1 Introduction

2 System Model

The system model considers N_{B} BSs, each has N_{T} transmit antennas and K users with a single receive antenna. The set $\mathcal{U}_b$ denotes the set of users linked to BS b where $b \in \{1, 2, \ldots, N_{\mathrm{B}}\}$. The set $\mathcal{U}$ given by $\mathcal{U}_b \subset \mathcal{U}$ represents the total users in the system as represented by $|\,\mathcal{U}\,| = K$. The set $\mathcal{B}_k \subset \mathcal{B} = \{1, 2, \ldots, N_{\mathrm{B}}\}$ represents the set of BSs transmitting for the user k. The cardinality of the set $|\,\mathcal{B}_k\,| \geq 1$ represents the coordinated transmission from more than one BS in the system. The users are statically assigned to a BS in the single BS transmission scenario based on the pathloss measures. The received signal y_k of the user k consisting of both inter cell and intra cell interference is given by

$$y_k = \underbrace{\sum_{b\,\in\,\mathcal{B}_k} \mathbf{h}_{b,k}\mathbf{x}_{b,k}}_{\text{desired term}} \;+\; \overbrace{\sum_{b\,\in\,\mathcal{B}_k} \mathbf{h}_{b,k} \sum_{i\,\in\,\mathcal{U}_b\,\setminus\,k} \mathbf{x}_{b,i}}^{\text{intra cell interference terms}} \;+\; \overbrace{\sum_{c\,\in\,\mathcal{B}\,\setminus\,\mathcal{B}_k} \mathbf{h}_{c,k} \sum_{j\,\in\,\mathcal{U}_c} \mathbf{x}_{c,j}}^{\text{inter cell interference terms}} \;+\; n_k \tag{1}$$

where the vector $\mathbf{x}_{b,k} \in \mathbb{C}^{N_{\mathrm{T}}}$ represents the transmitted symbol from the BS b to user k, $n_k \sim \mathcal{CN}(0, N_0)$, and $\mathbf{h}_{b,k} \in \mathbb{C}^{1\times N_{\mathrm{T}}}$ denotes the channel (including pathloss) between the BS b to the user k.

The transmitted symbol $\mathbf{x}_{b,k}$ for the user k from BS b is given by $\mathbf{x}_{b,k} = \mathbf{m}_{b,k}\, d_k$ where $\mathbf{m}_{b,k}$ is the precoder used by the BS b for user k and d_k denotes the data meant for user k with $\mathbf{E}[\,|\,d\,|^2] = 1$. The total power used by the transmitter is given by

$$\sum_{k\,\in\,\mathcal{U}_b} \mathrm{Tr}\left(\mathbf{x}_{b,k}\,\mathbf{x}_{b,k}^{\mathrm{H}}\right) \leq \mathrm{P}_t \tag{2}$$

The precoding scheme is based on weighted minimum mean squared error (W-MMSE) scheme discussed in [1] or by combined zero-forcing (CZF) scheme by stacking the channel of users in the transmission set of all BSs in $\mathcal{B}$ as discussed in [2]. Once precoders are defined, power allocation is performed over precoders designed by CZF scheme to either maximize the sum capacity or to minimize the expected queue size. W-MMSE based precoding scheme is also analyzed in this paper to bring out the performance comparison. W-MMSE scheme provides joint design of precoder and power allocation for each users in the network.

Scheduling schemes are compared with the well established schemes discussed in [3, 4] which aims at selecting least correlated users for the transmission set for MU-MIMO. The selection process is performed in an iterative manner by selecting the first user based on the channel norm.

$$\begin{aligned} \mathbf{N} &= \mathbf{I} - \mathbf{U}(\,\mathbf{U}^{\mathrm{H}}\mathbf{U}\,)^{-1}\mathbf{U}^{\mathrm{H}} \\ \mathbf{U} &= \left[\,\mathbf{h}_{b,x}^{\mathrm{T}}\;\mathbf{h}_{b,y}^{\mathrm{T}}\,\right]\;\forall\, x, y \,\in\, \mathcal{S}_b \text{ where } b \,\in\, \mathcal{B} \end{aligned} \tag{3}$$

The successive users are chosen by selecting the user with the highest projection gain over the null space formed by the channel vectors of the already chosen users at BS b as given by

$$j = \arg\max_{i} \| \mathbf{N}^{\mathrm{H}} \mathbf{h}_{b,i}^{\mathrm{T}} \| \; \forall\, i \in \mathcal{U}_b. \tag{4}$$

The user j is then selected as the next user in the transmission set and the process is repeated until the transmission user set $|\mathcal{S}_b| = N_{\mathrm{T}}$. The selection schemes discussed in [5, 6] performs similar to the one discussed earlier but with different interpretation which is based on maximizing the volume formed by the user channel vectors. The overall complex multiplications involved is given as

$$\approx \left(N_{\mathrm{T}}^3 (N_{\mathrm{T}} - 1) + N_{\mathrm{T}}^2 \right) \, |\mathcal{U}_b| + \sum_{i=1}^{N_{\mathrm{T}}-1} 2iN_{\mathrm{T}}^2 + i^2 (\, i + N_{\mathrm{T}} \,) \tag{5}$$

where the terms inside summation are meant for the null space calculations.

The scheduling schemes discussed are not limited to single receive antenna; it can be extended for multi antenna by treating each spatial streams as virtual users using singular value decomposition (SVD) over the channel matrix of users in $\mathcal{U}$. The precoder design using W-MMSE scheme is straight forward with the transmission user set selected based on scheduling schemes as it optimizes both transmit and receive beamformers jointly. The zero-forcing precoding is performed in an iterative manner by fixing the receive beamformers using MMSE receivers as discussed in [7].

3 Single BS User Scheduling

The user selection scheme for MU-MIMO transmission performs better when the user $k \in \mathcal{U}_b$ channels are uncorrelated. The uncorrelated channel constraint helps in decoupling the users data streams with the help of precoders thereby providing interference free transmission. The selection of users with two different objective is studied in this section namely, capacity achieving and fairness based queue size reduction. The channel represented by $\mathbf{h}_{b,k}$ is given by $\mathbf{h}_k$ by dropping the subscript corresponds to BS $b = 1$.

3.1 Max-Throughput based User Scheduling

The selection methods with the objective of maximizing the overall throughput is considered in this section. The algorithms mentioned here are classified based on the performance and complexity. The complexity involved is lowered by reducing the operations involved in calculating the metric used for comparison.

3.1.1 Eigen vector based User Selection

The main objective of capacity achieving MU-MIMO transmission is to select the users whose channel vectors are uncorrelated. Uncorrelated channels can be de-coupled easily to facilitate interference free transmission with the use of precoders. The measure of uncorrelated vectors can be obtained by finding determinant of the inner product of stacked channel vectors. The capacity achieving user selection is carried out by forming a matrix $\mathbf{M}$

of size $K \times K$ where $K = |\mathcal{U}_1| = |\mathcal{U}|$ which describes the number of users in a given BS.

The matrix $\mathbf{M}$ represents the priority matrix which quantifies the coexistence of a certain user with other users in the system. The metric which quantifies the measure of coexistence should depend on the channel vector which decides the precoder design for interference free transmission. The ideal metric to quantify the coexistence is the correlation between the users channel vector which is given by the determinant of the inner product of the stacked channel vectors. The matrix $\mathbf{M}$ which quantifies the significance of coexistence is symmetric since the inner product measure is commutative. Once the matrix $\mathbf{M}$ is formulated, the relative priorities are given by the Eigen vector of the dominant Eigen value of the matrix $\mathbf{M}$ [8].

With the above reasoning, the matrix $\mathbf{M}$ is populated with the metric which holds information about the coexistence measure between the users corresponding to the i^{th} row and j^{th} column. The metric also includes the channel vectors of users already selected for transmission at a given scheduling resource. The diagonal elements of the matrix $\mathbf{M}$ or self existence metric is given by

$$\mathbf{T} = \left[\mathbf{T}_{\text{o}} \, \mathbf{h}_i^{\text{T}} \right] \tag{6}$$

$$M_{i,i} = \det \left(\mathbf{T}^{\text{H}} \, \mathbf{T} \right) \tag{7}$$

and the metric used at non diagonal entries or coexistence metric represents the measure of area or volume subtended by the channel vector of users at i^{th} row and j^{th} column together with channel vectors already selected in the previous iteration as given by

$$\mathbf{T} = \left[\mathbf{T}_{\text{o}} \, \mathbf{h}_i^{\text{T}} \, \mathbf{h}_j^{\text{T}} \right] \tag{8}$$

$$M_{i,j} = \det \left(\mathbf{T}^{\text{H}} \, \mathbf{T} \right) \tag{9}$$

where $\mathbf{T}_{\text{o}}$ denotes the stacked channel vectors of users belonging to the transmission set $\mathcal{S}$, $|\mathcal{S}| \leq N_{\text{T}}$ at a given scheduling instant. The set $\mathcal{S}$ is initialized with $\emptyset$ and users are included in the set incrementally. Once the matrix $\mathbf{M}$ is populated, a user is selected based on the method followed in analytic hierarchy process (AHP) as in [8] which is briefed here for continuity. The matrix $\mathbf{M} \succeq 0$, Eigen value decomposition (EVD) decomposes the matrix into $\mathbf{P}\,\mathbf{D}\,\mathbf{P}^{\text{H}}$ where $\mathbf{P}$ is a unitary matrix consists of Eigen vectors of $\mathbf{M}$ and $\mathbf{D}$ is a diagonal matrix having Eigen values on its diagonal entries.

Let $D_{k,k}$ represents the maximum Eigen value at k^{th} index and the corresponding Eigen vector is given by $\mathbf{p}_k$. Since each entry $p_{i,k} \, \forall i \in \{1, 2, \ldots, K\}$ has unequal gains which corresponds to the significance of the entry in achieving the corresponding Eigen gain, user i is selected with

$$i = \arg \max_{\text{i}} \| \, p_{i,k} \, \|_2 \tag{10}$$

$$\mathcal{S} = \{ \, \mathcal{S} \cup i \, \} . \tag{11}$$

Selection is carried out in an iterative manner with the stopping criterion given by $|\mathcal{S}| = N_{\text{T}}$ as detailed in Algorithm. 1. Precoding is performed over the channels of the users in $\mathcal{S}$ which forms the transmission set. Precoding is either zero-forcing (ZF) with water-filling power allocation (WF-PA) as in [9] or weighted sum rate maximization using weighted minimum mean squared error (W-MMSE) as given in [1] with the power constraint as given in (2).

Algorithm 1: Eigen Vector based User Selection

```
Input: h_k ∀ k ∈ U
Data: S = ∅, M, T_o = [ ]
while | S | ≤ N_T do
    foreach M_{i,j}, where i, j ∈ {1, 2, ..., K} do
        formulate M_{i,j} using (7) and (9)
    end
    perform EVD as M = P D P^T
    select user i using (10)
    S = { S ∪ i }, T_o = [ T_o h_i ]
end
```

The complex multiplications involved in the selection process is given as

$$\approx \left\{ \sum_{i=1}^{N_{\mathrm{T}}-1} i N_{\mathrm{T}}^2 + i^3 + 2N_{\mathrm{T}} i^2 + \left((i+1)N_{\mathrm{T}}^2 + \frac{(i+1)^3}{3} \right) \right\} \frac{|\,\mathcal{U}_b\,|(|\,\mathcal{U}_b\,|+1)}{2} + \Delta N_{\mathrm{T}} \tag{12}$$

where $\frac{(i+1)^3}{3}$ provides the rough calculation of the determinant of a matrix of size i and Δ refers to the complexity involved in calculating SVD of the matrix sized $|\,\mathcal{U}_b\,|$. The factor $\frac{|\,\mathcal{U}_b\,|(|\,\mathcal{U}_b\,|+1)}{2}$ arises from the PSD of matrix $\mathbf{M}$ which requires only the earlier complexity instead of $|\,\mathcal{U}_b\,|^3$.

3.1.2 Selection based on Reduced Null Space Gain

The objective of capacity achieving user selection and the complexity involved in implementing the same is considered. The proposed method provides an alternative way to achieve the null space calculations involved in the selection procedure. User selection based on QR based decomposition requires null space formulation in order to find the orthogonal subspace for the given vector channel were discussed in [4, 7, 10, 11].

The user channels are projected on to the null space of the existing users channel vectors in $\mathcal{S}$ where $\mathcal{S}$ represents the users selected for the current scheduling instant. The null space is approximated by the product of the vertical projection distance from the given channel vector to the existing users channel vectors in $\mathcal{S}$. The channel vector $\mathbf{h}_i\,\forall\, i \in \mathcal{U}$ is projected on to the matrix $\mathbf{U}$ formed by stacking the normalized channel vectors of the users in $\mathcal{S}$.

$$\mathbf{U} = \left[\frac{\mathbf{h}_i^{\mathrm{T}}}{\|\,\mathbf{h}_i\,\|}, \ldots, \frac{\mathbf{h}_j^{\mathrm{T}}}{\|\,\mathbf{h}_j\,\|} \right], \;\forall\, i, j \in \mathcal{S} \tag{13}$$

In order to find the vertical projection distance, channel vector of the users in $\mathcal{U}$ is projected on $\mathbf{U}$ is denoted by $\mathbf{g}$. The vector $\mathbf{g}$ contains the projection gains of the given channel vector onto the normalized channel of the users in $\mathcal{S}$ as

$$\mathbf{g} = \mathbf{U}^{\mathrm{H}}\mathbf{h}_i^{\mathrm{T}},\, \forall\, i \in \mathcal{U} \tag{14}$$

The vertical projection distance is given by subtracting the norm of channel vector from the projection metric $\mathbf{g}$. The null space projection is then obtained by multiplying the vertical projection distance from all the unit vectors from the given channel vector as

$$m_i = \prod_{l=1}^{|\mathcal{S}|} \left(\| \mathbf{h}_i \| - | g_l | \right), \forall i \in \mathcal{U} \tag{15}$$

where g_l represents l^{th} element in $\mathbf{g}$ and m_i denotes the null space projection of user i on to the existing user set $\mathcal{S}$.

The product measures the gain achieved by projecting the vector $\mathbf{h}_i$ over the null space of the vectors formed by the channel vectors of users in the set $\mathcal{S}$. The product yields '0' when the vector $\mathbf{h}_i$ is in the direction of the unit channel vectors in $\mathbf{U}$. The product will not provide '0' when it is collinear with any two vectors in $\mathbf{U}$ as given by null space. Even though this method provides an approximation for null space, performance achieved by this scheme is closer to that of the selection scheme achieved by null space based projections discussed in [3, 7, 12].

The approximation of (14) and (15) in the current scheme provides the approximation for the null space projection discussed in (3) providing noticeable reduction in the complexity involved in calculating the metric for each users. The distance metric uses $(\| \mathbf{h}_i \| - | g_l |)$ with g_1 being the projection over the unit direction vector. The metric is different from the ideal vertical projection distance which is measured by the distance which is always higher. Since the metric is based on the absolute distance, it provides conservative estimate following the inequality $\| \mathbf{h}_i \| \sin(\theta) \geq 2 \| \mathbf{h}_i \| \sin^2\left(\frac{\theta}{2}\right)$ over the interval $\theta \in \left[0, \frac{\pi}{2}\right]$. The The pseudo code is briefed in the Algorithm. 2.

Algorithm 2: Selection based on Reduced Null Space Gain

```
Input: h_k ∀ k ∈ U
Data: S = ∅, U = [ ]
while | S | ≤ N_T do
    foreach i ∈ {1, 2, ..., K} do
        formulate U using (13)
        calculate m_i using (14), (15)
    end
    select user k = arg max_i m_i
    S = { S ∪ k }
end
```

The performance of this scheme is equivalent to the QR based scheme with the 2×1 system where the null space calculation in (3) is identical to (15). The complex multiplications involved in the search algorithm is given by

$$\approx |\mathcal{U}_b| \left(N_{\mathrm{T}}^2 + \sum_{i=1}^{N_{\mathrm{T}}} (i-1) N_{\mathrm{T}}^2 \right) + \sum_{i=1}^{N_{\mathrm{T}}-1} N_{\mathrm{T}}^2 \tag{16}$$

which involves the norm metric calculation instead of null space calculation as in (5).

3.2 Queue based User Scheduling

In this section, we discuss the selection schemes which considers the queue backlogs of each user and aims at minimizing it for each users in the set $\mathcal{U}$. Even though the users are selected based on the objective of minimizing the queues, power allocation based on water-filling (WF-PA) for zero-forcing (ZF) precoders has the objective of maximizing the sum capacity. The WF-PA scheme is replaced with the queue based power allocation scheme which uses queue weighted sum rate maximization objective (QW-PA).

The W-MMSE based precoding scheme is also analyzed in this section for the precoder design with the expected queue minimizing objective. The following section discusses two selection schemes with the objective of reducing the queues namely weighted user selection and percentile proportional fair scheduling scheme.

3.2.1 Queue weighted User scheduling

User selection which has the objective of reducing the expected queue of each user select users with higher queue backlogs. The selection strategy based on queue and the channel condition is discussed extensively in [13]. The queue based selection based on Lyapunov drift reduction along with power reduction is also discussed. The performance of those schemes are well suited for single user MIMO (SU-MIMO) where the entire spatial dimension is allotted for a single user.

In case of MU-MIMO, the users sharing spatial dimension will share the power and moreover interfere with each other unless precoder decouples the transmitted data. In order to decouple the transmitted data, the channel of the users in $\mathcal{S}$ should be uncorrelated spatially. Once precoders are designed, power allocation for each users is performed based on queues and the channel condition in order to minimize the expected queue length for each users.

Let $\mathbf{Q}_i(n)$ represents the backlog packets at n^{th} instant and $\mathbf{b}_i(n)$ represents the transmission bits at n^{th} instant for i^{th} user. The Lyapunov formulation is based on minimizing the queue variation at n and $n+1$ instant. The queue update at $n+1$ is given by

$$\mathbf{Q}(n+1) = \min\{\,\mathbf{Q}(n) - \mathbf{b}(n), 0\,\} + \lambda(n) \tag{17}$$

and the Lyapunov drift minimization is given by

$$\begin{aligned} L(\mathbf{Q}(n+1)) - L(\mathbf{Q}(n)) &= \mathbf{Q}(n+1)^2 - \mathbf{Q}(n)^2 && (18) \\ &= [\,\min\{\,\mathbf{Q}(n) - \mathbf{b}(n), 0\,\} + \lambda(n)\,]^2 - \mathbf{Q}(n)^2 && (19) \end{aligned}$$

where (19) is optimized to minimize the Lyapunov drift as discussed in [14].

Upon reducing the formulation in (19) with the minimizing objective, the drift minimizing user scheduling for the set $\mathcal{S}$ is given by

$$\max_{\mathcal{S}\in\mathcal{U}} \sum_{i\in\mathcal{S}} \mathbf{Q}_i(n)\,\mathbf{b}_i(n) \tag{20}$$

$$\text{subject to,} \quad |\,\mathcal{S}\,| \leq N_{\mathrm{T}}. \tag{21}$$

The above problem is difficult to solve due to the constraint (21) which casts it as a combinatorial search problem. The suboptimal solution is proposed by considering user selection in an incremental manner as seen in Sect. 3.1. Since $\mathbf{b}(n)$ is given by $\log\left\{1+\frac{\|\,\mathbf{h}_i\,\|^2}{N_0}\right\}$ at n^{th} instant, the optimization function for the user i is given by

$$\max_{\forall\, i\in\mathcal{U}} \mathbf{Q}_i\, \log\left\{1+\frac{P_{\mathrm{i}}\,\|\,\mathbf{h}_i\,\|^2}{N_0}\right\} \tag{22}$$

$$\max_{\forall\, i\in\mathcal{U}} \mathbf{Q}_i\, [\; 2\,\log\{P_{\mathrm{i}}\,\|\,\mathbf{h}_i\,\|\} - \log\{N_0\}\;] \tag{23}$$

where (23) is based on high-SNR approximation. Since N_0 is constant and can be dropped, the approximated objective is given by $\max_i \mathbf{Q}_i\,\mathbf{h}_i$ where $\log\{x\}\approx x$. The sharing of total transmission power among $\mathcal{S}$ is not known during user selection, the optimization objective includes only the channel norm with path loss in the final formulation.

The queue weighted user selection is performed as briefed in [3, 11] with queue weighted channel vectors. This scheduling provides queue stability and reduces the expected queue size for each user [13]. If precoder is based on ZF, power allocation needs to be performed with the objective of maximizing the sum rate with queue as weights to achieve the above discussed optimization objective.

3.2.2 Percentile proportional fair scheduling

The queue stability can be achieved by scheduling users based on their respective queue backlogs and their instantaneous channel state information. Proportional fair scheduling achieves this objective by weighing the rate fairness metric $\mathbf{r}_i/\mathbf{R}_i$ with the queue backlogs $\mathbf{Q}_i$, where $\mathbf{r}_i$, $\mathbf{R}_i$ represents the instantaneous and average rate of user i. The variants of proportional fair scheduling for multi carrier transmissions with different objectives were discussed in [15].

The fairness objective discussed in [15] perform well for single antenna transmission where the users are separated orthogonally by time or frequency. In case of MU-MIMO transmission, the performance are limited due to the co-channel interference when users are multiplexed in the spatial dimension. In order to provide scheduling fairness and interference free transmission, users selected based on fairness metric should also have uncorrelated channel vectors for efficient decoupling of transmission using precoders.

The proposed scheme achieves the above mentioned objective by considering two level user selection strategy. In the first level, the transmission set $\mathcal{S}$ is selected based on the sorted proportional fair metric with queue as the priority factor $f_i = \mathbf{r}_i\mathbf{Q}_i/\mathbf{R}_i$ where f_i is the i^{th} user fairness metric of the vector $\mathbf{f}$. Let $\mathbf{f}_\pi$ represents the sorted fairness metric in the descending order. The user set $\mathcal{S}_\pi$ is populated with x percentile users based on sorted fairness metric $\mathbf{f}_\pi$ as

$$\mathcal{S}_\pi = \{\, i \mid \arg_i \ f_{\pi,i} \geq f_{\pi,x}, \forall\, i \,\in \mathcal{U} \,\} \tag{24}$$

where $x = 50$ corresponds to 50% ile user selection.

The second level of user selection is performed over the subset $\mathcal{S}_\pi$ which has users with better fairness metric over $\mathcal{S}_\pi^{\mathrm{c}} = \mathcal{U} \setminus \mathcal{S}_\pi$. The user selection is carried out based on queue weighted user scheduling discussed in Sect. 3.2.1 over the reduced user set $\mathcal{S}_\pi$. The selection procedure achieves fairness objective from the initial user selection and the spatial separation based on the weighted queue user scheduling scheme.

4 Multi BS User Scheduling

The selection of users to maximize the system throughput with multiple BS is considered here. The scheduling schemes based on the user set over which the selection is carried out for multi-BS scenario is addressed in this section.

4.1 Static User Scheduling

This section discusses the user selection strategy performed over the set $\mathcal{U}_b \subset \mathcal{U}$ which represents the users linked to BS b. The selection performed over the associated user set $\mathcal{U}_b, \forall\, b \,\in \mathcal{B}$ provides the transmission set $\mathcal{S}_b$ of BS b which needs to be signalled to all BSs in the set $\{\mathcal{B} \setminus b\}$. The selection is carried out in an iterative manner with the limited signaling between the cooperating BSs to maximize the overall sum throughput.

In multi-BS scenario, the available spatial dimension N_T is shared between the transmission of information to the desired cell users and minimizing the interference created to the neighboring cells. This is performed by sharing the available spatial freedom N_T over the cooperating BSs in $\mathcal{B}$ by dividing the spatial dimension over $|\,\mathcal{B}\,|$ as given by $\left\lfloor \frac{N_\mathrm{T}}{N_\mathrm{B}} \right\rfloor$ with the assumption $N_\mathrm{T} \geq N_\mathrm{B}$.

Let $\mathcal{S}_b$ and $\mathcal{S}_{b,\mathrm{H}}$ be initialized with $\emptyset$, $\forall\, b \,\in \mathcal{B}$ where the later represents the user set selected during previous iteration for BS b. The scheduling is performed by a centralized server which updates the transmission user set or distributively by each BS by exchanging the transmission user set $\mathcal{S}_b$ between them. To begin with, let us consider BS $\hat{b}$ select the users from $\mathcal{U}_{\hat{b}}$ while satisfying the cardinality constraint mentioned earlier. The selected user set $\mathcal{S}_{\hat{b}}$ is broadcasted to all BS in $\mathcal{B}$ to perform their respective selections. The exchanged information $\mathcal{S}_{\hat{b}}$ is used by all BS in $\mathcal{B}$ to select the users for transmission set $\mathcal{S}_j$ of BS j using projection metric.

$$\begin{aligned} d_k &= \det\left(\mathbf{T}^\mathrm{H}\,\mathbf{T}\right) \\ \mathbf{T} &= \left[\,\mathbf{T}_\mathrm{X}\;\mathbf{T}_\mathrm{D}\;\mathbf{h}_{j,i}^\mathrm{T}\,\right], \forall\, i \,\in \mathcal{U}_j \end{aligned} \tag{25}$$

where $\mathbf{T}$ is estimated using $\mathbf{T}_\mathrm{X}$ which corresponds to the stacked channel vectors between neighboring users and the BS j and $\mathbf{T}_\mathrm{D}$ represents the stacked channel vectors of users already selected for transmission in the same BS j as given by

$$\begin{aligned} \mathbf{T}_{\mathrm{D}} &= \left[\mathbf{h}_{j,k}^{\mathrm{T}} \dots \right], \forall\, k \in \mathcal{S}_j \\ \mathbf{T}_{\mathrm{X}} &= \left[\mathbf{h}_{j,k}^{\mathrm{T}} \dots \right], \forall\, k \in \bigcup_{\forall\, b \in \mathcal{B} \setminus j} \mathcal{S}_b. \end{aligned} \tag{26}$$

The user set $\mathcal{S}_j$ is updated with the user having maximum metric $\hat{k} = \arg\max_k d_k$ as $\mathcal{S}_j = \{ \mathcal{S}_j \cup \hat{k} \}$ till the constraint $|\mathcal{S}_j| = \left\lfloor \frac{N_{\mathrm{T}}}{N_{\mathrm{B}}} \right\rfloor$ is satisfied. This process is performed over all BSs in a sequential manner to complete one iteration. The second and further iterations are carried out for each BS k by setting $\mathcal{S}_k = \emptyset$ and the new set of transmission user set $\mathcal{S}_k$ is updated by using (25) until the sets are converged. The convergence is guaranteed since the search is over the closet set of channel vectors. The Algorithm 3 describes the procedure of user selection for multi-BS in an iterative manner.

Algorithm 3: Static user scheduling

```
Input: h_{b,i} ∀ i ∈ U_b, ∀ b ∈ B
Data: iteration index i = 0, S_b = ∅, S_{b,H} = ∅, ∀ b ∈ B
while ( ⋃_b S_{b,H} ≠ ⋃_b S_b ) & ( ∑_b |S_{b,H}| ≠ 0 ), ∀ b ∈ B do
    i = i + 1
    foreach b ∈ B do
        S_b = ∅
        foreach k ∈ U_b do
            evaluate d_k using (25) and (26)
        end
        find user k̂ = arg max_k d_k
        S_b̂ = { S_b̂ ∪ k̂ }
    end
    S_{b,H} = S_b, ∀ b ∈ B
end
```

Precoder design is performed jointly over all BSs in the system over the transmission user set $\mathcal{S}_k \,\forall\, k \in \mathcal{B}$ using either WMMSE [1] or ZF of the stacked channel vectors of all users. The gain achieved by this selection scheme is equal to that of overloaded precoder design using WMMSE method with all users in the precoder design procedure.

4.2 Coordinate User Scheduling

This section discusses the coordinate transmission scheme for users in the cell-edge by associating users in $\mathcal{U}$ to BSs in $\mathcal{B}$ based on instantaneous channel realizations. Users with proportional rate fairness has been dealt in [16] where many variants of user selection for coordinate transmission are proposed. The current work addresses the user selection scheme for coordinate transmission by selecting users for $\mathcal{S}_b \,\forall\, b \in \mathcal{B}$ in an iterative manner.

Let $\mathcal{S}_b$ and $\mathcal{S}_{b,\mathrm{H}}$ are initialized to $\emptyset$, $\forall\, b \in \mathcal{B}$ where $\mathcal{S}_{k,\mathrm{H}}$ represents the transmission user set from $(i-1)^{\text{th}}$ iteration for BS b. The user selection is carried out in a centralized server with the transmission set $|\mathcal{S}_b| \leq \left\lfloor \frac{N_{\mathrm{T}}}{N_{\mathrm{B}}} \right\rfloor$ for each BS $b \in \mathcal{B}$. The first user is selected by sorting the channel gains of each users in $\mathcal{U}$ with each BSs in $\mathcal{B}$ and the user $\hat{k}$ with highest norm is assigned to the corresponding BS $\hat{b}$.

$$\mathcal{S}_{\hat{b}} = \{ \mathcal{S}_{\hat{b}} \cup \hat{k} \} \tag{27}$$

where $\{\hat{b}, \hat{k}\} = \arg\max_{b,k} \| \mathbf{h}_{b,k} \|$ represents the selection criterion for the first user. The second and there after will be selected based on finding the user whose channel vector is not on the same subspace ($\mathbf{h}_{\hat{b},k}, \forall k$) for users linking to BS $\hat{b}$ or on to the null space ($\mathbf{h}_{b,\hat{k}}$) for channel between users with BSs in $\{\mathcal{B} \setminus \hat{b}\}$. The set $\mathcal{S}_j$ for BS j is updated using (28) as

$$\begin{aligned} D_{j,i} &= \det\left(\mathbf{T}^{\mathrm{H}}\, \mathbf{T}\right) \\ \mathbf{T} &= \left[\mathbf{T}_{\mathrm{X}}\, \mathbf{T}_{\mathrm{D}}\, \mathbf{h}_{j,i}^{\mathrm{T}}\right], \forall i \in \mathcal{U} \end{aligned} \tag{28}$$

where $D_{j,i}$ corresponds the the metric evaluated for the user i when linked to BS j and $\mathbf{T}$ is evaluated using the following (29)

$$\begin{aligned} \mathbf{T}_{\mathrm{D}} &= \left[\mathbf{h}_{j,k}^{\mathrm{T}} \ldots\right], \forall k \in \mathcal{S}_j \\ \mathbf{T}_{\mathrm{X}} &= \left[\mathbf{h}_{j,k}^{\mathrm{T}} \ldots\right], \forall k \in \bigcup_{\forall b \in \mathcal{B} \setminus j} \mathcal{S}_b \end{aligned} \tag{29}$$

where $\mathbf{T}_{\mathrm{X}}$ corresponds to the stacked channel vectors of the channel between the current BS to the neighboring BSs users and $\mathbf{T}_{\mathrm{D}}$ represents the stacked channel vectors of users in the transmission set of the current BS. The selection and association is then carried out by selecting the maximizing metric over the entire $\mathbf{D}$ matrix as

$$\{\hat{b}, \hat{k}\} = \arg\max_{i,j} \quad D_{i,j} \;\; \forall i \in \mathcal{U} \text{ and } \forall j \in \mathcal{B} \tag{30}$$

and $\mathcal{S}_{\hat{b}}$ is updated using (27).

Once the users are selected for each BSs with the constraint on $|\mathcal{S}_b| \forall b \in \mathcal{B}$, the selection is performed again in an iterative manner until the transmission user set converges. This procedure is explained briefly in Algorithm. 4

This scheme provides significant improvement over the earlier scheduling scheme discussed in section 4.1 based on static user assignment. The improvement is significant only when the pathloss between users $\mathcal{U}$ and $\mathcal{B}$ are comparable and the users are few in number. As the user count increases over each BS, the coordinate gain vanishes since the gain from multi-user diversity performs significantly better.

5 Numerical Results

The power allocation is performed after ZF precoding at the transmitter to maximize the system capacity or to provide power allocation to reduce the expected queue size of each users in $\mathcal{U}$. The combined ZF for multi-BS and ZF for single BS transmission provides stacked channel inversion of user channels in $\mathcal{S}_b \forall b \in \mathcal{B}$ and $\mathcal{S}$ respectively. The power allocation after precoding is performed to achieve the objective of either maximizing sum rate or to provide queue minimization over each users.

The power allocation to maximize the sum capacity objective is mainly based on the water-filling technique which allocates more power to the users with better channel gains over the rest and also depends on the ambient

Algorithm 4: Coordinate user scheduling

```
Input: h_{b,i} ∀ i ∈ U_b, ∀ b ∈ B
Data: iteration index i = 0, S_b = ∅, S_{b,H} = ∅, ∀ b ∈ B
while ( ⋃_b S_{b,H} ≠ ⋃_b S_b ) & ( ∑_b | S_{b,H} | ≠ 0 ), ∀ b ∈ B do
    i = i + 1
    foreach b ∈ B do
        S_b = ∅
        foreach k ∈ U do
            evaluate D_{b,k} using (28), (29)
        end
        continue = true
        while continue do
            find user k̂ and BS b̂ using (30)
            if | S_b̂ | ≤ ⌊ N_T / N_B ⌋ then
                S_b̂ = { S_b̂ ∪ k̂ }
                continue = false
            else
                D_{b̂,k̂} = 0
                continue = true
            end
        end
    end
    S_{b,H} = S_b, ∀ b ∈ B
end
```

noise at the receiver. The power allocation discussed in [9] is given by the following equation

$$p_i = \left[\frac{1}{\lambda} - \frac{\mathrm{N}_0}{\| \mathbf{h}_i^{\dagger} \|^2} \right]^+ \tag{31}$$

where p_i denotes the power assigned for the precoder of user i, $\mathbf{h}_i^{\dagger}$ represents the equivalent zero-forced channel vector for i^{th} user in $\mathcal{S}$, $[x]^+$ represents $\max\{0, x\}$ and λ is the Lagrange multiplier for the constraint (2). In case of multi-BS precoding scheme, WF based power allocation is performed over the precoders belonging to the user set $\mathcal{S}_b$ independently.

The power allocation to reduce the expected queue size is based on weighted sum rate where the weights are based on the current backlogged packets of users in the transmission set. The weighted sum rate maximization is performed in the similar manner using lagrange multiplier for the constraint (2). The queue weighted power allocation strategy is given by the following equation

$$p_i = \left[\frac{\mathbf{Q}_i}{\lambda} - \frac{\mathrm{N}_0}{\| \mathbf{h}_i^{\dagger} \|^2} \right]^+ \tag{32}$$

where $\mathbf{Q}_i$ corresponds the backlogged packets of user i and the remaining variables are similar to (31). This power allocation scheme provides significant reduction in the queue size of each users thereby providing fairness among the users.

The proportional fair scheme provides fairness in terms of user selection based on the average service rate. The precoding scheme and power allocation which mainly seeks for capacity maximizing objective will be greedy

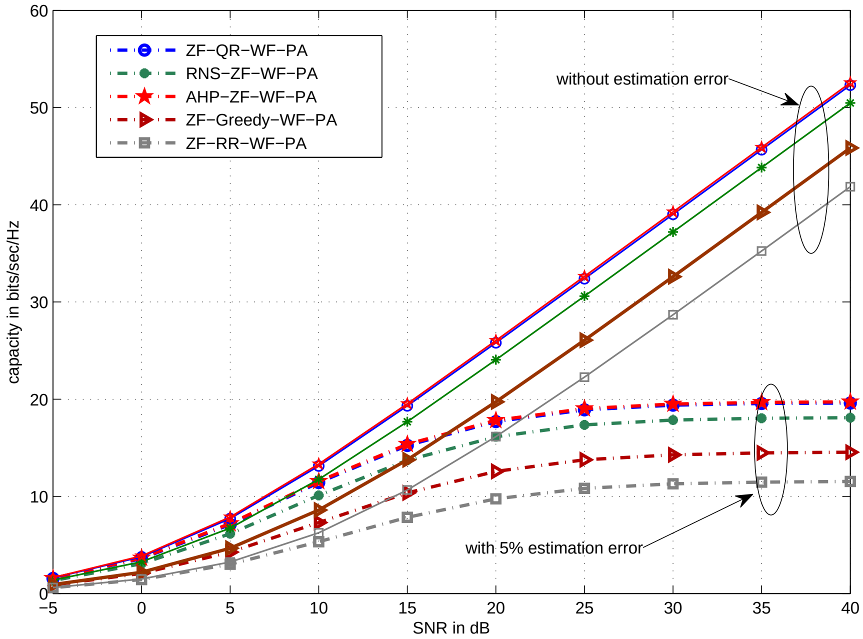

Figure 1: Sum capacity for $|\mathcal{U}| = 20$, $N_{\mathrm{T}} = 4$, $N_{\mathrm{R}} = 1$

in nature thereby negates the effect of fairness. The queue based power allocation based on weighted sum rate provides improved fairness among the users since it depends on the queue backlogs also into account while performing power allocation.

5.1 Single BS-US

This section deals two complementary scheduling strategies in detail with the performance plots to justify the claims. The first section discusses the capacity maximizing schemes discussed in section 3.1 with the assumption of zero pathloss between users and BS. The zero pathloss claim is assumed to bring out the scheduling schemes selection performance to maximize the sum capacity without which the all capacity seeking selection performs relatively closer with significant bias towards low pathloss users.

Fig. 1 compares the performance of capacity achieving user selection schemes discussed in section 3.1 to the QR based user selection scheme discussed in [7, 5]. The AHP based user selection provides marginal but still noticeable performance improvement over the existing QR based algorithm. The gain is mainly attributed to the selection mechanism which takes pair-wise channel correlation metric in to account. This additional information is not of great value as it can be seen but the complexity involved in this is significantly large.

Fig. 1 also shows the performance of reduced null space (RNS) scheme which performs closer to the existing QR algorithm but with huge reduction in the complexity involved in the metric calculation. Since QR based schemes requires matrix inverse computation for null space, RNS scheme provides a sub-optimal alternative to achieve the same. Fig. 2 compares the complexity involved in performing various schemes discussed so far. The complexity involved in RNS user selection scheme is the least among the scheduling schemes based on channel correlation discussed here. The complexity can further be reduced by saving the earlier results in the memory

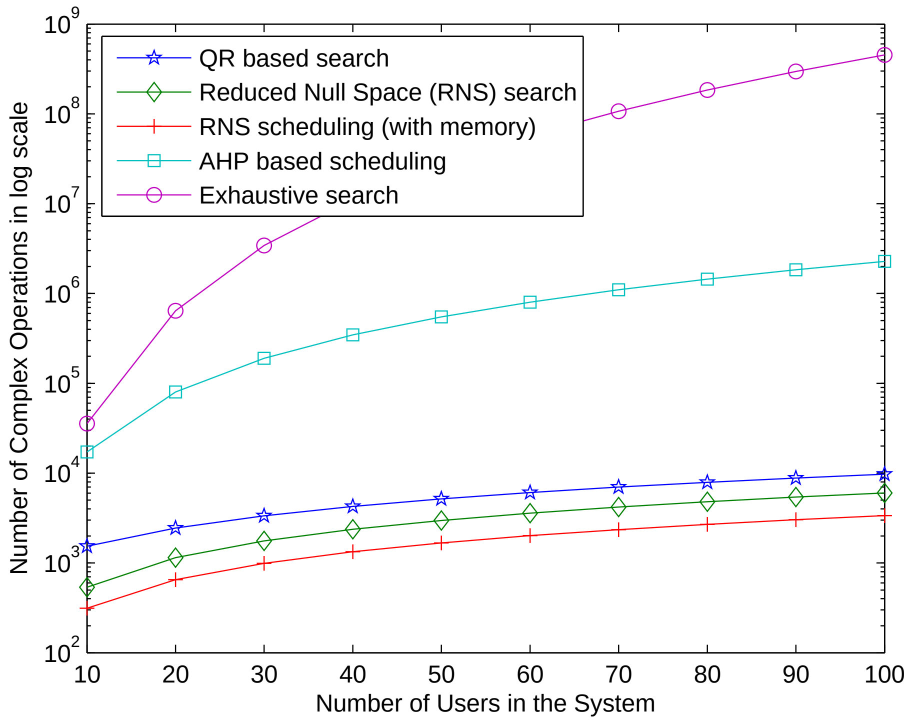

Figure 2: Scaling of complexity over users for $N_{\mathrm{T}} = 4$, $N_{\mathrm{R}} = 1$ system

(with some increase in the storage memory) to avoid redundant calculations over each iterations.

Fig. 1 depicts the performance degradation of above mentioned schemes with 5% estimation error modeled using Gaussian error. The estimation error degrades the performance by altering the precoders from being the perfect channel inverse there by creating interference among the transmitted streams. This effect is more pronounced at higher SNR as the capacity is starts saturating since the noise component is dominated by the interference in comparison with the AWGN.

The selection schemes discussed so far addresses the capacity achieving user selection schemes. Figs. 3 and 4 plots the performance of the selection schemes which minimizes the mean queue size and fairness among the users in the system. The users are associated with the uniform pathloss ranging in $[-30, 0]$ to analyze the fairness performance of user selection schemes. The power allocation scheme discussed earlier is used to achieve the objectives of fairness and queue backlogs reduction when the precoders are designed based on zero-forcing scheme. W-MMSE based scheme provides inherent power allocation for all the precoders based on the queue backlogs by maximizing the queue weighted sum rate objective. Fig. 3 shows that the mean queue size is reduced by all schemes except few fairness based selection schemes like proportional fair (PF) and percentile PF (PPF) scheduling.

The PPF scheme provides significantly closer performance with the PF scheduling with better reduction in the average queue size. The computation of PPF scheduling is greatly reduced in comparison with other scheduling schemes since the QR based search is performed over 50%ile users based on the PF metric as given by (24). The reduced user set allows users with PF metric significantly high to be searched for uncorrelated channels to form the set $\mathcal{S}$ thereby achieving fairness among users.

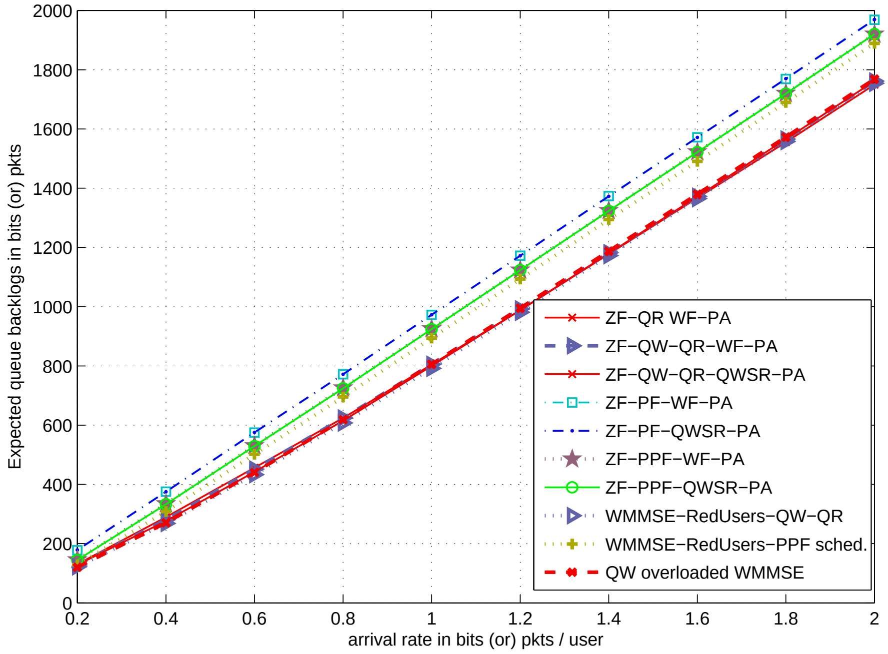

Figure 3: Expected queue size $|\mathcal{U}| = 50$, $N_{\mathrm{T}} = 4$, $N_{\mathrm{R}} = 1$, $\mathcal{U}(0, -30)$ over 1000 slots at 15dB SNR

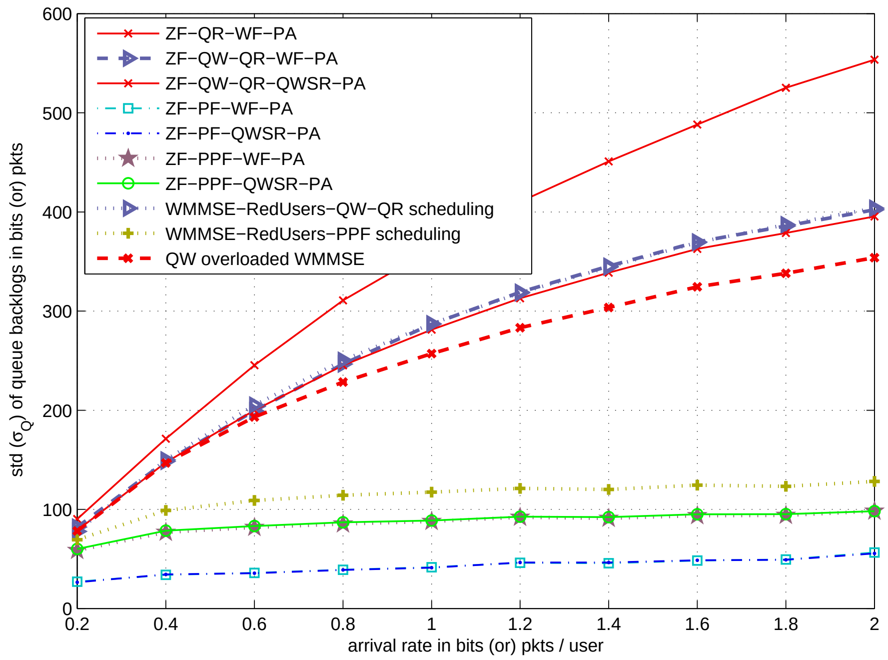

Figure 4: Standard deviation of backlogged packets $|\mathcal{U}| = 50$, $N_{\mathrm{T}} = 4$, $N_{\mathrm{R}} = 1$, $\mathcal{U}(0, -30)$ over 1000 slots at 15dB SNR

The performance of queue weighted user selection with ZF precoding and queue weighted WF scheme (32) provides better fairness in comparison with the capacity achieving user selection and queue weighted selection with WF power allocation scheme (31) for the similar backlogged packet size. The fairness in the ZF scheme is achieved by the use of queue weighted power allocation strategy (32). It is evident that the simple WF scheme (31) achieves fairness among user service in MU-MIMO with the queue weighted (QW) QR based scheme.

The overloaded W-MMSE case is plotted in Figs. 3 and 4 along with other schemes for the comparison purpose. The plot shows that the performance of the overloaded W-MMSE scheme provides significantly fair selection in comparison with the ZF counterparts. The reduced W-MMSE scheme with the selection based on QW QR scheme provides fairness in comparison with the overloaded W-MMSE scheme by greatly reducing the computational complexity involved.

5.2 Multi BS-US

This section discusses the user selection schemes discussed in sections 4.1 and 4.2. The users are assumed to be located at the cell-edge and the objective is to maximize the achievable capacity of the system. The precoding schemes considered here are based on combined zero-forcing with water-filling (CZF-WF) and W-MMSE based joint precoding and scheduling design discussed in [1]. The W-MMSE scheme performs joint precoding and scheduling of users with the maximizing sum capacity objective. The ZF scheme performs only precoder design for the users in $\mathcal{S}_b$ which mandates the users to be selected in way to maximize the sum capacity. The selection of users for multi-BS system needs to be exhaustive in order to achieve the capacity attaining user set $\mathcal{S}_b$. The static user scheduling (SUS) and coordinated user scheduling (CUS) provides one such approach in which users can be selected with minimal complexity in comparison with the exhaustive user search. The user selection schemes discussed earlier provides the better transmission set compared with the conventional user selection scheme for sum rate objective.

Fig. 5 shows the comparison of the above mentioned schemes using CZF-WF and W-MMSE precoding designs. The CZF-WF based design is equivalent to W-MMSE scheme at higher SNR when the transmission user set $\mathcal{S}_b$ is defined. The reduced W-MMSE scheme is performed over the transmission user set $\mathcal{S}_b \, \forall \, b \in \mathcal{B}$ in contrast with the overloaded W-MMSE scheme performing over the entire user set $\mathcal{U}_b$. The performance of the overloaded W-MMSE scheme degrades as the SNR increases to that of reduced W-MMSE scheme with SUS based user selection for the same stopping resolution ϵ as defined in [1]. Even though overloaded W-MMSE performs better at lower SNRs, the reduced W-MMSE using SUS based user selection provides significantly improved performance at higher SNR owing to the fact of reduced variables involved in the optimization problem. The variable reduction is mainly achieved by the selection scheme performed using SUS.

Fig. 5 plots the performance of CUS scheme as well in comparison with SUS schemes using W-MMSE over reduced user set obtained by the selection algorithms. The CUS scheme provides improved performance by selecting users over the entire user set $\mathcal{U}$ for all BSs in $\mathcal{B}$ instead selecting over $\mathcal{U}_b$ for BS $b, \forall \, b \in \mathcal{B}$. The CUS scheme avails the additional multi-user selection diversity over conventional SUS based schemes to achieve

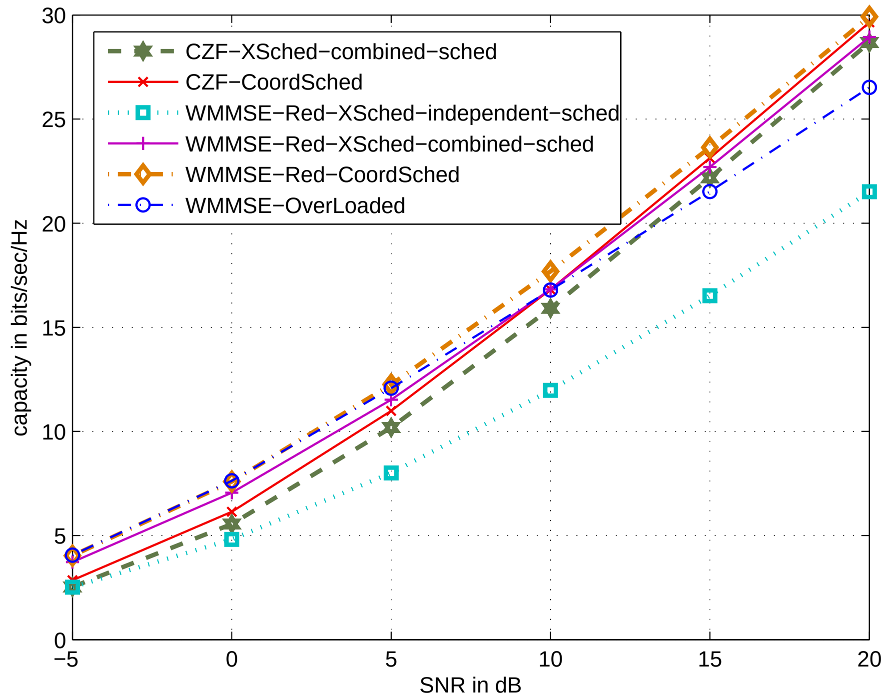

Figure 5: Sum capacity for $|\mathcal{B}| = 2$, $|\mathcal{U}_k| = 25$, $N_{\mathrm{T}} = 4$, $N_{\mathrm{R}} = 1$

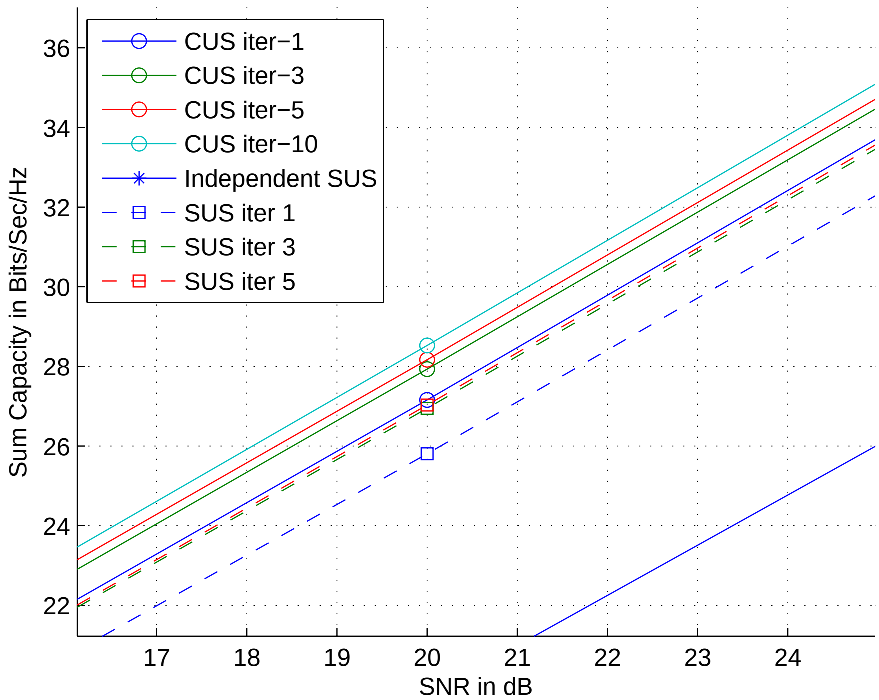

Figure 6: Iterative sum capacity for $|\mathcal{B}| = 2$, $|\mathcal{U}| = 20$, $N_{\mathrm{T}} = 4$, $N_{\mathrm{R}} = 1$

significant performance improvement. The CUS performance is limited by the number of users available for selection since more users in $\mathcal{U}_b$, more diverse users in the set thereby SUS also enjoys more multi-user diversity performing closer to CUS scheme. To achieve any gain from CUS scheme over SUS scheme, the user set $\mathcal{U}$ over which the scheduling is performed should be at the cell-edge thereby availing the instantaneous fading for scheduling users.

Fig. 6 depicts the iterative performance improvement of CUS and SUS schemes for different iteration count using sum capacity as the metric. Since the search space is bounded by the user sets $\mathcal{U}$, $\mathcal{U}_b$, the performance at each iteration will be monotonic in terms of sum rate objective. The selected set at each iterations will be at least good as the earlier transmission user set $\mathcal{S}_b$ but not inferior in terms of sum rate metric as shown in Fig. 6.

6 Conclusion

References

[1] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted mmse approach to distributed sum-utility maximization for a mimo interfering broadcast channel," *IEEE Transactions on Signal Processing*, vol. 59, pp. 4331 –4340, sept. 2011.

[2] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser *MIMO* channels," *IEEE Transactions on Signal Processing*, vol. 52, pp. 461 – 471, feb. 2004.

[3] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 528 – 541, march 2006.

[4] X. Zhang, E. Jorswieck, B. Ottersten, and A. Paulraj, "User selection schemes in multiple antenna broadcast channels with guaranteed performance," in *IEEE 8th Workshop on Signal Processing Advances in Wireless Communications, 2007. SPAWC 2007*, pp. 1–5, IEEE, 2007.

[5] L. Jin, X. Gu, and Z. Hu, "A novel volume-based scheduling scheme for multiuser multiple-input multiple-output downlink system," in *IEEE Conference on Radio and Wireless Symposium (RWS), 2010*, pp. 448–451, IEEE, 2010.

[6] K. Ko and J. Lee, "Determinant-based *MIMO* scheduling with re-duced pilot overhead," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 71, 2012.

[7] A. Tolli and M. Juntti, "Scheduling for multiuser mimo downlink with linear processing," in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC) 2005.*, vol. 1, pp. 156 –160, sept. 2005.

[8] T. Saaty, "Decision making with the analytic hierarchy process," *International Journal of Services Sciences*, vol. 1, no. 1, pp. 83–98, 2008.

[9] D. Tse and P. Viswanath, *Fundamentals of wireless communication.* Cambridge university press, 2005.

[10] L.-N. Tran, M. Bengtsson, and B. Ottersten, "Iterative precoder design and user scheduling for block-diagonalized systems," *IEEE Transactions on Signal Processing*, vol. 60, pp. 3726 –3739, july 2012.

[11] L. Sun and M. McKay, "Eigenmode transmission for the *MIMO* broadcast channel with semi-orthogonal user selection," in *Global Telecommunications Conference, 2009. GLOBECOM 2009*, pp. 1–6, IEEE, 2009.

[12] W. Li, Y. Ke, H. Hanying, and C. Weijia, "Low complexity user selection for *MU-MIMO* downlink based on *MSMSE* transceiver," in *International Conference on Signal Processing Systems (ICSPS), 2010 2nd*, vol. 2, pp. V2–702 –V2–706, july 2010.

[13] M. Neely, "Stability and probability 1 convergence for queueing networks via lyapunov optimization," *Journal of Applied Mathematics*, vol. 2012, 2012.

[14] M. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.

[15] M. Ibnkahla, *Adaptation and Cross Layer Design in Wireless Networks.* CRC press, 2008.

[16] A. Tolli, H. Pennanen, and P. Komulainen, "Sinr balancing with coordinated multi-cell transmission," in *IEEE Wireless Communications and Networking Conference, (WCNC) 2009.*, pp. 1 –6, april 2009.