Adaptive Resource Allocation for Multiuser MIMO Systems with Transmit Group MMSE

Vincent K. N. Lau, *Senior Member, IEEE*

***Abstract*—In this paper, we propose an integrated cross-layer optimization framework for resource allocation over a multi-user MIMO system using *Transmit Group MMSE* (Tx-GMMSE) at the base station. We adopt the full-rate full-diversity orthogonal STBC in the multi-user Tx-GMMSE physical layer allowing a flexible tradeoff of *spatial multiplexing* (for spectral efficiency) and *full rate spatial diversity* (for protection against packet outage) with various number of antenna groupings. The proposed cross-layer framework, which determines the user selection, rate selection, power selection as well as mode selection (spatial diversity or spatial multiplexing) for each user, can introduce robustness to the potential packet transmission errors even at moderate to large CSIT errors.**

***Index Terms*—Transmit group MMSE, cross-layer MIMO.**

I. INTRODUCTION

CROSS-layer designs for multi-user MIMO systems have been shown to be very effective in boosting spectral efficiency since they exploit the *multi-user selection diversity* and spatial multiplexing gain in the time varying wireless fading channels. However, most of the cross-layer designs rely on perfect channel state information (CSIT) at the base station, a very difficult to achieve in practice especially when the number of antennas N_T and the number of users K are large. When CSIT is subject to time delays at the base station, the instantaneous mutual information is unknown to the base station, resulting in channel outage events and thus potential packet errors. For instance, conventional cross-layer designs mostly utilize Transmit-Zero Forcing (Tx-ZF) for spatial multiplexing to maximize system capacity [6]. This is reasonable if the CSIT is perfect because packet outage can be avoided by rate adaptation. However, as we shall illustrate, the potential packet error is a significant factor of performance degradation in cross-layer systems with outdated CSIT and packet outage must be considered in the scheduling algorithm. As a result, naive cross-layer designs based on Tx-ZF for MIMO channels need to be enhanced to account for the outdated CSIT. Specifically, both the physical layer processing (Tx-ZF) and the resource allocation algorithm have to be enhanced so as to cater for outdated CSIT scenarios.

In this paper, we shall propose an integrated cross-layer design framework over an enhanced multi-user MIMO physical layer processing, namely the *Transmit Group MMSE* (Tx-GMMSE) for robust performance with outdated CSIT. Conventional system performance measure such as the ergodic capacity may not be suitable here because they failed to account for potential packet errors. To take into consideration of the potential packet errors, we shall design our system to maximize the *average system goodput*, which is defined as the total average b/s/Hz *successfully delivered* to the mobiles. Group zero-forcing was originally proposed in [7] as a receiver detection algorithm for point-to-point MIMO link for flexible tradeoff between spatial diversity and spatial multiplexing [5]. In this paper, we shall extend the idea to consider group MMSE at the transmitter. In the Tx-GMMSE architecture, the spatial streams are partitioned into a number of groups. Within each group, orthogonal space time block code is applied to achieve spatial diversity gain and between groups, spatial multiplexing is adopted to increase the spectral efficiency. In fact, spatial diversity does not bring us any advantage in a MIMO link for ergodic channels (fast fading channels) or slow fading channels with perfect CSIT because there will be virtually no packet errors in both cases (as long as the channel code is strong enough). However, when we have outdated CSIT in multi-user MIMO systems, *spatial diversity* may be important because it offers additional protection against potential *packet errors* especially at large CSIT errors.

While the Tx-GMMSE physical layer offers a flexible tradeoff between spatial diversity and spatial multiplexing, it is critical to have a properly designed resource allocation algorithm to determine the best *operating mode* (spatial diversity or spatial multiplexing) for each user in every scheduling slot. Specifically, based on the outdated CSIT on every fading slot, the proposed cross-layer framework determines the *scheduled data rates* $\{r_1, ..., r_K\}$, the *power allocations* $\{p_1, .., p_K\}$, the *set of admitted users* $\mathcal{A}$ as well as the *spatial multiplexing/diversity mode* of the selected users so as to achieve robust performance against packet errors. The optimal search for the user set and the spatial multiplexing/diversity mode involve complex combinatorial search which has exponential order of complexity with respect to N_T. As a result, we propose a *double greedy-based search algorithm* which achieves close-to-optimal performance at a much lower complexity.

This paper is organized as follows. In section II, we shall outline the multi-user MIMO system models, including the channel model, the CSIT error model, multi-user Tx-GMMSE processing, as well as the cross-layer scheduling model. In section III, we shall formulate the cross-layer scheduling problem with outdated CSIT into a mixed convex and combinatorial optimization problem. We shall outline both the optimal and sub-optimal solutions for the cross-layer scheduling problem. In section IV, numerical results are presented and discussed. Finally, we give a brief summary in section V.

Manuscript received February 22, 2007; revised July 11, 2007; accepted November 3, 2007. The associate editor coordinating the review of this paper and approving it for publication was A. Gulliver.

V. Lau is with the Department of Electrical and Computer Engineering, Hong Kong University of Science and Technology, HK (e-mail: eeknlau@ee.ust.hk).

This work is supported by RGC under contract HKUST7104/06E.
Digital Object Identifier 10.1109/TWC.2009.070219

II. Multiuser MIMO System Model

We shall adopt the following notations throughout this paper: $\mathbf{X}$ denotes a matrix, $\mathbf{x}$ denotes a column vector, $\mathbf{x}^\dagger$ denotes a row vector (where $(.)^\dagger$ denotes transpose) and x denotes a scalar. Furthermore, $(.)^H$ denotes the conjugate transpose operation, $(.)^*$ denotes the conjugate operation and matrix norm refers to Fobreneous norm.

A. MIMO Channel Model

We consider a communication system with K mobile users and a base station over a frequency flat fading channel. We assume the base station is equipped with N_T transmit antennas and each of the K mobile users is equipped with $M_k, k = 1, 2, .., K$ receive antennas respectively. The channel fading coefficients between the base station and the k-th mobile are modelled as i.i.d. complex Gaussian random processes (with zero mean and unit variance) and they are characterized by the $M_k \times N_T$ dimension channel matrix, $\mathbf{H}_k^H$. The system is targeted for low mobility users and therefore, the channel fading remains quasi-static within a scheduling time slot.

Let $\mathbf{y}_{k,t}$ be the $M_k \times 1$ received signal of the k-th mobile at the t-th symbol. The $\sum_k M_k \times 1$ dimension vector of the aggregate received signals $\mathbf{y}_t$ from all the K mobiles is given by:

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_{1,t} \\ \vdots \\ \mathbf{y}_{K,t} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1^H \\ \vdots \\ \mathbf{H}_K^H \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{z}_{1,t} \\ \vdots \\ \mathbf{z}_{K,t} \end{bmatrix} \tag{1}$$

where $\mathbf{x}$ is the $N_T \times 1$ transmit symbol from the base station to the K mobiles and $\mathbf{z}_{k,t}$ is the $M_k \times 1$ complex Gaussian channel noise vector with unit covariance $\mathbf{I}$.

B. CSIT Error Model

We consider a TDD system where the CSIT is estimated using the uplink dedicated pilots of the K mobiles. Due to duplexing delay[1], the estimated CSIT of the k-th user at the $t-$th scheduling slot is given by $\hat{\mathbf{H}}_k[t] = \mathbf{H}_k[t-\tau]$ where τ is the CSIT delay. Hence, the estimated CSIT $\mathbf{H}_k[t]$ can be expressed as [12]:

$$\begin{aligned} \mathbf{H}_k[t] &= \sqrt{1-\sigma_e^2(\tau)}\mathbf{H}_k[t-\tau] + \sigma_e[\tau]\Delta\mathbf{H}_k \\ &= \sqrt{1-\sigma_e^2(\tau)}\widehat{\mathbf{H}}_k[t] + \sigma_e[\tau]\Delta\mathbf{H}_k \end{aligned} \tag{2}$$

where $\mathbf{H}_k[t]$ is the actual CSIT at time t and $\Delta\mathbf{H}_k$ is the CSIT error matrix with components given by i.i.d. Gaussian distribution variables (zero mean, variance $\sigma_e^2(\tau)$). $\sigma_e^2(\tau)$ indicates the CSIT quality. When $\sigma_e^2 = 0$, we have perfect CSIT and when $\sigma_e^2 = 1$, we have no CSIT. For illustration purpose, we assume Jake's model for Doppler spread and the CSIT error variance σ_e^2 is given by:

$$\sigma_e^2(\tau) = 1 - J_0^2\left(\frac{2\pi f_c}{c}v\tau\right)$$

where $J_0(.)$ is the zeroth order Bessel function, c is the speed of light, v is the speed of the mobile and f_c is the carrier frequency.

[1]Duplexing delay refers to the delay due to the switching between UL and DL frames in TDD systems.

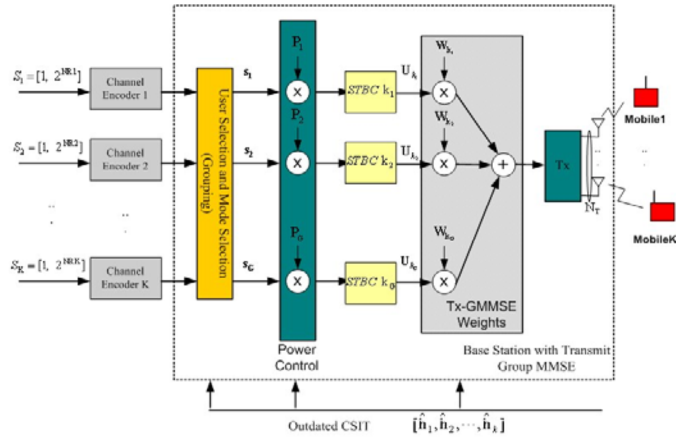

Fig. 1. Illustration of the transmit group MMSE (Tx-GMMSE) processing at the multi-user MIMO base station.

C. Multiuser Multi-antenna Tx-GMMSE Processing

Fig. 1 illustrates the block diagram of the proposed Tx-GMMSE processing. For easy notation, we assume the set of selected users $\mathcal{A} = \{1, 2, .., G\}$. The N_T *spatial channels* are partitioned into G groups with group $g \in \mathcal{A}$ consuming $n_g \le N_T$ spatial channels. Within the g-th group, the $n_g \le N_T$ spatial channels are used to carry codewords from *space time block code* (STBC) such as the Alamouti codes [8]. Between the G groups, G independent codewords from G different users are carried. Note that when $n_1 = n_2 = \cdots = n_G = 1$ and the Tx-GMMSE reduces to the conventional Tx-MMSE processing. By adjusting the spatial channel groupings $\{n_1, ..., n_G\}$, we can achieve a flexible tradeoff between spatial multiplexing across different users (large G) and spatial diversity for each user (small G). Let T be the time span of the STBC. The $n_g \times T$ dimensional STBC codewords of the g-th group, $\mathbf{U}_g$, is given by:

$$\mathbf{U}_g = \begin{bmatrix} \mathbf{A}_\mathbf{1}^{(\mathbf{g})}\mathbf{s}_g & \cdots & \mathbf{A}_\mathbf{T}^{(\mathbf{g})}\mathbf{s}_g \end{bmatrix} + \begin{bmatrix} \mathbf{B}_\mathbf{1}^{(\mathbf{g})}\mathbf{s}_g^* & \cdots & \mathbf{B}_\mathbf{T}^{(\mathbf{g})}\mathbf{s}_g^* \end{bmatrix} \tag{3}$$

where $(.)^*$ denotes the conjugate operation, $\mathbf{A}_t^{(g)}$ and $\mathbf{B}_t^{(g)}$ are the $g-$th STBC encoding matrices (dimension $n_g \times n_g$) for $t \in [1, T]$ and $\mathbf{s}_g$ is the $n_g \times 1$ dimensional complex transmitted symbols at the input of the $g-$th STBC. As illustrated in Fig. 1, each of the $n_g \times T$ STBC codeword $\mathbf{U}_g$ is scaled by the power allocation $\sqrt{p_g} \ge 0$. $\{p_1, .., p_G\}$ is the transmit power allocation on the G groups such that the total transmit power constraint is satisfied:

$$\sum_{g=1}^{G} p_g = P_0 \tag{4}$$

During the t-th symbol duration (within a STBC), each of the $n_g \times 1$ code symbol of the STBC, $\mathbf{u}_g^{(t)}$, is multiplied by a $N_T \times n_g$ dimensional complex weight $\mathbf{W}_g$ for $g = 1, .., G$

to produce the $N_T \times 1$ transmit symbol $\mathbf{x}^{(t)}$ given by:

$$\mathbf{x}^{(t)} = \underbrace{\left[\begin{array}{ccc} \mathbf{W}_1 & \cdots & \mathbf{W}_G \end{array}\right]}_{\mathbf{W}} \underbrace{\left[\begin{array}{ccc} \sqrt{p_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{p_G} \end{array}\right]}_{\Lambda} \underbrace{\left[\begin{array}{c} \mathbf{u}_1^{(t)} \\ \vdots \\ \mathbf{u}_G^{(t)} \end{array}\right]}_{\mathbf{u}^{(\mathbf{t})}} \quad \text{for } t \in [1,T] \tag{5}$$

where $\mathbf{W}$ is the $N_T \times N$ *aggregate Tx-GMMSE* weights for the G groups, $\mathbf{u}^{(t)}$ is the $N \times 1$ aggregate STBC code symbol during the t-th symbol duration. For power normalization, we assume that $\mathcal{E}[\mathbf{u}_g^{(t)} \mathbf{u}_g^{(t)H}] = \frac{1}{n_g}\mathbf{I}_{n_g}$.

D. Orthogonal STBC

Note that in the proposed Tx-GMMSE scheme, full-rate and full-diversity STBC is used in each group to realize the spatial diversity so as to reduce the packet error probability. The full diversity order is important because it can fully exploit the available spatial dimensions to protect the packet transmission from channel outage. On the other hand, the full rate requirement in the STBC is also important because otherwise, there will be goodput penalty on choosing higher diversity order. A very appealing choice on the family of STBC is *orthogonal STBC* (OSTBC) as it provides full-diversity and allows low complexity decoding. Unfortunately, there does not exist a family of OSTBC with full-rate and full-diversity over various numbers of transmit antennas ($n_g = 2, 3, 4, 5, ...$) [14]. Hence, when the antenna grouping involves more than two antennas, we shall adopt a family of full-rate full-diversity quasi-orthogonal STBC (QOSTBC) [15]. The QOSTBC offers full-rate full diversity performance over different numbers of antennas in the grouping and therefore, allow very efficient and flexible tradeoff. The $M_k \times T$ dimensional received signal of the k-th mobile user $\mathbf{Y}_k^\dagger = [\mathbf{y}_{k,1}..\mathbf{y}_{k,T}]$ is given by:

$$\mathbf{Y}_k^\dagger = \underbrace{\sqrt{p_k}\mathbf{H}_k^H \mathbf{W}_k \mathbf{U}_k}_{\text{Information}} + \underbrace{\sum_{g \neq k} \sqrt{p_g}\mathbf{H}_k^H \mathbf{W}_g \mathbf{U}_g}_{\text{Multiuser Interference}} + \mathbf{Z}_k^\dagger \tag{6}$$

where the first term contains the desired STBC codewords targeted to the k-th mobile, the middle term represents the *multi-user interference* due to simultaneous transmission of independent STBC codewords over the G spatial channels and the $\mathbf{Z}_k$ corresponds to the $T \times M_k$ dimensional complex channel noise.

E. Selection of the Tx-GMMSE Weights

Ideally, we would like to design $\mathbf{W}_k$ to completely eliminate the multiuser interference terms in (6). However, this is not possible since the base station has only an estimate knowledge of the $N_T \times \sum_k M_k$ dimensional estimated aggregate CSIT, denoted by $\hat{\mathbf{H}} = [\hat{\mathbf{H}}_1 \cdots \hat{\mathbf{H}}_G]$. In this paper, we choose the weight based on Tx-GMMSE approach where the total normalized MSE error J_t (at any symbol instance) is minimized for a given user set $\mathcal{A}$ and the given spatial channel groupings $\{\mathcal{N}_1, ..., \mathcal{N}_G\}$. The spatial antenna groupings of user k, $\mathcal{N}_k$, is the set of selected received antenna index of the k-th user (with cardinality $|\mathcal{N}_k| = n_k \leq M_k$). For illustration, we assume the selected user set is $\mathcal{A} = \{1, 2, .., G\}$ and the selected spatial channel grouping of user k is $\mathcal{N}_k = \{1, ..., n_k\}$ as illustrations. Let $\widetilde{\mathbf{y}}_{k,t} = [\mathbf{y}_{k,t}(1)...\mathbf{y}_{k,t}(n_k)]^\dagger$ be the received signal vector (with elements corresponding to the received antenna index $\mathcal{N}_k$) and $\widetilde{\mathbf{y}}_t = [\widetilde{\mathbf{y}}_{1,t}^\dagger ... \widetilde{\mathbf{y}}_{G,t}^\dagger]^\dagger$ be the overall aggregate received signal vector (with elements corresponding to $\mathcal{N}_1, .., \mathcal{N}_G$). The total conditional MSE at the t-th symbol duration is given by (7), where $\mathbf{w}_{k,j}$ is the j-th column of the $n_T \times n_k$ dimensional MMSE weight $\mathbf{W}_k$, $\hat{\mathbf{h}}_{k,j}$ is the j-th column of the estimated CSIT $\hat{\mathbf{H}}_k$. Hence, the optimal MMSE weights $\mathbf{w}_{k,j}$ can be obtained by standard optimization technique in a decoupled manner and is given by:

$$\mathbf{w}_{k,j} = \sqrt{n_T(1-\sigma_e^2)} \left(\sum_{i \in \mathcal{N}_k} \hat{\mathbf{h}}_{k,i} \hat{\mathbf{h}}_{k,i}^H + \lambda_k \mathbf{I} \right)^{-1} \hat{\mathbf{h}}_{k,j}. \tag{8}$$

for $k \in \mathcal{A}$ and $j \in \mathcal{N}_k$.

F. Packet Error Model and System Goodput

Assume the STBC is constructed such that $\mathbf{s}_g$ and $\mathbf{s}_g^*$ do not occur at the same time slot [8], [15]. From (6) and (3), the received signal $\mathbf{y}_{k,t}$ of the $k-$th user at the t-th symbol duration (within a STBC codeword $t \in [1, T]$) can be expressed as:

$$\begin{aligned} \overline{\mathbf{y}}_{k,t} = & \underbrace{\sqrt{p_k}(\mathbf{H}_k^H \mathbf{W}_k \mathbf{A}_t^{(k)} + \mathbf{H}_k^\dagger \mathbf{W}_k^* \mathbf{B}_t^{(k)*})\mathbf{s}_k}_{\text{Information}} \\ & + \underbrace{\sum_{j \in \mathcal{A}/k} \sqrt{p_j}(\mathbf{H}_k^H \mathbf{W}_j \mathbf{A}_t^{(j)} + \mathbf{H}_k^\dagger \mathbf{W}_j^* \mathbf{B}_t^{(j)*})\mathbf{s}_j}_{\text{Multiuser Interference}, \mathbf{v}_{k,t}} + \overline{\mathbf{z}}_{k,j} \end{aligned} \tag{9}$$

such that:

$$\overline{\mathbf{y}}_{k,t}\{\overline{\mathbf{z}}_{k,t}\} = \begin{cases} \mathbf{y}_{k,t}\{\mathbf{z}_{k,t}\} & \text{if} \mathbf{B}_t^{(k)} = \mathbf{0}_{n_k \times n_k} \\ \mathbf{y}_{k,t}^*\{\mathbf{z}_{k,t}^*\} & \text{if} \mathbf{A}_t^{(k)} = \mathbf{0}_{n_k \times n_k} \end{cases} \quad \forall g \in [1, G], \tag{10}$$

Let $\mathbf{v}_{k,t}$ denotes the $M_k \times 1$ residual multiuser interference due to outdated knowledge of CSIT. Collecting all the $T \times M_k$ observations of the STBC for user k, $\overline{\mathbf{y}}_k = [\overline{\mathbf{y}}_{k,1}^\dagger, ..., \overline{\mathbf{y}}_{k,T}^\dagger]^\dagger$, we have

$$\overline{\mathbf{y}}_k = \sqrt{p_k}\mathcal{H}_k \mathbf{s}_k + \sum_{j \in \mathcal{A} \setminus k} \sqrt{p_j}\mathcal{H}_j \mathbf{s}_j + \overline{\mathbf{z}}_k$$

where

$$\mathcal{H}_j = \begin{pmatrix} \mathbf{H}_k^H \mathbf{W}_j \mathbf{A}_1^{(j)} + \mathbf{H}_k^\dagger \mathbf{W}_j^* \mathbf{B}_1^{(j)*} \\ \vdots \\ \mathbf{H}_k^H \mathbf{W}_j \mathbf{A}_T^{(j)} + \mathbf{H}_k^\dagger \mathbf{W}_j^* \mathbf{B}_T^{(j)*} \end{pmatrix} \tag{11}$$

The structure of quasi-orthogonal STBC design [15] gives the following property,

$$\mathcal{H}_k^H \mathcal{H}_k = \begin{pmatrix} \|\mathbf{H}_k^H \mathbf{W}_k\|^2 & 0 & \cdots & * \\ 0 & \|\mathbf{H}_k^H \mathbf{W}_k\|^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ * & 0 & \cdots & \|\mathbf{H}_k^H \mathbf{W}_k\|^2 \end{pmatrix} \tag{12}$$

where '$*$' denotes a small nonzero element which occurs only at anti-diagonal and $'*'$ is much smaller than the main

$$J_t(\mathbf{W}) = \sum_{k\in\mathcal{A}} p_k \sum_{j\in\mathcal{N}_k}\left(|\hat{\mathbf{h}}_{k,j}^H \mathbf{w}_{k,j} - \sqrt{n_T(1-\sigma_e^2)}|^2 + \sum_{j'\in\mathcal{N}_k, j'\neq j} |\hat{\mathbf{h}}_{k,j'}^H \mathbf{w}_{k,j}|^2 \right) + \sigma_e^2 P_0 + |\mathcal{A}| \tag{7}$$

diagonal terms. Note that $\mathcal{H}_k^H\mathcal{H}_k$ will be exactly diagonal for orthogonal STBC. Hence, for simplicity, we can ignore the anti-diagonal term and the output observations after orthogonal STBC decoder[2] is given by:

$$\overline{\psi}_k = \frac{\mathcal{H}_k^H}{\|\mathcal{H}_k\|}\overline{\mathbf{y}}_k = \sqrt{p_k}\frac{\mathcal{H}_k^H\mathcal{H}_k}{\|\mathcal{H}_k\|}\mathbf{s}_k + \sum_{j\in\mathcal{A}\backslash k}\sqrt{p_j}\frac{\mathcal{H}_k^H\mathcal{H}_j}{\|\mathcal{H}_k\|}\mathbf{s}_j + \frac{\mathcal{H}_k^H\overline{\mathbf{z}}_k}{\|\mathcal{H}_k\|} \tag{13}$$

Using conventional orthogonal STBC detection scheme where the n_k observations in $\overline{\psi}_k$ are detected separately, the total instantaneous mutual information (bits per channel use) of the k-th user conditional on CSIR $\mathcal{H}_k$ is given by:

$$C_k \approx \log_2\left(1 + \frac{\frac{p_k}{n_k}\|\mathbf{H}_k^H\mathbf{W}_k\|^2}{1+\sum_{j\neq k}\frac{p_j}{n_j}\|\mathbf{H}_k^H\mathbf{W}_j\|^2}\right) \tag{14}$$

In order to capture the potential packet errors into the system performance measure, we shall consider the system goodput (b/s/Hz successfully delivered to the mobile stations) as our performance measure instead of conventional ergodic capacity. In general, packet error is contributed by two factors, namely the *channel noise* and the *channel outage*. In the former case, packet error happens because of non-ideal channel coding and finite block length of the channel codes. This factor can be reduced by using a stronger channel code and a longer block length. However, in the latter case, the effect is systematic and cannot be eliminated by simply using a stronger code or longer block length. This is because the instantaneous mutual information[3] for the k-th user, C_k, between the STBC inputs, $\mathbf{s}_k$, at the base station and the received signals $\overline{\mathbf{y}}_k$ is a function of actual CSI $\mathbf{H}_k$, which is unknown to the base station. Hence, the packet will be corrupted whenever the scheduled data rate exceeds the instantaneous mutual information. In practice, for reasonable block length (such as 8K byte) and strong coding (such as LDPC), Shannon's capacity C_k can be approached to within 0.05 dB for a target FER of 10^{-3}. Hence, we shall model the packet error solely by the probability that the scheduled data rate exceeding the instantaneous mutual information (i.e. packet error due to the channel outage only).

To take into consideration of the potential packet errors, we define the *instantaneous goodput* of the k-th user as:

$$\rho_k = r_k \mathbf{1}[C_k \geq r_k] \tag{15}$$

where $\mathbf{1}(E)$ is the indicator function which is equal to 1 if the event E is true and 0 otherwise. Hence, the instantaneous goodput measures the actual b/s/Hz successfully delivered to user k in the current fading slot. Similarly, the *average system goodput* is defined as the total average b/s/Hz successfully delivered to the K mobiles (averaged over multiple time slots) and is given by:

$$\begin{aligned} U_{thp}(\bar{\rho}_1,...,\bar{\rho}_K) &= \mathcal{E}[\sum_{k=1}^K \rho_k] = \mathcal{E}\left\{\sum_{k=1}^K r_k \Pr\left[r_k < C_k|\hat{\mathbf{H}}\right]\right\} \\ &= \mathcal{E}_{\hat{\mathbf{H}}}[\mathcal{G}(r_1,...,r_K)] \end{aligned} \tag{16}$$

where $\mathcal{G}(r_1,...,r_K)$ is the *conditional system goodput* given by:

$$\begin{aligned} \mathcal{G}(r_1,...,r_K) &= \sum_{k=1}^K r_k \Pr\left[r_k < C_k|\hat{\mathbf{H}}\right] \\ &= \sum_{k=1}^K r_k\left[1 - P_{out}(r_k|\hat{\mathbf{H}})\right] \end{aligned} \tag{17}$$

and $P_{out}(r_k|\hat{\mathbf{H}})$ is the conditional packet outage probability (conditioned on the CSIT $\hat{\mathbf{H}}$).

III. Cross Layer Scheduling Problem Formulation

While the Tx-GMMSE physical layer offers flexible transmission modes (spatial diversity/multiplexing), it is critical to have a properly designed cross-layer scheduler to take advantage of this flexibility to achieve robust performance with outdated CSIT. At the beginning of scheduling time slot, the cross-layer scheduler determines the admitted user set $\mathcal{A}$ (set of user indices who are selected in the current time slot), the power allocation $\{p_1,...,p_K\}$, the rate allocation $\{r_1,...,r_K\}$ as well as the *mode selection* $\{\mathcal{N}_1,...,\mathcal{N}_G\}$ for the selected users. The scheduling results are broadcasted to the mobile users through a common downlink channel. Afterwards, the payload from the selected users are transmitted at the scheduled rate and power in the downlink as illustrated in Fig. 1. In this section, we shall propose an integrated framework for the design of downlink cross-layer scheduling with Tx-GMMSE in the presence of outdated CSIT.

A. The Scheduling Problem Formulation

The cross-layer scheduling problem with outdated CSIT can be cast into the following optimization problem:

Problem 1 (Cross Layer Optimization Problem): Given any realization of the estimated CSIT from all mobile users $\hat{\mathbf{H}}$, determine the optimal admitted user set $\mathcal{A}$, the optimal power allocation $\{p_k\}$, the optimal rate allocation $\{r_k\}$ as well as the optimal mode selection (spatial multiplexing/diversity) set $\mathcal{N}$ so as to maximize the conditional system goodput in (17), $\mathcal{G}(r_1,...,r_K)$ subject to the transmit power constraint $\sum_{k\in\mathcal{A}} p_k \leq P_0$ and PER constraint $P_{out}(r_k|\hat{\mathbf{H}}) = \epsilon$.

B. Optimization Solution

Observe that the optimization parameters in problem 1 include real variables $(p_1,p_2,...,p_K)$ and $(r_1,r_2,...,r_K)$ as well as discrete variables $\mathcal{A}$ and $\mathcal{N}$. Hence, the solution involves a mixed analytical optimization (w.r.t. p_k, r_k) and combinatorial search (w.r.t. $\mathcal{A}$ and $\mathcal{N}$). The solution is expressed into the following two steps.

[2] In conventional orthogonal STBC detection, the vector channel outputs $\overline{\mathbf{y}}_k$ are pre-multiplied by the CSIR $\mathcal{H}_k^H$ to form the decoupled sufficient statistics with respect to the information symbols.

[3] The STBC, together with the MIMO channel, transforms a vector channel into a SISO *super-channel* for user k and we the mutual information between the STBC inputs and the received signals represents the maximum achievable rate for error free transmission if strong coding is concatenated onto the STBC.

1) Step 1: Analytical Optimization: In this step, we shall focus on solving for $\{p_k\}$ and $\{r_k\}$ for a given admitted user set $\mathcal{A}$ and the associated spatial stream grouping $\mathcal{N}$. To obtain the closed-form expressions on the scheduled data rate and power allocation, we need to work out the conditional outage probability. $P_{out}(r_k|\hat{\mathbf{H}})$ can be expressed as:

$$P_{out}(r_k|\hat{\mathbf{H}}) = \Pr\left(S_k < \Lambda_k|\hat{\mathbf{H}}\right) \quad (18)$$

where

$$S_k = \frac{p_k}{n_k}\|\mathbf{H}_k^H\mathbf{W}_k\|^2 - \sum_{j\neq k}\zeta_j\|\mathbf{H}_k^H\mathbf{W}_j\|^2, \quad (19)$$

$\Lambda_k = (2^{r_k}-1)n_k$ and $\zeta_j = \frac{p_j\Lambda_k}{n_j}$. In general, obtaining the distribution of S_k is not easy because conditioned on the CSIT $\hat{\mathbf{H}}$, the random variables $\|\mathbf{H}_k^H\mathbf{W}_k\|^2$ and $\|\mathbf{H}_k^H\mathbf{W}_j\|^2$ in S_k are correlated. In addition, the terms $\|\mathbf{H}_k^H\mathbf{W}_j\|^2$ are also correlated for different j. Yet, we found from the following lemma that S_k converges to non-central chi-square random variable as N_T increases. Hence, the outage probability can be approximated by chi-square distribution as below (which is asymptotically exact).

*Definition 1 (**Regular Power Allocation Policy**):* A power allocation policy $\mathcal{P} = \{p_i : \forall i \in \mathcal{A}\}$ is said to be **regular** if $\mathcal{E}[p_i^m] \leq \mathcal{O}(1/N^m)$ for $m = 2, 4$ where $\mathcal{O}(.)$ denotes asymptotic upper bound.

Lemma 1: [**Asymptotic Packet Outage Distribution**] Let $\sum_{k\in\mathcal{A}}|\mathcal{N}_k|/N_T = L$ be the *SDMA loading factor* and for sufficiently large N_T, the conditional packet outage probability of the k-th user (conditioned on the CSIT realization $\hat{\mathbf{H}}$), $P_{out}(r_k|\hat{\mathbf{H}})$, converges to:

$$P_{out}(r_k|\hat{\mathbf{H}}) \to F_{\chi^2_{2n_kM_k}}\left(\frac{\Lambda_k(\overline{\mu_I}+1)}{p_k\|\hat{\mathbf{H}}_k^H\mathbf{W}_k\|^2/n_k}\right) \quad (20)$$

with probability one if the power allocation $\{p_i : i \in \mathcal{A}\}$ is regular where $F_{\chi^2_{2n_kM_k}}(y)$ is cdf of non-central chi-square of $2n_kM_k$ degrees of freedom, non-central parameter $s = 1$ and variance $\sigma_e^2/\|\hat{\mathbf{H}}_k^H\mathbf{W}_k\|^2$, $\overline{\mu_I}$ is given by

$$\overline{\mu_I} = P_0M_k\sigma_e^2 + N_T(1-\sigma_e^2)P_0/\sum_j n_j,$$

and $\overline{B}$ is a deterministic constant given by $\overline{B} = \int_0^\infty \frac{s}{(s+\tilde{\lambda})^2}dG^*(s)$ where s is the eigenvalues of $\left(\sum_{j\neq k}\widetilde{\hat{\mathbf{h}}_j}^*\widetilde{\hat{\mathbf{h}}_j}\right)$ and $G^*(s)$ is the limiting empirical distribution of s [10].

The proof of this Lemma is given in *Appendix A*. Note that the extra robustness against packet outage (due to CSIT errors) by assigning n_k spatial subchannels to user k is reflected in the extra degrees of freedom in the non-central chi-square distribution in (20). The *regularity* requirement of power allocation implies that there is no single user being allocated *exceptionally large power* on average. For homogeneous users, the average power allocation among all users should be regular by symmetry. Otherwise, that means some users are exceptionally weak. In the following, we shall assume that the power allocation policy is regular and obtain the asymptotically optimal power and rate allocation solutions from (20).

Using (20) for the conditional packet outage probability, the outage constraint $P_{out} = \epsilon$ becomes:

$$P_{out} = \epsilon \Longleftrightarrow \Lambda_k(1+\overline{\mu_I})/p_k\|\hat{\mathbf{H}}_k^H\mathbf{W}_k\|^2/n_k = \varphi(\epsilon) \quad (21)$$

where $\varphi(\epsilon) = F^{-1}_{\chi^2_{2n_kM_k}}(\epsilon)$. Hence, the outage requirement is satisfied if and only if $r_k = \log_2(1+p_k\gamma_k)$ where $\gamma_k = \frac{\|\hat{\mathbf{H}}_k^H\mathbf{W}_k\|^2\varphi(\epsilon)}{n_k(1+\overline{\mu_I})}$ is the average SIR of user k (per unit received power). Using standard optimization techniques, we have the optimal rate and power allocation given by:

$$p_k^* = \left(\frac{1-\epsilon}{\mu} - \frac{1}{\gamma_k}\right)^+ \quad (22)$$

and

$$r_k^* = \left[\log_2\left(\frac{(1-\epsilon)\gamma_k}{\mu}\right)\right]^+ \quad (23)$$

where $\mu > 0$ is the Lagrange multiplier chosen to satisfy $\sum_{k\in\mathcal{A}}p_k = P_0$.

2) Step 2: Combinatorial Optimization: The remaining optimization parameters are the admitted user set $\mathcal{A}$ and the corresponding spatial multiplexing/diversity modes $\mathcal{N}$. The optimal solution can be obtained using the exhaustive search method over all possible combinations of $\mathcal{A}$ and $\mathcal{N}$ has two potential issues. Firstly, the search complexity is huge with a total search space of $\mathcal{O}(K^{N_T})$. Secondly, the search objective $\mathcal{G}(r_1, ..., r_K)$ is a function of $\mathcal{A}$, $\mathcal{N}$ and the Lagrandge multiplier μ. For notation convenience, we denote the search objective as $\mathcal{G}(\mu, \mathcal{A}, \mathcal{N})$. To maintain the same transmit power constraint P_0, the Lagrange multiplier μ will be changed as we change $\mathcal{A}$ and $\mathcal{N}$. Given any $\mathcal{A}$ and $\mathcal{N}$, we need to compute the Tx-GMMSE weight as well as the Lagrange multipler μ in order to evaluate $\mathcal{G}(\mu, \mathcal{A}, \mathcal{N})$. This results in a further increase in the search complexity. Since the scheduling algorithm is a realtime process per time slot, the exhaustive search solution is not suitable for implementation. In this section, we shall propose a low complexity suboptimal algorithm (the *double greedy-based selection algorithm* (DGBSA)) to determine the user selection $\mathcal{A}$ and the mode selection $\mathcal{N}$. We shall first establish an essential lemma before elaborating the DGBSA algorithm.

*Lemma 2 (**Equivalent Search Metric**):* For any given P_0, if $\mathcal{A}^*$ and $\mathcal{N}^*$ maximize the conditional goodput $\mathcal{G}(\mu, \mathcal{A}, \mathcal{N})$ then all the power $\{p_k\}$ are positive for all $k \in \mathcal{A}^*$ and $f(\mathcal{A}^*, \mathcal{N}^*) \geq f(\mathcal{A}, \mathcal{N})$ for any $(\mathcal{A}, \mathcal{N}) \neq (\mathcal{A}^*, \mathcal{N}^*)$ where $f(.)$ is given by:

$$f(\mathcal{A}, \mathcal{N}) = \prod_{k\in\mathcal{A}^*}\frac{\gamma_k}{|\mathcal{A}|^*}\left(P_0 + \sum_{k\in\mathcal{A}^*}\frac{1}{\gamma_k}\right)$$

and

$$\gamma_k = \sqrt{2}\varphi_k\sigma_{\Delta H}\|\hat{\mathbf{h}}_k^\dagger\mathbf{W}_k\| + (n_k\sigma_{\Delta H}^2 + \|\hat{\mathbf{h}}_k^\dagger\mathbf{W}_k\|^2).$$

On the other hand, if $\mathcal{A}^*, \mathcal{N}^*$ maximizes $f(\mathcal{A}, \mathcal{N})$ and $\mu < \gamma_k$ for all $k \in \mathcal{A}^*$, then $\mathcal{A}^*, \mathcal{N}^*$ maximizes $\mathcal{G}(\mu, \mathcal{A}, \mathcal{N})$.

Please refer to appendix B for the proof. In other words, the equivalent search metric for $\mathcal{A}$ and $\mathcal{N}$ can be replaced by $f(\mathcal{A}, \mathcal{N})$, which is free from the Lagrange multiplier, for sufficiently large P_0. Define $f^*(\mathcal{A}) = \max_{\mathcal{N}} f(\mathcal{A}, \mathcal{N})$ and $f^{**} = \max_{\mathcal{A}} f^*(\mathcal{A})$. Based on lemma 2, we shall elaborate the DGBSA below.

- **Step 1: Level I Initialize.** Set $\mathcal{A}^{(0)} = \emptyset$, $\mathcal{N}^{(0)} = \emptyset$ and $k = 1$.
- **Step 2: Search for $\mathcal{A}$.** We have $i^* = \arg\max_{i \in [1,K]/\mathcal{A}^{(k-1)}} f^*(\mathcal{A}^{(k-1)} \bigcup \{i\})$, $\mathcal{A}^{(k)} = \mathcal{A}^{(k-1)} \bigcup \{i^*\}$ and $\mathcal{N}^{(k)} = \arg\max_{\mathcal{N}} f(\mathcal{A}^{(k)}, \mathcal{N})$. To evaluate $f^*(\mathcal{A}^{(k-1)} \bigcup \{i\})$, we perform the Level II Greedy Algorithm as below:
  - **Step 2a: Level II Initialize.** Given the admitted user set from Level I $\mathcal{A}^{(k-1)} \bigcup \{i\}$, let k_j be the j-th user in $\mathcal{A}^{(k-1)}$. For a given $\mathcal{N}^{(k-1)}$, set $\mathcal{B}^{(0)} = \mathcal{N}^{(k-1)}$ and $m = 1$.
  - **Step 2b: Search for $\mathcal{N}$.** Denoting $\mathcal{B}(j)$ as the number of spatial streams allocated to user $j \in \mathcal{A}^{(k-1)}$ for any *mode selection set* $\mathcal{B}$. If $\mathcal{B}^{(m)} = \emptyset$, set $m = N_T$, $\mathcal{B}^{(m)}(i) = N_T$ and proceed to step 2c. Otherwise, set $\mathcal{B}_j^{(m)} = \mathcal{B}^{(m-1)}/\{j\}|\{i\}$ which denotes the operation of removing a spatial channel from user $j \in \mathcal{A}^{(k-1)}$ and assigning it to user i. i.e.,
  $$\mathcal{B}_j^{(m)}(j) = \mathcal{B}^{(m-1)}(j) - 1, \quad \mathcal{B}_j^{(m)}(i) = \mathcal{B}^{(m-1)}(i) + 1$$
  We have $j^* = \arg\max_{\mathcal{B}_j^{(m)}} f(\mathcal{A}^{(k-1)} \bigcup \{i\}, \mathcal{B}_j^{(m)})$ and $\mathcal{B}^{(m)} = \mathcal{B}_{j^*}^{(m)}$.
  - **Step 2c: Level II Termination.** If $m = N_T$ then $\mathcal{N}^{(k)} = \mathcal{B}^{(N_T)}$ and terminate with $f^*(\mathcal{A}^{(k-1)} \bigcup \{i\}) \approx f(\mathcal{A}^{(k-1)} \bigcup \{i\}, \mathcal{N}^{(k)})$. If $f(\mathcal{A}^{(k-1)} \bigcup \{i\}, \mathcal{B}^{(m)}) < f(\mathcal{A}^{(k-1)} \bigcup \{i\}, \mathcal{B}^{(m-1)})$, then $\mathcal{N}^{(k)} = \mathcal{B}^{(m-1)}$ and terminate with $f^*(\mathcal{A}^{(k-1)} \bigcup \{i\}) \approx f(\mathcal{A}^{(k-1)} \bigcup \{i\}, \mathcal{N}^{(k)})$. Otherwise, set $m = m + 1$ and goto step 2b.
- **Step3: Level I Termination.** If $k = K$ or $f^*(\mathcal{A}^{(k)}) < f^*(\mathcal{A}^{(k-1)})$, then $\mathcal{A}^* = \mathcal{A}^{(k-1)}$, $\mathcal{N}^* = \mathcal{N}^{(k-1)}$ and terminate. Otherwise, set $k = k + 1$ and repeat step 2.

The complexity of the DGBSA is of $O(KN_T^2)$, which is much lower than exhaustive search $O(K^{N_T})$.

IV. Simulation Results

In this section, we will evaluate the system performance in terms of sum of total user throughput using computer simulations. We shall compare the performance of the proposed robust cross-layer scheduler (based on Tx-GZF) with various reference baselines elaborated below.

- **Baseline 1: Regular Tx-ZF + Naive cross-layer.** In baseline 1, we consider the regular ZF-Tx with *naive* cross-layer scheduler (designed for perfect CSIT). The spatial streams are separated by regular Tx-ZF based on the imperfect CSIT $\hat{\mathbf{H}}$ at the physical layer and the naive cross layer scheduler allocate user selection, rate adaptation, power adaptation by using the imperfect CSIT as if they were perfect.
- **Baseline 2: Regular Tx-ZF + Robust cross-layer.** Baseline 2 is defined as the system with regular Tx-ZF and a robust cross-layer scheduler (matched to the outdated CSIT). Similar to baseline 1, the spatial streams are separated by Tx-ZF based on the outdated CSIT. However, we shall apply our proposed CSIT-error matched cross-layer design over the regular Tx-ZF physical layer.
- **Baseline 3: Tx-GMMSE + round robin scheduling.** Baseline 3 is defined as the system with Tx-GMMSE processing and simple round robin scheduling.

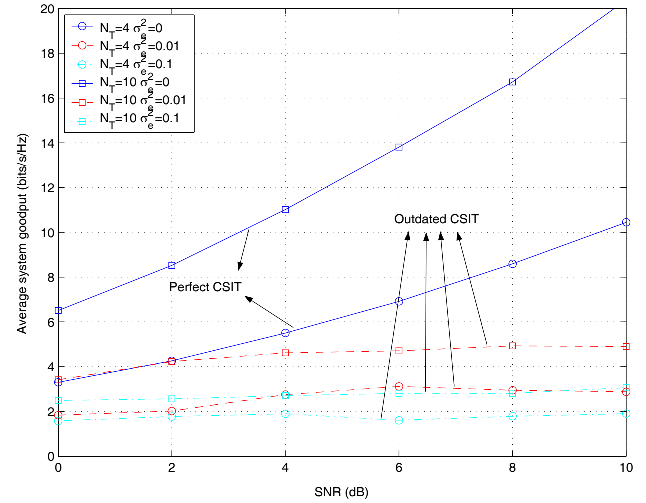

Fig. 2. System goodput (b/s/Hz) versus SNR of the baseline 1 system (Tx-ZF + naive scheduling) for $N_T = 4, 10$ and $\sigma_e^2 = 0, 0.01, 0.1$, at a target outage probability of $P_{out} = 0.01$. The loss of goodput is obvious at a moderate CSIT error $\sigma_e^2 = 0.01$. Performance saturates when SNR increases at CSIT error $\sigma_e^2 = 0.1$.

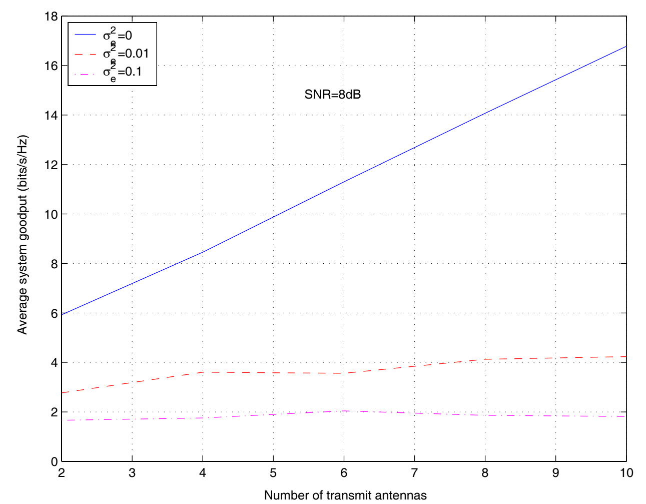

Fig. 3. Average goodput performance vs number of transmit antennas of the baseline 1 system for $\sigma_e^2 = 0, 0.01, 0.1$ and SNR=8dB, at a target outage probability of $P_{out} = 0.01$. Increasing the number of transmit antenna cannot improve the systems goodput anymore even if there is small CSIT error $\sigma_e^2 = 0.01$. This is due to packet errors when the scheduled rate exceeds the instantaneous mutual information.

A. Performance of the Naive Scheduler with CSIT Errors

Fig. 2 illustrates the average system goodput (b/s/Hz) versus SNR for the baseline 1 system with $N_T = 4, 10$ and CSIT errors $\sigma_e^2 = 0.01, 0.1$. Fig. 3 illustrates the average system goodput (b/s/Hz) versus N_T for the baseline 1 system at SNR=8dB. We observe that even with small CSIT error ($\sigma_e^2 = 0.01$), system goodput degrades significantly compared to the perfect CSIT result. In particular, the system goodput does not scale with either the SNR or N_T (spatial multiplexing gain). Hence, these results highlight the importance of designing the cross-layer scheduler matching to the CSIT errors.

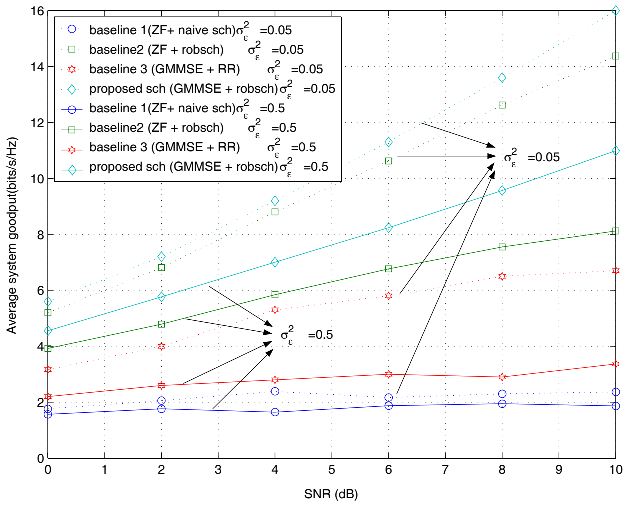

Fig. 4. Performance comparison of the proposed design and baseline 1, 2, 3 systems for $N_T = 8$ and $\sigma_e^2 = 0.05, 0.5$ at a target outage probability of $P_{out} = 0.002$. *robsch* stands for robust scheduling. Baseline 1 system fails to improve goodput. The proposed design and baseline 2 system are more robust CSIT errors. The advantage of Tx-GMMSE processing over Tx-ZF processing is observed, especially for large CSIT error and/or large number of transmit antennas.

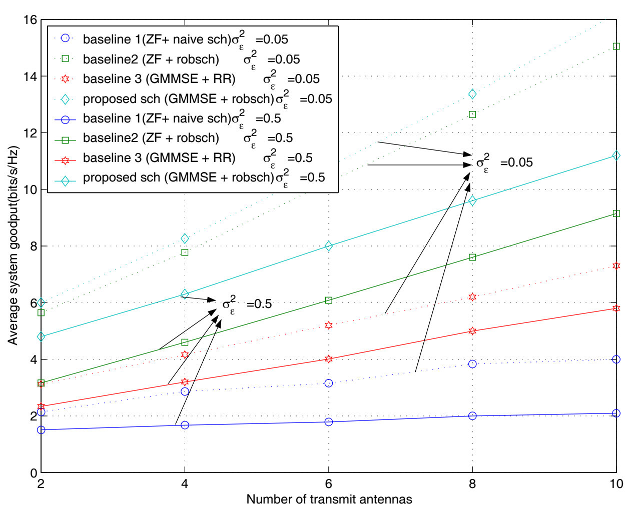

Fig. 5. Average goodput performance vs number of transmit antennas of the proposed design and other 3 baselines for $\sigma_e^2 = 0.05, 0.5$ and SNR=8dB at a target outage probability of $P_{out} = 0.002$. *robsch* stands for robust scheduling.

B. Performance of the Proposed Design (GMMSE + Robust Scheduler) with CSIT Errors

Fig. 4 illustrates the average system goodput versus the SNR for the proposed design, the baseline 1, baseline 2 and baseline 3 systems. Comparing the proposed design and the the baseline 2 system (Tx-ZF + robust scheduler), we can observe that the proposed Tx-GMMSE cross-layer scheduler is more robust to the CSIT error due to the spatial diversity, which is important to protect the packets. By comparing our proposed scheduler with baseline 3 (Tx-GMMSE + round robin), it is seen that high gain is achieved due to the robust cross-layer scheduler matching to the imperfect CSIT error. In addition, comparing the proposed design with the baseline 1 system (Tx-ZF + naive scheduler), we observe the

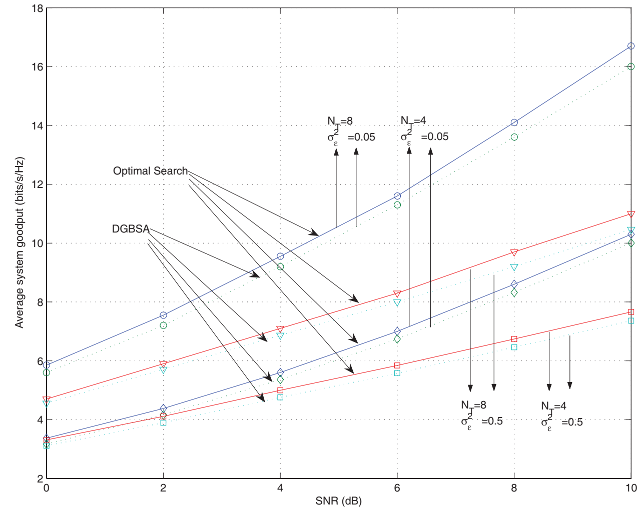

Fig. 6. Performance comparison of Tx-GMMSE based genetic scheduler with optimal scheduler for $N_T = 8$ at CSIT error $\sigma_e^2 = 0.05, 0.5$. The target outage probability is $P_{out} = 0.002$. DGBSA scheduler has near optimal performance but low complexity.

significant performance gain due to both the robust physical layer processing (Tx-GMMSE) as well as robust cross-layer scheduling. Finally, comparing the performance between baseline 2 and baseline 3, we can see the importance of cross-layer gain (multiuser diversity) even with moderate to large CSIT errors. Fig. 5 plots the average goodput vs N_T performance of the four systems at CSIT error $\sigma_e^2 = 0.05, 0.5$ and SNR=8dB. There is a significant gain in system goodput of the proposed design due to spatial multiplexing and multiuser diversity gains.

C. Performance of the DGBSA

Fig. 6 shows the average goodput performance of the sub-optimal DGBSA. We observe that the DGBSA search achieves close-to-optimal performance at different number of transmit antennas and at various CSIT errors. Yet, the complexity of the DGBSA search is around 100 times lower than that of the optimal search for $K = 10$ and $N_T = 4$.

V. Conclusion

In this paper, we propose the Tx-GMMSE physical layer processing as well as the corresponding cross-layer scheduling algorithm for multi-user MIMO systems with outdated CSIT. On the physical layer processing, the Tx-GMMSE introduces robustness for the potential packet outage because it allows a flexible tradeoff of spatial multiplexing and spatial diversity. Coupled with the cross-layer designs matching the CSIT errors, we show that the proposed Tx-GMMSE achieves very robust performance at moderate to large CSIT errors. Tx-GMMSE offers additional performance gains over regular ZF physical layer at moderate to large CSIT errors even if the corresponding cross-layer components are matched to the imperfect CSIT. This demonstrates the robustness of Tx-GMMSE with respect to packet outage.

Appendix A
Proof of the Lemma 1

For easy notation, assume $\mathcal{A} = \{1, 2, .., k, .., G\}$ and define $\widetilde{\mathbf{H}} = \mathbf{H}/\sqrt{N_T(1-\sigma_e^2)}$ as the normalized CSI for

$\sigma_e^2 < 1$. From (18), the conditional outage probability is given by $P_{out}(r_k|\hat{\mathbf{H}}) = \Pr\left[S_k < \Lambda_k/N_T(1-\sigma_e^2)|\hat{\mathbf{H}}\right]$ where $\Lambda_k = (2_k^r - 1)$, and S_k is the normalized random variable given by:

$$S_k = p_k\|\mathbf{X}_k\|^2/n_k - \Lambda_k \sum_{j\in\mathcal{A}, j\neq k} p_j\|\mathbf{X}_j\|^2/n_j \tag{24}$$

where $\mathbf{X}_j = \mathbf{W}_j^H\widetilde{\mathbf{H}}_k = \mathbf{W}_j^H\widetilde{\hat{\mathbf{H}}}_k + \mathbf{W}_j^H\widetilde{\Delta\mathbf{H}}_k$ for $j \in \mathcal{A}$. $\mathbf{X}_j$ is a complex Gaussian $n_j \times M_k$ matrix (conditioned on CSIT $\widetilde{\hat{\mathbf{H}}}$) with conditional mean $\mu_{X_k} = \mathbf{W}_j^H\widetilde{\hat{\mathbf{H}}}_k$. The conditional covariance of the i-th column of $\mathbf{X}_j$ is given by $\mathcal{E}\left[\mathbf{X}_j(i)\mathbf{X}_j(i)^H|\hat{\mathbf{H}}\right] = \sigma_e^2/(N_T(1-\sigma_e^2))\mathbf{I}_{n_j}$ for $i = 1, 2, .., M_k$. Furthermore, the conditional covariance between any two different columns in $\mathbf{X}_j$ is given by $\mathbf{0}$ $(n_j \times n_j)$. Hence, $\|\mathbf{X}_k\|^2 = \|\widetilde{\hat{\mathbf{H}}}_k^H\mathbf{W}_k\|^2$ in (24) is non-central distributed (conditioned on $\hat{\mathbf{H}}$).

Next, we shall look at the distribution of the term $\sum_{j\in\mathcal{A},j\neq k} p_j\|\mathbf{X}_j\|^2/n_j$ in (24). Notice that $\{\mathbf{X}_j\}_{j\neq k}$ are mutually correlated and we have $\mathcal{E}\left[\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_j^H|\hat{\mathbf{H}}\right] = \frac{M_k\sigma_e^2}{N_T(1-\sigma_e^2)}\mathbf{W}_i^H\mathbf{W}_j$ where $\widetilde{\mathbf{X}}_j = \mathbf{X}_j - \mu_{X_j}$. Define the spatial interference I_k as:

$$\begin{aligned} I_k &= \sum_{j\neq k} p_j\|\mathbf{X}_j\|^2/n_j \\ &= \underbrace{\sum_{j\neq k} p_j\|\mu_{X_i}\|^2/n_j}_{I_A} + \underbrace{2\Re tr\left(\sum_{j\neq k} p_j\mu_{X_j}^*\widetilde{\mathbf{X}}_j/n_j\right)}_{I_B} \\ &\quad + \underbrace{\sum_{j\neq k} p_j\|\widetilde{\mathbf{X}}_j\|^2/n_j}_{I_C} \end{aligned} \tag{25}$$

Since conditioned on the CSIT, I_A is a constant, we shall analyze the convergence of I_B, and I_C respectively below. Let $N = \sum_k n_k$ and $L = N/N_T$.

Lemma 3 (Convergence of I_B): If $L > 0$ and $\{p_i\}$ is regular, we have $I_B = 2\Re tr\left(\sum_{j\neq k} p_j\mu_{X_j}^*\widetilde{\mathbf{X}}_j/n_j\right)$ drops faster than $\mathcal{O}(1/N_T)$ for almost all realizations of CSIT. That is $\Pr[I_B \leq \mathcal{O}(1/N_T^{1+\delta})] \to 1$ for some $\delta > 0$ as N_T increases.

Proof: Observe that $\mathcal{E}[I_B]$ is given by:

$$\begin{aligned} \mathcal{E}[I_B] &= 2\Re tr\sum_{j\neq k}\mathcal{E}_{\hat{\mathbf{H}}}\left[p_j\mu_{X_j}^H\mathcal{E}[\widetilde{\mathbf{X_j}}|\hat{\mathbf{H}}]/n_j\right] \\ &= 2\Re tr\sum_{j\neq k}\mathcal{E}_{\hat{\mathbf{H}}}\left[p_j\mu_{X_j}^H\mathbf{0}/n_j\right] = 0. \end{aligned}$$

Using matrix inversion lemma and from [10], we have

$$\begin{aligned} \mathcal{E}[|\mu_{X_i}(n,m)|^4] &= \mathcal{E}\left[\frac{|\widetilde{\hat{\mathbf{h}}}_{k,m}^H\mathbf{M}_{k,m}^{-1}\widetilde{\hat{\mathbf{h}}}_{i,n}|^4}{|1+\widetilde{\hat{\mathbf{h}}}_{k,m}^H\mathbf{M}_{k,m}^{-1}\widetilde{\hat{\mathbf{h}}}_{k,m}|^4}\right] \\ &\leq \mathcal{E}[|\widetilde{\hat{\mathbf{h}}}_{k,m}\mathbf{M}_{k,m}^{-1}\widetilde{\hat{\mathbf{h}}}_{i,n}|^4] \underbrace{\leq}_{(a)} \mathcal{O}(1/N_T^2) \end{aligned} \tag{26}$$

and

$$\begin{aligned} \mathcal{E}\left[\left|\mathbf{W}_i^H(n)\mathbf{W}_j(n')\right|^4\right] &\leq \mathcal{E}\left[\left|\frac{\widetilde{\hat{\mathbf{h}}}_{i,n}^H\mathbf{M}_{i,n}^{-2}\widetilde{\hat{\mathbf{h}}}_{j,n'}}{1+\widetilde{\hat{\mathbf{h}}}_{i,n}^H\mathbf{M}_{i,n}^{-1}\widetilde{\hat{\mathbf{h}}}_{i,n}}\right|^4\right] \\ &\leq \mathcal{E}\left[\left|\widetilde{\hat{\mathbf{h}}}_{i,n}^H\mathbf{M}_{i,n}^{-2}\widetilde{\hat{\mathbf{h}}}_{j,n'}\right|^4\right] \underbrace{\leq}_{(a)} \mathcal{O}(1/N_T^2) \end{aligned} \tag{27}$$

where $\mathbf{W}_i(n)$ denotes the n-th column of $\mathbf{W}_i$, $\mathbf{M}_{k,m} = \left(\sum_{i\neq k}\widetilde{\hat{\mathbf{H}}}_i\widetilde{\hat{\mathbf{H}}}_i^H + \sum_{n'\neq m}\widetilde{\hat{\mathbf{h}}}_{k,n'}\widetilde{\hat{\mathbf{h}}}_{k,n'}^H + \widetilde{\lambda_{k,m}}\mathbf{I}\right)$ and $\mu_{X_i}(n,m)$ denotes the $(n,m)-$th element of the matrix μ_{X_i}. Similarly, we have $\mathcal{E}[|\mu_{X_i}(n,m)|^8] \leq \mathcal{O}(1/N_T^4)$. Consider the first expression at the top of the next page, where (b), (c) are due to Cauchy-Swartz inequality and (d) is due to (27) and (26). Hence, $\Pr[I_B \leq \epsilon] \geq 1 - \frac{\mathrm{Var}(I_B)}{\epsilon^2} \geq 1 - \mathcal{O}(1/N_T^{2.5}\epsilon^2)$ for any $\epsilon > 0$. Set $\epsilon = 1/N_T^{1+\delta}$ for some $\delta \in (0, 0.5)$, we have $\Pr[I_B \leq \mathcal{O}(1/N_T^{1+\delta})] \to 1$. ■

Lemma 4 (Convergence of I_C): If $L > 0$ and $\{p_i\}$ is regular, we have $I_C = \sum_{j\neq k} p_j\|\widetilde{\mathbf{X}}_j\|^2/n_j$ converges to $\overline{I_C} = P_0 M_k\sigma_e^2/(N_T(1-\sigma_e^2))$ in probability as n_T increases.

Proof: Consider the second expression at the top of the next page, where (b) is due to Cauchy-Swartz inequality and (c) is from (27). Since $\mathrm{Var}(I_C)$ drops faster than $\mathcal{E}^2[I_C] = \mathcal{O}(1/N_T^2)$, we have I_C converges to $\mathcal{E}[I_C]$ in probability as N_T increases. ■

Using Lemmas 3 and 4, we have for almost all realizations of CSIT, $N_T(1-\sigma_e^2)I_k \to P_0 M_k\sigma_e^2 + N_T(1-\sigma_e^2)\sum_{j\neq k} p_j\|\mu_{X_j}\|^2/n_j \approx \overline{\mu_I} = P_0 M_k\sigma_e^2 + N_T(1-\sigma_e^2)P_0/\sum_j n_j$ [10]. Therefore, we have $S_k N_T(1-\sigma_e^2) \to N_T(1-\sigma_e^2)p_k\|\mathbf{X}_k\|^2/n_k - \Lambda_k\overline{\mu_I}$. Since $\|\mathbf{X}_k\|^2$ is a non-central chi-square random variable with $2n_k M_k$ degrees of freedom, noncentral parameter $s = \|\widetilde{\hat{\mathbf{H}}}_k^H\mathbf{W}_k\|^2$. Hence, the result is proven.

Appendix B
Proof of Lemma 2

For the first part, let $(\mathcal{A}^*, \mathcal{N}^*)$ maximizes the goodput $\mathcal{G}(\mu, \mathcal{A}^*, \mathcal{N}^*)$ for some μ chosen to satisfy the power constraint P_0. We claim that all the power allocation $\{p_k\}$ for $k \in \mathcal{A}^*$ must be positive. This is because otherwise, say there exists a user $j \in \mathcal{A}^*$ who has zero power. Consider another admitted user set $\mathcal{A}' = \mathcal{A}^*/\{j\}$. Since $\mathcal{A}' \subset \mathcal{A}^*$, the interference space of the Tx-GMMSE weight over $\mathcal{A}'$ has one less dimension and hence, $\|\hat{\mathbf{H}}_k^H\mathbf{W}_k\|$ is larger for all $k \in \mathcal{A}'$. Moreover, the power allocation for $(\mathcal{A}', \mathcal{N}'$ is the same as that for $(\mathcal{A}^*, \mathcal{N}^*)$. Hence, $\mathcal{G}(\mu, \mathcal{A}', \mathcal{N}')$ is larger than $\mathcal{G}(\mu, \mathcal{A}^*, \mathcal{N}^*)$ and this contradicts the assumption that $(\mathcal{A}^*, \mathcal{N}^*)$ maximizes the goodput.

Since $\{p_k\}$ are all positive for $k \in \mathcal{A}^*$, we have the Lagrange multiplier $\mu < \gamma_k$ for all $k \in \mathcal{A}^*$ and is given by:

$$\frac{1-\epsilon}{\mu} = \frac{1}{G}\left(P_0 + \sum_{k\in\mathcal{A}^*}\frac{1}{\gamma_k}\right)$$

$$
\begin{aligned}
Var(I_B) &= \mathcal{E}[|I_B|^2] \leq \mathcal{E}\left[\sum_{i,j}(p_i p_j/n_i n_j) tr\left(\mu_{X_i}^H \widetilde{\mathbf{X}}_i + \widetilde{\mathbf{X}}_i^H \mu_{X_i}\right)\left(\mu_{X_j}^H \widetilde{\mathbf{X}}_j + \widetilde{\mathbf{X}}_j^H \mu_{X_j}\right)\right] \\
&= \sum_{i,j} \mathcal{E}_{\hat{\mathbf{H}}}\left[(p_i p_j/n_i n_j) tr\left(\mu_{X_j}\mu_{X_i}^H \mathcal{E}[\widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_j^H|\hat{\mathbf{H}}] + \mathcal{E}[\widetilde{X}_j \widetilde{X}_i^H|\hat{\mathbf{H}}]\mu_{X_i}\mu_{X_j}^H\right)\right] \\
&\leq \frac{2M_k\sigma_e^2}{N_T(1-\sigma_e^2)}\sum_{i,j}\sum_{m=1}^{M_k}\sum_{n=1}^{n_i}\sum_{n'=1}^{n_j}\mathcal{E}\left[(p_i p_j/n_i n_j)|\mu_{X_i}(n,m)\mu_{X_j}(n',m)||\mathbf{W}_i^H(n)\mathbf{W}_j(n')|\right] \\
&\underbrace{\leq}_{(b)} \mathcal{O}(1/N_T)\sum_{i,m,n}\sqrt{\mathcal{E}[p_i^4]\mathcal{E}[|\mu_{X_i}(n,m)|^4]} + \mathcal{O}(1/N_T)\sum_{i,j,m,n\neq n'}\sqrt{\mathcal{E}[p_i^2 p_j^2]\mathcal{E}[|\mu_{X_i}(n,m)\mu_{X_j}(n',m)|^2|\mathbf{W}_i^H(n)\mathbf{W}_j(n')|^2]} \\
&\underbrace{\leq}_{(c)} \mathcal{O}(1/N_T)\sum_{i,m,n}\mathcal{O}(1/N_T^3) + \mathcal{O}(1/N_T)\sum_{i,j,m,n\neq n'}\sqrt{\sqrt{\mathcal{E}[p_i^4]\mathcal{E}[p_j^4]}\sqrt{\sqrt{\mathcal{E}[|\mu_{X_i}(n,m)|^8]\mathcal{E}[\mu_{X_j}(n',m)|^8]}\mathcal{E}[|\mathbf{W}_i^H(n)\mathbf{W}_j(n')|^4]}} \\
&\underbrace{\leq}_{(d)} \mathcal{O}(1/N_T^3) + \mathcal{O}(1/N_T)\sum_{i,j,m,n\neq n'}\mathcal{O}(1/N_T^{3.5}) \leq \mathcal{O}(1/N_T^{2.5})
\end{aligned}
$$

$$
\begin{aligned}
Var(I_C) &= \sum_{i,j}\sum_{n=1}^{n_i}\sum_{n'=1}^{n_j}\mathcal{E}\left[p_i p_j \frac{\sigma_e^4}{N_T^2(1-\sigma_e^2)^2}\left|\mathbf{W}_i^H(n)\mathbf{W}_j(n')\right|^2/n_i n_j\right] \\
&\underbrace{\leq}_{(b)} \mathcal{O}(1/N_T^2)\sum_{i,j,n,n'}\sqrt{\mathcal{E}\left[\left|\mathbf{W}_i^H(n)\mathbf{W}_j(n')\right|^2\right]}\sqrt{\sqrt{\mathcal{E}[p_i^4]}\sqrt{\mathcal{E}[p_j^4]}} \underbrace{\leq}_{(c)} \mathcal{O}(1/N_T^3) + \mathcal{O}(1/N_T^4) \leq \mathcal{O}(1/N_T^3)
\end{aligned}
$$

where $G = |\mathcal{A}^*|$. Hence, the conditional goodput is given by:

$$
\mathcal{G}(\mathcal{A}^*, \mathcal{N}^*) = \epsilon \log_2\left(\prod_{k\in\mathcal{A}^*}\frac{\gamma_k}{|\mathcal{A}|^*}\left(P_0 + \sum_{k\in\mathcal{A}^*}\frac{1}{\gamma_k}\right)\right). \quad (28)
$$

Hence, $(\mathcal{A}^*, \mathcal{N}^*)$ maximizes $f(\mathcal{A}^*, \mathcal{N}^*)$.

For the second part, assume $(\mathcal{A}^*, \mathcal{N}^*)$ maximizes $f(\mathcal{A}^*, \mathcal{N}^*)$. If $\mu < \gamma_k$ for all $k \in \mathcal{A}^*$, we have $\{p_k\}$ are positive for all $k \in \mathcal{A}^*$. From (28), we have $(\mathcal{A}^*, \mathcal{N}^*)$ maximizes $\mathcal{G}(\mathcal{A}, \mathcal{N})$ as well.

References

[1] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Processing,* vol. 52, pp. 1179–1197, May 2004.

[2] T. Svantesson and B. D. Rao, "Capacity of spatio-temporally structured MIMO channels with estimation errors," in *Proc. ICASSP,* Philadelphia, PA, Mar. 2005, pp. 401–404.

[3] V. K. Lau, Y. Liu, and T. A. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. Commun.*, vol. 52, pp. 62–70, Jan. 2004.

[4] K. N. Lau and Y. Kwok, *Channel Adaptive Technologies and Crosslayer Designs for Wireless systems with Multiple Antennas*. Hoboken, NJ: John Wiley, Sons, Inc., 2006.

[5] L. Zheng and D. N. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.

[6] Q. H. Spencer, A. L. Swindlehurst, and M. Harrdt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 461–472, Feb. 2004.

[7] S. Sfar, L. Dai, and K. B. Letaief, "Optimal diversity-multiplexing tradeoff with group detection for MIMO systems," *IEEE Trans. Commun.*, vol. 53, no. 7, pp. 1178–1190, July 2005.

[8] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University, 2001.

[10] J. Zhang, E. Chong, and D. Tse, "Output MAI distributions of linear MMSE multiuser receivers in CDMA systems," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1456–1467, Mar. 2001.

[11] N. Jindal, "Finite rate feedback MIMO broadcast channels," *Information Theory and Applications (ITA) Inaugural Workshop (UCSD)*, 2006.

[12] P. Kyritsi, R. A. Valenzuela, and D. C. Cox, "Channel and capacity estimation errors," *IEEE Commun. Lett.*, vol. 6, no. 12, pp. 517–519, Dec. 2002.

[13] S. Sandhu and A. Paulraj, "Space-time block codes: a capacity perspective," *IEEE Commun. Lett.*, vol. 4, pp. 384–386, Dec. 2000.

[14] X.-B. Liang and X.-G. Xia, "On the nonexistence of rate-one generalized complex orthogonal designs," *IEEE Trans. Inform. Theory,* vol. 49, pp. 2984–2988, Nov. 2003.

[15] N. Sharma and C. B. Constantinos, "Full-rate full-diversity linear quasi-orthogonal space-time codes for any number of transmit antennas," *EURASIP J. Applied Signal Processing*, pp. 1246–1256, Aug, 2004.

[16] M. Sanchez Fernandez and A. Lozano, "Doppler sensitivity of link reciprocity in TDD MIMO systems," in *Proc. IEEE Globecom*, Nov, 2005.

Vincent Lau obtained B.Eng (Distinction 1st Hons) from the University of Hong Kong (1989-1992) and Ph.D. from the Cambridge University (1995-1997). He was with HK Telecom (PCCW) as system engineer from 1992-1995 and Bell Labs - Lucent Technologies as member of technical staff from 1997-2003. He joined the Department of ECE, Hong Kong University of Science and Technology (HKUST) as Associate Professor. His current research focus includes robust cross layer scheduling for MIMO/OFDM wireless systems with imperfect channel state information, communication theory with limited feedback as well as cross layer scheduling for users with heterogeneous delay requirements.