Asymptotic Analysis of SDMA Systems with Near-Orthogonal User Scheduling (NEOUS) under Imperfect CSIT

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

***Abstract*—In this paper, we focus on the asymptotic cross layer analysis of multi-antenna systems with transmit MMSE (Tx-MMSE) beamforming, *near orthogonal scheduling* and outdated CSIT. To capture the effect of the potential packet outage, we introduce the average *system goodput*, which measures the average b/s/Hz delivered to the mobiles successfully, as the system performance objective. We derive closed-form expressions for the optimal power and rate allocations as well as a low complexity *near orthogonal user scheduling* (NEOUS) algorithm to solve the cross-layer optimization problem. We derive the asymptotic order of growth in system goodput for general CSIT error variance σ^2 and found that for sufficiently large n_T (number of antennas at the base station) and K (number of users) where $K = g^{-1}(n_T)$ for some strictly increasing function $g(x) = o(x)$, the the system goodput grows in the order of $n_T \log[(1-\sigma^2)\log K]$ when $\sigma^2 < 1$. This is the same order of growth as the optimal order of growth in broadcast channels with perfect CSIT and hence, the NEOUS is *order-optimal*. On the other hand, we need exponentially larger K to compensate for the penalty in multiuser diversity gain due to CSIT errors.**

***Index Terms*—SDMA, Cross-Layer Analysis, Imperfect CSI**

I. Introduction

For multiuser multi-antenna systems, it is shown [1] that, by selecting a set of users with the best channel condition at each scheduling slot, the system spectral efficiency can be substantially improved due to the spatial multiplexing gain and multi-user selection diversity. The optimality of transmit zero-forcing beamforming (Tx-ZFBF) (as a result of cross-layer scheduling) has been shown asymptotically for large number of users in [2]. However, in all these works, the system performance is based on *ergodic capacity* and the channel state knowledge at the base station (CSIT) is assumed to be perfect. However, in practice, the CSIT can never be perfect due to either the CSIT estimation noise in Time Divison Duplexing (TDD) systems or the outdatedness of CSIT due to duplexing delay. When the CSIT is imperfect, there will be potential packet transmission error (packet outage) when the scheduled data rate exceeds the instantaneous mutual information. This happens even if powerful error correction coding is applied because with imperfect CSIT, the instantaneous mutual information is not known at the base station and appears as a random variable. In order to capture the penalty of potential packet errors, we shall define *system goodput*, which measures the average b/s/Hz *successfully* delivered to the mobiles, as our performance measure.

The cross-layer design with outdated CSIT is a relatively new topic. In [3], a multi-user downlink zero-forcing based scheduling is analyzed using limited feedback. In [4], an opportunistic scheduling approach is proposed with rate feedbacks from the mobiles. In [5], precoding of MISO system is studied under partial and analog feedbacks. Yet, in all these cases, due to the perfect (but partial) feedback[1] assumption, packet error is not an issue as long as the error correction code is sufficiently strong and hence, these works also considered ergodic capacity as the performance objective. As far as we are aware, the following are some open fundamental questions remained to be answered.

- Due to the CSIT errors, there will always be mutual interference between the spatial streams and it is not clear whether spatial multiplexing will do any good especially at high SNR.
- What is the asymptotic multi-user diversity gain for SDMA system when we have CSIT errors?
- How sensitive would the multi-user diversity gain in SDMA systems be with respect to the CSIT errors.

In this paper, we shall focus on the cross-layer design and asymptotic analysis in SDMA system with imperfect CSIT. We shall focus on transmit MMSE(Tx-MMSE) processing at the base station. Tx-MMSE has been investigated in [6] for multi-user systems. However, perfect CSIT is assumed and no user selection and rate allocation is allowed. In this paper, we shall formulate the cross-layer design with imperfect CSIT as a mixed combinatorial and real optimization problem. Using random matrix theory, we shall show that the conditional packet outage probability converges (in probability) to non-central chi-square cdf as n_T increases and closed form solutions for the rate and power adaptations (that maximize the system goodput) can be obtained. We propose a low complexity *near orthogonal user scheduling* (NEOUS) to solve the combinatorial optimization. We derive the asymptotic order of growth in system goodput for general CSIT error σ^2 and found that for sufficiently large n_T and $K = g^{-1}(n_T)$ for some strictly increasing function $g(x) = o(x)$, the system goodput grows in the order of $n_T \log[(1-\sigma^2)\log K]$ when

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

Manuscript received Mar 11, 2007; revised Sept 12, 2007.
This work is supported under RGC funding 615606.

[1] Partial feedback here refers to the limited feedback. Perfect feedback here refers to the assumption that there is no feedback errors or feedback delay in the limited feedback.

$\sigma^2 < 1$. This is the same order of growth as the optimal order of growth in broadcast channels with perfect CSIT and hence, the NEOUS is *order-optimal*. However, we need exponentially larger K to compensate for the penalty in multiuser diversity gain due to CSIT errors.

The rest of the paper is organized as follows. In Section II, we outline the system model and the Tx-MMSE processing. In Section III, we define *system goodput* and formulate the cross layer design as an optimization problem. In Section IV, we shall give closed-form solution for rate and power adaptation and the low-complexity *near orthogonal user scheduling*. In section V, we shall analyze the asymptotic system goodput for large number of users and large number of antennas. In Section VI, we conclude with a summary of results.

II. System Model

In this paper, we shall adopt the following convention. $\mathbf{X}$ denotes a matrix and $\mathbf{x}$ denotes a vector. $\mathbf{X}^\dagger$ denotes matrix transpose and $\mathbf{X}^H$ denotes matrix hermitian.

A. Channel Model and Outdated CSIT Model

We consider a downlink transmission in a multi-user system with a base station having n_T transmit antennas and K mobile terminals having one receive antenna. We are interested to study the case when $K > n_T$ so that cross layer scheduling becomes important. The channel fading between different users and different antenna are modeled as independent identically distributed (i.i.d.) complex Gaussian process with zero-mean and unit variance. We consider slow fading channels where the fading is *quasi-static* within a scheduling time slot. The signal received by a user k, y_k, can be written as:

$$y_k = \mathbf{h}_k^H \mathbf{x} + z_k, \quad k = 1, ..., K \tag{1}$$

where $\mathbf{x}$ is the $n_T \times 1$-dimension transmit symbol from the base station, $\mathbf{h}_k$ is the $n_T \times 1$-dimension channel fading matrix of the k-th user, z_k is the additive white Gaussian noise with zero-mean and unit variance.

We consider TDD systems where the CSIT is obtained by channel reciprocity from estimation of uplink pilots. Consider the case where the CSIT error is due to the estimation noise on the uplink pilot, the MMSE estimator of the CSIT $\widehat{\mathbf{h}}_k$ of user k at the base station is given by[7], [8], [9]:

$$\widehat{\mathbf{h}}_k = \mathbf{h}_k + \Delta\mathbf{h}_k, \tag{2}$$

where $\Delta\mathbf{h}_k = \frac{\sqrt{E_p}}{1+E_p}\mathbf{z}_k^{pilot} - \frac{1}{1+E_p}\mathbf{h}_k$ is the $n_T \times 1$ dimension CSIT error (or MMSE error), E_p is the uplink pilot SNR, $\mathbf{z}_k^{pilot}$ is the AWGN noise in the received samples of the uplink pilots (zero-mean unit variance). As a result, $\Delta\mathbf{h}_k$ is zero-mean complex Gaussian distributed with covariance $\sigma^2\mathbf{I}$ where $\sigma^2 = \frac{1}{1+E_p}$ is the CSIT error variance[2]. Furthermore, $\mathcal{E}[\Delta\mathbf{h}_k^H \widehat{\mathbf{h}}_k] = \mathbf{0}$ due to the orthogonality principle of MMSE. Hence, σ^2 is a parameter which represents the CSIT quality. When $\sigma^2 = 0$ (or $E_p \to \infty$), we have perfect CSIT. When $\sigma^2 = 1$ (or $E_p \to 0$), we have $\mathcal{E}[\widehat{\mathbf{h}}_k^H \widehat{\mathbf{h}}] = \mathbf{0}$ and this is equivalent to no CSIT.

In practice, a relatively strong downlink pilot channel is available from the base station because the downlink pilot can be shared among all the K users. Hence, the CSIR estimation error is insignificant relative to that of the CSIT and for simplicity, we shall assume the CSIR is known perfectly at mobile terminals.

[2] In fact, the same model in (2) can be used to describe the outdated CSIT due to delay τ [10]. In that case, the error variance $\sigma^2 = 1 - J_0^2(2\pi f_d \tau)$ where f_d is the Doppler spread and J_0 is the zeroth-order Bessel function of first kind.

B. Multi-antenna Base Station Processing

Since there are n_T spatial degrees of freedom in a base station with n_T transmit antennas, we consider full *spatial division multiplexing* (SDM) where n_T users are selected from the set of K users to transmit at each time slot. Define $\mathcal{A}$ to be the set of the n_T selected users for transmissions. For easy notation, we assume $\mathcal{A} = \{1, 2, .., n_T\}$ below as illustrations. Each information streams is encoded and modulated separately. The modulated symbols from the n_T streams are assigned with transmit power $\{p_1, .., p_{n_T}\}$ and the linear precoding weights $\{\mathbf{w}_1, .., \mathbf{w}_{n_T}\}$ where $p_k \geq 0$ is the average transmit power and $\mathbf{w}_k$ is the $n_T \times 1$ complex linear beamforming weight for the k-th user. As a result, the received signal at the k-th mobile after linear pre-processing at the base station is given by:

$$y_k = \mathbf{h}_k^H \sum_{i \in \mathcal{A}} \sqrt{p_i}\mathbf{w}_i u_i + z_k \tag{3}$$

where u_i is the encoded information symbol for the i-th user. Note that both p_i, $\mathbf{w}_i$ and $\mathcal{A}$ are functions of CSIT $\hat{\mathbf{h}}$ and power allocation is subject to the average power constraint $\sum_{i=1}^{K} p_i \leq P_0$.

Transmit MMSE (Tx-MMSE) Processing: In this paper, we shall focus on *Transmit MMSE* processing with imperfect CSIT for more robust performance against the CSIT errors. Let $\psi = \mathcal{E}[\|\hat{\mathbf{h}}_k\|^2]$ be the average norm of the CSIT. The Tx-MMSE weights $\{\mathbf{w}_k : k \in \mathcal{A}\}$ are selected to minimize the total *normalized mean-square error* J given by:

$$J(\mathbf{W}) = tr\left[\mathcal{E}[\left(\mathbf{y} - \sqrt{\psi}\Lambda\mathbf{u}\right)\left(\mathbf{y} - \sqrt{\psi}\Lambda\mathbf{u}\right)^H |\hat{\mathbf{H}}]\right] \tag{4}$$

where $\mathbf{y} = [y_1, ..., y_{n_T}]^\dagger$ and $\mathbf{U} = [u_1, ..., u_{n_T}]^\dagger$ are the $n_T \times 1$ vectors of received signals and encoded symbols for the selected users, $\mathbf{\Lambda}$ is the $n_T \times n_T$ diagonal matrix of the square-root of transmit powers $\sqrt{p_k}$. $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1 ... \hat{\mathbf{h}}_{n_T}]$ is the aggregate estimated CSIT matrix for all users. Hence, the optimal MMSE weights $\mathbf{w}_k$ can be obtained by standard optimization technique by:

$$\partial J / \partial \mathbf{w}_k = \left(\sum_j \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^H\right)\mathbf{w}_k - \sqrt{\psi}\hat{\mathbf{h}}_k + \lambda_k \mathbf{w}_k = \mathbf{0}$$

where λ_k is a Lagrandge multiplier for the constraint $\|\mathbf{w}_k\|^2 = 1$ and the MMSE weight is given by:

$$\mathbf{w}_k = \sqrt{\psi}\left(\sum_j \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^H + \lambda_k \mathbf{I}\right)^{-1} \hat{\mathbf{h}}_k$$
$$= \left(\sum_{j \in \mathcal{A}} \hat{\tilde{\mathbf{h}}}_j \hat{\tilde{\mathbf{h}}}_j^H + \overline{\lambda_k}\mathbf{I}\right)^{-1} \hat{\tilde{\mathbf{h}}}_k \quad \forall k \in \mathcal{A} \tag{5}$$

where $\hat{\tilde{\mathbf{h}}}_k = \hat{\mathbf{h}}/\sqrt{\psi}$ is the normalized CSIT so that $\mathcal{E}\left[\|\hat{\tilde{\mathbf{h}}}_k\|^2\right] = 1$ and $\overline{\lambda}_k$ is the Lagrandge multiplier for $\|\mathbf{w}\|^2 = 1$.

From Fig. 1, the adaptive parameters at the base station include the power allocation $\{p_k : k \in \mathcal{A}\}$, the rate allocation $\{r_k : k \in \mathcal{A}\}$ and the user selection $\mathcal{A}$. We shall formulate the cross-layer design as an optimization problem in the next section.

III. Problem Formulation

In this section, we shall first define an appropriate optimization objective, namely the *system goodput*, that measures the b/s/Hz successfully delivered to the mobiles. Afterwards, we shall elaborate on the optimization problem formulation.

A. Instantaneous Mutual Information and System Goodput

With outdated CSIT, the Tx-MMSE processing cannot completely eliminate multi-user interference and the received signal at k-th user is given by:

$$y_k = \underbrace{\sqrt{p_k}\mathbf{h}_k^H \mathbf{w}_k u_k}_{\text{Signal Term}} + \underbrace{\sum_{j \neq k, j \in \mathcal{A}} \sqrt{p_j}\mathbf{h}_k^H \mathbf{w}_j u_j}_{\text{Multi-user Interference Term}_{v_k}} + \underbrace{z_k}_{Noise} \tag{6}$$

where $\mathbf{h}_k$ is the actual CSI, and v_k denotes the residual multiuser interference after Tx-MMSE processing. Note that the receiver at the k-th mobile has perfect CSIR, $\mathbf{h}_k$. However, the base station only has knowledge of the imperfect CSIT, $\hat{\mathbf{h}}_k$.

Given the CSIR $\mathbf{h}_k$ at the $k-$th mobile, the instantaneous mutual information can be expressed as:

$$C_k(\mathbf{h}_k, \Delta\mathbf{h}_k) = \log_2\left(1 + \frac{p_k|\mathbf{h}_k^H \mathbf{w}_k|^2}{1 + \sum_{j \neq k, j \in \mathcal{A}} p_j |\hat{\mathbf{h}}_k^H \mathbf{w}_j|^2}\right) \tag{7}$$

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 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 and 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 $C_k(\mathbf{h}_k)$ is unknown to the base station and the packet will be corrupted whenever the scheduled data rate r_k exceeds $C_k(\mathbf{h}_k)$. 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, for simplicity, 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).

We first define the instantaneous goodput of a packet transmission for user k as

$$\rho_k = r_k \mathbf{1}(r_k \leq C_k) \tag{8}$$

where $1(.)$ is an indicator function which is 1 when the event is true and 0 otherwise. The *average total 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:

$$U_{thp}(\mathcal{A}, \mathcal{P}, \mathcal{R}) = \mathcal{E}[\sum_{k=1}^K \rho_k]$$
$$= \mathcal{E}_{\hat{\mathbf{H}}}\left\{\sum_{k=1}^K r_k \underbrace{\Pr[r_k \leq C_k|\hat{\mathbf{H}}]}_{\text{Conditional outage probability}_{P_{out}}}\right\} \tag{9}$$

where $\mathcal{R} = \{(r_1, ..., r_K)[\widehat{\mathbf{H}}] \in \Re_+^K : \forall \widehat{\mathbf{H}} \in \mathcal{C}^{K \times n_T}\}$ is the rate allocation policy (which is the set of "actions" for all possible CSIT realizations). Similarly, $\mathcal{P} = \{(p_1, .., p_K)[\widehat{\mathbf{H}}] \in \Re_+^K : \forall \widehat{\mathbf{H}} \in \mathcal{C}^{K \times n_T}\}$ is the power allocation policy and $\Omega = \{\mathcal{A}[\widehat{\mathbf{H}}] : \forall \widehat{\mathbf{H}} \in \mathcal{C}^{K \times n_T}\}$ is the user selection policy.

B. Cross-Layer Design Optimization

The cross-layer design is to select the optimal power, rate and user selection policies to maximize the total average system goodput $U_{thp}(\Omega, \mathcal{R}, \mathcal{P})$ at a target FER probability ϵ. This is summarized by the following.

Problem 1 (Cross-Layer Optimization Problem)**.** *The optimal power allocation policy $\mathcal{P}$, rate allocation policy $\mathcal{R}$ and user selection policy Ω are given by:*

$$(\mathcal{P}^*, \mathcal{R}^*, \Omega^*) = \arg\max_{\mathcal{P}, \mathcal{R}, \Omega} U_{thp}(\Omega, \mathcal{R}, \mathcal{P})$$

such that $P_{out} = \Pr\left[r_k > C_k|\hat{\mathbf{H}}\right] = \epsilon$.

From (9), $U_{thp}(\mathcal{A}, \mathcal{P}, \mathcal{R}) = \mathcal{E}_{\hat{\mathbf{H}}}\left\{\sum_{k=1}^K r_k \Pr[r_k \leq C_k|\hat{\mathbf{H}}]\right\}$ and hence, optimization w.r.t. policies (set of actions for all CSIT realizations) is equivalent to optimization w.r.t. the actions for a given CSIT realization. As a result, Problem 1 is equivalent to

$$(\mathbf{p}^*, \mathbf{r}^*, \mathcal{A}^*) = \arg\max_{\mathbf{p}, \mathbf{r}, \mathcal{A}} \sum_{k=1}^K r_k \Pr[r_k \leq C_k|\hat{\mathbf{H}}] \tag{10}$$

such that $P_{out} = \Pr\left[r_k > C_k|\hat{\mathbf{H}}\right] = \epsilon$ where $\mathbf{p} = \{p_1, ..., p_K\}$ and $\mathbf{r} = \{r_1, ..., r_K\}$. Strictly speaking, the variables $\mathcal{A}$ and $\mathbf{p}$ are redundant because $p_k = 0 \rightarrow k \notin \mathcal{A}$ and hence, the optimization problem can be formulated by optimizing w.r.t. $\mathbf{p}$ and $\mathbf{r}$ only. However, in this case, the solution will be very complicated and no closed-form solution is possible because the MMSE weights will be coupled with the power action $\{bfp\}$. In order to obtain closed form solutions, we introduce a redundant variable $\mathcal{A}$ so that the problem is transformed into a mixed combinatorial and real optimization. Given $\mathcal{A}$, the MMSE weights can be determined and this allows closed-form solutions for the power and rate. After solving for power and rate, we still have to optimize w.r.t. $\mathcal{A}$ which involves combinatorial search. The solution will be elaborated in the following section.

IV. Solutions of the Optimization Problem

The optimization in (10) is a mixed real $(\mathbf{p}, \mathbf{r})$ and combinatorial ($\mathcal{A}$) optimization. We shall solve it in two steps. In step 1, we fix a given user selection $\mathcal{A}$ and obtain optimal power and rate allocations for the given $\mathcal{A}$. In step 2, we shall perform combinatorial search on $\mathcal{A}$. For any given admitted user set $\mathcal{A}$, the Lagrangian function of the optimization problem 1 is given by:

$$\begin{aligned}L(\{p_k\}, \{r_k\}; \beta, \xi, \{\kappa_k\}, \{\alpha_k\}) &= \sum_k r_k \Pr[r_k \le C_k|\hat{\mathbf{H}}] \\ &-\beta\left(\Pr[r_k \le C_k|\hat{\mathbf{H}}] - \epsilon\right) - \xi\left(\sum_k p_k - P_0\right) \\ &- \sum_k \kappa_k r_k - \sum_k \alpha_k p_k \end{aligned} \quad (11)$$

where β, ξ, $\kappa_k \ge 0$ and $\alpha_k \ge 0$ are the Lagrangian multipliers corresponding to the constraints $P_{out} = \epsilon$, $\sum_k p_k = P_0$, $r_k \ge 0$ and $p_k \ge 0$ respectively.

A. Conditional Packet Outage Probability

One obstacle in solving for $\{p_k\}$ and $\{r_k\}$ is that we need to have closed form expression for the conditional outage probability $P_{out}(r_k, p_k, \mathcal{A}, \hat{\mathbf{H}})$. The conditional outage probability can be expressed as $P_{out}(r_k, p_k, \mathcal{A}, \hat{\mathbf{H}}) = \Pr\left[S_k < \Lambda_k/\psi|\hat{\mathbf{H}}\right]$ where $\Lambda_k = (2^{r_k} - 1)$, and S_k is the normalized random variable given by:

$$S_k = p_k|\widetilde{\mathbf{h}}_k^H \mathbf{w}_k|^2 - \Lambda_k \sum_{j\neq k} p_j|\widetilde{\mathbf{h}_k^H}\mathbf{w}_j|^2 \quad (12)$$

where $\widetilde{\mathbf{h}}_k = \mathbf{h}/\sqrt{\psi}$ is the normalized CSI. Conditioned on the CSIT $\hat{\mathbf{H}}$, the CSI $\widetilde{\mathbf{h}}_k$ is a Gaussian random vector with mean $\hat{\widetilde{\mathbf{h}}}_k$ and covariance $(\sigma^2/\psi)\mathbf{I}$. Hence, S_k is a indefinite quadratic form of Gaussian random variables. The cdf of the indefinite quadratic form is very tedious and is virtually impossible to optimize on the expression. Yet, using random matrix theory, we show that S_k converges to non-central chi-square random variable in probability as n_T increases. The result is summarized in the following Theorem.

Theorem 1 (Conditional Convergence of Packet Outage Probability)**.** *If the power allocation policy $\mathcal{P}$ is* regular, *then for almost all*[3] *given user selection $\mathcal{A}$ with CSIT $\hat{\mathbf{H}}$, the conditional outage probability P_{out} converges in probability to non-central chi-square cdf:*

$$P_{out}(r_k, p_k, \mathcal{A}, \hat{\mathbf{H}}) = F_{\chi^2(s^2,\sigma^2)}\left(\frac{\Lambda_k(\mu_I(\hat{\mathbf{H}}) + 1)}{p_k}\right). \quad (13)$$

where $F_{\chi^2(s^2,\sigma^2)}(y)$ is the c.d.f. of non-central chi-square distribution with 2 degrees of freedom (non-central parameter $s^2 = |\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2$ and variance σ^2) and $\mu_I(\hat{\mathbf{H}}) = P_0\sigma^2 + \sum_{j\neq k} p_j|\hat{\mathbf{h}}_k^H\mathbf{w}_j|^2$ is a constant.

Please refer to appendix for the proof. Figure 4 illustrates the conditional average outage probability for user 1 versus number of antennas n_T and SNR = 10dB. Observe that the simulated packet outage probability matches the approximated packet outage probability using non-central chi-square distribution as in (13) very well for moderate number of antennas. Hence, we can approximate the P_{out} by its asymptotic distribution (non-central chi-square) using Theorem 1.

B. Closed-form Solutions for Power and Rate Allocations

In this section, we shall derive simple closed-form solution for power and rate allocations. From the constraint $P_{out} = \epsilon$ in Prob. 1, we have:

$$\begin{aligned}P_{out} = \epsilon &\Longleftrightarrow \Lambda_k(\mu_I(\hat{\mathbf{H}}) + 1)/p_k = \varphi_k(\epsilon) \\ &\Longleftrightarrow \Lambda_k = p_k\varphi_k(\epsilon)/(\mu_I(\hat{\mathbf{H}}) + 1)\end{aligned} \quad (14)$$

where $\varphi_k(\epsilon) = F^{-1}_{\chi^2(s^2,\sigma^2)}(\epsilon)$. Based on NEOUS algorithm (elaborated in the next section), the selected CSIT will have CSIT square-norms (ψ) bounded by $n_T(1-\sigma^2)(B_-, B_+)$ and $|\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_j| \le \|\hat{\mathbf{h}}_i\|\|\hat{\mathbf{h}}_j\|\theta$. Hence, we have

$$|\hat{\mathbf{h}}_k^H\mathbf{w}_j|^2 = \left|\hat{\mathbf{h}}_k^H\mathbf{U}^H\left(\Lambda + \overline{\lambda_j}\mathbf{I}\right)^{-1}\mathbf{U}\hat{\mathbf{h}}_j\right|^2/\psi \le \frac{|\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_j|^2}{\psi\overline{\lambda}_j^2} \le \psi\theta^2/\overline{\lambda_j}^2 \quad (15)$$

where $\sum_{i\in\mathcal{A}} \hat{\widetilde{\mathbf{h}}}_i\hat{\widetilde{\mathbf{h}}}_i^H = \mathbf{U}\Lambda\mathbf{U}^H$ and therefore, $\mu_I(\hat{\mathbf{H}}) \le \sigma^2P_0 + P_0\theta^2\psi/\widetilde{\lambda}^2$ where $\widetilde{\lambda} = \min_{j\in\mathcal{A}}\overline{\lambda}_j$ is a constant. From (14), the constraint $P_{out} = \epsilon$ in Problem 1 is satisfied if and only if $r_k = \log_2(1 + p_k\gamma_k)$ where

$$\gamma_k = \varphi_k(\epsilon)/(\mu_I(\hat{\mathbf{H}}) + 1) \ge \widetilde{\gamma}_k = \frac{\varphi_k(\epsilon)}{1 + P_0\sigma^2 + P_0\theta^2\psi/\widetilde{\lambda}^2} \quad (16)$$

is the average SINR of user k (per unit received power). Hence, we shall optimize the lower bound of the conditional goodput based on the lower bound of SINR in (16). In the next section, we shall illustrate that average system goodput achieved by the power and rate allocations obtained, together with the NEOUS scheduling, has the same order of growth as in the optimal dirty paper coding (DPC) system and hence, the proposed solution is in fact *order-optimal*.

[3]"Almost all" here refers to with probability 1.

Substituting the constraint $P_{out} = \epsilon$ into the Lagrangian function in (11), Lagrangian function becomes:

$$L(\{p_k\}, \{r_k\}; \beta, \xi, \{\kappa_k\}, \{\alpha_k\}) = (1-\epsilon)\sum_k r_k$$
$$-\beta\left(r_k - \log_2(1 + p_k\widetilde{\gamma_k})\right) - \xi\left(\sum_k p_k - P_0\right)$$
$$-\sum_k \kappa_k \nu_k - \sum_k \alpha_k p_k.$$

The optimizing rate and power allocation solution (r_k^*, p_k^*) can be obtained using standard optimization techniques and they are given by:

$$p_k^* = \left(\frac{1-\epsilon}{\xi} - \frac{1}{\widetilde{\gamma_k}}\right)^+ \text{ and } r_k^* = \left[\log_2\left(\frac{(1-\epsilon)\widetilde{\gamma_k}}{\xi}\right)\right]^+ \quad (17)$$

where $\xi > 0$ is the Lagrange multiplier chosen to satisfy the constraint $\sum_k p_k = P_0$.

C. Near-Orthogonal User Scheduling (NEOUS) for $\mathcal{A}$

After we have solved for the power and rate allocations, the remaining variable in the optimization is the user set $\mathcal{A}$ which is combinatorial. The optimal solution for $\mathcal{A}$ involves exhaustive search over all possible combinations and the complexity is exponential in K and is not feasible for moderate K and n_T. In this section, we propose a *near-orthogonal user scheduling* (NEOUS) algorithm which is of much lower complexity and can be shown to be asymptotically optimal for large K. To reduce the search complexity, we shall limit the search to a smaller set of users while ensuring that a user set that will be found (from the restricted search) is close to optimal with high probability. We first have the following definition.

Definition 1 (Near Orthogonal Set)**.** *A* near orthogonal set $\mathcal{S}_{\theta,n_T}$ *is defined to be the collection of user sets* $\mathcal{A}$ *such that users in* $\mathcal{A}$ *are* near orthogonal *from each other. That is:*

$$\mathcal{S}_{\theta,n_T} = \Big\{\mathcal{A} \subset \{1,2,..,K\} : |\hat{\mathbf{h}}_i^H\hat{\mathbf{h}}_j| \le \|\hat{\mathbf{h}}_i\|\|\hat{\mathbf{h}}_j\|\theta$$
$$\textit{and } n_T(1-\sigma^2)B_- \le \|\hat{\mathbf{h}}_i\|^2 \le n_T(1-\sigma^2)B_+,$$
$$\forall i \ne j \in \mathcal{A} \textit{ and } |\mathcal{A}| = n_T\Big\} \quad (18)$$

Note that the near orthogonal set $\mathcal{S}_{\theta,n_T}$ is parameterized by (θ, B_-, B_+). As $\theta \to 0$, the CSITs in any $\mathcal{A} \in \mathcal{S}_{\theta,n_T}$ are increasingly orthogonal. In addition, the CSITs in $\mathcal{A} \in \mathcal{S}_{\theta,n_T}$ will have norms bounded between $n_T(1-\sigma^2)(B_-, B_+)$. Hence, instead of doing exhaustive search for $\mathcal{A}$ over all possible combinations, we can simply pick any $\mathcal{A} \in \mathcal{S}_{\theta,n_T}$ because from the definition of $\mathcal{S}_{\theta,n_T}$, any member of $\mathcal{S}_{\theta,n_T}$ will be "good candidate". Intuitively, near orthogonal vectors allow Tx-MMSE beamforming to perform well[4].

[4]Alternatively, one can further improve the performance by searching the best member in $\mathcal{S}_{\theta,n_T}$. The extra complexity will be smaller than the original exhaustive search because $\mathcal{S}_{\theta,n_T}$ is of much smaller cardinality (compared with the set of all subsets of $\{1,2,..,K\}$ in exhaustive search). Yet, we shall illustrate that both ways (with or without further search in $\mathcal{S}_{\theta,n_T}$) are *order optimal* w.r.t. cross-layer gain.

Let $P_s = \Pr[|\mathcal{S}_{\theta,n_T}| > 0]$ be the probability that the *near orthogonal set* $\mathcal{S}_{\theta,n_T}$ is non-empty. In order for the NEOUS algorithm to work, we have to make sure there is high probability that the near orthogonal set $\mathcal{S}_{\theta,n_T}$ is non-empty (i.e. $P_s \to 1$). As K increases, we would like to find out how fast (B_-, B_+) can scale with respect to K so that P_s is still close to 1. In fact, using similar approach as in [11], we can show that if (B_-, B_+) increases at most on the order of $\log K$, then asymptotically as $K \to \infty$, we still have $P_s \to 1$. Specifically, the result is summarized below.

Asymptotic Existence Probability of $\mathcal{S}_{\theta,n_T}$: For any given $\delta > 0$, if $B_- = \Theta(\log K)$, $B_+ = \Theta(\log K)$, $\theta = \Theta(1/\sqrt{n_T})$ and $K = g^{-1}(n_T)$ for some strictly increasing function $g(x) = o(x)$, then there exists some constant $K_0(\delta,\theta) > 0$ such that $P_s = \Pr[|\mathcal{S}_{\theta,n_T}| > 0] > 1-\delta$ for all $K > K_0(\delta,\theta,n_T)$. Note that $g(x) = o(x)$ is the *small o* notation meaning that $\lim_{x\to\infty}\frac{g(x)}{x} = 0$ ($g(x)$ is asymptotically smaller than x). For example, $g(x) = \sqrt{x}$ ($K = n_T^2$) or $g(x) = \log(x)$ ($K = exp(n_T)$) are possible choices for $g(x)$.

D. Summary of the NEOUS Scheduling and Power/Rate Allocation Solution

Figure 2 illustrates the top level flow chart of the proposed cross-layer scheduling solution for multi-antenna system with outdated CSIT. This is elaborated in the following steps.

- **Step 1:** For a given set of CSIT $\{\hat{\mathbf{h}}_1, ..., \hat{\mathbf{h}}_K\}$, Initialize $\mathcal{A} = \emptyset$ and $k = 1$.
- **Step 2:** If $n_T(1-\sigma^2)B_- \le \|\hat{\mathbf{h}}_k\|^2 \le n_T(1-\sigma^2)B_+$, then $\mathcal{A} \to \mathcal{A}\bigcup\{k\}$ and go to step 3. Otherwise, $k \to k+1$ and goto step 2.
- **Step 3:** For $k \notin \mathcal{A}$, if $|\mathcal{A}| < n_T$, $n_T(1-\sigma^2)B_- \le \|\hat{\mathbf{h}}_k\|^2 \le n_T(1-\sigma^2)B_+$ and $|\hat{\mathbf{h}}_k^H\hat{\mathbf{h}}_j| \le \|\hat{\mathbf{h}}_k\|\|\hat{\mathbf{h}}_j\|\theta$ for all $j \in \mathcal{A}$, then $\mathcal{A} \to \mathcal{A}\bigcup\{k\}$ and goto step 3. Otherwise, if $|\mathcal{A}| = n_T$, the goto step 4. Else, randomly form a $\mathcal{A}$ with $|\mathcal{A}| = n_T$ and goto step 4.
- **Step 4:** For the given $\mathcal{A}$, calculate the Tx-MMSE weights according to (5). Calculate the rate and power allocation using (17). The algorithm is completed.

V. Asymptotic Analysis on the System Goodput and Cross-Layer Gains

In this section, we shall obtain a lower bound on the average system goodput of the multi-antenna system based on Tx-MMSE beamforming, the power and rate allocations in (17) as well as the NEOUS scheduling for asymptotically large K and n_T. To simplify the notation, we shall assume that $K = g^{-1}(n_T)$ for some strictly increasing function[5] $g(x) = o(x)$. Hence, we shall consider limit on n_T only with the understanding that as n_T scales, K also scales to infinity (at a faster rate than n_T). We shall show that the lower bound of the average system goodput grows at the same rate as that of the optimal DPC performance and hence, show that the proposed cross-layer solution is in fact *order optimal*.

Due to the NEOUS scheduling, there is probability one that the CSIT of the selected user set $\mathcal{A}$ have norms

[5]For example, $g(x) = x^\alpha$ for $0 < \alpha < 1$ and $g(x) = \log(x)$ are possible choice of $g(x)$.

$\|\hat{\mathbf{h}}_k\|^2 \geq n_T(1-\sigma^2)B_- = \Theta(n_T(1-\sigma^2)\log K)$. Normalizing the CSIT $\hat{\mathbf{h}}_k$ with $\sqrt{n_T(1-\sigma^2)B_-}$ and define $\hat{\mathbf{g}}_k = \hat{\mathbf{h}}_k/\sqrt{n_T(1-\sigma^2)B_-}$, the non-central parameter $s = |\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2$ can be expressed as:

$$\begin{aligned} s = |\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2 &\geq n_T(1-\sigma^2)B_- \left| \hat{\mathbf{g}}_k^H \left(\sum_{i=1}^{n_T} \hat{\mathbf{g}}_i \hat{\mathbf{g}}_i^H + \overline{\lambda_k}\mathbf{I} \right)^{-1} \hat{\mathbf{g}}_k \right|^2 \\ &= n_T(1-\sigma^2)B_- \left| \hat{\mathbf{g}}_k^H \left(\mathbf{M}_k + \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H \right)^{-1} \hat{\mathbf{g}}_k \right|^2 \\ &= n_T(1-\sigma^2)B_- \frac{|\hat{\mathbf{g}}_k^H \mathbf{M}_k^{-1} \hat{\mathbf{g}}_k|^2}{(1+|\hat{\mathbf{g}}_k^H \mathbf{M}_k^{-1}\hat{\mathbf{g}}_k|)^2} \end{aligned} \quad (19)$$

where $\mathbf{M}_k = \left(\sum_{i\neq k} \hat{\mathbf{g}}_i \hat{\mathbf{g}}_i^H + \overline{\lambda_k}\mathbf{I}\right)$. From [12], we have $|\hat{\mathbf{g}}_k^H \mathbf{M}_k^{-1} \hat{\mathbf{g}}_k|$ converges almost surely to a constant $b_g = \int_0^\infty \left[\frac{1}{\omega + \overline{\lambda}}\right] dG^*(\omega)$ as $n_T \to \infty$ where ω is the eigenvalue of $\sum_{i\neq k} \hat{\mathbf{g}}_i \hat{\mathbf{g}}_i^H$ and $G^*(\omega)$ is the limiting empirical distribution of ω (whose Stieljas transform is given by (29)). Hence, we have almost surely, the non-central parameter s^2 is lower bounded by:

$$s^2 \geq n_T(1-\sigma^2)B_- \frac{b_g^2}{(1+b_g)^2} = \Theta(n_T(1-\sigma^2)\log K) \quad (20)$$

Hence, the *scaling factor* $\varphi_k(\epsilon)$ in (17) is given by:

$$\varphi_k(\epsilon) = \begin{cases} \Theta(s^2) = \Theta(n_T(1-\sigma^2)\log K) & \text{if } 0 \leq \sigma^2 < 1, \\ \Theta(F^{-1}_{\chi^2(0,\sigma^2)}(\epsilon)) & \text{if } \sigma^2 \approx 1. \end{cases} \quad (21)$$

for small ϵ and large n_T (and hence, large K as $K = g^{-1}(n_T)$) where $F^{-1}_{\chi^2(0,\sigma^2)}(\epsilon)$ is the inverse cdf of the central chi-square distribution. Figure 3 verifies the order of growth of $\varphi_k(\epsilon)$ with respect to the non-central parameter s^2 in (21).

On the other hand, for large n_T, $\overline{\lambda}$ converges to the solution of the fixed point equation in (30) which is of constant order. Finally, from Section IV-C, we have $\theta = \Theta(1/\sqrt{n_T})$. As a result, the lower bound of the SINR $\widetilde{\gamma}_k$ is of the order $\Theta\left(\frac{\varphi_k(\epsilon)}{1+P_0\sigma^2 c_1 + c_2 P_0(1-\sigma^2)}\right)$ for some constants c_1 and c_2. Substituting the power and rate allocation solutions in (17) into the system goodput, we have

$$\begin{aligned} U_{thp} &\geq \mathcal{E}_{\hat{\mathbf{H}}}\left[\sum_{k\in\mathcal{A}} r_k(1-\epsilon)\right] = (1-\epsilon)\mathcal{E}_{\hat{\mathbf{H}}}\left[\sum_{k\in\mathcal{A}}\left[\log_2\left(\frac{(1-\epsilon)\widetilde{\gamma}_k}{\xi}\right)\right]^+\right] \\ &\underbrace{=}_{(a)} (1-\epsilon)\mathcal{E}_{\hat{\mathbf{H}}}\left[\sum_{k\in\mathcal{A}}\Theta\left(\log_2\left(1+\frac{P_0\varphi_k(\epsilon)}{n_T(1+\sigma^2 P_0 c_1 + c_2 P_0(1-\sigma^2))}\right)\right)\right] \\ &= (1-\epsilon)n_T\Theta\left(\log_2\left(1+\frac{P_0\overline{\varphi(\epsilon,n_T,K,\sigma^2)}}{n_T(1+\sigma^2 P_0 c_1 + c_2 P_0(1-\sigma^2))}\right)\right) \end{aligned} \quad (22)$$

where (a) is due to the fact that from (17) and the constraint $\sum_{k\in\mathcal{A}} p_k^* = P_0$, we have $(1-\epsilon)/\xi = \Theta(\frac{P_0}{n_T} + \frac{1}{n_T}\sum_k \frac{1}{\widetilde{\gamma}_k})$ and $\overline{\varphi(\epsilon, n_T, K, \sigma^2)} = F^{-1}_{\chi^2(s^2 = n_T(1-\sigma^2)\log K, \sigma^2)}(\epsilon)$.

A. Numerical Results and Discussions

- **Multi-user Diversity Gain:** In general, for $\sigma^2 < 1$, large n_T (and hence, large K), large P_0 and small target packet outage probability ϵ, the multi-user diversity gain on the average system goodput of our proposed cross layer design is of the order $\Theta(n_T \log[\overline{\varphi(\epsilon, n_T, K, \sigma^2)}/n_T]) = \Theta(n_T\log[(1-\sigma^2)\log K])$. This demonstrates that with proper cross-layer design, spatial multiplexing gain n_T is still important for the total system goodput even when there is CSIT error. Figure 5 illustrates the average system goodput versus number of users (K) for $n_T = 10$, SNR=10dB and CSIT errors of $\sigma^2 = 0.01, 0.1, 0.5$. We see that the order of growth with respect to K matched closely with the asymptotic expression in (22) $\log\left((1-\sigma^2)\log K\right)$ even for large CSIT errors $\sigma^2 = 0.5$. On the other hand, the effect of CSIT errors on the system goodput is that exponentially larger K is needed to compensate for the penalty on system goodput due to CSIT errors.
- **Performance at Large CSIT Errors:** When the CSIT error $\sigma^2 \to 1$, the factor $\overline{\varphi(\epsilon, n_T, K, \sigma^2)} \to c_3 = \Theta(F^{-1}_{\chi^2(0,\sigma^2)}(\epsilon))$ which is a constant independent of n_T and K. Hence, the average system goodput in (22) approaches $(1-\epsilon)n_T\Theta\left(\log_2\left(1+\frac{c_3}{n_T(\sigma^2 c_1)}\right)\right)$ for large P_0 and hence, the multi-user diversity gain essentially vanished. The average goodput scales at the order $\Theta(n_T\log(1+c/n_T))$ for some constant c. Hence, when we have large CSIT errors, spatial multiplexing does not offer much gain in system goodput (as illustrated in figure 6) due to mutual interference between the spatial channels.
- **Performance at Small CSIT Errors:** When the CSIT error $\sigma^2 \approx 0$, the factor $\overline{\varphi(\epsilon, n_T, K, \sigma^2)} \to \Theta(n_T \log K)$ and the average system goodput in (22) approaches $(1-\epsilon)n_T\Theta\left(\log_2\left(1+\frac{\log K}{c_2}\right)\right)$ for large P_0. Hence, for small CSIT error σ^2, the optimal order of growth of the multiuser diversity gain (with respect to K) in system goodput ($\log\log K$) can be maintained using the proposed cross-layer design. This is the same order of growth with respect to K as the DPC processing with perfect CSIT. Furthermore, the system goodput grows linearly as n_T (as illustrated in figure 6) and this indicates that spatial multiplexing is effective to increase the system goodput for large K even with CSIT errors.
- **Comparison of MMSE versus other baseline references** Figure 7 illustrate the benchmarking results of the proposed robust cross-layer MMSE based scheduler. We compared the system goodput versus K of the proposed scheduler against the cross-layer ZF-based scheduler, TDMA-based scheduler (selecting one user at a time) as well as the opportunistic scheduler[4]. We showed that the proposed MMSE-based scheduler out-performs the others in both small and large CSIT errors ($\sigma^2 = 0.02, 0.1$). Furthermore, the sensitivity of the proposed MMSE-based solution (w.r.t. CSIT errors) is significantly less than that of the ZF-based scheduler.

VI. Conclusion

In this paper, we propose a cross-layer design with *near-orthogonal user selection* and dynamic power and rate allocation with Tx-MMSE processing for multi-antenna systems with outdated CSIT. The CSIT error is incorporated in the cross-layer design to maximize the system goodput. Using random matrix theory, we derived closed form expressions for the optimal power and rate allocations as well as the asymptotic system goodput. We found that asymptotically for sufficiently large n_T and $K = g^{-1}(n_T)$ for some strictly increasing function $g(x) = o(x)$, the multiuser diversity gain on the system goodput grows at the order of $n_T \log[(1-\sigma^2)\log K]$ (where σ^2 is the CSIT error variance) when $\sigma^2 < 1$. Hence, the proposed cross-layer design is *order-optimal* w.r.t. K despite the CSIT errors. On the other hand, when $\sigma^2 \approx 1$, the multi-user diversity gain vanished and spatial multiplexing cannot give any benefit due to spatial interference.

References

[1] D. Gesbert, M. Shafi, D. Shiu, P. Smith, and A. Nagiub, “From theory to practice:an overview of mimo space-time coded wireless systems,” *IEEE Journal of Selected Areas on Communications*, vol. 21, pp. 281–302, April 2003.

[2] T. Yoo and A. Goldsmith, “On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming,” *IEEE JSAC Special Issue on 4G Wireless Systems, 2006*, Mar. 2006.

[3] N. Jindal, “Finite rate feedback mimo broadcast channels,” *Information Theory and Applications (ITA), inaugural workshop, UCSD*, 2006.

[4] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Trans. on Info. Theory*, vol. 48, pp. 1277–1294, Jun. 2002.

[5] D. L. Ruyet and B. Ozbek, “Partial and analog feedback for miso precoding systems,” *IEEE Proceedings on International Conference on Communications (ICC)*, June 2007.

[6] R. Bendlin, M. Joham, J. A. Nossek, and Y.-F. Huang, “Iterative mmse transmit and receive filter design for frequency selective mu-miso systems,” *IEEE Proceedings on International Conference on Communications (ICC)*, pp. 281–302, June 2007.

[7] P. Kyritsi, R. Valenzuela, and D. Cox, “Channel and capacity estimation errors,” *IEEE Comm. Letters*, vol. 6, pp. 517–519, Dec. 2002.

[8] G. Caire, “Mimo downlink joint processing and scheduling: a survey of classical and recent results,” *Information Theory and Applications (ITA), inaugural workshop, UCSD*, 2006.

[9] S. Haykin, *Adaptive Filter Theory*. Prentice Hall International Editions, third ed., 1998.

[10] A. Liavas, “Tomlinson-harashima precoding with partial channel knowledge,” *IEEE Trans. on Communications*, vol. 53, pp. 5–9, Jan. 2005.

[11] C. Swannack, E. Uysal-Biyikoglu, and G. Wornell, “Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel,” *Proc. Allerton Conf. Communications, Control and Computing, 42nd*, 2004.

[12] J. Zhang, E. Chong, and D. Tse, “Output mai distributions of linear mmse multiuser receivers in cdma systems,” *IEEE Trans. on Infom. Theory*, vol. 47, pp. 1128–1144, Mar. 2001.

[13] J. Silverstein, “Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices,” *J. Multivariate Anal*, vol. 55, pp. 331–339, 1995.

Appendix A
Proof of Lemma 1

From (5), the Tx-MMSE weight $\mathbf{w}_i$ is given by

$$\begin{aligned}
\mathbf{w}_i &= \left(\sum_j \hat{\tilde{\mathbf{h}}}_j \hat{\tilde{\mathbf{h}}}_j^H + \widetilde{\lambda}\mathbf{I}\right)^{-1} \hat{\tilde{\mathbf{h}}}_i \\
&= \left(\underbrace{\left(\sum_{j\neq i} \hat{\tilde{\mathbf{h}}}_j \hat{\tilde{\mathbf{h}}}_j^H + \overline{\lambda_i}\mathbf{I}\right)}_{\mathbf{M}_i} + \hat{\tilde{\mathbf{h}}}_i \hat{\tilde{\mathbf{h}}}_i^H\right)^{-1} \hat{\tilde{\mathbf{h}}}_i^H \\
&\underbrace{=}_{(a)} \left(\mathbf{M}_i^{-1} - \frac{\mathbf{M}_i^{-1}\hat{\tilde{\mathbf{h}}}_i \hat{\tilde{\mathbf{h}}}_i^H \mathbf{M}_i^{-1}}{1 + \hat{\tilde{\mathbf{h}}}_i^H \mathbf{M}_i^{-1}\hat{\tilde{\mathbf{h}}}_i}\right)\hat{\tilde{\mathbf{h}}}_i
\end{aligned} \tag{23}$$

where (a) is by matrix inversion lemma and $\widetilde{\lambda}$ is the Lagrandge multiplier chosen to satisfy $\|\mathbf{w}_i\|^2 = 1$. Hence, we have

$$\|\mathbf{w}_i\|^2 = \left|\frac{\hat{\tilde{\mathbf{h}}}_i^H \mathbf{M}_i^{-2}\hat{\tilde{\mathbf{h}}}_i}{1 + \hat{\tilde{\mathbf{h}}}_i^H \mathbf{M}_i^{-1}\hat{\tilde{\mathbf{h}}}_i}\right|^4 \tag{24}$$

Consider the term in (24),

$$\begin{aligned}
\hat{\tilde{\mathbf{h}}}_i^H \mathbf{M}_i^{-2}\hat{\tilde{\mathbf{h}}}_i &= \left(\mathbf{U}_i\hat{\tilde{\mathbf{h}}}_i\right)^H \operatorname{diag}\left(\frac{1}{(\overline{\lambda}+s_1)^2}, .., \frac{1}{(\overline{\lambda}+s_{n_T})^2}\right) \\
\left(\mathbf{U}_i\hat{\tilde{\mathbf{h}}}_i\right) &= \sum_n \frac{1}{(\overline{\lambda}+s_n)^2}|\mathbf{g}_n|^2
\end{aligned} \tag{25}$$

where $\mathbf{U}_i$ is the eigenmatrix of $\mathbf{M}_i$ and $\mathbf{g} = \left(\mathbf{U}_i\hat{\tilde{\mathbf{h}}}_i\right)$ is identically distributed as $\hat{\tilde{\mathbf{h}}}_i$ with $\mathcal{E}[|\mathbf{g}_n|^2] = 1/n_T$. From [12], the term $\hat{\tilde{\mathbf{h}}}_i^H \mathbf{M}_i^{-1}\hat{\tilde{\mathbf{h}}}_i \to a_g = \int_0^\infty \frac{1}{s+\overline{\lambda}} dG^*(s)$ almost surely as $n_T \to \infty$ where $G^*(s)$ is the limiting distribution of the eigenvalues s with Stiejlas transform given by (29). Hence, the constraint $\|\mathbf{w}_i\|^2 = 1$ implies

$$\begin{aligned}
&\|\mathbf{w}_i\|^2 = 1 \\
&\Rightarrow \mathcal{E}[\sqrt{\|\mathbf{w}_i\|}] = 1 \Rightarrow \mathcal{E}\left[\frac{1}{1+a_g}\sum_n \frac{1}{(\overline{\lambda}+s_n)^2}|\mathbf{g}_n|^2\right] \\
&= \frac{1}{1+a_g}\mathcal{E}\left[\sum_n \frac{1}{(\overline{\lambda}+s_n)^2}\mathcal{E}[|\mathbf{g}_n|^2]\right] \\
&= \frac{1}{1+a_g}\mathcal{E}\left[\frac{1}{n_T}\sum_n \frac{1}{(\overline{\lambda}+s_n)^2}\right] \\
&= \frac{1}{1+a_g}\int_0^\infty \frac{1}{(\overline{\lambda}+s)^2} dG^*(s) = 1
\end{aligned}$$

Appendix B
Conditional Convergence of I_B

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

$$\begin{aligned}
\mathcal{E}[I_B] &= 2\Re\sum_{j\neq k}\mathcal{E}_{\hat{\mathbf{H}}}\left[p_j\mu_{X_j}^*\mathcal{E}[\widetilde{X_j}|\hat{\mathbf{H}}]\right] \\
&= 2\Re\sum_{j\neq k}\mathcal{E}_{\hat{\mathbf{H}}}\left[p_j\mu_{X_j}^* 0\right] = 0.
\end{aligned} \tag{26}$$

From (23), we have

$$\begin{aligned}\mathcal{E}[|\mu_{X_i}|^4] &= \mathcal{E}\left[\frac{|\hat{\tilde{\mathbf{h}}}_k^H \mathbf{M}_k^{-1}\hat{\tilde{\mathbf{h}}}_i|^4}{|1+\hat{\tilde{\mathbf{h}}}_k^H \mathbf{M}_k^{-1}\hat{\tilde{\mathbf{h}}}_k|^4}\right] \\ &\leq \mathcal{E}[|\hat{\tilde{\mathbf{h}}}_k^H \mathbf{M}_k^{-1}\hat{\tilde{\mathbf{h}}}_i|^4] \underbrace{\leq}_{(a)} \mathcal{O}(1/n_T^2)\end{aligned} \tag{27}$$

and

$$\begin{aligned}\mathcal{E}\left[\left|\mathbf{w}_i^H\mathbf{w}_j\right|^4\right] &\leq \mathcal{E}\left[\left|\frac{\hat{\tilde{\mathbf{h}}}_i^H \mathbf{M}_i^{-2}\hat{\tilde{\mathbf{h}}}_j}{1+\hat{\tilde{\mathbf{h}}}_i^H \mathbf{M}_i^{-1}\hat{\tilde{\mathbf{h}}}_i}\right|^4\right] \\ &\leq \mathcal{E}\left[\left|\hat{\tilde{\mathbf{h}}}_i^H \mathbf{M}_i^{-2}\hat{\tilde{\mathbf{h}}}_j\right|^4\right] \underbrace{\leq}_{a} \mathcal{O}(1/n_T^2)\end{aligned} \tag{28}$$

where (a) in both equations are from [12]. Similarly, we have $\mathcal{E}[|\mu_{X_i}|^8] \leq \mathcal{O}(1/n_T^4)$. Consider

$$\begin{aligned}&Var(I_B) = \mathcal{E}[|I_B|^2] \\ &\leq \mathcal{E}\left[\sum_{i,j} p_i p_j \left(\mu_{X_i}^* \widetilde{X}_i + \mu_{X_i}\widetilde{X}_i^*\right)\left(\mu_{X_j}^* \widetilde{X}_j + \mu_{X_j}\widetilde{X}_j^*\right)\right] \\ &= \sum_{i,j}\mathcal{E}_{\hat{\mathbf{H}}}\left[p_i p_j\left(\mu_{X_i}^*\mu_{X_j}\mathcal{E}[\widetilde{X}_i\widetilde{X}_j^*|\hat{\mathbf{H}}] + \mu_{X_i}\mu_{X_j}^*\mathcal{E}[\widetilde{X}_i^*\widetilde{X}_j|\hat{\mathbf{H}}]\right)\right] \\ &\leq \frac{\sigma^2}{n_T(1-\sigma^2)}\sum_{i,j}\mathcal{E}\left[p_i p_j|\mu_{X_i}\mu_{X_j}||\mathbf{w}_j^H\mathbf{w}_i|\right] \\ &\underbrace{\leq}_{(b)} \mathcal{O}(1/n_T)\sum_i\sqrt{\mathcal{E}[p_i^4]\mathcal{E}[|\mu_{X_i}|^4]} + \mathcal{O}(1/n_T)\sum_{i\neq j}\sqrt{\mathcal{E}[p_i^2p_j^2]\mathcal{E}[|\mu_{X_i}\mu_{X_j}|^2|\mathbf{w}_j^H\mathbf{w}_i|^2]} \\ &\underbrace{\leq}_{(c)} \mathcal{O}(1/n_T)\sum_i\mathcal{O}(1/n_T^3) + \mathcal{O}(1/n_T)\sum_{i\neq j}\sqrt{\sqrt{\mathcal{E}[p_i^4]\mathcal{E}[p_j^4]}\sqrt{\sqrt{\mathcal{E}[|\mu_{X_i}|^8]\mathcal{E}[|\mu_{X_j}|^8]}\mathcal{E}[|\mathbf{w}_j^H\mathbf{w}_i|^4]}} \\ &\underbrace{\leq}_{(d)} \mathcal{O}(1/n_T^3) + \mathcal{O}(1/n_T)\sum_{i\neq j}\mathcal{O}(1/n_T^{3.5}) \leq \mathcal{O}(1/n_T^{2.5})\end{aligned}$$

where (b), (c) are due to Cauchy-Swartz inequality and (d) is due to (28) and (27). 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 $\Pr[I_B \leq \mathcal{O}(1/n_T^{1+\delta})] \to 1$.

Appendix C
Conditional Convergence of I_C

Consider

$$\begin{aligned}Var(I_C) &= \sum_{i,j}\mathcal{E}\left[p_i p_j \frac{\sigma^4}{n_T^2(1-\sigma^2)^2}\left|\mathbf{w}_i^H\mathbf{w}_j\right|^2\right] \\ &\underbrace{\leq}_{(a)} \mathcal{O}(1/n_T^2)\sum_{i,j}\sqrt{\mathcal{E}\left[\left|\mathbf{w}_i^H\mathbf{w}_j\right|^2\right]}\sqrt{\mathcal{E}[p_i^2p_j^2]} \\ &\underbrace{\leq}_{(b)} \mathcal{O}(1/n_T^2)\sum_{i,j}\sqrt{\mathcal{E}\left[\left|\mathbf{w}_i^*\mathbf{w}_j\right|^2\right]}\sqrt{\sqrt{\mathcal{E}[p_i^4]}\sqrt{\mathcal{E}[p_j^4]}} \\ &\underbrace{\leq}_{(c)} \mathcal{O}(1/n_T^3) + \mathcal{O}(1/n_T^4)\sum_{i\neq j}\mathcal{O}(1/n_T) \leq \mathcal{O}(1/n_T^3)\end{aligned}$$

where (a) and (b) are due to Cauchy-Swartz inequality and (c) is from (28). Since 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.

Appendix D
Proof of Theorem 1

The asymptotic convergence result hinges on the fact that the limiting empirical distribution of the eigenvalues (ω) of a large random matrix $\left(\sum_{j\neq k}\hat{\tilde{\mathbf{h}}}_j\hat{\tilde{\mathbf{h}}}_j^H\right)$ is deterministic and converges weakly to $G^*(\omega)$. In general, $G^*(\omega)$ does not have a closed-form solution. Its Stieltjes transform[13], defined as $m(z) = \int \frac{1}{\omega - z}dG^*(\omega)$, satisfies:

$$m(z) = \left[-z + L\int\frac{1}{1+m(z)}dG^*(z)\right]^{-1} \quad \forall z \in \mathcal{C}^+ \equiv \{z : \Im\{z\} > 0\} \tag{29}$$

Based on this, we shall first introduce several important lemmas which will be helpful to obtain the asymptotic packet outage expression.

Lemma 1. *Asymptotically for large n_T, the Lagrandge multiplier $\overline{\lambda}_k$ for the MMSE weights in (5) is a solution of the following fixed point equation in $\overline{\lambda}$.*

$$\int_0^\infty \frac{1}{(\omega+\overline{\lambda})^2}dG^*(\omega) = \left(1 + \int_0^\infty\frac{1}{\omega+\overline{\lambda}}dG^*(\omega)\right) \tag{30}$$

Proof 1. *From (5), the Tx-MMSE weight $\mathbf{w}_i$ is given by*

$$\begin{aligned}\mathbf{w}_i &= \left(\sum_j \hat{\tilde{\mathbf{h}}}_j\hat{\tilde{\mathbf{h}}}_j^H + \widetilde{\lambda}\mathbf{I}\right)^{-1}\hat{\tilde{\mathbf{h}}}_i \\ &= \left(\underbrace{\left(\sum_{j\neq i}\hat{\tilde{\mathbf{h}}}_j\hat{\tilde{\mathbf{h}}}_j^H + \overline{\lambda}_i\mathbf{I}\right)}_{\mathbf{M}_i} + \hat{\tilde{\mathbf{h}}}_i\hat{\tilde{\mathbf{h}}}_i^H\right)^{-1}\hat{\tilde{\mathbf{h}}}_i^H \\ &\underbrace{=}_{(a)} \left(\mathbf{M}_i^{-1} - \frac{\mathbf{M}_i^{-1}\hat{\tilde{\mathbf{h}}}_i\hat{\tilde{\mathbf{h}}}_i^H\mathbf{M}_i^{-1}}{1+\hat{\tilde{\mathbf{h}}}_i^H\mathbf{M}_i^{-1}\hat{\tilde{\mathbf{h}}}_i}\right)\hat{\tilde{\mathbf{h}}}_i\end{aligned} \tag{31}$$

where (a) is by matrix inversion lemma and $\widetilde{\lambda}$ is the Lagrandge multiplier chosen to satisfy $\|\mathbf{w}_i\|^2 = 1$. Hence, we have

$$\|\mathbf{w}_i\|^2 = \left|\frac{\hat{\tilde{\mathbf{h}}}_i^H\mathbf{M}_i^{-2}\hat{\tilde{\mathbf{h}}}_i}{1+\hat{\tilde{\mathbf{h}}}_i^H\mathbf{M}_i^{-1}\hat{\tilde{\mathbf{h}}}_i}\right|^4 \tag{32}$$

Consider the term in (32),

$$\hat{\tilde{\mathbf{h}}}_i^H\mathbf{M}_i^{-2}\hat{\tilde{\mathbf{h}}}_i = \left(\mathbf{U}_i\hat{\tilde{\mathbf{h}}}_i\right)^H diag\left(\frac{1}{(\overline{\lambda}+s_1)^2}, .., \frac{1}{(\overline{\lambda}+s_{n_T})^2}\right)\left(\mathbf{U}_i\hat{\tilde{\mathbf{h}}}_i\right) = \sum_n\frac{1}{(\overline{\lambda}+s_n)^2}|\mathbf{g}_n|^2$$

where $\mathbf{U}_i$ is the eigenmatrix of $\mathbf{M}_i$ and $\mathbf{g} = \left(\mathbf{U}_i\hat{\tilde{\mathbf{h}}}_i\right)$ is identically distributed as $\hat{\tilde{\mathbf{h}}}_i$ with $\mathcal{E}[|\mathbf{g}_n|^2] = 1/n_T$. From [12], *the term* $\hat{\tilde{\mathbf{h}}}_i^H \mathbf{M}_i^{-1} \hat{\tilde{\mathbf{h}}}_i \rightarrow a_g = \int_0^\infty \frac{1}{s+\overline{\lambda}} dG^*(s)$ *almost surely as* $n_T \rightarrow \infty$ *where* $G^*(s)$ *is the limiting distribution of the eigenvalues* s *with Stiejlas transform given by (29). Hence, the constraint* $\|\mathbf{w}_i\|^2 = 1$ *implies*

$$
\begin{aligned}
&\|\mathbf{w}_i\|^2 = 1 \\
&\quad \Rightarrow \mathcal{E}[\sqrt{\|\mathbf{w}_i\|}] \\
&\quad = 1 \Rightarrow \mathcal{E}\left[\frac{1}{1+a_g}\sum_n \frac{1}{(\overline{\lambda}+s_n)^2}|\mathbf{g}_n|^2\right] = \frac{1}{1+a_g}\mathcal{E}\left[\sum_n \frac{}{(\overline{\lambda}}\right. \\
&\quad = \frac{1}{1+a_g}\mathcal{E}\left[\frac{1}{n_T}\sum_n \frac{1}{(\overline{\lambda}+s_n)^2}\right] \\
&\quad = \frac{1}{1+a_g}\int_0^\infty \frac{1}{(\overline{\lambda}+s)^2} dG^*(s) = 1
\end{aligned}
$$

Define the spatial interference I_k as:

$$
I_k = \sum_{j \neq k} p_j |\widetilde{\mathbf{h}}_k^H \mathbf{w}_j|^2 = \underbrace{\sum_{j \neq k} p_j |\mu_{X_i}|^2}_{I_A} + \underbrace{2\Re\left(\sum_{j \neq k} p_j \mu_{X_j}^* \widetilde{X}_j\right)}_{I_B} + \sum_{j \neq} \qquad (33)
$$

where $X_j = \widetilde{\mathbf{h}}_k^H \mathbf{w}_j$, $\mu_{X_j} = \hat{\tilde{\mathbf{h}}}_k^H \mathbf{w}_j$ and $\widetilde{X}_j = X_j - \mu_{X_j}$. We would like to show that for almost all realizations of CSIT, the spatial interference I_k converges to a deterministic constant as n_T increases. We first have the following definitions and Lemmas.

Definition 2 (Asymptotic Upper Bound $\mathcal{O}(.)$)**.** $f(n) = \mathcal{O}(g(n))$ *if there exists* $M > 0$ *and* $n_0 > 0$ *such that* $|f(n)| \leq M|g(n)|$ *for all* $n > n_0$*. In other words,* $g(n)/f(n) < M$ *for some* $M > 0$ *as* $n \rightarrow \infty$*.*

Definition 3 (Asymptotic Tight Bound $\Theta(.)$)**.** $f(n) = \Theta(g(n))$ *if there exists* $M_u > 0$*,* $M_l > 0$ *and* $n_0 > 0$ *such that* $M_l|g(n)| \leq |f(n)| \leq M_u|g(n)|$ *for all* $n > n_0$*. In other words,* $M_l < g(n)/f(n) < M_u$ *for some* $M_u > 0$ *and* $M_l > 0$ *as* $n \rightarrow \infty$*.*

Definition 4 (Regular Power Allocation Policy)**.** *A power allocation policy* $\mathcal{P} = \{p_i\}$ *is said to be* regular *if* $\mathcal{E}[p_i^2] = \mathcal{O}(1/n_T^2)$ *and* $\mathcal{E}[p_i^4] = \mathcal{O}(1/n_T^4)$ *for all* $i \in \{1, 2, .., K\}$*.*

The regularity requirement in Definition 4 implies that there is no single user being allocated *exceptionally large power* on average. This is a mild condition because for large n_T, the fluctuation in the channel quality among users are limited.

Lemma 2 (Conditional Convergence of I_B)**.** *If* $\mathcal{P} = \{p_i\}$ *is regular, we have* $I_B = 2\Re\left(\sum_{j \neq k} p_j \mu_{X_j}^* \widetilde{X}_j\right)$ *drops faster than* $\mathcal{O}(1/n_T)$ *for almost all CSIT realizations. Precisely, we have* $\Pr[I_B \leq \mathcal{O}(1/n_T^{1+\delta})] \rightarrow 1$ *for some* $\delta > 0$ *as* n_T *increases.*

Proof 2. *Please refer to appendix B.*

Lemma 3 (Conditional Convergence of I_C)**.** *If* $\mathcal{P} = \{p_i\}$ *is regular, we have* $I_C = \sum_{j \neq k} p_j |\widetilde{X}_j|^2$ *converges to* $\overline{I_C} = P_0\sigma^2/\psi$ *in probability as* n_T *increases.*

Proof 3. *Please refer to appendix C.*

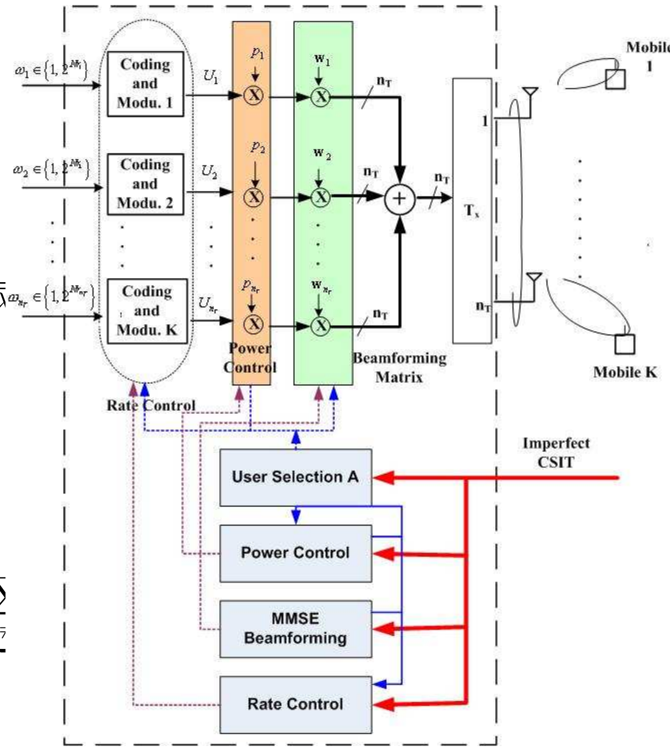

Fig. 1. Downlink transmit MMSE strategy with isolated encoding per spatial stream. n_T independent streams of information carrying message indices $\{\omega_1, ..., \omega_{n_T}\}$ are spatially multiplexed at the base station using MMSE beam-forming.

Using Lemmas 2 and 3, we have for almost all CSIT realizations, $I_k \rightarrow \overline{I}_k = P_0\sigma^2/\psi + \sum_{j \neq k} p_j |\mu_{X_j}|^2$ (conditioned on CSIT) in probability. Therefore, we have $S_k \rightarrow p_k|X_k|^2 - \Lambda_k \overline{I}_k$. Since X_k is a complex Gaussian random variable (conditioned on CSIT) with non-zero mean, $\psi|X_k|^2$ is a non-central chi-square random variable with 2 degrees of freedom, noncentral parameter $s = |\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2$. Hence, the conditional outage probability can be expressed as:

$$
P_{out}(r_k, p_k, \mathcal{A}, \hat{\mathbf{H}}) = \Pr\left[S_k\psi < \Lambda_k|\hat{\mathbf{H}}\right] = F_{\chi^2(s^2,\sigma^2)}\left(\frac{\Lambda_k(\mu_I(\hat{\mathbf{H}})+1)}{p_k}\right.
$$

where $F_{\chi^2(s^2,\sigma^2)}(y)$ is the c.d.f. of non-central chi-square distribution with non-central parameter $s^2 = |\hat{\mathbf{h}}_k^H \mathbf{w}_k|^2$ and variance σ^2.

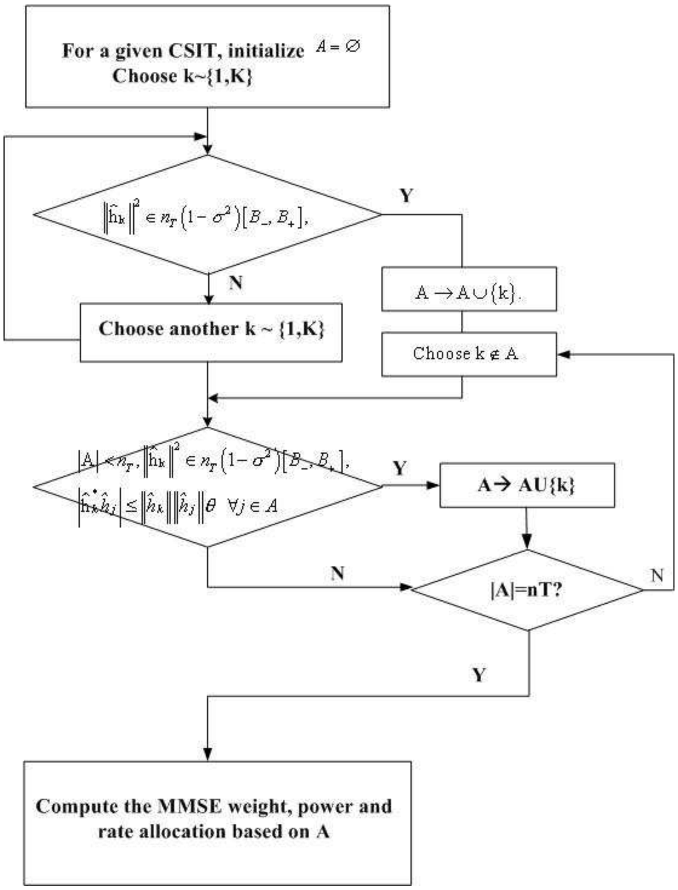

Fig. 2. A flow chart of the NEOUS cross layer scheduling algorithm.

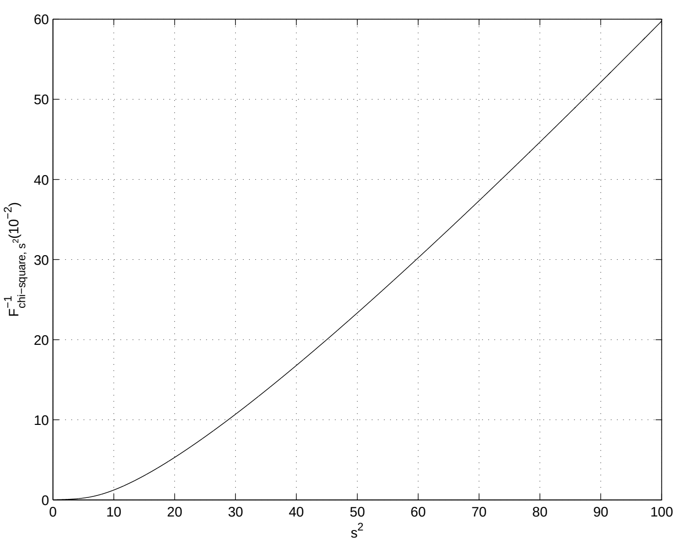

Fig. 3. Order of growth of $\varphi_k(\epsilon)$ with respect to the non-central parameter s^2 for $\epsilon = 10^{-2}$.

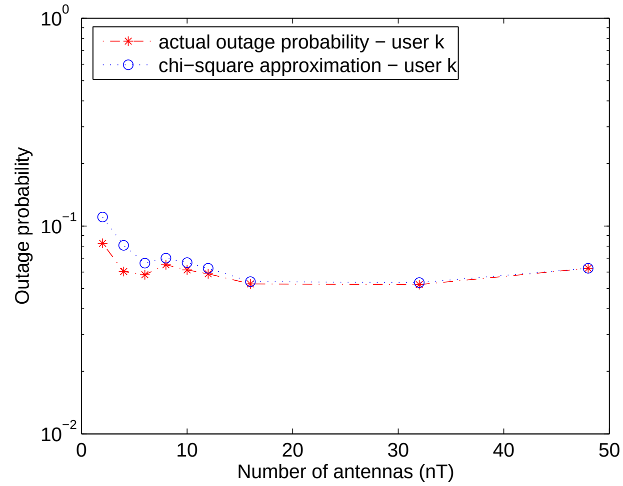

Fig. 4. Conditional average packet outage probability versus number of antennas n_T for CSIT errors $\sigma^2 = 0.05$ and SNR = 10dB.

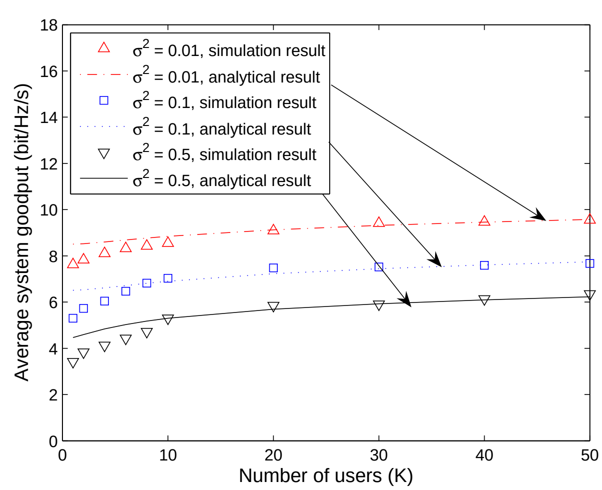

Fig. 5. Average system goodput versus number of users K for $n_T = 10$, CSIT errors $\sigma^2 = 0.01, 0.1, 0.5$ and SNR = 10dB.

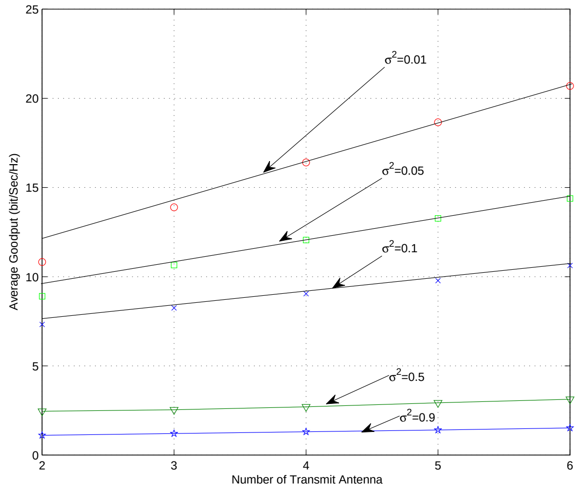

Fig. 6. Average system goodput versus number of transmit antennas at the base station n_T for $K = 50$, CSIT errors $\sigma^2 = 0.01, 0.05, 0.1, 0.5, 0.9$ and SNR = 12dB. Solid line represents analytical expression and marker represents simulation results.

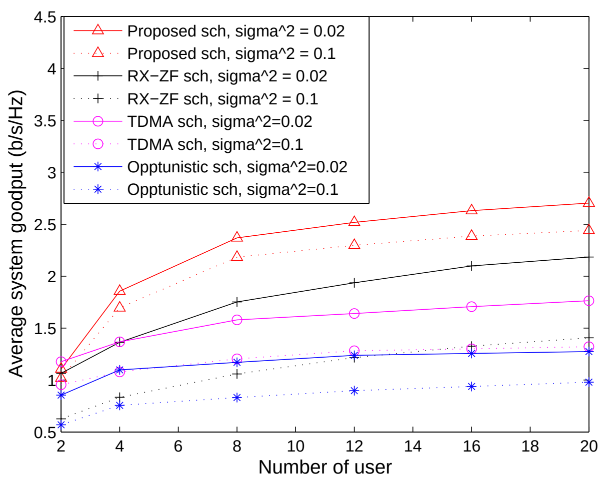

Fig. 7. Average system goodput versus K of the proposed cross-layer MMSE, cross-layer ZF, TDMA-based scheduler (selecting one user at a time) and opportunistic scheduler for $n_T = 4$, CSIT errors $\sigma^2 = 0.02, 0.1$ and SNR = 10dB.