R. W. Heath, Jr., "Opportunistic Feedback for Downlink Multiuser Diversity," *IEEE Communications Letters*, vol. 9, no. 10, Oct. 2005.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[7] IEEE Std 802.16e-2005 and IEEE 802.16-2004/Cor 1-2005. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. Dec. 2005.

[8] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, pp. 33–37, Oct. 1996.

[9] J. Mo and J. Walrand, "Fair End-to-End Window-Based Congestion Control," *IEEE/ACM Trans. Networking*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[10] Sem. Borst, "User-Level Performance of Channel-Aware Scheduling Algorithm in Wireless Data Networks," in *Proc. IEEE INFOCOM'03*, San Francisco, Mar. 2003.

[11] "Guidelines for the evaluation of radio transmission technologies for IMT-2000," *Recommendation ITU-R M.1225*, 1997.

[12] M. Pätzold, *Mobile Fading Channels.* Baffins Lane, England: John Wiley and Sons, 2002

[13] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared system," *DEC TR-301*.