Multi-User Cooperative Base Station Systems With Joint Precoding and Beamforming

Wibowo Hardjawana, Branka Vucetic, Fellow, IEEE, and Yonghui Li, Senior Member, IEEE

Abstract—Interference among multiple base stations that co-exist in the same location limits the capacity of wireless networks. In this paper, we propose a method to design a single/multi-stream multi-user multiple-input multiple-output (MIMO) cooperative downlink transmission scheme employing precoding and beamforming under both the per base station (BS) and total BS power constraints. The per BS and total BS power constraints are the constraints where the power for each BS and all BSs is limited to a particular value, respectively. The algorithm eliminates the interference and achieves symbol-error rate (SER) fairness among different users. To eliminate the interference, Tomlinson–Harashima precoding (THP) is used to cancel part of the interference while the transmit–receive antenna weights cancel the remaining part. An iterative method based on the uplink–downlink duality principle is used to generate the transmit–receive antenna weights. The algorithm provides an equal signal-to-interference-plus-noise-ratio (SINR) across all users and reduces its computational complexity by trading off the complexity for a slight performance degradation. The proposed methods are extended to the scenario when the receiver does not have complete channel state informations (CSIs). Lastly, we propose a new method of selecting the user precoding order, which has a much lower complexity than VBLAST ordering but with almost the same performance. The simulation results show that the proposed schemes considerably outperform existing nonlinear and linear cooperative transmission schemes in terms of SER performance and approach an interference free performance.

Index Terms—Base station (BS) cooperation, beamforming, cooperative communications, multi-user multiple-input multiple-output (MIMO), precoding.

I. INTRODUCTION

The spectral efficiency of existing cellular networks [1] and wireless local area networks (WLANs) [2] is limited by interference. In cellular mobile networks, the dominant interference comes from adjacent cells [1], while in coworking WLANs [2], the interference from other networks, operating in the same area, is a major limiting factor [2]. This is especially true when the users are located near the cell edges where the interference from the adjacent cells is strong. In this paper, we consider a cooperative base station system to eliminate the coworking interference in WLAN or other wireless networks.

In the proposed cooperative transmission scheme, multiple BSs share information about the transmitted messages to their respective users and wireless channels via a backbone network. Each base station (BS) can either transmit a single symbol stream or multiple symbol streams to its respective mobile station (MS). Individual BSs and MSs are equipped with multiple transmit and receive antennas, respectively. Each BS transmitter uses the transmitted signal information from other BSs and wireless channel conditions to precode its own signal. The precoded signal for each BS is broadcast through all BS transmit antennas in the same frequency band and time slots. The precoding operation and transmit–receive antenna coefficients are chosen in such a way as to minimize the interference coming from other BS transmissions.

Most of the published papers in this area such as [3]–[6] considered only a multi-user multiple-input single-output (MISO) system with a single receive antenna. In [3] and [4], the authors introduced the uplink–downlink duality concept to facilitate the calculation of transmit weights and downlink power. It has been shown that the maximum downlink signal-to-interference-plus-noise-ratio (SINR) can be designed to be equal to the maximum uplink SINR under the same total available power but with different power allocations in the downlink and uplink channels. An algorithm to find the transmit weights and transmission powers that maximize the individual downlink SINR was proposed by solving its dual uplink equivalence. A different approach was proposed in [6] where a combination of a Zero Forcing (ZF) method, that determines transmit weights by forcing part of the interference to zero, and dirty paper coding (DPC) [7] was used to suppress interference from other users. A more practical approach than [6] was considered in [5] where DPC is replaced with Tomlinson–Harashima–Precoding (THP) [8], [9]. These algorithms, however, only consider single receive antenna scenarios, which are not directly applicable to MIMO systems.

An extension of a multi-user MIMO system with a single receive antenna to a multi-user MIMO system with multiple transmit–receive antennas has been considered by several researchers. In [10], the authors used the uplink–downlink duality concept introduced in [3], [4] to design the downlink transmit–receive weights and downlink transmission powers in a multi-user MIMO system with multiple transmit–receive antennas. However, the convergence of the proposed algorithm...
cannot be guaranteed and it does not work when more than one symbol stream is transmitted to each user. In [11], the authors studied a ZF method for a multi-stream multi-user MIMO system, where transmit–receive antenna weights are jointly optimized by a ZF diagonalization technique. The water-filling power allocation method is then applied to allocate power to each user. The method in [11] is then significantly improved in [12] by iteratively finding the transmit–receive weights. Nonlinear methods utilizing a combination of the ZF method with DPC and a combination of the ZF method with THP [8], [9] for a multi-user MIMO system was considered in [13] and [14], respectively, where a ZF precoder was used to eliminate part of the interlink interference and DPC or THP are used to cancel the remaining interference. However, in these schemes the symbol-error-rate (SER) performance varies from user to user. This SER variation is not desirable for practical cooperative MIMO systems as the operators expect that the users have the similar performance.

In this paper, we exploit the uplink–downlink duality concept [1], [3], [4] to design a cooperative transmission scheme employing nonlinear precoding and beamforming for the downlink of a multi-user single/multi-stream MIMO system. In this algorithm, THP [8], [9] cancels part of the interference. The remaining interference is eliminated by using joint transmit–receive antennas weights. We first propose an iterative method to derive the transmit–receive antenna weights and the downlink power such that individual signal-to-interference-plus-noise-ratio (SINR) for each user is maximized. In other words, the downlink SINR is designed to be equal to the maximum achievable uplink SINR under the same total available power constraint. The transmit–receive antenna weights and the downlink power allocation are optimized based on the alternating optimization method in [15]. We investigate two types of power constraints. The first one is a total BS power constraint, where the total power for all BSs is constrained to a particular value. The second one is a per BS power constraint, where the power for each BS is constrained to a particular value. We derive expressions for the power allocation under these two power constraints.

We then propose a simplification of this algorithm by simply setting the virtual uplink power for all links used when uplink–downlink duality concept is used to be equal. By doing this, we effectively eliminate the iteration step required to find the transmit–receive antenna weights. Furthermore, we then assume that the receiver does not have a complete channel state information (CSI). Specifically, each MS receiver only knows its own CSI; thus, the joint design of transmit–receive antenna weights is not possible. Results show that the performance degradation due to the partial CSI is relatively small compared to the full CSI case. Lastly, we propose a new method of ordering precoded users referred to as adaptive precoding order (APO). Users are ordered according to the number of directions where interference coming from for each users. The concept of matrix condition number [16] is employed to indicate the number of directions of the interference for each user. In addition, APO has a much lower complexity than the best known ordering method [17]. We also show that by using the proposed method, we can manage the complexity of the proposed algorithm by trading it off with the performance. The complexity of the proposed scheme is significantly lower than existing scheme under the same SER performance. This feature is important when we want to accommodate a large number of users transmitting at the same time. Simulation results show that the proposed algorithms are significantly superior to the existing methods and are at most 3 dB away from an interference free channel under the general configuration. The proposed methods can be used to improve the performance and the capacity of co-working WLANs systems and cellular mobile systems.

The contribution of the paper is first, unlike [11], [13], and [14], we incorporate the effect of the receiver noise in the proposed algorithm. This consequently, boost the SINR for each user, leading to a better performance when the receiver noise is not negligible.

The second contribution is to apply the concept of uplink–downlink duality to a multi-stream multi-user multi-antenna scenario. The published papers, such as [3], and [18]–[20] consider either a multi-stream single-user or a single-stream multi-user system with a single receive antenna scenario. Thus, unlike the proposed method in this paper, these approaches cannot be used for a multi-stream multi-user MIMO transmission.

The third contribution is a novel adaptive precoding order (APO). The proposed APO is simple to implement and can be used when there are a large number of MSs. It has a much complexity compared to the standard ordering method, referred to as VBLAST ordering. [17]. The significant complexity reduction makes APO a much more practical solution.

The remainder of this paper is organized as follows. Section II presents the system model. The design of the joint transmit–receive weights is introduced in Section III. Section IV shows how to extend the proposed method to work when full CSI is not available at the MSs. Section V presents the new ordering method. Section VI compare the complexity of the proposed algorithms and APO with other existing schemes and VBLAST, respectively. Section VII shows the simulation results. The notations used in the paper are as follows. We use boldface lower case letters to denote vectors and boldface upper case letters to denote matrices. The superscripts $H$, $T$, $I$, and $\text{Diag}(a_1, \ldots, a_N)$ denote the conjugate transpose, transpose, an identity matrix, and a diagonal matrix, respectively. $\mathbb{C}^{a \times b}$ indicates a complex matrix with $a$ rows and $b$ columns. $\| x \|$ denotes the Euclidean distance and the absolute value, respectively. The operation to extract the lower triangular components of matrix $A$ to its other components to zero is defined as $\text{LoT}(A)$. The operation to extract the upper triangular components of matrix $A$ and to set its other components to zero is defined as $\text{UpT}(A)$. The operation to extract the diagonal components of matrix $A$ and to set its other components to zero is defined as $\text{DiT}(A)$. The component of matrix $A$ located in row $\mathbf{l}$ and column $\mathbf{f}$ is denoted by $[A]_{\mathbf{l}\mathbf{f}}$. Lastly, we define $\mathbf{1}$ as a column vector with all entries equal to 1.

II. SYSTEM MODEL

In this paper, we consider a multi-user MIMO system, where $K$ BSs transmit to $K$ MSs. Each BS and MS are equipped with

Authorized licensed use limited to: UNIVERSITY OF SYDNEY. Downloaded on January 10, 2010 at 23:06 from IEEE Xplore. Restrictions apply.
$N_{\text{BS}}$ and $N_{\text{MS}}$ antennas, for $k = 1, \ldots, K$, respectively. All BSs cooperate with each other to transmit $S$ symbol streams to their respective MSs via $N_{\text{BS}} = \sum_{k=1}^{K} N_{\text{BS}}$ antennas. Each of these transmissions is defined as a link. In a practical cellular network or a WLAN, if FDMA or OFDMA multiple access is used, a frequency or time slot can only be allocated to one user at a given time instant. Therefore, only one user can be served by a base station at any one time or frequency slot. Thus, to simplify the analysis, we consider a scenario where each BS transmits at a given time slot to a single mobile station. In a practical system, the proposed scheme enables a BS to communicate with multiple MSs through frequency and time division multiple access. The proposed method aims to enable $K$ such base stations deployed by different network operators in the same location to communicate simultaneously with $K$ respective mobile stations, using the same frequency band at a given time slot.

### A. Transmitter Structure

The proposed transmitter structure with joint precoding and beamforming is shown in Fig. 1(a). Let $x_k = [x_{1,k} \cdots x_{S,k}]^T$ represents the modulated signal vector, consisting of $K$-ary QAM (2-ary Quadrature Amplitude Modulation) modulated symbols, where $x_{s,k}$ is the $s$th modulated symbol stream from BS $k$ intended for MS $k$. Thus, we have a multi-stream transmission where $S$ symbol streams are transmitted from BS $k$ to MS $k$ simultaneously. The constellation points for $Q$-ary QAM are drawn from the signal set $A = \{\pm 1 \pm j, \ldots, \pm \sqrt{Q} \pm j\sqrt{Q}\}$. The modulated symbols for $K$ MSs can be then written as $\mathbf{x} = [\mathbf{x}_1^T \cdots \mathbf{x}_K^T]^T$. The transmitted symbols for each user are first permuted by a block diagonal permutation matrix $\mathbf{M}_{\text{perm}} = [\mathbf{m}_1 \cdots \mathbf{m}_K]$, where $\mathbf{m}_i = [0 \ldots 1_{S(i-1)+i} \ldots 1_{S} \ldots 0]^T$, $i = 1, \ldots, K$, is a $K \times 1$ vector with its elements $S(i-1)+i$ to $iS$ set to 1 and its other elements set to 0. The permutation operation is done by changing the location of $\mathbf{m}_i$ in $\mathbf{M}_{\text{perm}}$. By doing this, we have $K$ possible $\mathbf{M}_{\text{perm}}$. The operation of selecting $\mathbf{M}_{\text{perm}}$ is referred to as the adaptive precoding order (APO). The APO adaptively selects the precoding order of $\mathbf{x}$ that maximizes the minimum SINR of $K$ users. It selects a suitable permutation matrix $\mathbf{M}_{\text{perm}}$ to permute $\mathbf{x}$. Let $\mathbf{u} = \mathbf{M}_{\text{perm}} \mathbf{x} = [\mathbf{u}_1^T \cdots \mathbf{u}_j^T \cdots \mathbf{u}_K^T]^T$ be the permuted transmitted symbol vector, where $\mathbf{u}_j = [u_{1,j} \cdots u_{n,j} \cdots u_{S,j}]^T$. Thus, after the APO, $\mathbf{x}_k$ for MS $k$ is permuted into $\mathbf{u}_j$, which will be transmitted in link $j$. We will explain the APO in more details later in Section V.

Let us assume initially that we do not use the THP scheme. Thus, we omit THP precoding and decoding in Fig. 1. The SINR equalization module then allocates powers to each symbol in $\mathbf{u}$ in such a way that the received SINRs for all $KS$ symbols are equal. This is done by multiplying $\mathbf{u}$ with the matrix $\mathbf{P} = \text{Diag}([P_1, \ldots, P_K])$, $\mathbf{P}_j = \text{Diag}([\sqrt{p_{1,j}}, \ldots, \sqrt{p_{S,j}}])$, where $p_{k,j}$ is the downlink power allocated to the $k$th symbol in link $j$, denoted by $u_{n,j}$. The signal for $K$ links is then given as $\mathbf{P} \mathbf{u}$.

The interference in each link needs to be suppressed by multiplying the signal from each link by the transmit antenna weights of all BSs, $\mathbf{T} \in \mathbb{C}^{N_{\text{BS}} \times KS}$ and by the receive antenna weights matrix at the receiver of link $j$, $\mathbf{R}_j \in \mathbb{C}^{N_{\text{MSj}} \times S}$. The transmitted signal, thus, is given as $\mathbf{x}_T = \mathbf{T} \mathbf{P} \mathbf{u}$.

### B. Receiver Structure

The receiver for each link is shown in Fig. 1(b). Note that there is no cooperation among the receivers. Let $\mathbf{y}_j \in \mathbb{C}^{N_{\text{MSj}}}$ represents the received signal matrix for link $j$. The received signal at the $j$th receiver is

\[
\mathbf{y}_j = \mathbf{R}_j \mathbf{x}_T = \mathbf{R}_j \mathbf{T} \mathbf{P} \mathbf{u}.
\]
signal matrix for $K$ links, denoted by $\mathbf{y} = [\mathbf{y}_1^T \cdots \mathbf{y}_K^T]^T$, can be written as

$$\mathbf{y} = \mathbf{H} \mathbf{P} \mathbf{u} + \mathbf{N}$$  \hspace{1cm} (1)

where $\mathbf{H} = [\mathbf{H}_1 \cdots \mathbf{H}_j \cdots \mathbf{H}_K]^T$, with $\mathbf{H}_j \in \mathbb{C}^{N_{MS,j} \times N_{BS}}$ is the channel matrix for link $j$. $\mathbf{N} = [\mathbf{n}_1^T \cdots \mathbf{n}_j^T]^T$ and $\mathbf{T} = [\mathbf{T}_1 \cdots \mathbf{T}_K]$. $\mathbf{n}_j \in \mathbb{C}^{N_{MS,j} \times 1}$ is the noise vector for link $j$. The transmit weights for link $j$ is defined as $\mathbf{t}_j = [t_{1,j} \cdots t_{S,j}] \in \mathbb{C}^{N_{BS} \times S}$, where $t_{s,j}$ is the transmit weights vector for the $s$th symbol transmitted in link $j$. After multiplying by $\mathbf{y}$ by the receive weights matrix $\mathbf{R}$, the received signal vector becomes

$$\mathbf{y} = \mathbf{R} \mathbf{H} \mathbf{P} \mathbf{u} + \mathbf{R} \mathbf{N}$$  \hspace{1cm} (2)

where $\mathbf{R} = \text{Diag}(\mathbf{R}_1^H, \cdots, \mathbf{R}_K^H)$ is a block diagonal matrix with $\mathbf{R}_1^H, \cdots, \mathbf{R}_K^H$ as its block diagonal components. The received signal at MSs is defined as $\mathbf{y} = [\mathbf{y}_1 \cdots \mathbf{y}_j \cdots \mathbf{y}_K]^T$, $\mathbf{y}_{s,j} = [y_{s,1,j} \cdots y_{s,S,j}]^T$, $\mathbf{y}_{s,j}$ is the received signal at the input of the THP decoder for $s$th symbol transmitted in link $j$. The received weights matrix for all symbols is defined as $\mathbf{R}_j = [\mathbf{r}_{1,j} \cdots \mathbf{r}_{S,j}]$, where $\mathbf{r}_{s,j}$ is the receive antenna weights for $s$th symbol transmitted in link $j$. The received signal $\mathbf{v}_j$ can be further written as

$$\mathbf{v}_j = \mathbf{R} \mathbf{H} \mathbf{P} \mathbf{u}_j + \mathbf{R} \mathbf{N}$$  \hspace{1cm} (3)

where $\mathbf{D} = \text{Diag}(\mathbf{R}_1^H, \cdots, \mathbf{R}_K^H)$, $\mathbf{B} = \text{U} \text{Diag}(\mathbf{R}_1^H, \cdots, \mathbf{R}_K^H)$, and $\mathbf{F} = \text{LoI}(\mathbf{R}_1^H)$, $\mathbf{D} \mathbf{P} \mathbf{u}_j$ is a vector of scaled replicas of the transmitted symbols for $K$ links.

We define interlink interference as the interference between symbol streams in different links and inter-stream interference as the interference between symbol streams in the same link. $\mathbf{FP}$ is defined as the front-channel interference matrix, since the rows $j = 1, \ldots, K$ of $\mathbf{FP}$ represent the interlink interference caused by links $1, \ldots, j-1$ and interstream interference caused by symbol streams $1, \ldots, s-1$ transmitted in link $j$ to symbol stream $s$ in link $j$. Similarly, $\mathbf{BP}$ is defined as the rear-channel interference matrix, since the rows $j = 1, \ldots, K$ of $\mathbf{BP}$ represent the interlink interference caused by rear links $j+1, \ldots, K$ and the inter-stream interference caused by symbols $s+1, \ldots, S$ in link $j$ to symbol stream $s$ in link $j$.

### C. THP Precoding Structure

Note that here, the transmit–receive weights for symbol stream $s$ in link $j$ are used to suppress the interference coming from streams transmitted in link $1, \ldots, j-1, j+1, \ldots, K$ and symbols $1, \ldots, s-1, s+1, \ldots, S$ transmitted in link $j$. To further improve system performance, we pre-subtract some of the interference prior to transmission. Because of this, the transmit–receive weights matrix needs to suppress less interference. In this paper, we will use the THP scheme proposed in [8], [9] and choose to pre-subtract the front-channel interference $\mathbf{FP}$. Thus, prior to the power allocation module, THP precodes $\mathbf{u}$ into $\mathbf{v} = [\mathbf{v}_1 \cdots \mathbf{v}_j \cdots \mathbf{v}_K]^T \in \mathbb{C}^{KS}$, where

$$\mathbf{v}_j = [v_{1,j} \cdots v_{S,j}]^T.$$  \hspace{1cm} (4)

where $(\mathbf{D} \mathbf{F})^{-1}$ is used to normalize the front-channel interference with respect to $\mathbf{u}$, $\mathbf{d} = [d_{1,1} \cdots d_{S,K}]^T$, $d_{s,j} = 2\sqrt{M} \Delta$ and $\Delta$ is a complex number whose real and imaginary parts are suitable integers selected to ensure the real and imaginary parts of $v_{s,j}$ are constrained into $(\sqrt{M}, \sqrt{M})$. Here, the integers for $\Delta$ can be found by an exhaustive search across all integers [21]. $\mathbf{d}$ is an offset to ensure the energy of $\mathbf{v}$ to lie between $(\sqrt{M}, \sqrt{M})$, since the value of $\mathbf{v}$ after pre-subtraction of the front-channel interference can be very large and exceed $(\sqrt{M}, \sqrt{M})$.

Note that if $d_{s,j}$ is selected as above, adding $d_{s,j}$ to $u_{s,j}$ is equivalent to performing a modulo operation on $d_{s,j} + u_{s,j}$ [1], [14], [21]

$$u_{s,j} = \text{mod}_{\sqrt{M}}(\tilde{u}_{s,j}) = \text{mod}_{\sqrt{M}}(d_{s,j} + u_{s,j})$$  \hspace{1cm} (5)

$s = 1, \ldots, S, j = 1, \ldots, K$, where the modulo operation is defined as [5]

$$\text{mod}_{\sqrt{M}}(u_{s,j}) = u_{s,j} - \sqrt{M} \left[ \left( u_{s,j} + \sqrt{M/2} \right) / \sqrt{M} \right]$$  \hspace{1cm} (6)

for $s = 1, \ldots, S$ and $j = 1, \ldots, K$. In addition, if we apply (6) to (4), the operation in (6) actually maps $\mathbf{v}$ into the interval of $(-\sqrt{M}, \sqrt{M})$ [21]. Thus, by using the modulo operation, we implicitly find $\mathbf{d}$ that forces the THP precoded symbols $\mathbf{v}$ to lie within this interval. In addition as mentioned in [21], the variance of $\mathbf{v}$ is $E[v_{s,j}^2] = 2M/3$ and it is distributed uniformly in $(-\sqrt{M}, \sqrt{M})$. Note that $E[v_{s,j}^2] = 2(M-1)/3$. Thus, there is a power enhancement of $M/(M-1)$ due to THP. This enhancement will need to be taken into consideration when designing the transmit–receive weights and power allocation. In general, the transmit symbol energy is normalized, i.e., $E[v_{s,j}^2] = 1$. To ensure this, we scale down the $d_{s,j}$ and $r_{s,j}$ by $3/(2M)$. This is discussed in detail in [21]. By taking this into consideration, the operation of the THP precoder in (4) can then be rewritten as

$$[v_j]_j = \left\{ \begin{array}{ll} \sqrt{2M-1}[u_j], & j = 1 \\
\text{mod}_{\sqrt{M}} \left( \frac{3}{2M} [a_j]_j - a_j \right), & j = 2, \ldots, KS. \end{array} \right.$$  \hspace{1cm} (7)

where $a_j = \sum_{j=1}^S |\mathbf{M}_{THP}|_{j,j}[\mathbf{v}_j]$ and $M' = \sqrt{3/(2M)} \sqrt{M}$. $|\mathbf{M}_{THP}|_{j,j}$ denotes the $(j,j)$th component of $\mathbf{M}_{THP}$ and $[\mathbf{a}]_j$ is the $j$th component of vector $\mathbf{a}$. Note that the first line of (7) comes from the fact that when pre-coding $[v_j]_j$, there is no front-stream interference to be canceled. Thus, we can simply set $[v_j]_j$ as the normalized $a_{j,1}$.

Since now we are using the THP scheme, we are transmitting THP precoded symbol streams, $\mathbf{v}$ instead of $\mathbf{u}$. After THP precoder, the received signal $\mathbf{y}$ in (3) can be rewritten by replacing $\mathbf{u}$ by $\mathbf{v}$. By using (4) and (3), the received signal now becomes

$$\mathbf{y} = (\mathbf{D} + \mathbf{F}) \mathbf{P} \mathbf{v} + \mathbf{B} \mathbf{P} \mathbf{v} + \mathbf{R} = \mathbf{DP} \mathbf{v} + \mathbf{BP} \mathbf{v} + \mathbf{RN}.$$  \hspace{1cm} (8)
The estimates of the transmitted symbols for link \( j \), denoted by \( \hat{u}_{s,j} = [\hat{u}_{1,j} \ldots \hat{u}_{S,j}]^T \), can be recovered from \( y_{s,j} \) by applying an element-wise modulo operator in (6) to each \( y_{s,j} \), as

\[
\hat{u}_{s,j} = \begin{cases} 
\frac{y_{s,j}}{\text{mod}_{M}(y_{s,j})}, & s = 1, j = 1 \\
\text{otherwise}
\end{cases}
\]  

(9)

where \( \hat{u}_{s,j} \) is the estimate of \( u_{s,j} \). Here, the effect of offset vector \( d \) on the desired transmitted signal is removed at MS \( j \). The effect of offset \( u_{s,j} \) on the desired transmitted signal is removed at MS \( j \). Thus, MSs need to know who will be scheduled in the first link. This information can be appended in the signal preamble of wireless systems such as cellular networks, WiMax or WLANs. However, if high modulation rates are used, (e.g., \( M = 16, 64 \)), the power enhancement of THP is negligible.

The THP performance loss is given as \( M/(M - 1) \) [21] and decreases as the modulation rate increases. Thus, the THP performance loss for \( M = 16 \) and \( M = 64 \) are 0.28 and 0.07 dB, respectively. Thus, if a higher modulation rate is used, the operation is performed to the first stream to be transmitted.

III. JOINT TRANSMIT–RECEIVE ANTENNA WEIGHTS OPTIMIZATION

A. Algorithm I

In this section, we propose a joint transmit–receive antenna weights optimization method to cancel the rear-channel interference, while maximizing the SINR for each link and maintaining the same SER for all links at all times. To do this, we use the fact that 1) \( E[u_{s,j}^H r_{s,j}^H] = E[u_{s,j}^H u_{s,j}^H] = 1 \), \( s = 1, \ldots, S \), \( j = 1, \ldots, K \) and 2) the effect of offset vector \( d \) on the received signals is completely removed by the THP decoder modulo operation. By using (8), the received downlink SINR for the \( s \)th transmitted symbol in link \( j \) can then be written as

\[
\text{SINR}_{s,j}^{\text{down}} = \frac{p_{s,j} r_{s,j}^H \hat{H}_j t_{s,j} (\hat{H}_j t_{s,j})^H r_{s,j}}{z}
\]  

(10)

where

\[
z = \sum_{k=1}^{S} \sum_{i=j+1}^{K} p_{k,i} \| r_{k,i}^H \hat{H}_j t_{s,j} \|^2 + \sum_{k=1}^{S} p_{s,j} \| r_{s,j}^H \hat{H}_j t_{s,j} \|^2 + 1
\]  

(11)

\( \hat{H}_j = H_j / \sqrt{\sigma_j} \) and \( \sigma_j \) are the interference term, normalized channel matrix and the MS receiver noise for link \( j \), respectively. The operation of maximizing the minimum SINR for each symbol, while maintaining it equal for all links, can be formulated as follows:

\[
\min_{\mathbf{t}, \mathbf{p}} \max_{s,j} \text{SINR}_{s,j}^{\text{down}}
\]

subject to

\[
(1) \ t_{s,j}^H t_{s,j} = 1, (2) \ r_{s,j}^H r_{s,j} = 1
\]

\[
(3) \ 1^T \mathbf{p} = \mathbf{P}_{\text{max}}
\]  

(12)

for \( j = 1, \ldots, K \), \( s = 1, \ldots, S \). \( P_{\text{max}} \) and \( \mathbf{p} = \mathbf{p}^{2 \times 1} \), where \( \mathbf{1} \) is a vector \([1 \ldots 1]^T\) with size \( K \), \( S \), are the total BS power constraint at the cooperative transmitter and the set of downlink powers assigned to each symbol stream, respectively. Here, the objective of (12) is to maximize the minimum SINR for each stream. The first, second, and third constraints in (12) are to ensure that the transmit–receive weights are unitary vectors and the sum of the power allocated to each link does not exceed the maximum power available at the transmitter. These constraints will bound the possible solutions for \( \mathbf{R}, \mathbf{T}, \) and \( \mathbf{P} \) and ensure the convergence of (12) to a solution. Thus, the optimal solution is reached when all links attain equal SINR [3], [22], denoted by \( \text{SINR}^{\text{down}} \).

This optimization problem, however, is difficult to solve as the transmit weights and the powers for the links in (10) are entangled with each other. In other words, when the transmit weights and the power for one of the users are changed, the design of the transmit weights for other users will be affected. To solve (12), we use the uplink–downlink duality concept described in [1], [3], [4], and [23]. The authors have shown that the duality concept, we can decouple the design of the transmit weights at BSs. The duality concept helps to remove the coupling between transmit weights and downlink power by swapping the roles of the transmitter and receiver and designing the transmit weights as if they are receive weights which we refer to as the virtual receive weights. Here, the transmission from each MS to BSs constitutes one link. Thus, the interference coming to BSs from MSs does not depend on the virtual receive weights since they are located at the receiver side. That means we can design these weights individually so that the received SINR of an individual link is maximized without affecting the performance of other links. Thus, we can design the transmit weights for each user without affecting the transmit weights for other users. This process simplifies the beamforming design considerably. However, the optimum power allocations in the downlink and uplink channels are different. To apply the duality concept, we create a virtual uplink and swap the role of the transmitter and the receiver. In the virtual uplink, the receiver of a MS acts as a transmitter. The MS previously ordered in link \( j \), now transmits \( S \) virtual symbol streams \( \hat{u}_j = [u_{1,j} \ldots u_{S,j}]^T \) in link \( j \). We let the virtual received signals at the BSs be \( \hat{y}_j = [\hat{y}_1 \ldots \hat{y}_K]^T \), where \( \hat{y}_j = [\hat{y}_{1,j} \ldots \hat{y}_{K,j}]^T \) and \( \hat{y}_{s,j} \) is the \( s \)th virtual uplink received symbol transmitted in link \( j \). Let the transmitted symbols for \( K \) links be \( \hat{u} = [u_1 \ldots u_K]^T \). Here, we want to use the normalized channel matrix \( \hat{H}_j \) to represent the channel. Thus we need to scale the virtual uplink signal by multiplying the \( j \)th link with the inverse of the receiver noise, \( 1 / \sqrt{\sigma_j} \). Note that this scaling will not alter the solution. By using (8), the virtual received uplink signal can then be written as

\[
\hat{y}_{\text{up}} = \text{Diag} \left( \frac{1}{\sqrt{\sigma_1}}, \ldots, \frac{1}{\sqrt{\sigma_K}} \right) \cdot (DQ^H + B^H Q + \text{Diag} (T_1^H, \ldots, T_K^H) \hat{N})
\]  

(13)
where \( Q = \text{Diag}(Q_1 \ldots Q_K) \), \( Q_j = \text{Diag}(q_{1,j} \ldots q_{s,j}) \). \( q_{s,j} \) denotes the virtual uplink power allocated to \( s \)-th virtual uplink symbol in link \( j \). \( \mathbf{N} = [\mathbf{n}_1 \ldots \mathbf{n}_K]^T \), where \( \mathbf{n}_j \in \mathbb{C}^{N_{\text{tx}}_j} \) is the virtual receiver noise for link \( j \) at BSs, modeled as an AWGN with a zero mean and the variance \( \sigma_j^2 \). The virtual uplink configuration is shown by dashed arrows in Fig. 1. By using (13), the virtual uplink SINR for each symbol in each link, can be written as

\[
\text{SINR}_{s,j}^{\text{up}} = \frac{q_{s,j} t_{s,j}^H \mathbf{H}_{s,j}^H r_{s,j} (\mathbf{H}_{s,j}^H r_{s,j})^H t_{s,j}}{z}
\]  

(14)

where

\[
z = \sum_{k=1}^{S-1} \sum_{i=1}^{K} q_{k,i} \left\| t_{s,j}^H \mathbf{H}_{i}^H r_{s,j} \right\|^2 + \sum_{k=1}^{S-1} q_{k,i} \left\| t_{s,j}^H \mathbf{H}_{k}^H r_{s,j} \right\|^2 + 1
\]  

(15)

for \( s = 1, \ldots, S \), \( j = 1, \ldots, K \). There are three terms in the numerator of (14). The first, second, and third terms denote the interlink interference, the inter-stream interference and the normalized AWGN noise, respectively. The optimization problem can then be written as

\[
\begin{align*}
\max_{\mathbf{R}, \mathbf{T}} & \min_{s,j} \text{SINR}_{s,j}^{\text{up}} \\
\text{subject to} & \quad (1) \ t_{s,j}^H t_{s,j} = 1, \quad (2) \ t_{s,j}^H r_{s,j} = 1 \\
& \quad (3) \ t^H q = P_{\text{max}}
\end{align*}
\]  

(16)

for \( j = 1, \ldots, K \), \( s = 1, \ldots, S \) and \( q = Q^2 \mathbf{1} \). Here, the optimal solution is reached when all links attain equal SINR, denoted by \( \text{SINR}^{\text{up}} \).

Since it is very difficult to solve (16) directly, here we propose an iterative solution that splits the problem into a two-step optimization problem. The first step is to solve \( \mathbf{R} \) and \( \mathbf{T} \), when the interlink interference power to link \( j \), \( q_{s,j} \), \( s = 1, \ldots, S \), \( i = 1, \ldots, K \) are fixed to certain values, while setting \( q_{s,j} = 1, s = 1, \ldots, S \). For each link, we first find transmission spaces that have the minimum interlink interference and then use this transmission space to design the transmit–receive weights within each link. The second step is to solve \( \mathbf{q} \) in a way that equalizes SINR for all links under fixed \( \mathbf{R} \) and \( \mathbf{T} \). The process is then repeated until the SINR reaches a floor. We say that the solutions converge when the uplink SINR will not be further improved with iterations. In Appendix A, we show that the uplink SINRs, \( \text{SINR}^{\text{up}}_{s,j} \), \( j = 1, \ldots, K \), \( s = 1, \ldots, S \), obtained by solving (16), using the proposed 2-Step Optimization, is in fact increasing as the number of iterations increases. Thus here, we will stop the iteration process when there is no more improvement in uplink SINR as iteration proceeds. The two-step optimization process is illustrated in Fig. 2.

1) First Step: In the first step, we create a transmission space that has the minimum sum of the interlink interference and receiver noise. We need to find orthonormal vectors that maximize (14) without the inter-stream interference term. Here, the objective is to find a transmission space for each user so that the interference between users is minimized. The transmit and receive weights for each user are then selected from the transmission space. Note that we can neglect the inter-stream interference since we can cancel the inter-stream interference in each user’s transmission space without affecting the performance of other users. By denoting these orthonormal vectors as \( \mathbf{T}_j = \{ \mathbf{t}_{1,j} \ldots \mathbf{t}_{S,j} \} \), we can write this problem as

\[
\max_{\mathbf{T}_j} \frac{\text{trace} \left( \mathbf{H}_j^H \mathbf{R}_j \right) \left( \mathbf{H}_j^H \mathbf{R}_j \right)^H}{\mathbf{T}_j^H \mathbf{R}_j \mathbf{T}_j}
\]  

(17)

where

\[
\mathbf{R}_j = \sum_{i=1}^{S} \mathbf{H}_i^H \mathbf{R}_i \left( \mathbf{H}_j^H \mathbf{R}_j \right)^H + \sigma_j^2 \mathbf{I}
\]  

(18)

is the summation of the interlink interference and the equivalent AWGN noise in link \( j \). Note that since the interference power \( q_{s,j} \) for each link \( j \) in the first step is fixed, (17) becomes a standard generalized eigenvalue problem and can be solved using standard methods [16], [24]

\[
\mathbf{H}_j^H \mathbf{R}_j \mathbf{T}_j = \lambda_j \mathbf{H}_j^H \mathbf{R}_j \mathbf{T}_j
\]  

(19)

where \( \lambda_j = \text{diag}(\lambda_1 \ldots \lambda_S) \). \( \lambda_j \) is the \( j \)-th largest eigenvalue of \( \mathbf{H}_j^H \mathbf{R}_j \mathbf{H}_j \). \( \mathbf{T}_j \) denotes the \( S \) eigenvectors associated with these eigenvalues and represents the solution to the optimization problem of (17).

We then project the channel \( \mathbf{H}_j \) into the transmission space of link \( j \), \( \mathbf{T}_j \) to obtain an effective channel \( \tilde{\mathbf{H}}_j \)

\[
\tilde{\mathbf{H}}_j = \mathbf{H}_j \mathbf{T}_j, \quad j = 1, \ldots, K.
\]  

(20)

Here, by transmitting along \( \tilde{\mathbf{H}}_j \) we could ensure that (17) is maximized.

To cancel the inter-stream interference, while still maintaining an equal SINR across \( S \) symbols within each link and the lower triangular structure required by THP [14], we use the Geometric Mean Decomposition (GMD) method [20], [25]. The GMD is used to decompose \( \tilde{\mathbf{H}}_j \) into a lower triangular matrix with an equal diagonal component. The process is as follows. We first decompose \( \tilde{\mathbf{H}}_j \) by using the Singular Value Decomposition (SVD) [16] as follows:

\[
\tilde{\mathbf{H}}_j \mathbf{T}_j = [\mathbf{U}_S \mathbf{V}_S] \begin{pmatrix} \mathbf{D}_S & 0 \\ 0 & \mathbf{D}_S \end{pmatrix} \begin{pmatrix} \mathbf{V}_S & 0 \end{pmatrix}^T
\]  

(21)

where \( \mathbf{U}_S \) and \( \mathbf{V}_S \) consist of the first \( S \) left and right singular vectors of \( \tilde{\mathbf{H}}_j \mathbf{T}_j \). \( \mathbf{D}_S \) is a diagonal matrix with entries being the first \( S \) non-negative square roots of the eigenvalues of \( \tilde{\mathbf{H}}_j \mathbf{T}_j \). The GMD takes \( \mathbf{U}_S \), \( \mathbf{V}_S \) and \( \mathbf{D}_S \) as inputs and produces \( \bar{\mathbf{U}}_j \), \( \bar{\mathbf{V}}_j \) and \( \bar{\mathbf{D}}_S \). The GMD transforms \( \mathbf{D}_S \) into a lower triangular
matrix with equal diagonal entries, \( \mathbf{D}_S \), by rotating \( \mathbf{U}_S \) and \( \mathbf{V}_S \). This is given as

\[
\mathbf{T}_j = \mathbf{U}_j \mathbf{D}_S \mathbf{V}_S^H. \tag{22}
\]

The goal of our transmit–receive design is to create a lower triangular structure within each link. Thus, by using (22), the unitary transmit–receive weights matrix (or a single vector for a single-stream transmission) can be written as

\[
\mathbf{T}_j = \begin{cases} \frac{\bar{t}_{i,j}}{||\mathbf{t}_{i,j}||}, & S = 1 \\ \mathbf{T}_j \mathbf{V}_S D_i T^{-\frac{1}{2}} \left( \mathbf{V}_S^H \mathbf{T}_j \mathbf{V}_S \right)^{-1}, & S > 1 \end{cases} \tag{23}
\]

and

\[
\mathbf{R}_j = \bar{\mathbf{U}}_j, \quad S = 1, \ldots, K. \tag{24}
\]

It is obvious that if \( S = 1, [\mathbf{V}_S \mathbf{V}_0]^T \) in (21) is a scalar since \( \mathbf{H}_j \in \mathbb{C}^{N_{MS} \times K} \). Thus, we can use \( \bar{t}_{i,j} \) and \( \mathbf{U}_1 \) directly as the transmit–receive weights in (23) and (24).

To solve (17), however, we also need to know \( \mathbf{R}_{i=1,\ldots,j-1} \). We utilize the fact that link 1 does not have any interference coming from other links. Thus, we start by designing the transmission space and the transmit–receive weights for the first link. We then design the transmission space and the transmit–receive weights for the second link by treating the first link as interference and so on.

2) Second Step: In the second step, we use \( \mathbf{R} \) and \( \mathbf{T} \) obtained in the first step to find \( \mathbf{q} \) so that the uplink SINRs for all links are equalized. Let \( [\mathbf{M}]_{i,j} \), \( i = 1, \ldots, K \) denote the \( (i, j) \)th component of matrix \( \mathbf{M} \), where \( \mathbf{M} = D_i T (\mathbf{R}_0 \mathbf{T}) + U_j T (\bar{\mathbf{R}} \mathbf{T}), \mathbf{H} = [\mathbf{H}_1 \ldots \mathbf{H}_K]^T. \) By rearranging (14), we can write the uplink SINR for each stream as

\[
z = \frac{\bar{a}_s t_{s,j}^H \mathbf{H}^H t_{s,j}}{\text{SINR}^{UP}_{s,j}} \tag{25}
\]

for \( s = 1, \ldots, S, j = 1, \ldots, K \), where \( z \) is as defined in (15) and \( \bar{a} = (M - 1)/M \) if \( j = S = 1 \) and \( a = 1 \) otherwise. The resulting SINR for stream 1 in link 1 is always \( MF/(M-1) \) higher than the resulting SINR for other streams due to the TBP power enhancement. The power enhancement caused by THP [21] is taken into consideration by using \( \bar{a} \). By using (25), the uplink SINRs for all links can then be written in a matrix format as

\[
\begin{bmatrix} \mathbf{A}^{-1} B^H \mathbf{q} + \mathbf{A}^{-1} \end{bmatrix} = \frac{-\mathbf{q}}{\text{SINR}^{UP}} \tag{26}
\]

where \( \mathbf{A} = \text{Diag}(\bar{a}_1, \ldots, 1) \) is a \((K \times K)\) diagonal matrix, \( \mathbf{A} = D_i T (\mathbf{M}), \mathbf{B} = U_j T (\mathbf{M}), \) and \( \mathbf{M} \) is a \((K \times K)\) matrix with its component \([\mathbf{M}]_{i,j} \) equal to \([\mathbf{M}]_{i,j}^H t_{s,j}^H t_{s,j} \). By multiplying both sides of (26) with \( 1^H \), we obtain [3]

\[
\frac{1}{P_{\text{max}}} (1^H \mathbf{A}^{-1} \mathbf{A}^{-1} B^H \mathbf{q} + 1^H \mathbf{A}^{-1} \mathbf{A}^{-1} 1^H) = \frac{1}{\text{SINR}^{UP}} \tag{27}
\]

where \( 1^H \mathbf{q} = 1^H \mathbf{p} = P_{\text{THP}}. \) By defining the extended uplink power vector as \( \mathbf{q}_e = \left[ \mathbf{q}^H 1^H \right]^T \), we can then combine (26) and (27) to obtain an equations matrix given as [3]

\[
\mathbf{Q} = \begin{pmatrix} \mathbf{A}^{-1} \mathbf{A}^{-1} B^H \mathbf{q} + \mathbf{A}^{-1} \mathbf{A}^{-1} \end{pmatrix}
\]

\[
\Psi \mathbf{q}_e = \frac{\mathbf{q}_e}{\text{SINR}^{UP}} \tag{28}
\]

Hence, the optimum virtual uplink power \( \mathbf{q} \) can be obtained by selecting \( \mathbf{q}_e \) that corresponds to the maximum eigenvalue of \( \Psi \). This is the only possible solution for (28) satisfying \( d_{s,j} \geq 0 \) for \( s = 1, \ldots, S, j = 1, \ldots, K \) and \( \text{SINR}^{UP} \geq 0 \). The proof is described in detail in Theorem 1 and 2 of [26]. We then repeat the process in the first step by using \( \mathbf{q} \) found in the second step.

By using (23), (24), and (28), we can calculate the transmit–receive weights and the maximum achievable SINR in all virtual uplinks, \( \text{SINR}^{UP} \). From the uplink–downlink duality concept [3], we know that this \( \text{SINR}^{UP} \) are also achievable in the downlink transmission. As a result, the achievable virtual uplink SINR for all users, \( \text{SINR}^{UP} \) in (16), is always equal to the achievable downlink SINR in (12), \( \text{SINR}^{UP} \), as long as the total power constraints for both the virtual uplink and downlink are equal. The proof is shown in Appendix B. Thus, \( \text{SINR}^{UP} \) is maximized in that sense.

We can then use \( \mathbf{R}, \mathbf{T} \), and \( \mathbf{q} \), obtained from the iterative process, to find the optimum downlink power \( \mathbf{p} \). The optimum downlink power can be written in terms of the transmit–receive weights and the virtual uplink power, given by

\[
\mathbf{p} = \hat{\mathbf{p}} \mathbf{q} \tag{30}
\]

where \( \hat{\mathbf{p}} = (\mathbf{A}^{-1} B / \text{SINR}^{UP} - \mathbf{A}^{-1} B^H \mathbf{q} + \mathbf{A}^{-1} \mathbf{A}^{-1} 1^H) \). The proof of (30) is shown in Appendix C.

The total BS power constraint used above is not very practical. This power constraint does not take into account the power constraint for each BS in a real system. A more practical power allocation that takes into account the power constraint for each BS, is also derived. We refer to this power constraint as the per BS power constraint. We denote the maximum power available at each BS \( j \) as \( P_{\text{max},j} \) and the transmit weights as \( [t_{s,j}^H \ldots t_{s,j}^H] = [1^H] \) where \( t_{s,j} \in \mathbb{C}^{N_{MS} \times 1} \) are the entries of \( t_{s,j} \) corresponding to BS \( j \). By using (10), the downlink SINRs for all links can then be written as

\[
\mathbf{A}^{-1} (\mathbf{A}^{-1} \mathbf{B}^H \mathbf{q} + \mathbf{A}^{-1} \mathbf{A}^{-1} 1^H) = \frac{\mathbf{p}}{\text{SINR}^{UP}} \tag{31}
\]

By using (31), the power constraint at each BS can be written as

\[
\text{SINR}^{UP}(j) \mathbf{A}^{-1} (\mathbf{A}^{-1} \mathbf{B}^H \mathbf{q} + \mathbf{A}^{-1} 1^H) = P_{\text{max},j} \mathbf{q}_e \tag{32}
\]

for \( j = 1, \ldots, K \), where \( \mathbf{q}_e = \left[ \mathbf{q}_e^H 1^H \right]^T \) is a \((K \times 1)\) vector. (32) comes from the fact that the power constraint for each BS is \( \mathbf{q}_e \mathbf{p} = P_{\text{max},j} \mathbf{q}_e \). Note that \( \mathbf{q}_e \leq 1 \) is a power scaling factor for BS \( j \) to ensure the power used at BS \( j \) is always less than and equal to \( P_{\text{max},j} \). By using (31) and (32)
and defining \( \mathbf{p}_{\text{ext}} = [\mathbf{p}^T \ a_1 \ldots a_K]^T \) as a \((K_S + K) \times 1\) vector, we can obtain an equations matrix given as

\[
\begin{pmatrix}
\mathbf{A}^{-1} \mathbf{A}^{-1} \mathbf{B} & \mathbf{A}^{-1} \mathbf{A}^{-1} \\
\mathbf{T} \mathbf{A}^{-1} \mathbf{A}^{-1} \mathbf{B} & \mathbf{T} \mathbf{A}^{-1} \mathbf{A}^{-1} \mathbf{B}
\end{pmatrix}
\mathbf{p}_{\text{ext}} = \frac{\mathbf{p}_{\text{ext}}}{\text{SINR}_{\text{down}}}
\]

(33)

where \( \mathbf{A} \) is a \( K_S \times K \) matrix with \( \mathbf{A}^{-1} 1/K \) in each of its column and \( \mathbf{T} = [[t_1/P_{\max,1}] \ldots (t_K/P_{\max,K})]^T \). The maximum downlink power \( \mathbf{p} \) can be obtained by selecting \( \mathbf{p}_{\text{ext}} \) corresponding to the maximum eigenvalue of the matrix in (33). Similar to (28), the solution is unique and the proof has been shown in [26]. The complete algorithm is shown in Table I, where \( i, \text{Maxit} \) represent the iteration number and the maximum number of iterations shown, respectively. We refer to the combination of the described Algorithm I, APO, and THP precoding as AII = APO = Full CSI.

### B. Algorithm II

In this section, we propose a simplification of Algorithm I. We refer to the simplified method as Algorithm II. In Algorithm II, we use the uplink–downlink duality concept to find \( \mathbf{R}_j \) and \( \mathbf{T}_j \) while setting the virtual uplink power for \( \mathbf{s}_k \) symbol in link \( j \) as defined in Section III-A. \( q_{kj} \) to be equal. For simplicity, we assume in this paper, \( P_{\max} = K_S \). Thus, we have \( \mathbf{Q} = \text{Diag}(Q_1 \ldots Q_K) = \text{Diag}(q_{1j} \ldots q_{kj}) = \mathbf{I} \). Here, by fixing the virtual uplink power \( q_j \), we do not need to find \( \mathbf{Q} \) and \( \mathbf{T}_j=1,\ldots,K \) iteratively as in Algorithm I. By doing this, we significantly reduce the complexity of Algorithm I since 1) we do not need to compute the optimal uplink power allocation \( q_j \) 2) there is no iterative process involved in Algorithm II.

By using the fact above, the uplink SINR in (14) can now be expressed as

\[
\text{SINR}_{\text{up}}^{\text{V}} = \frac{t_{s_kj}^H \mathbf{H}_{kj}^H \mathbf{r}_{s_kj} \mathbf{H}_{kj}^H t_{s_kj}}{z}
\]

(34)

where

\[
z = \sum_{i=1}^{S} \sum_{i=1}^{K_S} \left| t_{s_kj}^H \mathbf{H}_{kj}^H t_{l_{id}} \right|^2 + \sum_{i=1}^{K_S} \left| t_{s_kj}^H \mathbf{H}_{kj}^H t_{l_{id}} \right|^2 + 1.
\]

(35)

The optimization problem in (16) can then be written as

\[
\max_{\mathbf{t}_{s_kj}} \min_{\mathbf{r}_{s_kj}} \text{SINR}_{\text{up}}^{\text{V}}
\]

subject to

\[
(1) \ t_{s_kj}^H \mathbf{t}_{s_kj} = 1, \ (2) \ t_{s_kj}^H \mathbf{r}_{s_kj} = 1.
\]

(36)

for \( s = 1,\ldots,S \) and \( j = 1,\ldots,K \). Here, the objective is to maximize the minimum SINR for individual users. Here, like (12), the solution reaches the optimal value when all streams in the link attain an equal SINR. The process of constructing transmission space for each link \( j \) by using \( \mathbf{T}_j \) can then be written as in (17), where its \( \mathbf{R}_j \) is defined in (18). Once the transmission space is obtained we can find \( \mathbf{R}_j, \mathbf{T}_j, \) and \( \mathbf{p} \) by using the same procedure described in Section III-A. The complete algorithm is tabulated in Table II. We refer to the combination of Algorithm II, the APO, and THP precoding as AII = APO = Full CSI.

### IV. LIMITED CSI AT THE RECEIVER

Algorithm I and II work by jointly designing \( \mathbf{R}_j \) and \( \mathbf{T}_j \). This joint design process requires MSs to either 1) know the complete CSI, \( \mathbf{H}_{k=1,\ldots,K} \) in order to design \( \mathbf{R}_j \) or 2) receive the information about \( \mathbf{R}_j \) from BSs. This condition increases the network cost and reduces its spectral efficiency. In this section, we want to address the limitation of Algorithms I and II by trading off the network complexity/spectral efficiency with a slight performance degradation.

We aim to eliminate the requirement of complete CSI for the algorithms described in Section III by separating the design of \( \mathbf{R}_j \) and \( \mathbf{T}_j \). We assume that \( N_{BS_k} \geq S, k = 1,\ldots,K \) and use the following assumptions. 1) If the MS \( k \) knows its own \( \mathbf{H}_k \), we can independently design the receive weights for each symbol. 2) If BSs know all CSIs, \( \mathbf{H}_{k=1,\ldots,K} \), BSs know the receive weights used by MSs.

First, we describe how to design the receive weights for \( KS \) symbols. The receive weights for a link \( j \) denoted by \( [\mathbf{r}_{1,j} \ldots \mathbf{r}_{S,j}] \), can be designed as follows:

\[
[\mathbf{r}_{1,j} \ldots \mathbf{r}_{S,j}] = \begin{cases} 
\mathbf{r}_{\text{SVD}}(\mathbf{H}_j, S), & S = 1 \\
\mathbf{r}_{\text{GMD}}(\mathbf{H}_j, S), & S \neq 1
\end{cases}
\]

(37)

where \( \mathbf{r}_{\text{SVD}} \) and \( \mathbf{r}_{\text{GMD}} \) are the SVD [16] and GMD [20] operations to extract \( S \) left eigenvectors, respectively. Thus, the \( s \)th eigenvector is denoted as \( \mathbf{r}_{s,j} \). In Algorithms I and II, the transmit–receive weights are jointly designed. Thus, we can suppress interference by using (22). Here, however, the receive weights are fixed independently in (37) and the task of cancelling interference lies with the transmit weights. We create a transmission space for symbol stream \( s \) in link \( j \) that has the minimum interlink interference under fixed receive weights. This can be represented as

\[
\max_{\mathbf{t}_{s_kj}} \frac{\mathbf{t}_{s_kj}^H \mathbf{H}_{kj}^H \mathbf{r}_{s_kj} \mathbf{H}_{kj} \mathbf{t}_{s_kj}}{\mathbf{t}_{s_kj}^H \mathbf{R}_j \mathbf{t}_{s_kj}}
\]

(38)

Authorized licensed use limited to: UNIVERSITY OF SYDNEY. Downloaded on January 10, 2010 at 23:06 from IEEE Xplore. Restrictions apply.
where $\tilde{R}_j$ is

$$\tilde{R}_j = \sum_{i=1}^{j-1} \hat{A}_i^H R_i Q_i (\hat{A}_i^H R_i)^H + \sum_{i=1}^{s-1} d_{i,j} \hat{A}_i^H r_{i,j}^H \tilde{R}_j + I. \quad (39)$$

(38) is a standard generalized eigenvalue problem as in (19) and can be solved using standard methods [24]

$$\tilde{H}_j^H \tilde{R}_j \tilde{H}_j \tilde{t}_{s,j} = \Lambda_j \tilde{H}_j \tilde{t}_{s,j}. \quad (40)$$

Once all $t_{s,j}$, $s = 1, \ldots, S$, $j = 1, \ldots, K$, are obtained, we arrange these vectors as $\tilde{T}_j = [t_{1,j}, \ldots, t_{S,j}]$. Here, $\tilde{T}_j$ consist of the orthonormal vectors defining the transmission space for link $j$. The channel $H_j$ is then projected into the transmission space of link $j$, $T_j$, to obtain an effective channel $\tilde{H}_j$

$$\tilde{H}_j = \tilde{H}_j \tilde{T}_j, \quad j = 1, \ldots, K. \quad (41)$$

In the previous algorithms, since the transmit–receive weights are jointly designed, we can use left and right eigenvectors as in (23). However, since $R_j$ is fixed, we can only use the right eigenvectors to triangularize the channel, $\tilde{H}_j$. We use the QR decomposition [16] to find these eigenvectors and arrive at

$$R_j^H \tilde{A}_j \tilde{T}_j = D_S \tilde{V}_S^H \quad (42)$$

where $\tilde{D}_S$ and $\tilde{V}_S^H$ are a lower triangular matrix and a unitary matrix obtained by applying the QR decomposition [24] to $(R_j \tilde{H}_j \tilde{T}_j)^H$. The unitary transmit–receive weights matrix (or a vector for a single-stream transmission) can be written as

$$T_j = \begin{cases} \tilde{r}_{i,j}^H, & S = 1 \\ \tilde{T}_j \tilde{V}_S \text{DiT}^{-\frac{1}{2}} \left( \tilde{V}_S^H \tilde{T}_j \tilde{T}_j \tilde{V}_S \right), & S > 1. \end{cases} \quad (43)$$

For multi-stream transmission, we note that to find the transmit weights for each link, we need to solve (38) $S$ times per link. Here, we propose a simpler method to compute the transmission space in (38). We want to compute $T_j = [t_{1,j}, \ldots, t_{S,j}]$ in a single step. To do that, we rewrite (38) as

$$\begin{aligned} &\max_{T_j} \text{trace} \frac{\tilde{T}_j^H \tilde{H}_j \tilde{T}_j}{\tilde{T}_j^H \tilde{R}_j \tilde{T}_j} \quad (44) \\
\end{aligned}$$

where we use $\tilde{R}_j$ defined in (18) instead of (39).

We also observe that the number of receive weights available in link $j$, depends on the rank of $H_j$, $\tilde{S}$. Thus, for a single-stream transmission ($S = 1$), $r_{s,j}$ can be selected from $\tilde{S}$ possible receive weights. We want to utilize this fact to obtain a better performance for a single-stream transmission. We will select the combination of transmit and receive weights that gives the maximum gain

$$\max_{\tilde{r}_{s,j} \in \tilde{T}_j} \left| \left| \tilde{r}_{s,j} \tilde{H}_j \tilde{t}_{s,j} \right| \right|, \quad s = 1, \ldots, \tilde{S} \quad (45)$$

where each $t_{s,j}$ is computed by using (40) for a fixed $r_{s,j}$.
concentrated in a few directions. We use the concept of the matrix condition number to indicate the level of the interference which is defined as [24]

$$k(U) = \frac{\lambda_{\text{max}}(U)}{\lambda_{\text{min}}(U)}$$  \tag{46}

where $U$ is an arbitrary matrix. $\lambda_{\text{max}}$ and $\lambda_{\text{min}}$ are the maximum and the minimum eigenvalues of $U$, respectively. Thus, a large $k(U)$ indicates that the vectors in $U$ are concentrated in a single direction while a small $k(U)$ means that the vectors of $U$ are scattered. From Section III, we know that the beamforming weights for the user in link $j$ need to cancel interference in links $1, \ldots, j-1$. That means that the interference in the link $j$ contains more interference than links $1, \ldots, j-1$. To minimize the interference (maximize the minimum SINR), we need to schedule in link $j$, a user that has a higher condition number than the users scheduled in links $1, \ldots, j-1$. To apply this concept mentioned above, we first create an interference channel matrix for the link $j$, denoted by $\tilde{H}_j$, as follows:

$$\tilde{H}_j = [\tilde{h}_1 \ldots \tilde{h}_{j-1} \tilde{h}_{j+1} \ldots \tilde{h}_K]^T.$$  \tag{47}

Let $(c_1, \ldots, c_j, \ldots, c_K)$ be the adaptive precoding order (APO) for users prior to the THP, where $c_j$ denotes $k$th user ordered in link $j$. The precoding is selected by using

$$(c_1, \ldots, c_j, \ldots, c_K) = \text{sort} \left(k(\tilde{H}_1), \ldots, k(\tilde{H}_K)\right)$$  \tag{48}

where sort($\cdot$) is a function used to order the matrix condition in the ascending order. If user $k$ is ordered as the $j$th element of the $(k(\tilde{H}_1), \ldots, k(\tilde{H}_K))$, then $c_j = k$. To update the permutation matrix, $M_{\text{perm}}$, we simply replace the $k$th column of the permutation matrix $M_{\text{perm}}$ with $j$th column of $M_{\text{perm}}$, for $k = 1, \ldots, K$.

VI. ADVANTAGES OF THE PROPOSED OVER OTHER KNOWN SCHEMES

In this section, we discuss the advantages of the proposed scheme over other existing schemes. We compare the proposed methods with the linear precoding scheme in [11], [12] and nonlinear THP precoding schemes in [13], [14]. In order to have a fair comparison with the proposed method, a power allocation scheme that assigns power in such a way that SINRs for all links are equalized [27], is applied in these schemes. In the original schemes proposed in these papers, there is no power allocation in [12] while the water-filling power allocation used in [11], [13], and [14] tends to assign more power to stronger links and less power to weaker links. Under these two circumstances, the links will have unbalanced SINRs and the performance of the weaker link will dominate the overall system performance.

The main differences between the proposed algorithms and the schemes [11]–[14] are as follows.

1) the relaxation of the zero forcing and orthogonality constraint for the transmit weights.

2) Unlike [11], [12] where transmit–receive weights are used to suppress both the front-channel and rear-channel interference and [13], [14] where no iteration is performed to improve transmit–receive weights, the proposed method suppresses the rear-channel and front-channel interference using THP and transmit–receive weights, respectively. The transmit–receive weights design are then improved iteratively.

3) The effect of the receiver noise is taken into consideration when designing the transmit–receive weights and allocating power.

4) The proposed methods can work when the receiver for link $j$ only knows its own CSI, $H_j$.

5) Unlike [13], [14], and [11], [12], we do not require the constraint of $N_{\text{BS}} > (K-1)SN_{\text{MS}}$ and $N_{\text{BS}} \geq KS$ to be satisfied, respectively. Here, we assume that each MS has the same number of antennas $N_{\text{MS}} = N_{\text{MS}} = \ldots = N_{\text{MS}}$. The fifth difference is a definite advantage. For example, to support say five users where each user has a single data stream $S = 1$, and each user has four antennas $N_{\text{MS}} = 4$, the proposed method can use $\leq 5$ transmit antennas, while [13], [14], and [11], [12] need $\geq 12$ and $\geq 5$ transmit antennas, respectively.

The computational complexities in terms of the number of floating point operations (flops) for the proposed schemes and the schemes [11]–[14] are listed in Table V, where $L = 1$ when $S \neq 1$ and $L = N_{\text{MS}}$ when $S = 1$. In addition, $C = 1$ and $\tilde{C} = K$ for the proposed methods under the total BS and per BS power constraints, respectively. $\text{Maxit}_{\text{perm}}$ in Table V indicates The number of iterations for the scheme in [12]. To analyze this algorithm, we make a practical assumption that the number of transmit antennas is always greater than the number of receive antennas at each MS, i.e., $N_{\text{BS}} > N_{\text{MS}}$. This is always true in practical cooperative wireless networks. We then denote the complexity order of the proposed algorithms as $C_x$, where $x$ denotes which scheme is used. Now, from Table V, we can obtain the complexity order of our algorithms, $C_{\text{AI-APo-Full CSI}} = O(\text{Maxit} \cdot \text{max}(30K_{\text{BS}}^3,30(KS + \tilde{C})^3))$, $C_{\text{AI-APo-Limited CSI}} = O(\text{Maxit} \cdot \tilde{L} \cdot \text{max}(30K_{\text{BS}}^3,30(KS + \tilde{C}))$, and $C_{\text{AI-APo-Full CSI}} = O(\text{max}(30\tilde{L}K_{\text{BS}}^3,30(KS + \tilde{C}))$. By using Table V, the complexity order of methods in [11]–[14] is given as $O(18K_{\text{BS}}^3)$ and $O(9(K^3/2)K_{\text{BS}}^3)$, $O((9K^3/2)K_{\text{BS}}^3)$, and $O(\text{Maxit}_{\text{perm}}18K_{\text{BS}}^3)$, respectively. Thus, for large $K$, the proposed algorithms are more complex than schemes in [11] but less complex than [12]–[14]. However, as we will see in a later section, the performance of our proposed algorithms outperforms the schemes in [11]–[14].

Now, we compare the proposed ordering method referred to as APO, with the optimum ordering method referred to as VBLAST ordering [17] used in nonlinear methods [13], [14]. In [17], the authors proposed the idea of the Myopic Optimization method and prove that this ordering is optimal. With this ordering, to reach the maximum SINR, they only need to search $\approx (K^2/2)$ possible orderings. This ordering however, is too complex for the proposed interference cancellation schemes or any other known schemes as shown in Table V. The first term in the computational complexity for AI-APo-Full CSI, AII-APo-Full CSI, AI-APO-Limited CSI,
TABLE V  
COMPUTATIONAL COMPLEXITY OF LINEAR AND NONLINEAR PRECODING ALGORITHMS (IN FLOPS)

<table>
<thead>
<tr>
<th>Schemes</th>
<th>Computational Complexity in flops</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm I-APO-Full CSI</td>
<td>((K - 1)N_{BS}\overline{N}<em>{BS}^2 + \text{Maxit} \cdot (30K</em>{BS}^2 + 2K_{BS}N_{BS}N_{MS} + 2K_{BS}^2N_{BS}^2 + 4KSN_{BS}^2 + 8KSN_{BS}N_{BS} + 9K^2 + N_{BS}S + K^2 + 30KS + C^2 + 30K_{BS} + C^2 + K^2N_{MS}N_{BS}))</td>
</tr>
<tr>
<td>Algorithm II-APO-Full CSI</td>
<td>((K - 1)N_{BS}\overline{N}<em>{BS}^2 + (30K</em>{BS}^2 + 2K_{BS}N_{BS}N_{MS} + 2K_{BS}^2N_{BS}^2 + 4KSN_{BS}^2 + 8KSN_{BS}N_{BS} + 9K^2 + N_{BS}S + K^2 + 30KS + C^2 + 30K_{BS} + C^2 + K^2N_{MS}N_{BS}))</td>
</tr>
<tr>
<td>Algorithm I-APO-Limited CSI</td>
<td>((K - 1)N_{BS}\overline{N}<em>{BS}^2 + 8KSN</em>{BS}^2 + 9S^3 + N_{BS}S + 8^3 + 30K_{BS}^2N_{BS}N_{MS} + 2K_{BS}N_{BS}^2 + 2K_{BS}^2N_{BS} + 30KS + C^2 + K^2N_{MS}N_{BS})</td>
</tr>
<tr>
<td>Algorithm II-APO-Limited CSI</td>
<td>((K - 1)N_{BS}\overline{N}<em>{BS}^2 + 4KSN</em>{BS}^2 + 9S^3 + N_{BS}S + 8^3 + (30K_{BS}^2 + 2K_{BS}N_{BS}N_{MS} + 2K_{BS}N_{BS}^2 + 30KS + C^2 + K^2N_{MS}N_{BS}))</td>
</tr>
<tr>
<td>Pan et. al. [12]</td>
<td>Maxit \text{Pan}(4N_{BS}^2N_{BS} + 8N_{BS}^2N_{BS}^2 + 9S^3 + K^2N_{BS} + 8KN_{BS}^2 + 9KN_{BS}^2)</td>
</tr>
<tr>
<td>Spencer et. al. [11]</td>
<td>(8K\overline{N}<em>{BS}^2N</em>{BS} + 16KN_{BS}^2N_{BS}^2 + 18KN_{BS}^2 + (\sum_{k=1}^{K} 4^3S^2N_{BS} + 8S^3N_{BS} + 9KN_{BS}^2 + K^2N_{BS}N_{MS}N_{BS}))</td>
</tr>
<tr>
<td>Liu et. al. [14]</td>
<td>(\frac{K^2}{2}(4K^2N_{BS}N_{BS}^2 + 8KSN_{BS}^2N_{BS} + 9K^2N_{BS}^2 + N_{BS}N_{BS})K + \left(\sum_{k=1}^{K - 1} 3K^2N_{BS}^2N_{BS} - \frac{1}{3}N_{BS}N_{BS}^2 + 2N_{BS}^2N_{BS} + 2K^2N_{MS}N_{BS}</td>
</tr>
</tbody>
</table><p>ight)) |
| Foschini et. al. [13]       | (\frac{K^2}{2}(4K^2N_{BS}^2N_{BS} + 8KSN_{BS}^2N_{BS} + 9K^2N_{BS}^2 + (\sum_{k=1}^{K - 1} 3K^2N_{BS}^2N_{BS} - \frac{1}{3}N_{BS}N_{BS}^2 + 2N_{BS}^2N_{BS} + 2K^2N_{MS}N_{BS}) + K^2N_{MS}N_{BS})) |</p>

and All-APO-Limited CSI is the computational complexity of APO. Let \(C_{\text{r}}\) denote the complexity order of the proposed algorithms without considering the APO, where \(x\) again denotes the algorithm name. The complexity order of the proposed algorithms can be expressed as \((K - 1)N_{BS}\overline{N}_{BS}^2 + C_{\text{r}}\), where we again assume that each user/MS has \(N_{MS}\) receive antennas. The complexity order of the proposed algorithms when VBLAST ordering is used is \((K^2/2)C_{\text{r}}\). Thus, when the following condition must be met

\[ 30K\overline{N}_{BS}^2 > 2N_{MS}\overline{N}_{BS} \Rightarrow \overline{N}_{BS} > \frac{N_{MS}}{15K\overline{L}}. \]

This condition in (50) is very realistic. As an example, we have two users each receiving two symbol streams transmitted from three BSs each equipped with two antennas. By using (50), we could see that APO will always be less complex than VBLAST ordering as long as the number of receive antennas for each MS is less than 180. Thus, we can conclude that APO is always less complex than VBLAST ordering on any practical condition. In addition to the computational advantage mentioned above, as we will see in a later section, the APO is at most 1 dB worse than the optimal ordering at the SER = 10^{-4}. Thus, it is not wise to increase the computational complexity of the transmitter enormously just to achieve a very small gain.

VII. NUMERICAL RESULTS AND DISCUSSION

Monte Carlo simulations have been carried out to evaluate the performance of the four proposed algorithms. We investigate their performance and compare it with [11], [13], and [14] and with interference free performance. Here, an interference free performance is defined as the performance of any random single link \(i\) assuming there is no interference from other links at all. In this case, the received signal of the cooperative transmission system is given as \(y_i = \mathbf{h}_i^H \mathbf{t}_i + \mathbf{n}_i\), where \(\mathbf{t}_i\) and \(\mathbf{n}_i\) are the left and right eigenvectors associated with the maximum eigenvalue of \(\mathbf{H}_i^H\). To generate an interference-free performance for multi-stream transmission with \(S\) symbols transmitted in each link, we use the left and right eigenvectors associated with the largest eigenvalues of \(\mathbf{H}_i^H\) found by using the SVD. We then maximize the minimum SINR of \(S\) symbols by applying the power allocation method given in [27].

For convenience, we will use the notations \((N_{BS}, N_{BS}, K, S)\) in all figures to denote the number of transmit antennas of BSs, the number of receive antennas per MS, the number of users, and the number of symbols transmitted per user in the network, respectively. In the simulations, for simplicity, we assume \(N_{BS} = 1, \ldots, K = N_{BS}\) and the receiver noise for \(K\) links are equal as \(\sigma_1 = \sigma_2 = \ldots = \sigma_K\). A Rectangular 64-QAM (\(M = 64\)) modulation is used for all transmissions. To simulate the wireless channel, we set the elements of each \(\mathbf{H}_i\) channel matrix as a complex Gaussian variable with a zero mean and unit variance. In all simulations, we fix the signal-to-noise-ratio of each THP precoded symbol to be \(\text{SNR} = 2E[|v_{i,k}|^2]/\sigma^2\), where \(E[|v_{i,k}|^2]\) is normalized to 1, \(P_{\text{trans}} = KS\) and \(P_{\text{max}} = 1, \ldots, K = S\). In all simulation cases, Algorithms I and II and the comparison schemes [11]–[14] are all simulated under the total BS power constraint unless stated otherwise. Note that we set the number of iterations for the scheme in [12] to be 11, Maxit\text{Pan} = 11.

A. Convergence Behavior

Figs. 3 and 4 show the convergence characteristics of AI-APO-Full CSI for (6,2,3,1) and (6,2,3,2) systems. The cumulative distribution function of SINR\text{down} for a large number of channel realizations are shown for different number of iterations Maxit\text{Pan} = 1, \ldots, 5 at SINR is equal to 21, 24, and 27 dB, respectively. The number next to each plot indicates the number of iterations used. We could see that the downlink SINR, SINR\text{down} increases as the number of iterations increases. Only three iterations are required for the SINRs of the
two systems to converge. That means, there are no meaningful improvement can be obtained by increasing the number of iterations higher than 3. This is a definite advantage since it shows that the proposed method converges very fast. An interesting observation here is that the SINR convergence seems to be independent of system configurations. The three systems can be seen to converge at the same time after three iterations. As the required number of iterations does not depend on the system configuration, it is possible to then fix the number of iterations, \( M_{\text{MAX}} \). Note that, due to the space limitation, we are not able to show the cumulative distribution function of SINR\(_{\text{known}}\) for AI-APO-Full CSI under the per BS power constraint and AI-APO-Limited CSI under the total BS and per BS power constraints. Our simulation results however indicate the same trend as AI-APO-Full CSI. Thus, in all further simulations, we set \( M_{\text{MAX}} \) for the AI-APO-Full CSI and AI-APO-Limited CSI to 3.

**B. Average SER Performance**

Here, we studied the average SER performance of each link under the total BS and per BS power constraints. Note that from Sections III and IV, we know that the SER performance of each link will be identical, since the downlink SINR for each link is the same. Here, we show how we can manage the complexity order of the proposed methods by trading off the performance with complexity and more realistic assumptions such as a per BS constraint and limited CSI at the receiver. We first consider a (6,2,3,1) system. In Fig. 5, the SER performances for the four proposed algorithms are compared with schemes in [11]–[14] at SER = 10^{-4}. Note that all existing schemes [11]–[14] 1) require full complete CSIs at both the receiver and transmitter, 2) except [13] which works only for the total BS power constraint. To measure how much the computational complexity improvement of the proposed scheme over comparison schemes [11]–[14], we use \( \Delta C \), obtained by subtracting the computational complexity of other schemes with the proposed scheme and dividing it with the computational complexity of the proposed scheme.

As can be seen from Fig. 5, AI-APO-Full CSI and AI-APO-Limited CSI under the total BS power constraint significantly outperform the existing schemes [11]–[14] by at least 2 dB. Most importantly, the complexity order of AI-APO-Full CSI and AI-APO-Limited CSI under the total BS power constraint is approximately 120% and 560%, respectively, lower than the best comparison scheme [12]. Thus, we could achieve a better performance with the same complexity order. Note also that the performance of the proposed method under limited CSI is also better than [11]–[14] while the complexity order of the proposed schemes is managed to be 10% and 230% lower than [12] for AI-APO-Limited CSI and AI-APO-Limited CSI, respectively. Furthermore, the performances of AI-APO-Full CSI and AI-APO-Limited CSI under the per BS power constraint outperform [11], [13], [14] for all SNR values and [12] for a high SNR. This performance gain is achieved without increasing the...
complexity order of the proposed method. This is another definite advantage of our proposed method since 1) in a practical network each BS or antennas might have its own power constraint, and 2) these performance gains are achieved without increasing the complexity order of the proposed method. In addition, the performance of the best proposed scheme AI-APO-Full CSI under the per BS and total BS power constraints are 2 dB and 0.5 dB, respectively, away from an interference free performance. Apart from that there is no performance loss with AI-APO-Full CSI compared to AI-APO-Full CSI under the total BS power constraint.

Now, let us consider the performance of a (6,2,3,2) system. In this case, BSs have six transmit antennas and BSs transmit two symbols each to three users. The SER performance is shown in Fig. 6. We compare AI-APO-Full CSI and AI-APO-Full CSI with [14] under a total BS power constraint, since [14] is better than [11]–[13] in terms of SER performance. AI-APO-Full CSI and AI-APO-Full CSI outperform [14] by 10 dB and are only 3 dB away from an interference free performance at SER = 10^{-3}. Unfortunately, this improvement comes with a cost for AI-APO-Full CSI since its complexity is now 20% higher than for the scheme in [14]. The complexity of AI-APO-Full CSI is 140% lower than for the scheme in [14]. Here, we also note that the performance of AI-APO-limited CSI under the per BS power constraint in Fig. 6 is worse than for the scheme in [14]. This is expected since the per BS power constraint is more stringent than the total BS power constraint.

Notice also that the performance of AI-APO-Full CSI, AI-APO-Full CSI, and AI-APO-Limited CSI under a total BS power constraint in Figs. 5 and 6 differ by less than 1 dB. This is desirable since we only need to give up 1 dB for not requiring full CSI at each receiver. Apart from that, we could also see from Figs. 5 and 6 that the performance of AI-APO-Limited CSI is worse than AI-APO-Full CSI, AI-APO-Full CSI, and AI-APO-Limited CSI. The reason is because the transmit–receive weights for link $j$ is calculated under the fixed uplink power allocation matrix $Q = I$ and we do not perform joint optimization of weights and power allocation $Q$ as in Algorithm I. It is also interesting to note that in Figs. 5 and 6 is that when the per BS power constraint is used for Algorithm I and II, the performance is only degraded by less than 2 dB as compared to the performance of Algorithm I and II under the total BS power constraint for the two system configurations. Thus, the proposed method is able to take into account practical constraints of real systems. In addition, we also simulated the case where algorithm I applies VBLAST ordering under a total BS power constraint, denoted by AI-VBLAST-Full CSI. Even though this scheme outperforms AI-APO-Full CSI by 1 dB, its complexity is $K^2/2$ higher, making it less desirable for a practical implementation.

**VIII. Conclusion**

In this paper, we exploit the uplink–downlink duality concept to design a cooperative multi-stream multi-user MIMO downlink transmission scheme employing precoding and beamforming under a total BS and a per BS power constraints. THP and transmit–receive weights optimization are used to cancel the interference. SINR equalization and APO are applied to achieve SER fairness among different users and further improve the system performance, respectively. We proposed an iterative method to optimize the transmit–receive weights. Furthermore, we reduce the complexity significantly by allowing a slight performance degradation. We then extend these methods to work in a situation where the receiver only knows its own CSI. The error performance for three sets of system parameters $(N_{BS}, N_{MS}, K, S)$ is shown. The proposed methods outperform the existing schemes [11]–[14] by at least 2 dB and 10 dB for (6,2,3,1) and (6,2,3,2) systems when a total BS power constraint and full CSIs assumption at the receiver are applied. In addition, the proposed method for (6,2,3,1) and (6,2,3,2) systems is only 0.5 dB and 3 dB away from an interference-free performance. Unlike the existing schemes, the proposed method can be extended to work under more practical constraints such as a per BS power constraint and limited CSI with a slight performance degradation. The performance degradation for not requiring full CSI at the receiver and using a per BS power constraint is less than 1 dB and 2 dB, respectively. In addition, the application of APO to order users in precoding has been shown to degrade the performance of the proposed methods by at most 1 dB compared to the VBLAST ordering. The proposed APO, however, is shown to be significantly less complex than VBLAST ordering, when used in the proposed methods. Lastly, we have shown that by using the proposed method, we can manage the complexity of the proposed algorithm by trading it off with the performance. It is also shown the complexity of the proposed scheme is at least 100% lower than the complexity of the existing schemes with the best SER performance. Thus, the proposed method can be applied to improve the performance and capacity of coworking WLANs and cellular mobile networks.

**APPENDIX A**

**PROOF FOR TWO-STEP OPTIMIZATION**

The iterative solution proposed in Section III-A essentially splits the optimization problem in (16) into a two-step optimization. The first step is to solve $\mathbf{R}$ and $\mathbf{T}$, when the interlink in-
terference power to link \( j \) from link \( i \), \( q_{si, s', j} \), \( s = 1, \ldots, S, s' = 1, \ldots, K \) is fixed. Under this condition, (16) can be written as

\[
\hat{f}_1(q_{si, s', j}) = \max_{r_{si, s', j}, t_{si, s'}} \min_{s} \text{SINR}_{up}^{s, j}^{(a-1)}
\]

subject to

\[(1) t_{si, s', j}^H t_{si, s'} = 1, \quad (2) t_{si, s', j}^H r_{si, s'} = 1 \quad (51)\]

for \( j = 1, \ldots, K \) and \( s = 1, \ldots, S \). Here, the virtual uplink powers for link \( i \), \( q_{si, s', j} \), used to calculate \( \text{SINR}_{up}^{s, j} \), (14) are set to fixed values. We perform this optimization process by using a function, \( f_1(q_{si, s', j}) \), where \( r_{si, s', j} \) and \( t_{si, s'} \) are its optimization variables for a given \( q_{si, s', j} \), \( j = 1, \ldots, K, \ s = 1, \ldots, S \). The solution of (51) is obtained by first, finding transmission spaces that have the minimum interlink interference for each link by using (17). As it was mentioned in [28], this solution is optimum for a given \( q \) since it is equivalent to the scaled MMSE beamforming.

Thus, here we can conclude that (17) is optimum. These transmission spaces are then used to design the transmit–receive weights, within each link, to maximize the SINR by using GMD. Note that the GMD operation has been shown in [20] and [25] to also give a unique solution that maximizes the minimum achievable SINR for each stream within a single link for a given transmission space. Thus, its solution is also optimum for a given transmission space. Thus, we can conclude that the solution of (51) is optimum.

The second step is to solve \( q \) in a way that equalizes SINR for all links under fixed \( R \) and \( T \). Under this condition, the optimization problem can be written as

\[
\hat{f}_2(r_{si, s', j}, t_{si, s'}) = \max_{q_{si, s', j}} \min_{s} \text{SINR}_{up}^{r_{si, s', j}, t_{si, s'}}^{(a-1)}
\]

subject to

\[(1) 1^T q = P_{\text{max}} \quad (52)\]

for \( j = 1, \ldots, K, s = 1, \ldots, S \), and \( q = Q \Phi 1 \) as defined in Section III-A. Here, the transmit–receive weights for all users are fixed when solving (52). We denote this optimization process with a function, \( \hat{f}_2(r_{si, s', j}, t_{si, s'}) \) with \( q_{si, s', j} \) as its optimization variable, optimized for given \( r_{si, s', j} \) and \( t_{si, s'} \). To solve (52), we use (28). It has been shown in Theorem 1 and 2 of [26] that the solution of (28) is unique and optimum for a given transmit and receive weights.

To prove a monotonic convergence in the Two-Step Optimization, we first define SINR as a cost function of transmit–receive weights and power allocation vectors [3]

\[
\text{SINR}_{up}^{(a-1)} = f_3(R^{(a)}, T^{(a)}, q^{(a)}) \quad (53)
\]

where \( x \) is an auxiliary vector which has no physical meaning and is used only to facilitate the proofing. \( \Psi \) and \( \phi_e \) are as defined in (28). Given the power allocation vector \( q^{(a)} \) obtained by using (52) in iteration \( a \), we then compute in iteration \( a + 1 \), the optimum transmit–receive weights \( R^{(a+1)}, T^{(a+1)} \) by using (51). This gives the SINR value \( \text{SINR}_{up}^{(a+1)} \). That means

\[
f_3(R^{(a)}, T^{(a)}, q^{(a)}) \leq f_3(R^{(a+1)}, T^{(a+1)}, q^{(a)}) \quad (54)
\]

Therefore, the SINR value obtained at iteration \( a = 1 \) is always greater or equal to the SINR in iteration \( a \). This is so since \( R^{(a)} \) and \( T^{(a)} \) are not the optimum transmit–receive weights for iteration \( a + 1 \). Similarly, we have

\[
f_3(R^{(a)}, T^{(a)}, q^{(a-1)}) \leq f_3(R^{(a)}, T^{(a)}, q^{(a)}) \quad (55)
\]

for all links under fixed power. We denote this optimization process with a function, \( f_2(r_{si, s', j}, t_{si, s'}) \) with \( q_{si, s', j} \) as its optimization variable, optimized for given \( r_{si, s', j} \) and \( t_{si, s'} \). To solve (52), we use (28). It has been shown in Theorem 1 and 2 of [26] that the solution of (28) is unique and optimum for a given transmit and receive weights.
By equating the terms on the right hand side of (57), we can obtain

\[ p = \left( \frac{A}{\text{SINR}^\text{down}} - B \right)^{-1} \left( \frac{A}{\text{SINR}^\text{up}} - B^H \right) q = \hat{p} q. \]  

(58)

Thus, the downlink power can be obtained from (58) once the virtual uplink power and the transmit–receive weights are calculated.

**ACKNOWLEDGMENT**

The authors would like to thank the anonymous reviewers for their suggestions that helped to improve the quality of the paper.

**REFERENCES**


