CHAPTER 4

OPTIMAL SUBCARRIER ASSIGNING STRATEGY FOR DELAY SENSITIVE USERS

Orthogonal Frequency Division Multiplexing (OFDM) has been shown to provide excellent performance over frequency selective channel and has been proposed as the modulation and multiple access schemes for providing high speed data service in the next-generation wireless networks, e.g. IEEE 802.16 Wireless Metropolitan Area Network, IEEE (2004). Besides providing high speed data service, supporting the Quality of Service (QoS) requirements is also very important for the next generation networks which consist of mixed applications having heterogeneous nature, such as video conferencing, Internet gaming and multimedia streaming applications. These applications can have diverse definitions of QoS requirements.

4.1 INTRODUCTION

Conventional multi-user OFDM system, e.g. OFDM with Frequency Division Multiple Access (OFDM-FDMA) and OFDM with Time Division Multiple Access (OFDM-TDMA) allows only a single user to transmit data on all of the subcarriers or a fixed subset of subcarriers (Rohling and Gruneid 1997). However, such a fixed subcarrier allocation scheme fails to exploit the independence among users in the time varying wireless channel known as multi-user diversity. OFDMA with cross-layer scheduling exploits this multi-user diversity by carefully assigning multiple users to transmit simultaneously on different subcarriers for each OFDM symbol with optimal power and rate allocations, and as a result, the overall system throughput is increased significantly. There are quite a number of existing works on cross-layer scheduling design for OFDMA systems, such as Jang and Lee (2003), Song (2005a) and references therein. However, these cross-layer designs, while achieving throughput gain by exploiting multi-user diversity, are only based on a decoupled approach where source statistics, queue dynamics and application level requirements are decoupled and ignored from the physical layer information theoretical models. The negligence of traffic requirement aspects leads to inappropriate design from higher layer system performance perspective, particularly upon the provision of diverse QoS requirements in terms of delay. Hence these designs were suitable for delay insensitive applications only and not for multimedia applications where the heterogeneous classes with different delay requirements are involved.

Attempts on cross-layer scheduling design that incorporated both the source statistics and queue dynamics have been reported in Yeh and Cohen (2004), Kittipiyakul and Javidi (2004), Lau et al (2004) , Parag et al (2005) and Huang and Niu (2007). A simple on-off physical layer model was assumed in Kittipiyakul and Javidi (2004), and the multiple access channel model with homogeneous users was studied in Yeh and Cohen (2004) through combined information theory (Gallager 1968, Cover and Thomas 1991) and queuing theory (Bertsekas Gallager 1992) with the objective to minimize the average system delay. Cross-layer heuristic schedulers were also proposed in Lau et al (2004) and Parag et al (2005) for multi-user multiple antenna and OFDMA systems respectively. In Lau et al (2004), the authors presented a heuristic urgency based allocation policy for multi-user multiple antenna system (with Spatial Division Multiple Access (SDMA)) consisting of only two classes of users viz., delay-sensitive Voice over IP (VoIP) users and delay insensitive data users. In Parag et al (2005), heuristic schedulers have been designed for maximizing the system throughput while providing better fairness between users and packet error rate performance in an OFDMA system, in which the performance gain over fixed schedulers was demonstrated through some simulations. But no indication has been given how good these proposed heuristic allocation policy in Lau et al (2004) and Parag et al (2005) performs compared with the optimal performance and with other simulation settings. To sum up, all these designs have not provided explicit delay guarantee and were targeted for systems with homogeneous users only or limited to systems which has no delay requirement, in Lau et al (2004).

Furthermore, all the aforementioned designs for multi-user OFDM systems are derived based on the assumption that CSIT is perfect. However the CSIT at the Base Station (BS), which is obtained by channel estimation based on uplink pilots in TDD system, can never be perfect especially when the number of carriers is large, Ye et al (2002). There are also some recent publications addressing the effect of imperfect CSIT on power allocation and scheduler design in multi-user multiple antenna (SDMA) contexts such as Lau et al (2006). However, all these designs did not address the heterogeneous delay requirements of multimedia users as well as left out the queue dynamics in the Cross-Layer Design. A Cross-Layer Design with the consideration on both heterogeneous delay requirements and imperfectness of CSIT is therefore a very challenging and practical problem which has been rarely discussed so far.

With the adaptive subcarrier and power allocation considering cross-layer system dynamics (David et al 2007), the cross-layer scheduling has been shown to be very effective to boost the spectral efficiency of OFDM systems through multi-user diversity while guaranteeing the heterogeneous delay requirements of users based on the assumption that CSIT is perfect. However, it was shown that the cross-layer scheduler designed with perfect CSIT assumption is very sensitive to CSIT errors. To properly address this important problem, this section of the thesis focuses on delay-sensitive cross-layer scheduling design for OFDM systems consisting of users with mixed traffics and heterogeneous delay requirements while matching the imperfectness of CSIT. Specifically, a jointly optimal delay-sensitive subcarrier and rate allocation policy is proposed to maximize the total successfully deliverable data rate from the system in terms of throughput while satisfying the heterogeneous user delay requirements. The proposed optimization framework involves both information theory (to model the multi-user OFDM physical layer) and queuing theory (to model the delay dynamics). By transforming the delay constraints into the rate constraints, the delay-sensitive cross-layer scheduling problem can be formulated into a mixed convex and combinatorial optimization problem. Further, simulation results show that by considering CSIT error statistics in cross-layer design, the proposed scheme provides robust performance gain while satisfying heterogeneous user delay requirements even at high CSIT errors.

The rest of this chapter is organized as follows: Section 4.2 discusses about the CSIT, impact of CSIT error and the corresponding mathematical model. The system description including the system model, channel, Physical layer, MAC layer and Source models for Cross-layer scheduling in OFDM based wireless networks is given in section 4.3. The Cross-layer problem formulation for accounting the CSIT error in section 4.4 and the CSIT-error considerate cross-layer scheduling strategies are discussed in section 4.5. Finally the simulation results and the summary of this chapter are given in sections 4.6 and 4.7 respectively.

4.2 CHANNEL STATE INFORMATION AT THE TRANSMITTER (CSIT)

Channel adaptation techniques including power and rate allocations have been widely used to improve the spectral efficiency in data communication systems. However, most of the existing literature assumed perfect knowledge of CSIT. This assumption was reasonable in time-invariant wire line systems but it is impossible for wireless channels, which are randomly time-varying. The CSIT would inevitably have error. The value of

CSIT error depends on mode of transmission. For Frequency Division Duplexing (FDD) system, the forward and reverse links fade independently, and the transmitter acquires the CSI via a feedback channel. The channel estimate is usually quantized before feeding back by the receiver. The channel reciprocal exists in forward and reverse links, and the downlink CSIT at the BS is estimated from the uplink dedicated pilots sent by all the K mobiles. Since the BS downlink pilot can be shared by all K users, the pilot power is usually larger and the CSI at the mobiles is usually of a much smaller error variance compared with the CSIT at the BS. Hence for simplicity, the mobiles can be assumed to have perfect CSI. The CSIT error may arise from many origins. For instance, channel estimation error and quantization error results in partial CSIT, while feedback delay or duplexing delay with Doppler spread results in limited CSIT.

Most of literature (Xia et al 2004, David et al 2005) addressing CSIT error for multicarrier system mainly deal with FDD systems, where the partial CSIT is obtained from limited feedback. Several publications also address the imperfectness of CSIT in OFDM system, Leke and Cioffi (1998a, b), Souryal and Picholtz (2001) and Ye et al (2002). However, none of them has considered the impact of queuing dynamics and heterogeneous delay requirements in Cross-layer Scheduler design together.

4.2.1 Impact of CSIT Error

Taking an account of the imperfect CSIT is important in Cross-layer scheduling design. When the CSIT is perfect, the packet error rate probability could be close to zero if strong error correction coding is applied. However, in the presence of imperfectness in CSIT, there will always be finite packet error rate probability even if powerful channel coding is applied. This is because the scheduled data rate of a user based on imperfect CSIT may exceed the actual channel capacity and resulting in packet transmission outage. The packet outage will have a significant impact on the delay performance of the heterogeneous users.

The focus is made in this thesis on the obsolete CSIT, i.e. CSIT is imperfect but not partial. There are two types of imperfect CSIT, namely the limited CSIT and the obsolete CSIT. The Limited CSIT refers to the incomplete knowledge of CSI at the transmitter (such as limited CSI feedback). David et al (2005) discussed the power adaptation for OFDM system with limited CSIT to optimize the ergodic capacity. In contrast, obsolete CSIT refers to the delay from the CSI estimation time to CSI utilization time. Under obsolete CSIT, systematic packet errors occur whenever the scheduled data rate exceeds the instantaneous mutual information (namely channel outage) despite the use of strong channel coding. Conventional performance measure, such as ergodic capacity, is thus no longer meaningful since the penalty of packet errors was not accounted. It is therefore very important to control the packet error probability to a low level for reasonable system throughput and delay performance. Few works considered the obsolete CSIT in single user OFDM systems like Yao and Giannakis (2005). Only very few works had addressed both issues of delay sensitive applications and obsolete CSIT in cross-layer design of OFDMA system (David et al 2007, 2009).

4.2.2 Mathematical Model of CSIT Error

Assuming the TDD system with channel reciprocity, the downlink CSIT could be obtained by channel estimation based on the uplink preambles at BS. However, due to duplexing delay between uplink and downlink, the estimated downlink CSIT will be obsolete. For example, in the beginning of each scheduling slot, a Minimum Mean Square Error (MMSE) estimation on CSIT at time slot *m*, ie., $h_{j,i}(m)$ is performed based on an obsolete CSIT $h_{j,i}(m-D)$. Thus the estimated downlink CSIT in frequency domain $\{\hat{h}_{ij}\}$ for all users over all subcarriers at BS accounting the CSIT obsoleteness is modeled as:

$$\hat{h}_{ij} = h_{ij} + \Delta h_{ij} \tag{4.1}$$

where $\{\Delta h_{ij}\}$ are i.i.d. Gaussian random variables with zero mean and variance $\sigma_{\Delta H}{}^{2}$. Since MMSE estimation is assumed to obtain $\hat{h}_{ij}$, the parameters Δh_{ij} and $\hat{h}_{ij}$ are uncorrelated, i.e. $E\left[\Delta h_{ij}\ \hat{h}_{ij}\right] = 0$.

4.3 SYSTEM MODEL OF CROSS-LAYER SCHEDULING IN OFDM BASED MULTIUSER NETWORKS

The general cross-layer scheduling model of OFDM multi-user wireless systems is shown in Figure 4.1, where outdated CSIT and Queue States Information (QSI) are the inputs to the scheduler at BS. In this section, the OFDMA channel model, multi-user physical layer, source, and MAC layer models are discussed.

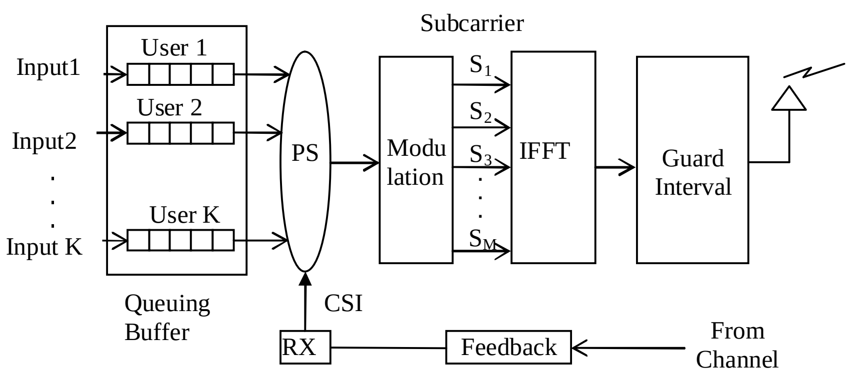

Figure 4.1 General Cross-Layer Scheduling Model of OFDM Based Multiuser Wireless System

4.3.1 System Description

The downlink system model of cross-layer scheduling in OFDM based multiuser networks is described in this section. The OFDMA systems would be assumed to have K users and F_N subcarriers. The conceptual Cross-layer scheduling model for multi-user downlink OFDM scheduler with heterogeneous application users are shown in Figure 4.1. Before the scheduling operation is performed, the cross-layer resource / packet scheduler (PS) first collects the QoS requirements of all users. In the beginning of each scheduling interval, the resource scheduler in the base station (BS) obtains CSI through the uplink dedicated pilots from all mobile users and collects QSI by observing number of backlogged packets in all these users' buffers. The PS then makes a scheduling decision based on this information and passes the resource allocation scheme to the OFDMA transmitter. The update process of state information of all users and also the scheduling decision process are made once every time slot. The subcarrier allocation and rate allocation decision made by the BS transmitter is assumed to be announced to individual mobile user through a separate control channel.

4.3.2 Downlink Channel Model

An OFDMA system containing K users with frequency selective channel model consisting of $L = [BW / \Delta f_c]$ resolvable paths is considered where Δf_c represent coherent bandwidth. The scheduling slot duration of 2 ms is considered to make a reasonable assumption of users with pedestrian mobility where the coherence time of the channel fading is around 20ms or more. Due to OFDMA, the N_F subcarriers are decoupled. Let i be the subcarrier index and j be the user index. The received symbol Y_{ij} at mobile user j on subcarrier i is given by

$$Y_{ij} = h_{ij} X_{ij} + Z_{ij} \quad (4.2)$$

where X_{ij} is the data symbol from the BS to user j on subcarrier i, h_{ij} is the complex channel gain of the i^{th} subcarrier for the j^{th} mobile which is zero mean complex Gaussian with unit variance and Z_{ij} is the zero mean complex Gaussian noise with unit variance. The transmit power allocated from the base station to user j through subcarrier i is given by $P_{ij} = E[|X_{ij}|^2]$. A subcarrier allocation strategy matrix is defined as $S_{N_F xK} = [s_{ij}]$, where the subcarrier allocation indicator s_{ij} =1 when user j is selected to occupy subcarrier i, otherwise s_{ij} =0.

4.3.3 Physical Layer Model

Assuming multiple transmission modes at the physical layer, where each mode represents a pair of a specific modulation format and forward error correcting code as per IEEE 802.16 standard specifications given by Liu et al (2006), is given in table 4.1. Based on the out-dated CSI estimated at the receiver in terms of SNR, the adaptive modulation and coding (AMC) scheme selects the modulation coding pair using the a_n , g_n , the transmission mode and packet size dependent constants and γ_{pn}, the SNR threshold defined under IEEE 802.11/ 16 standards. These values are transmitted to the AMC controller at the receiver to update the transmission mode. Efficient bandwidth utilization for a prescribed packet error rate performance at the physical layer can be accomplished with AMC schemes, which match transmission parameters to the time varying wireless channel conditions adaptively. The main objective of the AMC is to maximize the data rate by adjusting transmission parameters to channel variations, while maintaining a prescribed packet error rate $P_{0.}$ The Signal to Noise ratio (SNR) around the base station may vary depending upon the environment and location.

Table 4.1 Transmission Modes Adopted in IEEE 802.16 Standards

Mode	**1**	**2**	**3**	**4**	**5**	**6**
Modulation	QPSK	QPSK	16QAM	16QAM	64QAM	64QAM
RS code	(32,24,4)	(40,36,2)	(64,48,8)	(80,72,4)	(108,96,6)	(102,108,6)
CC Code Rate	2/3	5/6	2/3	5/6	3/4	5/6
Coding rate	1/2	3/4	½	3/4	2/3	3/4
R_n(bits/ symbol)	1.0	1.5	2.0	3.0	4.0	4.5
a_n	232.9242	140.7922	264.0330	208.5741	216.8218	220.7515
g_n	22.7925	8.2425	6.5750	2.7885	1.0675	0.8125
$\gamma_{pn}(dB)$	3.7164	5.9474	9.6598	12.3610	16.6996	17.9629

Let N be the total number of transmission modes available. Assuming fixed power transmission, the entire SNR range is partitioned into N + 1 non-overlapping consecutive interval, with boundary points denoted as $\{\gamma_n\}_{n=0}^{N+1}$ where γ denotes the received SNR. For every interval, a transmission mode is assigned. Based upon the estimated obsolete CSI at the transmitter, the users' connections are made to adapt for the assigned transmission mode. In this case, mode n is chosen, when

$$\gamma \in [\gamma_n, \gamma_{n+1}) \tag{4.3}$$

To avoid deep channel fading, no data are sent when $\gamma_0 \le \gamma < \gamma_1$, which corresponds to the mode 0 with rate R_0 = 0 (bits/symbol). The design objective of AMC is to determine the boundary points $\{\gamma_n\}_{n=0}^{N+1}$. For simplicity, the approximate instantaneous packet error rate (PER) is defined as,

$$PER_n(\gamma) \approx \begin{cases} 1, & \text{if } 0 < \gamma < \gamma_{pn} \\ a_n \exp(-g_n\gamma), & \text{if } \gamma \ge \gamma_{pn} \end{cases} \tag{4.4}$$

where n is the mode index and γ is the received SNR. The parameters a_n, g_n, and γ_{pn} are the mode fitting or dependent parameters.

The SNR region boundary (switching threshold) γ_n for the transmission mode n is the minimum SNR required to achieve P_0, the target error probability. Inverting the PER expression in Equation (4.4) leads to,

$$\begin{aligned} \gamma_0 &= 0, \\ \gamma_n &= \frac{1}{g_n}\ln\left(\frac{a_n}{P_{target}}\right), \qquad \text{where } n = 1,2,\ldots,N \\ \gamma_{N+1} &= +\infty \end{aligned} \tag{4.5}$$

By using the AMC, the users closer to the transmitter (high SNR region) get higher order modulation, those farther from the transmitter (low SNR region) get lower order modulation, ensuring the best performance for each user within the transmitter coverage.

In general, packet error contribution is determined by two factors, namely the channel noise and channel outage. The channel noise packet errors can be reduced by using a strong channel code and longer block length. But the packet error due to channel outage is a systematic one and can't be eliminated because the instantaneous mutual information between base station and user j in i^{th} subcarrier c_{ij}, is a function of actual CSI h_{ij}, which is unknown to the base station. The instantaneous mutual information is given as

$$C_{ij} = \max_{p(X_{ij})} I(X_{ij};Y_{ij}/h_{ij}) = \log_2(1 + p_{ij}\left|h_{ij}\right|^2/\sigma_z^2) \tag{4.6}$$

where σ_z^2 is the CSIT error variance.

Packets will be corrupted whenever scheduled data rate exceeds instantaneous mutual information. To take account of the packet error rate due to channel outage, the instantaneous throughput of the j^{th} user i.e., the total instantaneous data bits /s successfully delivered to user *j* is given as

$$th_j = \sum_{i=1}^{N_F} r_{ij} I[r_{ij} \le c_{ij}] \tag{4.7}$$

where $I[r_{ij} \le c_{ij}] = \begin{cases} 1, & if\ r_{ij} \le c_{ij} \\ 0, & if\ r_{ij} > c_{ij} \end{cases}$ is an indicator function, and r_{ij} is the scheduled data rate of the j^{th} user on the i^{th} subcarrier.

Hence, the average throughput of user j (measuring the average data rate successfully delivered to user *j* averaged over ergodic realizations of $H_{N_F xK} = \lfloor h_{ij} \rfloor$ and $\hat{H}_{N_F xK} = \left[\hat{h}_{ij} \right]$) is given by:

$$\begin{aligned}
\tilde{th}_j &= E_H \left[th_j \right] \\
&= E_{\hat{H}} \left\{ E_{H|\hat{H}} \left[\sum_{i=1}^{N_F} r_{ij} I\left[r_{ij} < c_{ij} \right] \right] \right\} \\
&= E_{\hat{H}} \left\{ \sum_{i=1}^{N_F} r_{ij}\, E_{H|\hat{H}}\, I\left[r_{ij} < c_{ij} \right] \right\} \\
&= E_{\hat{H}} \left\{ \sum_{i=1}^{N_F} r_{ij}\, \Pr\left[r_{ij} < c_{ij} \middle| \hat{H} \right] \right\} \\
&= E_{\hat{H}} \left\{ \sum_{i=1}^{N_F} r_{ij} \left(1 - P_{out,j} \right) \right\}
\end{aligned} \tag{4.8}$$

where $P_{out,j} = 1 - \Pr\left\lfloor r_{ij} < c_{ij} \middle| \hat{H} \right\rfloor$ is the packet outage probability on subcarrier *i* conditioned on the estimated CSIT realization $\hat{H}$.

4.3.4 Source Model

Packets from heterogeneous user applications come into each user j's buffer according to a Poisson process with independent data rate λ_j packets per time slot with packets of fixed size consisting of F bits without packet overflow. The nature of user j is characterized by a tuple [λ_j, T_j], where λ_j is the average packet arrival rate to user j and T_j is the delay constraint requirement by user j. User j with heavier traffic load will have a higher λ_j and more delay-sensitive user j will have stringent delay requirements T_j.

4.3.5 MAC Layer Model

The system dynamics are characterized by system state $x = (H, Q)$, which is composed of channel state $H = [\left|h_{ij}\right|^2]$ and buffer state $Q = [q_j]$, where h_{ij} is the complex channel gain of the ith subcarrier for the jth mobile user and q_j is the number of packets remaining in user j's buffer after getting service at time t. The MAC layer is responsible for the channel resource allocation at every fading block of information based on the current system state. At the beginning of each frame, the base station estimates the obsolete CSI at the transmitter. Based on that CSI estimation and the queue states obtained, the scheduler determines the subcarrier allocation from the policy $S_{N_F xK}[H, Q]$ for the selected users, in each scheduling slot.

4.4 CROSS-LAYER SCHEDULING PROBLEM FORMULATION

The Cross-Layer scheduling in multi-user OFDM wireless system is formulated for heterogeneous users in the presence of imperfect CSIT as a constrained optimization problem based on the system model introduced in previous section. The objective is to maximize average throughput while maintaining OFDMA physical layer constraints on subcarrier selection, transmission power and delay constraints. Specifically, the optimization problem is formulated as follows:

Find optimal rate and subcarrier allocation policies

$\left(R_{N_F XK}\left[\hat{H}, Q \right], S_{N_F XK}\left[\hat{H}, Q \right] \right)$ such that:

$$\max_{S,R} E\left(\sum_{i=1}^{N_F} \sum_{j=1}^{K} r_{ij} \left(1 - p_{out,i} \right) \right)$$

subject to (C1) : $s_{ij} \in \{0,1\}$,

(C2) : $\sum_{j=1}^{K} s_{ij} = 1$,

(C3) : $p_{ij} \geq 0$,

(C4) : $E\left[\frac{1}{N_F} \sum_{j=1}^{K} \sum_{i=1}^{N_F} s_{ij}\, p_{ij} \right] \leq P_{TOT}$, (4.9)

(C5) : $E\left[\tilde{W}_j \right] \leq T_j$,

(C6) : $P_{out,ij} = \varepsilon$.

where the expectation E[.] is taken over all system states and P_{TOT} is the average total power constraint, s_{ij} is the subcarrier allocation factor and p_{ij} is the power allocation factor.

In the optimization problem Equation (4.9), constraints (C1) and (C2) are used to ensure only one user can occupy a subcarrier at one time indicated by s_{ij}, the subcarrier allocation factor. (C3) is used to ensure transmit power would only take positive value indicated by p_{ij}, power allocation factor. (C4) is the average total power constraint, Constraint (C5) is the average delay constraint where $E\left[\tilde{W}_j\right]$ is the average system time (including waiting time and service time) of user j and (C6) is to ensure the outage probability $P_{out,ij}$ that satisfies a target outage probability ε for all subcarriers *i*, which is usually specified by application requirements.

Before solving optimization problem, a relationship between scheduled data rate r_j of user *j* and user *j*'s characteristic tuples $\left\lfloor \lambda_j, T_j \right\rfloor$ ([arrival rate, delay requirement]) has to be established first in order to transform constraint (C5) to a traceable form. The delay constraint is expressed in terms of physical layer parameters by applying a formula called modified Pollaczek-Khinchin formula (Bertsekas et al 1992). Three necessary and sufficient conditions for the constraint (C5) are given (David et al 2007) as:

1. A necessary and sufficient condition for the constraint (C5) is

$$E[X_j] + \frac{\lambda_j E[X_j^2] + \lambda_j E[X_j] (E[\overline{S_j}] / E[S_j]) t_s)}{2(1-\lambda_j E[X_j] / E[S_j])} \le T_j \qquad (4.10)$$

where the random variable X_j is the service time of the packet of user *j*, λ_j is the arrival rate of user *j*, T_j is the average delay requirement of user *j* and t_s is the duration of scheduling slot. Note that S_j and $\overline{S}_j$ are indicator variables for the availability and unavailability of the subcarrier for user *j* respectively, i.e. $\left(s_j(m)=1, \bar{s}_j(m)=0\right)$ if there is a subcarrier allocated to user *j* at time slot index *m* or $\left(s_j(m)=0, \bar{s}_j(m)=1\right)$ otherwise. In a practical OFDMA system, the number of subcarrier N_F is usually much greater than number of users *K* and thus there is always a subcarrier available for any particular user *j*, i.e., $E[S_j] = 1$ and $E[\overline{S}_j] = 0$. Notably, in this work, CSIT obsoleteness is additionally considered. In the evaluation of the whole transmission-retransmission duration X_j of each packet, the packet is assumed to be retransmitted immediately whenever outage occurs, and this retransmission will be repeated until it is successfully delivered (before transmission of other packets).

From the condition mentioned in Equation (4.10), the constraint (C5) is ready to be transformed to an equivalent rate constraint that directly relates scheduled data rate R_j of user j to the user characteristic tuple [λ_j, T_j] and also the packet size F.

(2) The second necessary and sufficient condition for the constraint (C5) when $T_j \to \infty$ is

$$E[S_j R_j] \geq F\lambda_j . \qquad (4.11)$$

This condition shows that average scheduled data rate $E[S_j R_j]$ of user j should be at least same as the bits arrival rate of user j's queue (even without any delay requirement) in order to guarantee the stability of the queue.

(3) The third necessary condition for the constraint (C5), which is called the equivalent rate constraint with outage consideration, is

$$E(s_j r_j)(1 - P_{out}) \geq \rho_j(\lambda_j, T_j, F),$$

where $$\rho_j(\lambda_j, T_j, F) = \frac{(2T_j\lambda_j + 2) + \sqrt{(2T_j\lambda_j + 2)^2 - 8T_j\lambda_j}}{4T_j} F \qquad (4.12)$$

4.5 CROSS-LAYER SCHEDULING STRATEGIES WITH CSIT ERROR

The optimization problem Equation (4.9) is a mixed combinatorial (with respect to $\{s_{ij}\}$) and convex optimization problem (with respect to $\{p_{ij}\}$). For each possible subcarrier allocation $\{s_{ij}\}$, the corresponding user data rate vector $(r_{11}, r_{12}, \cdots, r_{N_F K})$ is computed. Based on the computed data rate vector, the total system throughput $\sum_{i=1}^{N_F}\sum_{j=1}^{K} s_{ij} r_{ij}$ can be evaluated. The total system throughput is evaluated for all different cases by enumerating all possible combinations of $\{s_{ij}\}$ and the one that gives the largest average throughput will be the optimal solution. However, based on this exhaustive search approach for $\{s_{ij}\}$, the total search space is K^{N_F} which is not feasible for moderate N_F.

Using the condition (3) and Equation (4.11), optimization problem Equation (4.9) with the same constraints of (C1), (C2) and (C3) can be reformulated as follows:

$$\max_{S:\left\{s_{ij}\in\{0,1\},\sum_{j=1}^{K} s_{ij}=1\right\},\,P:\left\{p_{ij}\geq 0\right\}} E\left[\sum_{i=1}^{N_F}\sum_{j=1}^{K} s_{ij}(1-\varepsilon)\log_2\left(1+P_{ij}\phi_{ij}\left|\hat{h}_{ij}\right|^2\right)\right]$$

subject to (C4) : $E\left[\frac{1}{N_F}\sum_{i=1}^{N_F}\sum_{j=1}^{K} s_{ij} P_{ij}\right] \leq P_{TOT}$ (4.13)

(C5) : $E\left[\sum_{i=1}^{N_F} s_{ij}(1-\varepsilon)\log_2\left(1+P_{ij}\phi_{ij}\left|\hat{h}_{ij}\right|^2\right)\right] \geq \tilde{\rho}_j\left(\lambda_j, T_j, F\right)$ (4.14) where $\tilde{\rho}_j(\lambda_j, T_j, F) = \rho_j(\lambda_j, T_j, F)\left(\frac{1}{t_s} / \frac{BW}{N_F}\right)$ and BW is the total bandwidth of the OFDMA system.

In order to make the optimization problem more traceable, the integer constraint on $\{s_{ij}\}$ is relaxed to a time sharing factor between 0 and 1 and the variable $\overline{p}_{ij} = p_{ij} s_{ij}$ is introduced, so that the reformulated problem is a convex maximization problem with the following Lagrangian:

$$L = \sum_{j=1}^{K} \sum_{i=1}^{N_F} s_{ij} (1 - P_{out,ij}) \log_2 \left(1 + \frac{\tilde{p}_{ij} \phi_{ij} \left| \hat{h}_{ij} \right|^2}{s_{ij}} \right) - \mu \left(\sum_{j=1}^{K} \sum_{i=1}^{N_F} \tilde{p}_{ij} - N_F P_{TOT} \right)$$

$$+ \sum_{j=1}^{K} \gamma_j \left(s_{ij} (1 - P_{out,ij}) \log_2 \left(1 + \frac{\tilde{p}_{ij} \phi_{ij} \left| \hat{h}_{ij} \right|^2}{s_{ij}} \right) - \tilde{\rho}_j \right) + \sum_{i=1}^{N_F} \varphi_i \left(\sum_{j=1}^{N_F} s_{ij} - 1 \right) \qquad (4.15)$$

After finding the Karush-Kuhn-Tucker (KKT) condition through this Lagrangian, the jointly optimal subcarrier allocation and rate allocation is obtained.

Given the estimated CSIT realization $\hat{h}_{ij}$, the optimal subcarrier allocation $S_{opt}\left[\hat{H}\right] = \left[s_{ij}\right]$ can be decoupled between N_F subcarrier is given (Boyd and Vandenberghe 2004):

For $i = 1 : N_F$

$$j^* = \arg\max_{j \in [1,K]} (1 + \gamma_j)(1 - \varepsilon) \left(\log_2 \left(\frac{(1 + \gamma_j)(1 - \varepsilon)}{\mu} \phi_{ij} \left| \hat{h}_{ij} \right|^2 \right) \right)^+ - \mu \left(\frac{(1 + \gamma_j)(1 - \varepsilon)}{\mu} - \frac{1}{\phi_{ij} \left| \hat{h}_{ij} \right|^2} \right)$$

(4.16)

$$s_{ij} = \begin{cases} 1, & j = j^* \\ 0, & otherwise \end{cases}$$

END

The subcarrier allocation strategy Equation (4.16) can be decoupled between N_F subcarriers and thus a greedy algorithm with linear complexity is feasible with N_F x K requirements only.

4.6 SIMULATION RESULTS AND DISCUSSIONS

In the subsequent simulation analysis, from the literature of Jang et al (2003) and Song et al (2005b), the following naïve baseline schedulers (designed assuming the CSIT is perfect) are also investigated and compared.

1) Pure Opportunistic Scheduling: In subcarrier allocation part, each subcarrier *i* is assigned to the best user $j^* = \arg\max_{j \in [1,K]} \left|\hat{h}_{ij}\right|^2$.

2) Fixed Assignment Scheduling*:* Each subcarrier is always assigned to a fixed user and power is evenly distributed among subcarriers.

In both Opportunistic and Fixed assignment scheduling, the rate allocation is $r_{ij} = \left(1 - p_{ij}\left|\hat{h}_{ij}{}^2\right|\right)$.

4.6.1 Simulation Environment

For the purpose of the simulation, an OFDMA system with total system bandwidth of 5 MHz consisting of 256 subcarriers; each subcarrier with a bandwidth of 19 KHz and experiencing a channel fading is considered. The target outage probability of each subcarrier is set to $P_{out,ij}$ = 0.01. The duration of a scheduling slot is assumed to be 2ms. It is assumed that all users are of the same distance from the BS, and thus they are assumed to be homogeneous in terms of path loss. The channel fading between different users and different subcarriers is modelled as independent and identically distributed complex Gaussian with unit variance. Each packet consists of 1.024 kbits. We consider four classes of users in the system with arrival rates and delay requirements of each class being specified by (λ,T) ={(0.2, 2), (0.3,4), (0.4,100), (0.5,400), (0.6,1000)} (packets per time slot, time slots). Class 1 and Class 2 users represent delay sensitive traffic with heterogeneous delay requirements while Class 3 and Class 4 users represent delay insensitive applications with heterogeneous traffic loading. The system also contains some background users which are unclassed, having no delay constraint.

In this section, the performance of the proposed cross-layer scheduling with optimal subcarrier assigning strategy is compared with that of the naïve opportunistic scheduling and the conventional baseline reference – fixed power and subcarrier assignment scheduling. All schedulers under comparison have the same linear complexity in terms of number of subcarriers and users.

4.6.2 Throughput Performance of the Proposed Cross-Layer Scheduling with CSIT Error

Figure 4.2 illustrates the normalized average system throughput versus scheduled data traffic rate for transmission. The proposed scheduling provides substantial throughput enhancement over opportunistic and fixed assignment scheduling in the presence of CSIT error $(\sigma_{\Delta h}^{2}=0.05)$. It is observed that at low scheduled data rate traffic i.e., at moderate and high SNR regime (above 7 dB), the number of users waiting for transmission is less and the system throughput is higher when delay requirement of the users are more stringent. This is because in high SNR regime, the power levels are the same for all users and thus the optimal subcarrier allocation reduces to the conventional delay-insensitive scheduling policy. In low SNR regime (below 7 dB), when the scheduled data rate traffic is above 6 Mbps (heavy loading), the throughput performance is degrading because of the increment in the number of users waiting for transmission regardless of the value of the imposed delay constraint. This is because more urgent users with heavy traffic loading (higher arrival rate) will have higher transmission power and thus have higher chances of seizing subcarriers for transmission. It is also observed that the minimum required power to support all delay constraints of the user would increase as the delay requirements become more stringent.

Furthermore, the opportunistic scheduling performs adaptive subcarrier allocation based only on the estimated channel condition $\hat{H} = [\left|h_{ij}\right|^2]$ with single level power allocation, i.e., $p_{ij} = \left(1/\mu - 1/\left|\hat{h}_{ij}\right|^2\right)$, in which $(x)^+ = \max(0, x)$, while in fixed power and subcarrier assigning scheduling, each subcarrier is always assigned to a fixed user and power is evenly distributed among subcarriers.

From Figure 4.2, it can be seen that the opportunistic scheduling strategy achieves the better system throughput compared with the fixed subcarrier assigning scheduling.

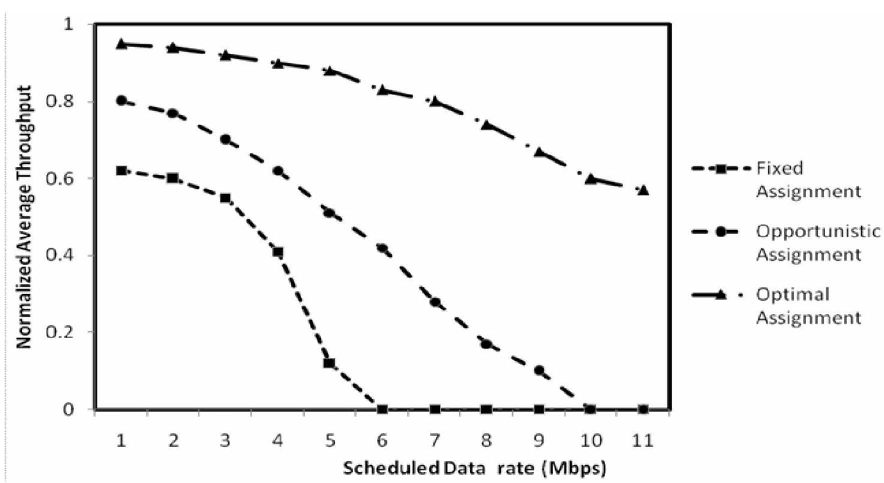

Figure 4.2 Normalized Average Throughput Versus Scheduled Data Rate with CSIT Error Variance = 0.05

This is because when the number of users is large, the multiuser diversity gain ensures that the effective SNR per subcarrier is high and hence, power adaptation provides marginal gains only. On the other hand, when the number of users is smaller, the power adaptation becomes more important.

But in both cases, there is significant throughput reduction in opportunistic and fixed subcarrier assigning strategies compared to the proposed optimal assigning strategy. This is because of the adaptation in subcarrier, rate and power assignment in the proposed scheduling. It shows the significance of the proposed cross-layer scheduling.

4.6.3 Delay Performance of the Proposed Cross-Layer Scheduling with CSIT Error

Figure 4.3 illustrates the average delay performance for different arrival rates of the user's packets with fixed CSIT error variance $(\sigma_{\Delta h}^2 = 0.05)$. In fixed assignment scheduling algorithm (Wong et al 1999), each subcarrier is always assigned to a fixed user and the power is evenly distributed among subcarriers. So the delay performance is constant for all arrivals rates. In opportunistic assigning scheduler (Jang and Lee 2003), each subcarrier is assigned to the user holding the best channel gain. As arrival rate of background user increases, the channel gain degrades significantly. Therefore, the delay performance of all users under opportunistic scheduling degrades significantly. It is also noted that by using the proposed scheduler, with the increase in traffic loading (number of packets per time slot), the traffic requirements of delay-sensitive users are satisfied, while the only price to be paid is an increased average delay for those delay insensitive users.

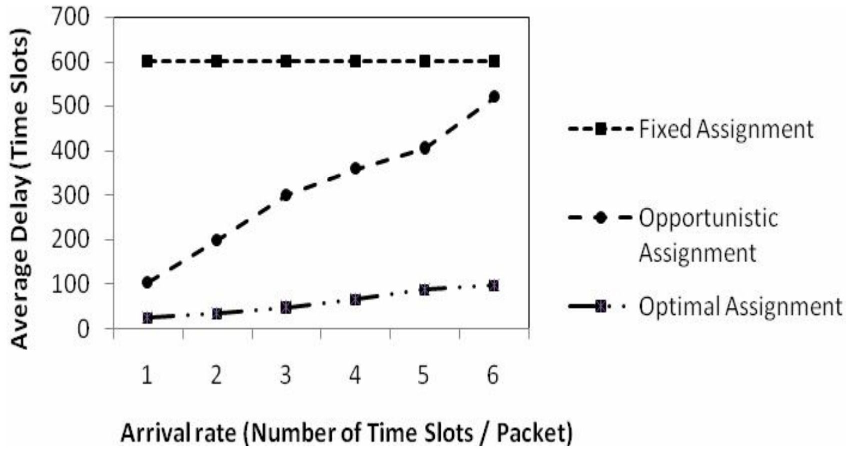

Figure 4.3 Average Delay in Time Slots Versus Arrival Rate in Number of Packets per Time Slot with CSIT Error Variance =0.05

Such characteristic of delay performance guarantee is important for serving bursty delay-sensitive real time heterogeneous traffic in the next-generation wireless networks. Figure 4.4 shows the average delay performance versus CSIT error variance with fixed arrival rate (λ=0.6).

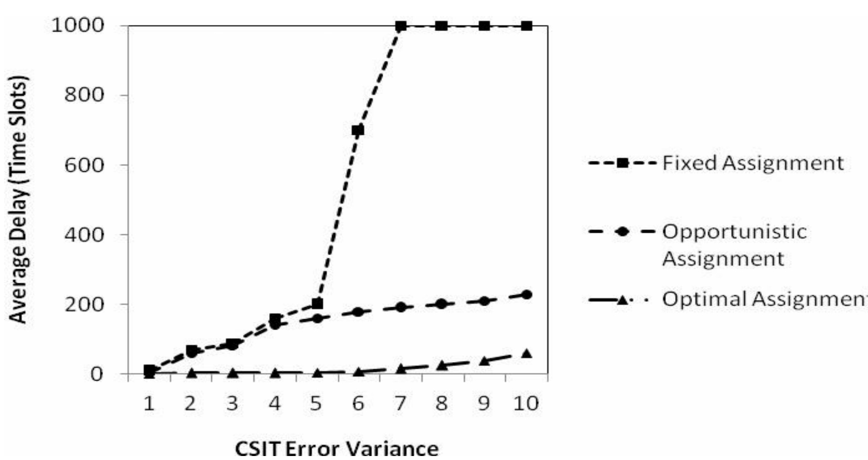

Figure 4.4 Average Delay Versus CSIT Error Variance for Different Schedulers with Arrival Rate λ = 0.6

Even in low CSIT error variance, the average delay of all users in the conventional fixed assigning scheduler degrades significantly (and hence it obviously cannot meet any delay requirements). Also it is shown that the delay performance of the proposed optimal subcarrier assigning cross-layer scheduler is very robust even at moderate to high CSIT errors, whereas the delay performance of fixed assigning strategy and opportunistic based schedulers is very sensitive even at low CSIT error variance. This robustness to CSIT errors introduced by the proposed scheduler is significant for practical implementation on OFDMA-TDD system in which obsoleteness of CSIT is often not negligible.

4.7 SUMMARY

In this chapter, an analytical framework is proposed for the design of optimal subcarrier assigning cross-layer scheduler for delay-sensitive users with heterogeneous delay requirements in the presence of obsolete CSIT in multiuser OFDM systems. The cross-layer design problem is formulated as an optimization problem with consideration of the imperfect CSIT, source statistics and queue dynamics of the OFDM systems. The optimal rate and subcarrier allocation solutions are obtained based on the optimization framework. The proposed cross-layer scheduler offers a nice balance of maximizing throughput and providing delay differentiation of heterogeneous users with robust performance even at moderate to large CSIT error. It is also shown that substantial throughput gain is retained by the proposed scheduler and all users' delay constraints are maintained regardless of the variation of traffic loadings and CSIT error.