Design and Analysis of Delay-Sensitive Cross-Layer OFDMA Systems with Outdated CSIT

David Shui Wing Hui and Vincent Kin Nang Lau, *Senior Member*, IEEE

Abstract— It is well known that cross-layer scheduling can boost the spectral efficiency of multi-user OFDMA systems through multi-user selection diversity but existing designs usually have two important assumptions – users are delay-insensitive and channel state information at the transmitter (CSIT) is perfect. In practice, users have heterogeneous delay requirements and CSIT usually becomes outdated in time varying channel, which in turns leads to systematic packet errors and hence results in significant degradation on the throughput and delay performance in the OFDMA systems. In this paper, a novel cross-layer design problem is formulated as a convex optimization problem, on which a delay-sensitive jointly optimal power, rate and subcarrier allocation scheme is proposed so as to maintain heterogeneous users' delay requirement as well as achieving a target packet outage probability through combining queueing theory and information theory. Furthermore, we obtain closed-form asymptotic performance of the proposed delay-sensitive scheduler. Unlike the well-known SNR gain of $\Theta(\log K)$ in conventional cross-layer scheduler with perfect CSIT, we demonstrate a cross-layer SNR gain of $\Theta\left(\left(1-\sigma_{\Delta H}^2\right)\log K\right)$ can still be achieved under heterogeneous delay constraints and outdated CSIT with error variance $\sigma_{\Delta H}^2$. Finally, simulation results show that our proposed delay-sensitive CSIT error considerate schemes provide robust system performance enhancement over conventional CSIT error inconsiderate opportunistic scheduler and naive queue length based MAX-Weight scheduler while satisfying heterogeneous delay requirements even at moderate to high CSIT errors.

Index Terms— Orthogonal Frequency Division Multiple Access (OFDMA), heterogeneous applications, delay-sensitive cross-layer scheduling, outdated channel state information (CSI)

I. Introduction

Orthogonal Frequency Division Multiple Access (OFDMA) has been proposed as an indispensable mean to support high data rates demand in many applications such as WLAN and WiMAX. Many recent efforts in OFDMA development have been devoted to cross-layer scheduling in OFDMA systems ([1, 2, 3, 4], extensions to nonconvex objective [5], and references therein) due to its promising gain through exploitation of multi-user diversity by carefully assigning multiple users to transmit simultaneously on different subcarriers for each OFDM symbol with optimal power and rate allocations. However, these cross-layer designs rely on two important assumptions – users are delay-insensitive and channel state information at the transmitter (CSIT) is perfect. These assumptions are usually impractical since next generation networks are expected to contain real time users of heterogeneous classes with different delay requirements. Moreover, due to the delay and resource limitation in feedback of channel states[1], CSIT obtained at base station (BS) will be outdated and imperfect.

In view of the importance of delay sensitivity of real-time users and burstiness of traffic arrival, enormous delay-sensitive packet scheduling algorithms [6, 7] were proposed assuming simple error-free static channels wherein the CSI of the multi-user wireless channels were ignored, and thus the potential cross-layer gain is unexploited. Subsequently, a surge of recent research efforts is devoted to cross-layer scheduling aiming at providing a synergy between channel dynamics in physical layer and queueing dynamics in MAC layer in various contexts, such as ON-OFF channel based OFDMA system [8] and Multiuser MIMO system [9]. Several surveys provide useful summary of these recent cross-layer efforts [9-11] - in particular, [12] and [13] proposed a cross-layer scheduling algorithm, called Longer Queue Higher Possible Rate (LQHPQ)[2] based on combined information theory [14] and queueing theory [15, 16] to minimize average system delay in multiaccess channels for homogeneous users. Regarding cross-layer performance analysis, only limited results are available. Notably [17] provides an asymptotic delay-power tradeoff of a simple threshold based scheduling policy in point-to-point (single user) fading channel. However, perfect CSIT was assumed in all these works.

On the other hand, several recent publications addressed the effect of imperfect CSIT on scheduler design. There are two types of "imperfect CSIT", namely the "*limited CSIT*" and the "*outdated CSIT*". Limited CSIT refers to the incomplete knowledge of CSI at the transmitter (such as *limited* CSI feedback). For example, [18] discussed the power adaptation for OFDM system with limited CSIT to optimize the ergodic capacity. In contrast, "*outdated* CSIT" refers to the delay from the CSI estimation time to CSIT utilization time. Under outdated CSIT, systematic packet errors occurs 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

---

[1] The feedback of CSI to BS can be explicit feedback (for FDD systems) or implicit feedback (for TDD systems). In both cases, there will be feedback delay or duplexing delay which causes outdatedness in the CSIT.

[2] Another name of LQHPQ scheduler is MAX-WEIGHT scheduler in the literature with queue length being the weight.

performance. To our best knowledge, there are only a few works considered the outdated CSIT [19, 20] in single user OFDM systems and none of the works had addressed both issues of delay-sensitive applications and outdated CSIT in cross-layer design of OFDMA system. In general, the followings are critical issues to be addressed for delay-sensitive cross-layer design with outdated CSIT:

- *Design joint rate, power and subcarrier allocation for heterogeneous traffic and delay requirements*: It needs a careful combination of queueing theory and information theory. A common approach was to model the problem as a markov decision process (MDP) problem [13, 17, 21] but the solutions obtained would be very complicated when realistic channel state model is incorporated and no closed-form analytical performance could be obtained [21][3].
- *Introduce robustness to cross-layer design with respect to CSIT errors / outdatedness:* The packet errors due to channel outage will have significant impact on both the throughput and delay performance (due to retransmission of corrupted packet), yet it was ignored in conventional cross-layer design. In this paper, we shall illustrate that the delay performance of the conventional opportunistic scheduler [1, 2, 3] and simple MAX-WEIGHT delay-sensitive cross-layer scheduler (designed for perfect CSIT) [12, 13] are very sensitive to the CSIT errors. It is thus critical to take the potential packet errors into account during the joint rate, power and subcarrier allocation policy design for robust performance.
- *Study impacts of delay requirements and CSIT errors on the asymptotic cross-layer SNR gain (which was well-known to be* $\Theta(\log K)$) *and cross-layer system throughput gain*

In this paper, we shall address both the heterogeneous delay requirements and CSIT outdateness simultaneously with regard to these aspects. Based on a modified M/G/1 queue model (with packet errors and retransmissions consideration), the delay-sensitive cross-layer design is modeled as a convex optimization problem. Closed-form delay-sensitive rate, power and subcarrier allocation solutions are derived. The optimal delay-sensitive power allocation strategy is found to be of multi-level water-filling structure where users with stringent delay constraint and packet error (outage) requirements having a higher "water-level". The optimal delay-sensitive subcarrier assignment in the presence of CSIT errors is shown to be decoupled between all N_F subcarriers and thus, it has linear complexity with respect to N_F. Furthermore, we obtain closed-form asymptotic performance of the proposed delay-sensitive scheduler. Compared with the well-known SNR gain of $\Theta(\log K)$ in conventional cross-layer scheduler with perfect CSIT, a cross-layer SNR gain of $\Theta\left(\left(1-\sigma_{\Delta H}^2\right)\log K\right)$ is still demonstrated under

---

[3] High complexity value /policy iteration algorithm is usually required upon MDP when no analytical closed-form resulted.

heterogeneous delay constraints and outdated CSIT with error variance $\sigma_{\Delta H}^2$. Finally, simulation results show, by considering CSIT error statistics in cross-layer design, our proposed scheme provides robust performance gain while satisfying heterogeneous user delay requirements even at high CSIT errors.

This paper is organized as follows. Section II describes the system model, including downlink channel model, multiuser physical layer model, source model and MAC layer model. Section III formulated the optimization problem. The corresponding optimal rate, power and subcarrier allocation policy is given in Section IV. Asymptotic cross-layer gain from the proposed scheduler is presented in Section V. Simulation results are studied in Section VI with conclusion finally presented in Section VII.

II. SYSTEM MODEL

The general cross-layer system model of multiuser wireless systems is shown in Figure 1 where outdated CSIT and queue states information (QSI) are the inputs to the scheduler at BS. Before we formulate the cross-layer design into an optimization problem, we shall elaborate the OFDMA channel model, the corresponding CSIT error model, multiuser physical layer, source, and MAC layer model.

A. Downlink Channel Model and CSIT estimation from Outdated CSIT

We consider a *K* users OFDMA system with frequency-selective channel model consisting of $L = \left\lfloor BW / \Delta f_c \right\rfloor = \left\lfloor \text{Signal Bandwidth} / \text{Coherent Bandwidth} \right\rfloor$ resolvable paths. For simplicity, uniform power-delay profile is adopted, i.e. each path has normalized power given by 1/*L*. Thus the channel impulse response between the BS and the *j*-th user at time slot *m*, $h_j(m)$, can be modeled through a *L*-tap delay line channel model, i.e. $h_j(m) = \sum_{l=0}^{L-1} h_{j,l}(m)\delta(m - l/W)$, where $\left\{h_{j,l}\right\}$ are modeled as independent identically distributed (i.i.d.) circularly symmetric complex Gaussian (CSCG) random variables with distribution $CN\left(0, 1/L\right)$. $h_{j,l}(m)$ is assumed to be quasi-static within each time slot *m*, but slowly time-varying across time slots according to Jakes' model such that $E[h_{j,l}(m)h_{j,\tilde{l}}^*(n)] = J_0(2\pi f_d \mid m - n \mid t_s)$ if $l = \tilde{l}$ (else 0), where $J_0(.)$ denotes zero order Bessel function of the first kind, t_s is scheduling slot duration, and f_d is Doppler spread of the channel (with $t_s \ll$ coherent time, i.e. $t_s \ll 1/f_d$). With N_F -point IFFT and FFT in the OFDMA system, the equivalent discrete channel model in the frequency domain (after the length-L cyclic prefix removal) is:

$$Y_{ij} = H_{ij} U_{ij} + Z_{ij} \tag{1}$$

where (i denotes subcarrier index and j denotes user index) Y_{ij} is the received symbol, U_{ij} is the data symbol from BS, Z_{ij} is the noise distributed with $CN\left(0,\sigma_z^2\right)$, $H_{ij}=\sum_{l=0}^{L-1} h_{j,l} e^{-j2\pi li/N_F}$ is the channel gain distributed with $CN(0,1)$ which is i.i.d. for different users, and but correlated within user *j*, where the correlation of the channel gain between the *i*-th and $\tilde{i}$ -th subcarriers of user *j* is given by $E\left[H_{ij}H_{\tilde{i}j}^H\right]=\frac{1}{L}\frac{1-e^{-2j\pi L(i-\tilde{i})/N_F}}{1-e^{-2j\pi(i-\tilde{i})/N_F}}$. The transmit power allocated from BS to user *j* through subcarrier *i* is given by $p_{ij}=E[\left|U_{ij}\right|^2]$. We define a subcarrier allocation strategy $S_{N_F\times K}=[s_{ij}]$, where $s_{ij}=1$ when user *j* is selected for subcarrier *i*, otherwise $s_{ij}=0$. The average total transmit power of BS is constrained by $E\left[\left(1/N_F\right)\sum_{j=1}^{K}\sum_{i=1}^{N_F} s_{ij}p_{ij}\right]\le P_{TOT}$, where P_{TOT} is available average power in BS.

Assume the system is using TDD 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 outdated. For example, in the beginning of each scheduling slot, a Minimum Mean Square Estimation (MMSE) on CSIT at time slot *m*, $h_{j,l}(m)$, is performed based on an outdated CSIT $h_{j,l}(m-D)$. Thus the estimated downlink CSIT in frequency domain $\left\{\hat{H}_{ij}\right\}$ for all users over all subcarriers at BS accounting the CSIT outdatedness is modeled as:

$$\hat{H}_{ij}=H_{ij}+\Delta H_{ij},\ \text{where } \left\{\Delta H_{ij}\right\} \text{ is the CSIT error with zero mean CSCG distribution.} \tag{2}$$

Suppose the uplink transmit power of user *j* (for CSIT estimation) is P , then the correlation of the CSIT error between the *i*-th and $\tilde{i}$ -th subcarriers of user *j* is given by (details are omitted for simplicity)

$$E\left[\Delta H_{ij}\left(\Delta H_{\tilde{i}j}\right)^*\right]=\frac{\sigma_{\Delta H}^2}{L}\frac{1-e^{-2j\pi L(i-\tilde{i})/N_F}}{1-e^{-2j\pi(i-\tilde{i})/N_F}},\ \text{where}\ \sigma_{\Delta H}^2=1-\frac{\left|J_0(2\pi f_d t_s D)\right|^2}{1+L\sigma_z^2/NP}.$$

B. Multi-user PHY Model for OFDMA Systems with Packet Outage and Goodput Modeling

We consider information theoretical capacity [14] as the abstraction of the multi-user physical layer in order to decouple the problem from specific implementation of coding and modulation schemes.

In general, packet error is contributed by two factors, namely the “channel noise” and the “channel outage”. In the former case, packet error is contributed by the effect of non-ideal channel coding and finite block length of the channel codes. This factor can be reduced by using a strong channel code (e.g. LDPC or turbo code) and longer block length (e.g. 10K bits), so that Shannon’s capacity can be achieved to within 0.05dB at 1% target Frame Error Rate. However, in the latter case, the effect is systematic and cannot be eliminated, because the instantaneous mutual information between BS and user j in i-th subcarrier, $c_{ij} = \max_{p(X_{ij})} I(U_{ij};Y_{ij} \mid H_{ij}) = \log_2(1+p_{ij} \mid H_{ij} \mid^2 / \sigma_z^2)$, is a function of actual CSI H_{ij}, which is unknown to BS. Packets will be corrupted whenever scheduled data rate exceeds instantaneous mutual information. And so probability of this event is the dominating factor of packet error.

To take account of the packet error due to channel outage, we define the instantaneous goodput of the j^{th} user (which measures the total instantaneous data bits/s/Hz successfully delivered to user j) as:

$$g_j = \sum_{i=1}^{N_F} r_{ij} I[r_{ij} \le c_{ij}], \text{ where } I(.) \text{ is the indicator function i.e. } I[r_{ij} \le c_{ij}] = \begin{cases} 1 & \text{if } r_{ij} \le c_{ij} \\ 0 & \text{if } r_{ij} > c_{ij} \end{cases}, \tag{3}$$

where r_{ij} is the scheduled data rate of the j^{th} user on the i^{th} subcarrier.

Hence, the average goodput of user j (averaged over ergodic realizations of $\{H_{ij}\}$ *and* $\{\hat{H}_{ij}\}$) is:

$$\bar{g}_j = E_H\left[g_j\right] = E_{\hat{H}}\left\{E_{H|\hat{H}}\left[\sum_{i=1}^{N_F} r_{ij} I[r_{ij} \le c_{ij}]\right]\right\} = E_{\hat{H}}\left\{\sum_{i=1}^{N_F} r_{ij} E_{H|\hat{H}}\left[I[r_{ij} \le c_{ij}]\right]\right\} = E_{\hat{H}}\left\{\sum_{i=1}^{N_F} r_{ij}\left(1-P_{out,ij}\right)\right\}. \tag{4}$$

$P_{out,ij} = 1-\Pr\left[r_{ij} \le c_{ij} \middle| \hat{H}\right]$ is packet outage probability for each subcarrier i conditioned on $\hat{H} = \{\hat{H}_{ij}\}$.

C. Source Model

We assumed packets come into each user j's buffer according to a Poisson process with independent rate λ_j with packets of fixed size F. Furthermore, we consider the scenario that mobile user applications are heterogeneous in nature in terms of their packets' arrival rates and delay requirements. User j is characterized by a tuple $[\lambda_j, T_j]$, where λ_j is the average arrival rate to user j and T_j is the average delay requirement imposed by user j. User j with heavier traffic loading will have a higher λ_j and more delay-sensitive user j application will have stringent delay requirement T_j.

D. MAC Layer Design Model from Cross-Layer Perspective

The system dynamics are characterized by system state $\chi = (\hat{H}_{N_F \times K}, Q_K)$, which composes of estimated CSIT $\hat{H}_{N_F \times K}$ from Physical Layer and queue state information (QSI) Q_K from MAC Layer users' buffer, where $Q_K = [q_j]$ is a $K \times 1$ vector with the j^{th} component denotes the number of packets remains in user j's buffer. The MAC layer is responsible for the cross-layer scheduling[4] at

[4] Cross-layer scheduling refers to resource allocations for different users, instead of flow control/scheduling.

every fading block based on current system state χ. At the beginning of every frame, the BS estimates the CSIT from dedicated uplink pilots and observes the current backlogs in the buffer. Based on CSIT and QSI obtained, the scheduler determines the subcarrier allocation from the policy $S_{N_F \times K}[\hat{H}, Q]$, the power allocation from the policy $P_{N_F \times K}[\hat{H}, Q]$ and the corresponding rate allocation from the policy $R_{N_F \times K}[\hat{H}, Q]$ for the selected users. The scheduling results are then broadcasted on downlink common channels to all mobile users before subsequent downlink packets transmissions at scheduled rates.

III. CROSS-LAYER PROBLEM FORMULATION

In this section, we shall formulate the OFDMA cross-layer design for heterogeneous users with outdated CSIT as a constrained optimization problem based on the system models in Section II.

To take into account of potential packet outage, we shall adopt total average system goodput, $\sum_{j=1}^{K} \bar{g}_j$ as our optimization objective. Specifically, the optimization problem is formulated as:

Cross-layer Formulation:

Find optimal rate, subcarrier and power allocation policies ($R_{N_F \times K}[\hat{H}, Q], S_{N_F \times K}[\hat{H}, Q], P_{N_F \times K}[\hat{H}, Q]$) s.t.:

$$\max_{S,P,R} E\left(\sum_{i=1}^{N_F} \sum_{j=1}^{K} r_{ij} \Pr[r_{ij} \le c_{ij} \mid \hat{H}]\right)$$

$$\textit{subject to } (C1): s_{ij} \in \{0,1\}, \quad (C2): \sum_{j=1}^{K} s_{ij} = 1, \quad (C3): p_{ij} \ge 0, \quad (C4): E[(1/N_F) \sum_{j=1}^{K} \sum_{i=1}^{N_F} s_{ij} p_{ij}] \le P_{TOT}, \quad (C5): P_{out,ij} = \varepsilon, \quad (C6): E[\tilde{W}_j] \le T_j \tag{5}$$

where the expectation $E[.]$ is taken over all system state and P_{TOT} is the average total power constraint.

In the optimization problem (5), constraints (C1) and (C2) are used to ensure only one user can occupy a subcarrier *i* at one time. (C3) is used to ensure transmit power would only take positive value, (C4) is the average total power constraint, (C5) ensures the outage probability $P_{out,ij}$ satisfies a target outage probability ε, which is usually specified by applications requirements and (C6) is the average delay constraint[5] where $E[\tilde{W}_j]$ is average system time (including waiting time and service time) of user *j*.

[5] Delay constraint could also be defined as delay outage requirement, i.e. each user *j* has a maximum allowable probability δ_j of exceeding a fixed delay requirement threshold T_j [$P(\tilde{W}_j > T_j) \le \delta_j$]. But under heavy traffic loading, this delay outage requirement can be rewritten as average delay constraint $E[\tilde{W}_j] \le T_j / \ln(1/\delta_j)$ [p.290, [16]] for GI/G/1 queue (hence also for M/G/1). Moreover, "average delay" is also widely adopted in the literature [8, 12, 13, 17] as a performance measure of the delay performance. In short, average delay is usually a good characterization of overall delay performance.

A. Relationship between scheduled data rate and average delay requirements

To solve optimization problem (5), delay constraint (C6) have to be expressed in terms of optimization variables - physical layer parameters. Our previous work (Lemma 1 [22]) shows that the queue for each user *j* in OFDMA system can be modeled as a modified M/G/1 queue (with non-selected time slot) and hence (C6), the constraint on average system time $E[\tilde{W}_j]$ of user *j*'s packet, can be rewritten as, $E[\tilde{W}_j] = E[X_j] + \dfrac{\lambda_j E[X_j^2] + \lambda_j E[X_j](E[\overline{S_j}]/E[S_j])(t_s)}{2\left(1-\lambda_j(E[X_j]/E[S_j])\right)} \le T_j$, where X_j is the effective transmission time of the packet of user *j* (retransmission time for packet outage is also accounted), λ_j and T_j are user *j*'s arrival rate and average delay requirement respectively, and S_j is an indicator variable of the availability of subcarrier for user *j*[6]. Notably, in this paper, CSIT outdateness 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)[7]. Based on the Lemma 1 [22], (C6) can be transformed to a constraint that directly relate CSIT-dependent scheduled data rate R_j of user *j* to its characteristic tuple $[\lambda_j, T_j]$, and also packet size *F*.

Corollary 1: A necessary condition for the constraint (C6) is

$$E(S_j R_j)(1-\varepsilon) \ge \rho_j(\lambda_j, T_j, F), \text{ where } \rho_j(\lambda_j, T_j, F) = (((2T_j\lambda_j+2)+\sqrt{(2T_j\lambda_j+2)^2-8T_j\lambda_j})\big/4T_j)F \quad (7)$$

The equivalent rate requirement for user *j*, in the subsequent context would be referred to, as $\tilde{\rho}_j(\lambda_j, T_j, F) = \rho_j(\lambda_j, T_j, F)/\left(t_s \times BW/N_F\right)$ (in bits/s/Hz).

IV. Scheduling Strategies

Optimization problem (5) is a mixed combinatorial (in $\{s_{ij}\}$) and convex (in $\{p_{ij}\}$) optimization

---

6 In practice $N_F \gg K$, so there is always a subcarrier available for any particular user *j*, i.e. $E[S_j]=1$ and $E[\overline{S}_j]=0$.

7 Since the target outage probability $P_{out} = P_{out,ij} = \varepsilon$ for each user *j*'s allocated subcarrier *i* is fixed for all time slots. As a result, before each time slot of successful transmission occurs, there are $P_{out}\big/(1-P_{out})$ outage slots on average (which is the mean of geometric distribution with parameter 1-P_{out}), i.e. the new transmission duration is $1+P_{out}\big/(1-P_{out})$ times compared to that of the duration with perfect CSIT. Thus, the average service time of user *j* (in terms of number of time slot), denoted as $E[X_j]$, can be calculated as, by the total service time of all packets average over the number of packets of user *j* that are ever served, and is given by $E[X_j] = \dfrac{E[S_j]F}{E[S_j R_j](1-P_{out})}$ as in the perfect CSIT case [22].

problem, which is difficult to be solved in general. We shall simplify it in the following ways.

A. Optimal Delay-sensitive Subcarrier, Power and Rate Allocation (matched to the CSIT errors)

Given any CSIT $\hat{H}_{ij}$, the actual CSI H_{ij} is Gaussian distributed with mean and variance given by $E_{H|\hat{H}}[H_{ij} \mid \hat{H}_{ij}] = \hat{H}_{ij}$ and $E_{H|\hat{H}}[(H_{ij} - \hat{H}_{ij})^*(H_{ij} - \hat{H}_{ij}) \mid \hat{H}_{ij}] = \sigma_{\Delta H}^2$ respectively. Hence, $|H_{ij}|^2 / (\sigma_{\Delta H}^2 / 2)$ is a noncentral chi-square random variable with 2 degrees of freedom and noncentric parameter $\theta = |\hat{H}_{ij}|^2 / \sigma_{\Delta H}^2$ and with c.d.f. $F_{\chi_2^2(\theta)}(x)$. To satisfy the target outage probability ε, the rate allocation policy is given by: $r_{ij} = \log_2(1 + p_{ij}\varphi_{ij} |\hat{H}_{ij}|^2 / \sigma_z^2)$, where $\varphi_{ij} = F_{\chi_2^2(\theta)}^{-1}(\varepsilon) / \theta$. (8)

To avoid high complexity in solving mixed integer and convex optimization problem, the Boolean constraint (C1) is further relaxed to a real number between [0,1] - $s_{ij} \in [0,1]$ is a sharing factor indicating fraction of time that user *j* have occupied the subcarrier *i*. Together with the definition of variable $\hat{p}_{ij} = p_{ij} s_{ij}$, problem (5) is reformulated as a convex optimization problem in $\left(\hat{p}_{ij}, s_{ij}\right)$. Using Lagrange Multiplier techniques [23], the following Lagrangian is obtained as:

$$L = \sum_{j=1}^{K} \left(1 + \gamma_j\right) \sum_{i=1}^{N_F} s_{ij}(1-\varepsilon) \log_2\left(1 + \frac{\hat{p}_{ij}\varphi_{ij} |\hat{H}_{ij}|^2}{\sigma_z^2 s_{ij}}\right) - \mu\left(\sum_{j=1}^{K}\sum_{i=1}^{N_F} \hat{p}_{ij} - N_F P_{ToT}\right) - \sum_{j=1}^{K} \tilde{\rho}_j + \sum_{i=1}^{N_F} \phi_i \left(\sum_{j=1}^{K} s_{ij} - 1\right) \quad (9)$$

where $\mu \geq 0, \gamma_j \geq 0, \phi_i$ are Lagrange multipliers. Using standard optimization techniques [23], we get the optimal power and subcarrier allocation stated in the following Theorem 1[8].

Theorem 1: Given CSIT realization $\hat{H} = \left\{\hat{H}_{ij}\right\}$, optimal subcarrier allocation $S_{opt}(\hat{H}) = [s_{ij}]$ is:

$$For\ i = 1 : N_F, \quad Perform\left(j^* = \arg\max_{j \in [1,K]} \Psi_{ij}\left(c_j, \varphi_{ij} |\hat{H}_{ij}|^2\right), s_{ij} = \begin{cases} 1, & j = j^* \\ 0, & otherwise \end{cases} \right), \quad END \qquad (10)$$

The corresponding optimal power allocation $P_{opt}(\hat{H}) = [p_{ij}]$ is given by:

$$p_{ij} = s_{ij}\left(c_j - 1 \Big/ \left(\varphi_{ij} |\hat{H}_{ij}|^2\right)\right)^+ \qquad (11)$$

where $(x)^+ = \max(0, x)$, $\Psi_{ij}\left(c_j, \varphi_{ij} |\hat{H}_{ij}|^2\right) = c_j \left(\log_2\left(c_j \varphi_{ij} |\hat{H}_{ij}|^2\right)\right)^+ - \left(c_j - 1 \Big/ \left(\varphi_{ij} |\hat{H}_{ij}|^2\right)\right)^+$ is a function of scaled channel gain $\varphi_{ij} |\hat{H}_{ij}|^2$ and $c_j = (1+\gamma_j)(1-\varepsilon) / \mu$ is the user *j*'s water-level.

[8] Without loss of generosity, we assume $\sigma_z^2 = 1$ (unit noise variance) in subsequent descriptions for presentation simplicity.

In Theorem 1[9], subcarrier allocation strategy (10) can be decoupled between N_F subcarriers, and thus a greedy algorithm with linear complexity is feasible - $N_F \times K$ only. The optimal power allocation $P_{opt}(\hat{H}) = [p_{ij}]$ expressed in (11) can be interpreted as a *multi-level water-filling strategy*[10] It means those users with urgent packets have to transmit at higher (urgency and outage target dependent) power level, while non-urgent users (with strict inequality $E[\tilde{W}_j] < T_j$) have same power level.

Note that some user requirement specifications may not lead to feasible solution in problem (5). The minimum required power $P_{\min,opt}$ to support delay constraints (or equivalent rate requirements $\tilde{\rho}_j$) for all users specified in problem (5) is given by $P_{\min,opt} = E[\left(1/N_F\right)\sum_{i=1}^{N_F}\sum_{j=1}^{K} s_{ij}(c_j - 1/(\varphi_{ij}\,|\,\hat{H}_{ij}\,|^2))^+]$, where c_j is user j's water-level by solving $E[\sum_{i=1}^{N_F} s_{ij}(\log_2(c_j\varphi_{ij}\,|\,\hat{H}_{ij}\,|^2))^+] = \tilde{\rho}_j,\ \forall j$.

V. Asymptotic Cross-Layer Gains

It is well-known that the cross-layer SNR gain (without delay constraint and under perfect CSIT) scales as $\Theta\left(\log(K)\right)$[11] for large K (see [3] for example). In this section, we shall study the asymptotic cross-layer gain of the proposed delay-sensitive scheduler and various existing schedulers under heterogeneous delay constraints and outdated CSIT. For illustration, we consider an OFDMA system with 2 classes of users (K_1 delay-sensitive Class 1 users and K_2 delay-insensitive Class 2 users).

Given average delay requirements (T_1, T_2) and arrival rates (λ_1, λ_2), (or equivalent rate requirements $\tilde{\rho}_1, \tilde{\rho}_2$) for class 1 and class 2 users, $P_{TOT} \ge P_{\min,opt}$ (where $P_{\min,opt}$ is minimum required power to satisfy delay constraints) and large number of users $K(=K_1+K_2)$, the following lemmas summarize the cross-layer gains and minimum power requirement with heterogeneous users and outdated CSIT.

Lemma 2: The conditional cross-layer SNR gains for Class 1 and Class 2 users are both

$$E[s_{ii}\varphi_{ii}\,|\,\hat{H}_{ii}\,|^2\,\big|s_{ii}=1,\, j\in Class_d] = \Theta\left(\left(1-\sigma_{\Delta H}^2\right)\log\left(K\right)\right),\ \forall d = 1, 2, \text{ for large } \left(1-\sigma_{\Delta H}^2\right)\log\left(K\right). \quad (12)$$

---

[9] Note that subcarrier allocation in (10), derived as an optimal solution to relaxed problem of (5) is also optimal solution to original problem (5). It is because $\Psi_{ij}(c_j, \varphi_{ij}\,|\,\hat{H}_{ij}\,|^2)$ are different for different user j with probability 1 since $\hat{H}_{ij}$ are i.i.d. for different j. As a result, s_{ij} is almost surely either 1 or 0 in (10).

[10] The methodology to find the water-levels described in our previous work [22] is also suitable for the current context.

In the subsequent analysis, the following naïve baseline schedulers (designed assuming the CSIT is perfect) will also be investigated and compared:

1) ***Pure opportunistic scheduler*** *[1,2,3]:* In subcarrier allocation part, each subcarrier *i* is assigned to the best user j^* through the rule - $j^* = \arg\max_{j \in [1,K]} |\hat{H}_{ij}|^2$. In power allocation part, it performs single level $1/\mu$ waterfillling, i.e. $p_{ij} = \left(1/\mu - 1/|\hat{H}_{ij}|^2\right)^+$ is power allocated to user *j* in subcarrier *i*.

2) ***Fixed assignment strategy*** *[1,2,3]:* Each subcarrier is always assigned to a fixed user, and power is evenly distributed among subcarriers. In both *1)* and *2)*, the rate allocation is $r_{ij} = \log_2\left(1 + p_{ij}|\hat{H}_{ij}|^2\right)$.

Lemma 3: The minimum required power (to satisfy the delay requirements) of proposed cross-layer scheduler $P_{\min,opt}$, pure opportunistic scheduler $P_{\min,OS}$ and fixed assignment $P_{min,fixed}$ are given by

$$P_{\min,opt} = \Theta\left(\frac{2^{(\tilde{\rho}_1 K_1 + \tilde{\rho}_2 K_2)/((1-\varepsilon)N_F)}}{\left(1-\sigma_{\Delta H}^2\right)\log(K_1+K_2)}\right), P_{\min,OS} = \Theta\left(\frac{2^{(\max_d \tilde{\rho}_d)2(K_1+K_2)/N_F}}{\left(1-\sigma_{\Delta H}^2\right)\log(K_1+K_2)}\right), \text{and } P_{\min,\text{fixed}} = \Theta\left(\frac{2^{(\max_d \tilde{\rho}_d)2(K_1+K_2)/N_F}-1}{\left(1-\sigma_{\Delta H}^2\right)}\right)$$

respectively, where $\tilde{\rho}_d(\lambda_d, T_d, F)$ is the equivalent rate constraint of class *d* user mentioned in (7).

Hence, the relative saving in minimum required power using the proposed cross-layer scheduler compared to opportunistic scheduler is $P_{\min,OS}/P_{\min,opt} = \Theta(2^{\left(\left(2(\max_d \tilde{\rho}_d) - \tilde{\rho}_1/(1-\varepsilon)\right)K_1 + \left(2(\max_d \tilde{\rho}_d) - \tilde{\rho}_2/(1-\varepsilon)\right)K_2\right)/N_F})$ and to fixed assignment is $P_{\min,fixed}/P_{\min,opt} = \Theta((2^{\left(\left(2(\max_d \tilde{\rho}_d) - \tilde{\rho}_1/(1-\varepsilon)\right)K_1 + \left(2(\max_d \tilde{\rho}_d) - \tilde{\rho}_2/(1-\varepsilon)\right)K_2\right)/N_F})\log K)$. Besides, the proposed CSIT-error considerate scheduler gives fastest gain in goodput per unit increase in power:

Lemma 4: Suppose total goodput at $P_{TOT} = P_{\min}$ and $P_{\min} + \Delta P$ are denoted by $G_{P_{\min}}$ and $G_{P_{\min}+\Delta P}$ respectively. The total goodput gain ($\Delta G = G_{P_{\min}+\Delta P} - G_{P_{\min}}$), under proposed cross-layer scheduler, pure opportunistic scheduler and fixed assignment, per ΔP increase in power over, $P_{\min,opt}, P_{\min,OS}$, and $P_{\min,fixed}$ are given by $(1-\varepsilon)N_F \log_2\left(1+\frac{\Delta P}{P_{\min,opt}}\right), \frac{1}{2}N_F \log_2\left(1+\frac{\Delta P}{P_{\min,OS}}\right)$, and $\frac{1}{2}N_F \log_2\left(1+\frac{\Delta P}{P_{\min,fixed}}\right)$

Proof: Proof of Lemma 2, Lemma 3 and Lemma 4 are presented in the Appendix A.

---

[11] $a_K = \Theta(b_K)$ if $\limsup_{K\to\infty} |a_K|/|b_K| < \infty$ and $\limsup_{K\to\infty} |b_K|/|a_K| < \infty$.

A. Asymptotic Performance Analysis results' interpretations

The intuition behind the cross-layer SNR gain (with outdated CSIT) in Lemma 2 is given as follows:

1) ***Impact of heterogeneous delay requirements***: The proposed scheduler can still retain the same order of multiuser diversity gain $\Theta(\log(K))$ as $K \to \infty$, even with heterogeneous delay constraints. It is because each subcarrier would be assigned to the best user from one class. This best user (such as from class *d*), is chosen according to pure opportunistic scheduler in single class scheduling [3], already achieves a $\Theta(\log(K_d))$ gain. Thus no matter the final selected user belongs to which class, the final multiuser diversity gain is better than or equal to the class *d*'s gain $\Theta\left(\log(K_d)\right)$ (i.e. $\Theta(\log(K)))$.

2) ***Impact of CSIT outdatedness***: Since the factor $F^{-1}_{\chi_2^2(\theta)}(\varepsilon)$ grows in the same rate as the noncentral parameter θ, $\varphi_{ij^*} = F^{-1}_{\chi_2^2(\theta)}(\varepsilon)/\theta$ does not affect the growth order of multiuser diversity gain, and hence $E[\varphi_{ij^*} \,|\, \hat{H}_{ij^*}|^2] = \Theta((1-\sigma^2_{\Delta H})\log(K))$ (Lemma 2). In one extreme ($\sigma^2_{\Delta H} = 0$, perfect CSIT), the multi-user diversity gain is given by $\Theta\left(\log(K)\right)$ as presented in [3]. In the other extreme ($\sigma^2_{\Delta H} \to 1$, no CSIT), the multiuser diversity gain still approaches $\Theta((1-\sigma^2_{\Delta H})\log(K))$ with large $(1-\sigma^2_{\Delta H})\log(K)$ (i.e. fast growth in *K*), but approaches $\Theta((\sigma^2_{\Delta H}/2)F^{-1}_{\chi_2^2(\eta/(\sigma^2_{\Delta H}/2))}(\varepsilon))$ (no multiuser SNR gain) with $(1-\sigma^2_{\Delta H})\log(K) \to \eta$ (i.e. limited growth in *K*). In general, for intermediate CSIT errors, the cross-layer SNR gain decreases linearly as $\sigma^2_{\Delta H}$ increases and exponentially more users *K* is needed to compensate for the penalty of poor CSIT quality $\sigma^2_{\Delta H}$.

Comparisons of cross-layer gains of the proposed scheduler over various schedulers are further summarized as follows (with derivation details to be presented in Appendix A):

3) ***Cross-Layer Gain Comparison:*** Table 1 and Figure 2 summarize the asymptotic cross-layer gains of various schedulers. The proposed scheduler is shown to have a huge $P_{\min}$ saving to support all users' delay constraints due to *1) multiuser diversity gain* (it corresponds to $P_{\min}$ saving in horizontal direction in Fig. 2, the saving is in a rate of $\log(K)$) and *2) proper handling of users' requirements* (corresponds to $P_{\min}$ saving in vertical direction in Fig. 2, i.e. $P_{\min}$ can be achieved to satisfy a lower total equivalent rate requirement - $\tilde{\rho}_1 K_1 + \tilde{\rho}_2 K_2$, in contrast to that of other schedulers - $\max\left\{\tilde{\rho}_1, \tilde{\rho}_2\right\}\left(K_1 + K_2\right)$); it also provides fastest goodput gain ΔG for every unit increases of transmit power ΔP due to *3) proper outage handling* (characterized by the slope in Fig. 2, that depends on the ratio between target probability $(1-\varepsilon)$ and $1/2$ non-outage probability in conventional schedulers).

VI. Simulation Results

In this section, we present the simulation results to illustrate the performance of the proposed cross-layer scheduler for OFDMA system with heterogeneous delay requirements in the presence of CSIT errors.

A. Simulation Model

In the simulation, an OFDMA system, with total system bandwidth of 1.024 MHz consisting of 64 subcarriers with 5 independent paths and 5 users, is considered (with two classes of users specified by arrival rates and delay requirements $(\boldsymbol{\lambda}, \mathbf{T}) = \{(\lambda_1, T_1), (\lambda_2, T_2)\}$ (in packets per time slot, time slots), and some unclassed users having no delay constraint (with requirements of 1000 time slots)). Scheduling slot duration would be 2ms. We assume all mobile users suffer the same path loss from BS. The target outage probability of each subcarrier is set to $P_{out,ij} = 0.01$. Each packet consists of 1.024 kbits and each point in the figures is simulated from 5000 independent realizations.

B. Simulation Results

In this section, we compare the performance of the proposed CSIT error considerate scheduler, with that of the conventional naïve opportunistic scheduler [1, 2, 3] and naïve MAX-WEIGHT[12] scheduler proposed in [12, 13] (which treats the outdated CSIT estimate as the perfect CSIT) and the conventional baseline reference – fixed power and subcarrier assignment. It is remarked that all schedulers under comparison have the same linear complexity in terms of number of subcarriers and users.

1) Goodput Comparison

Figure 3 shows the proposed optimal CSIT error inconsiderate scheduler provides substantial goodput enhancement over opportunistic scheduler [1, 2, 3] and MAX-WEIGHT scheduler [12, 13] and fixed assignment in the presence of CSIT error ($\sigma_{\Delta H}^2 = 0.05$). It also shows the impact on goodput performance of the proposed scheduler upon different CSIT errors. When $\sigma_{\Delta H}^2$ increases (from 0 to 0.05), there would be a small decrease in goodput, and the minimum required power supporting all

[12] In [12, 13], the author considered the QSI-CSI considerate MAX-WEIGHT scheduler, and named it as LQHPR scheduler. In current context, this scheduler would consider each subcarrier i to be assigned to the best user $j^* = \arg\max_{j\in[1,K]} q_j \,|\hat{H}_{ij}|^2$.

delay constraints would increases. (Noted that existing schedulers cannot even provide desired delay performance to classed users within average transmit power region shown in Figure 3 (see Table 1).)

2) Delay Performance Comparison w.r.t. background traffic loading and CSIT errors

Figure 4 illustrates average delay versus background traffic loading. It shows conventional schedulers cannot provide any delay guarantee even under small CSIT errors $\sigma_{\Delta H}^2 = 0.05$ ($P_{TOT} = 11dB$). Besides, as arrival rate of background user increases, the delay performance of all users under opportunistic scheduling and MAX-WEIGHT scheduling degrades significantly. On the other hand, the proposed CSIT error considerate scheduler can satisfy the delay requirements of users 1 and user 2 regardless of background users' loading at the expense of delay performance of delay-insensitive background users.

Figure 5 illustrates the average delay performance versus CSIT errors ($P_{TOT} = 15dB$). It shows, using CSIT error inconsiderate schedulers, the delay performance of all users degrade significantly even in very low $\sigma_{\Delta H}^2$ due to significant packet outage (refer to Table 1). In contrast, the proposed CSIT error considerate scheduler can achieve delay requirements of class 1 and 2 users even under high $\sigma_{\Delta H}^2$.

In short, performance gain in goodput and delay performance satisfaction shown in these simulation results matches well with the analyses (explanations) shown in Section V A.3 (with Fig. 2 illustration).

VII. Conclusion

In this paper, we have proposed an analytical framework on OFDMA cross-layer scheduler design for delay-sensitive users with heterogeneous delay requirements and outdated CSIT. We proposed a modified M/G/1 queue model (taking care of packet errors and retransmission) to transform the delay constraint into an equivalent rate constraint. The cross-layer design problem is formulated as a convex optimization problem which takes account of the outdated CSIT, source statistics and queue dynamics of the OFDMA systems. The optimal delay-sensitive power, rate and subcarrier allocations are obtained and the proposed cross-layer scheduler offers a nice balance between maximizing goodput and achieving delay requirements of heterogeneous delay-sensitive users. Asymptotic performance analysis showed large multiuser diversity (SNR) gain, goodput gain and minimum power requirement saving can be achieved by the proposed scheduler over conventional naïve baseline schemes through proper consideration of multiuser diversity, heterogeneous delay constraints and CSIT outdateness. Simulation results further showed that the proposed scheduler can achieve substantial goodput gain while satisfying the delay requirements of all delay-sensitive users under high CSIT error and traffic loading.

APPENDIX

A. Asymptotic Performance Analysis of the proposed scheduler over conventional schedulers

In this appendix, we investigate asymptotic performance of various schedulers under delay-sensitive cross-layer framework with imperfect Channel State Information at transmitter (CSIT).

1) *Asymptotic $P_{\min}$ requirements*

a) *Proposed scheduler (Proof of Lemma 2 and Lemma 3)*

Define the best user within each class d perceived by subcarrier i to be $j(d,i)=\arg\max_{j\in \mathrm{Class_d}}\left(c_j\varphi_{ij}\,|\hat{H}_{ij}|^2\right)^{c_j}$,

where $c_j=\left(1+\gamma_j\right)\left(1-\varepsilon\right)\big/\mu$ is the water-level of user j. Noted that water-levels of all users from the same class are the same, i.e. $c_j=c(1),\forall j\in Class1, c_{j'}=c(2),\forall j'\in Class2$ and let $c(1)>c(2)$.

The pdf of $\varphi_{ij(d,i)}\,|\hat{H}_{ij(d,i)}|^2$ for each class d is given by:

$$p(\varphi_{ij(d,i)}\,|\hat{H}_{ij(d,i)}|^2=\gamma)=[K_d(1-e^{-\frac{x}{1-\sigma_{\Delta H}^2}})^{K_d-1}e^{-\frac{x}{1-\sigma_{\Delta H}^2}}\left(1\bigg/\frac{d(\sigma_{\Delta H}^2/2)F^{-1}_{\chi_2^2(x/(\sigma_{\Delta H}^2/2))}(\varepsilon)}{dx}\right)]_{x:\gamma=(\sigma_{\Delta H}^2/2)F^{-1}_{\chi_2^2(x/(\sigma_{\Delta H}^2/2))}(\varepsilon)} \quad .(A.1)$$

The pdf of $s_{ij(1,i)}\varphi_{ij(1,i)}\,|\hat{H}_{ij(1,i)}|^2$ can be obtained as: (pdf of $s_{ij(2,i)}\varphi_{ij(2,i)}\,|\hat{H}_{ij(2,i)}|^2$ is similarly derived)

$$\begin{aligned}
&p(s_{ij(1,i)}\varphi_{ij(1,i)}\,|\hat{H}_{ij(1,i)}|^2=\gamma)\\
&=p(\varphi_{ij(1,i)}\,|\hat{H}_{ij(1,i)}|^2=\gamma)\Pr((c(1)\varphi_{ij(1,i)}\,|\hat{H}_{ij(1,i)}|^2)^{c(1)}>(c(2)\varphi_{ij(2,i)}\,|\hat{H}_{ij(2)}|^2)^{c(2)}\,\Big|\,\varphi_{ij(1,i)}\,|\hat{H}_{ij(1,i)}|^2=\gamma)+\Pr(s_{ij(1,i)}=0)\delta(\gamma)\\
&=K_1e^{-\frac{x}{1-\sigma_{\Delta H}^2}}(1-e^{-\frac{x}{1-\sigma_{\Delta H}^2}})^{K_1-1}(1-e^{-\frac{x_c}{1-\sigma_{\Delta H}^2}})^{K_2}\left(1\bigg/\frac{d(\sigma_{\Delta H}^2/2)F^{-1}_{\chi_2^2(x/(\sigma_{\Delta H}^2/2))}(\varepsilon)}{dx}\right)+\Pr(s_{ij(1,i)}=0)\delta(\gamma)
\end{aligned} \quad (A.2)$$

where $\delta(\gamma)$ is "delta function", $(c(1)(\sigma_{\Delta H}^2/2)F^{-1}_{\chi_2^2(x/(\sigma_{\Delta H}^2/2))}(\varepsilon))^{c(1)/c(2)}\big/c(2)=(\sigma_{\Delta H}^2/2)F^{-1}_{\chi_2^2(x_c/(\sigma_{\Delta H}^2/2))}(\varepsilon)$.

Moreover, as K_1 and K_2 is large, it can be shown that $c(1)/c(2)\to 1$, thus the conditional diversity gain for each class d (d = 1, 2) from multiuser diversity (i.e. the average SNR of class d selected user) is given as: (Notice that *d'* is used to denote the other class, i.e. if d = 1, then *d'* = 2, and vice versa)

$$\begin{aligned}
&E[s_{ij(d,i)}\varphi_{ij(d,i)}\,|\hat{H}_{ij(d,i)}|^2|\,s_{ij(d,i)}=1]=\frac{E[s_{ij(d,i)}\varphi_{ij(d,i)}\,|\hat{H}_{ij(d,i)}|^2]}{\Pr(s_{ij(d,i)}=1)}=\frac{\int_0^\infty K_d(\sigma_{\Delta H}^2/2)F^{-1}_{\chi_2^2(x/(\sigma_{\Delta H}^2/2))}(\varepsilon)e^{-\frac{x}{1-\sigma_{\Delta H}^2}}(1-e^{-\frac{x}{1-\sigma_{\Delta H}^2}})^{K_d-1}(1-e^{-\frac{x_c}{1-\sigma_{\Delta H}^2}})^{K_{d'}}dx}{\int_0^\infty K_de^{-\frac{x}{1-\sigma_{\Delta H}^2}}(1-e^{-\frac{x}{1-\sigma_{\Delta H}^2}})^{K_d-1}(1-e^{-\frac{x_c}{1-\sigma_{\Delta H}^2}})^{K_{d'}}dx}\\
&=\Theta\left((\sigma_{\Delta H}^2/2)F^{-1}_{\chi_2^2\left(\left(1-\sigma_{\Delta H}^2\right)\ln\left(K_1+K_2\right)/(\sigma_{\Delta H}^2/2)\right)}(\varepsilon)\right)=\begin{cases}\Theta\left(\left(\sigma_{\Delta H}^2/2\right)F^{-1}_{\chi_2^2\left(\eta/(\sigma_{\Delta H}^2/2)\right)}\left(\varepsilon\right)\right),\text{where }\left(1-\sigma_{\Delta H}^2\right)\log\left(K_1+K_2\right)\to\eta<\infty\text{ as }K_d\to\infty\\ \Theta\left(f\left(\varepsilon\right)\left(1-\sigma_{\Delta H}^2\right)\log\left(K_1+K_2\right)\right),\text{where }\left(1-\sigma_{\Delta H}^2\right)\log\left(K_1+K_2\right)\to\infty\text{ as }K_d\to\infty\end{cases}
\end{aligned} \quad .(A.3)$$

In most setting of ε, $f\left(\varepsilon\right)\to 1\ as\ K_1\to\infty, K_2\to\infty$ in (A.3). The above integral with c(1) = c(2) = 1 could be found in [3] (with only single class consideration).

The $P_{\min,opt}$ required satisfying delay constraints of all Class d (d = 1, 2) is calculated as follows:

$$\begin{cases} K_d \tilde{\rho}_d = (1-\varepsilon) E_H[\sum_{i=1}^{N_F} s_{ij(d,i)} \log_2(c(d)\varphi_{ij(d,i)} \mid \hat{H}_{ij(d,i)} \mid^2)], \forall d, \text{ where } \tilde{\rho}_d \text{ is the equivalent rate constraint of class } d \text{ users} \\ P_{\min,opt} = E[\frac{1}{N_F}\sum_{i=1}^{N_F}\sum_{j=1}^{K} s_{ij} p_{ij}] = \Pr[s_{ij(1,i)}=1]c(1) + \Pr[s_{ij(2,i)}=1]c(2), \text{ since selection process is independently identical } \forall i \end{cases} \quad (A.4)$$

For performance comparison with other existing schedulers, we will focus on the more interesting case when $\left(1-\sigma_{\Delta H}^2\right)\log\left(K_1+K_2\right) \to \infty$ in the subsequent description. By (A.4), we have

$$\sum_{d=1}^{2} K_d \tilde{\rho}_d = (1-\varepsilon) E\left[\sum_{i=1}^{N_F}\sum_{d=1}^{2} s_{ij(d,i)} \log_2(c(d)\varphi_{ij(d,i)} \mid \hat{H}_{ij(d,i)} \mid^2)\right] \cong (1-\varepsilon) N_F \log_2\left(P_{\min,opt}\Theta\left(\left(1-\sigma_{\Delta H}^2\right)\log\left(K_1+K_2\right)\right)\right) \quad (A.5)$$

As a result, $P_{\min,opt} = \Theta\left((2^{(\tilde{\rho}_1 K_1+\tilde{\rho}_2 K_2)/((1-\varepsilon)N_F)}) \big/ ((1-\sigma_{\Delta H}^2)\log(K_1+K_2))\right)$.

The following sub-sections *b)* and *c)* will give asymptotic minimum power requirements for various existing schedulers. Since they either improperly handled outage (when $\sigma_{\Delta H}^2 > 0$) or neglected delay differentiation/guarantee of different users' classes, the $P_{\min,(\text{existing})}$ used to satisfy delay constraints of all users occurs when each user's goodput is the same as the most stringent equivalent rate requirement.

b) Fixed assignment strategy

In fixed assignment strategy, $\left|\hat{H}_{ij}\right|^2$ is used as the actual $\left|H_{ij}\right|^2$. The $P_{\min,\text{fixed}}$ is calculated as

$$\max_d \tilde{\rho}_d = \left(\frac{N_F}{K_1+K_2}\right) E_{\hat{H}}[\log_2(1+P_{\min,fixed} \underbrace{|\hat{H}_{ij}|^2}_{E_{\hat{H}}\left[|\hat{H}_{ij}|^2\right]=\left(1-\sigma_{\Delta H}^2\right)}) \underbrace{E_{H|\hat{H}}\left[I[|\hat{H}_{ij}|^2 \le |H_{ij}|^2]\right]}_{E_{\hat{H}}\left[E_{H|\hat{H}}\left[I[|\hat{H}_{ij}|^2\le|H_{ij}|^2]\right]\right]=1/2}]. \quad (A.6)$$

Thus, $P_{\min,\text{fixed}} = \Theta\left(\left(2^{(\max_d \tilde{\rho}_d)2(K_1+K_2)/N_F} - 1\right) \Big/ \left(1-\sigma_{\Delta H}^2\right)\right)$ when CSIT is outdated, i.e. $\sigma_{\Delta H}^2 > 0$.

c) Pure opportunistic schedulers presented in [1, 2, 3]

For pure opportunistic scheduler, its $P_{\min,OS}$ can be calculated as follows:

$$\max_d \tilde{\rho}_d \cong E_{\hat{H}}\left[\sum_{i=1}^{N_F} s_{ij}\log_2(P_{\min,\text{OS}} \mid \hat{H}_{ij} \mid^2) E_{H|\hat{H}}\left[I[|\hat{H}_{ij}|^2 \le |H_{ij}|^2]\right]\right] \cong (\frac{N_F}{K_1+K_2})\frac{1}{2}\log_2(P_{\min,\text{OS}} \underbrace{E_{\hat{H}}\left[|\hat{H}_{ij}|^2 \middle| s_{ij}=1\right]}_{\Theta\left(\left(1-\sigma_{\Delta H}^2\right)\log(K_1+K_2)\right)}) \quad (A.7)$$

Thus, $P_{\min,OS} = \Theta\left(\left(2^{(\max_d \tilde{\rho}_d)2(K_1+K_2)/N_F}\right)\Big/\left(1-\sigma_{\Delta H}^2\right)\log(K_1+K_2)\right)$. The same $P_{\min}$ result could be observed when our proposed scheduler is operated under perfect CSIT assumption (when CSIT errors occur).

2) Asymptotic Goodput Performance

a) Proposed scheduler

From (A.5), the total goodput at $P_{\min,opt}$ is $G_{P_{\min,opt}} = (1-\varepsilon) N_F \log_2\left(P_{\min,opt}\Theta\left(\left(1-\sigma_{\Delta H}^2\right)\log\left(K_1+K_2\right)\right)\right)$.

The total goodput at $P_{\min,opt}+\Delta P$ can be calculated (similar to the result in (A.5)) as

$G_{P_{\min,opt}+\Delta P} = (1-\varepsilon) N_F \log_2\left(\left(P_{\min,opt}+\Delta P\right)\Theta\left(\left(1-\sigma_{\Delta H}^2\right)\log\left(K_1+K_2\right)\right)\right)$.

Thus the change in total goodput per ΔP unit change in total available power is given by:

$$\Delta G = G_{P_{\min,opt}+\Delta P} - G_{P_{\min,opt}} = (1-\varepsilon) N_F \log_2\left(1+\Delta P / P_{\min,opt}\right). \qquad \text{(A.8)}$$

b) Other existing schedulers

These schedulers' total goodput changes are $\Delta G = G_{P_{\min,(\text{existing})}+\Delta P} - G_{P_{\min,(\text{existing})}} = (1/2) N_F \log_2\left(1+\Delta P / P_{\min,(\text{existing})}\right)$.

In both cases – *2) a)* and *b)*, the total goodput is $\sum_{d=1}^{2} K_d \tilde{\rho}_d + \Delta G$, where $\Delta P = P_{ToT} - P_{\min}$.

REFERENCES

[1] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747-1758, Oct. 1999.

[2] J. H. Jang and K. B. Lee, "Transmit Power Adaptation for Multiuser OFDM Systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171-178, Feb. 2003.

[3] G. Song and Y. (G.) Li, "Cross-layer Optimization for OFDM wireless network–Part I: Theoretical framework," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 614-624, Mar. 2005.

[4] L. M. C. Hoo, B. Halder and J. M. Cioffi, "Multiuser Transmit Optimization for Multicarrier Broadcast Channels: Asymptotic FDMA Capacity Region and Algorithms," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 922-930, June 2004.

[5] W. Yu and R. Lui, "Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310-1322, July 2006.

[6] H. Zhang, "Service disciplines for guaranteed performance services in packet-switching networks," *IEEE Proc.*, vol. 83, no. 10, pp. 1374-1396, Oct. 1995.

[7] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the single-node case," *IEEE/ACM Trans. Network.*, vol.1, no.3, pp. 344-357, June 1993.

[8] S. Kittipiyakul and T. Javidi, "Resource Allocation in OFDMA: How Load-Balancing Maximizes Throughput When Water-Filling Fails," *UW Technical Report*, UWEETR-2004-0007, 2004.

[9] C. Anton-Haro, P. Svedman, M. Benqtsson, A. Alexiou and A. Gameiro, "Cross-layer scheduling for multi-user MIMO systems," *IEEE Commun. Mag.*, vol. 44, no. 9, pp. 39-45, Sept. 2006.

[10] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 74-80, Oct. 2003.

[11] X. Lin, N. B. Shroff and R. Srikant, "A Tutorial on Cross-Layer Optimization in Wireless Networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1452-1463, Aug. 2006.

[12] E. M. Yeh and A. S. Cohen, "Information Theory, Queueing, and Resource Allocation in Multi-user Fading Communications," *Proc. Information Sciences Syst. Conf.*, pp. 1396-1401, Mar. 2004.

[13] E. M. Yeh, *Multiaccess and Fading in Communication Networks*, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT, 2001.

[14] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, 1991.

[15] D. Bertsekas and R.G. Gallager, *Data Networks*, 2nd Ed., Prentice-Hall, 1992.

[16] Soren, *Asmussen, Applied Probability and Queues*, Springer, 2003.

[17] R. Berry and R. Gallager, "Communication Over Fading Channels with Delay Constraints," *IEEE Trans. on Info. Theory*, vol. 48, no. 5, pp. 1135-1149, May 2002.

[18] D. J. Love and R. W. Heath, Jr., "OFDM Power Loading Using Limited Feedback," *IEEE Trans. Wireless Commun.*, vol. 54, no. 5, pp. 1773-1780, Sept. 2005.

[19] Y. W. Yao and G. B. Giannakis, "Rate-Maximizing Power Allocation in OFDM Based on Partial Channel Knowledge," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1073-1083, May. 2005.

[20] A. Leke and J. M. Cioffi, "Multicarrier systems with imperfect channel knowledge," in *Proc. IEEE International Symposium on PIMRC*, vol. 2, pp. 549–553, Sep. 1998.

[21] A. K. Karmokar, 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. on Communication*, vol. 9, pp. 4315 - 4320, June 2006.

[22] D. S. W. Hui, Vincent K. N. Lau and W. H. Lam, "Cross-Layer Design for OFDMA Wireless Systems With Heterogeneous Delay Requirements," *IEEE Trans. Wireless Commun.*, Aug. 2007, to be published.

[23] S. Boyd and L. Vandenberghe, *Convex optimization*, 2004.

	Proposed scheduler	Fixed assignment strategy	Pure opportunistic scheduler
$P_{\min}$	$\Theta\left(\frac{2^{\kappa(\tilde{\rho}_1 K_1+\tilde{\rho}_2 K_2)/N_F}}{\left(1-\sigma_{\Delta H}^2\right)\log(K_1+K_2)}\right)$ $\kappa = 1-\varepsilon, \varepsilon \in (0, 1/2]$	$\Theta\left(\frac{2^{\kappa(\max_d \tilde{\rho}_d)(K_1+K_2)/N_F}-1}{\left(1-\sigma_{\Delta H}^2\right)}\right)$, $\kappa = 2$	$\Theta\left(\frac{2^{\kappa(\max_d \tilde{\rho}_d)(K_1+K_2)/N_F}-1}{\left(1-\sigma_{\Delta H}^2\right)\log(K_1+K_2)}\right)$, $\kappa = 2$
Goodput Gain (ΔG)	$\Delta G = (1-\varepsilon) N_F \log_2\left(1+\frac{\Delta P}{P_{\min,opt}}\right)$	$\Delta G = \frac{1}{2} N_F \log_2\left(1+\frac{\Delta P}{P_{\min,fixed}}\right)$	$\Delta G = \frac{1}{2} N_F \log_2\left(1+\frac{\Delta P}{P_{\min,OS}}\right)$

Table 1: Comparison of Cross-Layer Gains of the proposed scheduler versus fixed assignment strategy and pure opportunistic scheduler.

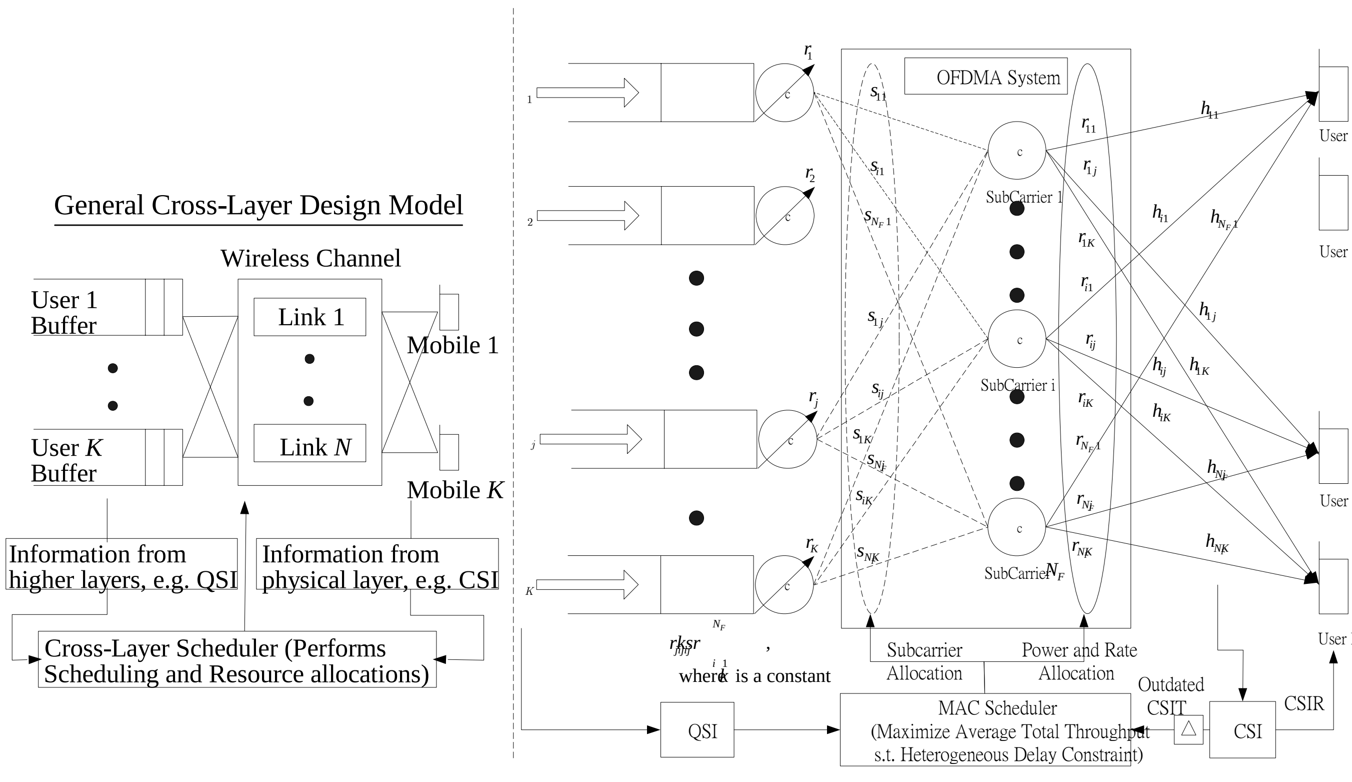

Figure 1. General Cross-Layer System Model (Left) and Cross-Layer Scheduling model under Conceptual Channel Model for OFDMA system with heterogeneous application users in the presence of imperfect CSIT (Right)

λ

λ

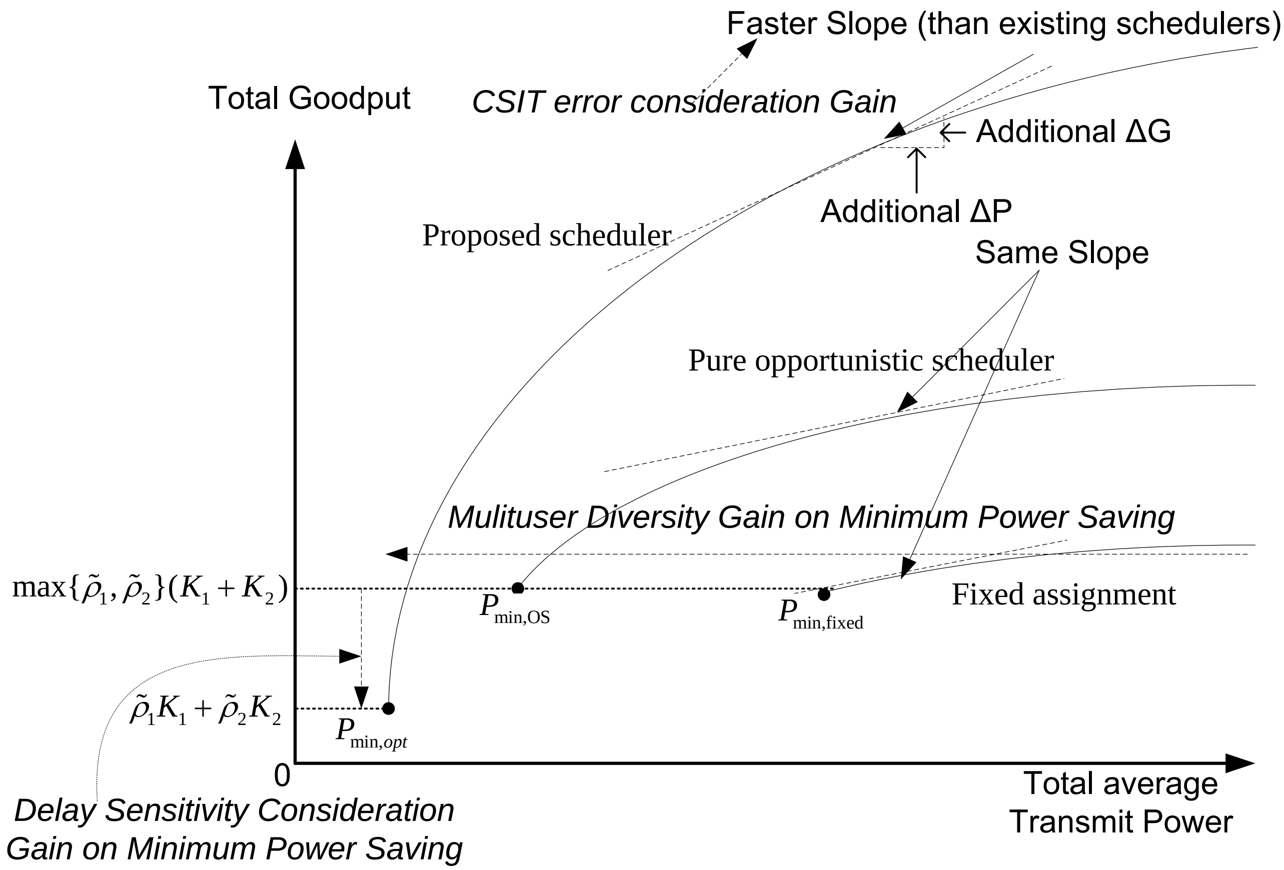

Figure 2. A graphical illustration of asymptotic performance comparison of various schedulers (in terms of order of growth of goodput performance per unit increase of transmit power)

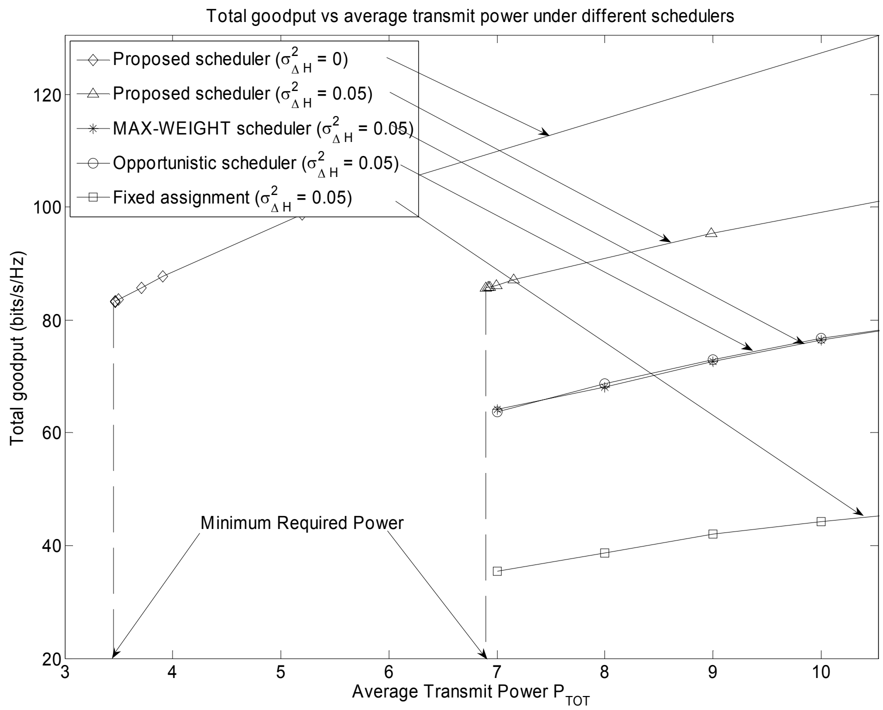

Figure 3. Total goodput vs. average transmit power with various schedulers – user 1, user 2 have delay constraint of $T_1 = T_2 = 2.5$. $\lambda_1 = \lambda_2 = \lambda_{unclassed} = 0.5$.

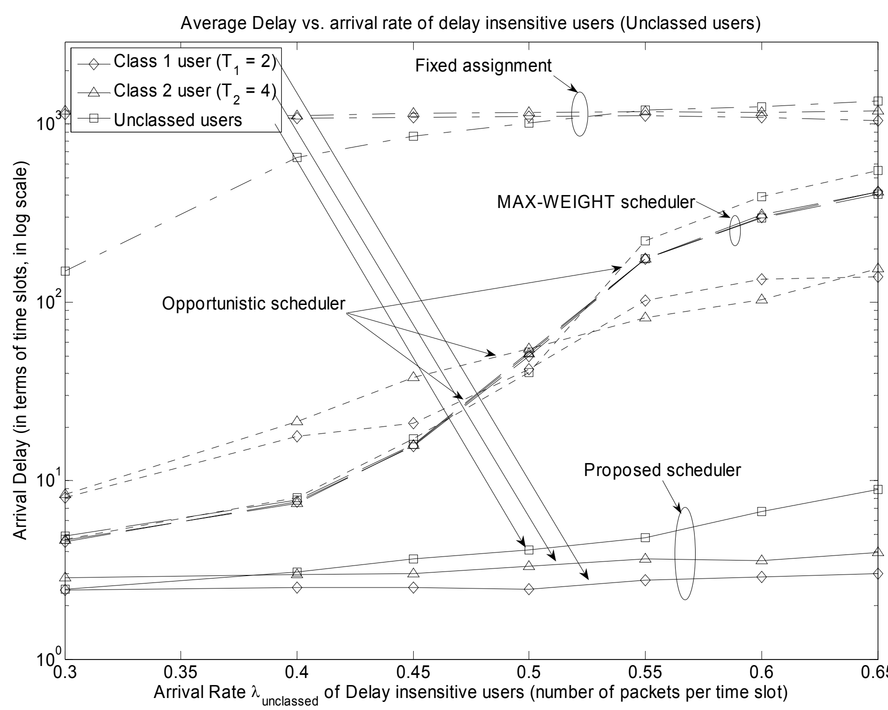

Figure 4. Average delay vs background traffic loading with various schedulers - user 1, user 2 have delay constraints of T_1= 2, T_2 = 4, respectively. $\lambda_1 = \lambda_2 = 0.5, \sigma^2_{\Delta H} = 0.05$.

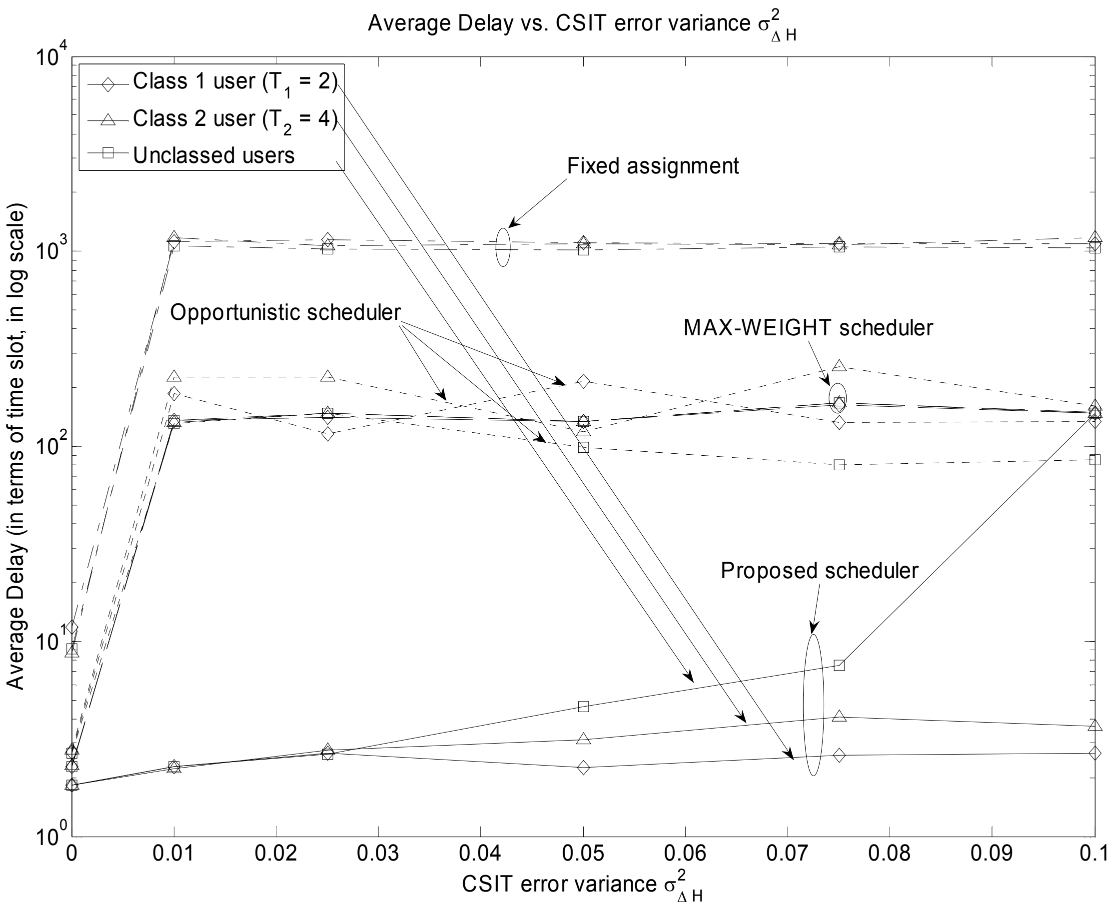

Figure 5. Average delay vs CSIT error variance with various schedulers - user 1, user 2 have delay constraint of T_1= 2, and T_2 = 4, respectively. $\lambda_1 = \lambda_2 = \lambda_{unclassed} = 0.6$ **.**