10

Analysis and Design of Tomlinson-Harashima Precoding for Multiuser MIMO Systems

Xiang Chen, Min Huang, Ming Zhao, Shidong Zhou and Jing Wang
Research Institute of Information Technology, Tsinghua University
Beijing, China

1. Introduction

The multiuser multiple-input-multiple-output (MIMO) downlink has attracted great research interests because of its potential of increasing the system capacity(Caire & Shamai, 2003; Vishwanath et al., 2003; Viswanath & Tse, 2003; Weingarten et al., 2006). Many transmitter precoding schemes have been reported in order to mitigate the cochannel interference (CCI) as well as exploiting the spatial multiplexing of the multiuser MIMO downlink. Tomlinson-Harashima precoding (THP) has become a promising scheme since the successive interference pre-cancelation structure makes THP outperform linear precoding schemes (Choi & Murch, 2004; Zhang et al., 2005) with only a small increase in complexity. Many THP schemes based on different criteria have been reported in the literature(Doostnejad et al., 2005; Joham et al., 2004; Mezghani et al., 2006; Schubert & Shi, 2005; Stankovic & Haardt, 2005; Windpassinger et al., 2004), in which one is the zero-forcing (ZF) criterion and the other is the minimum mean square error (MMSE) criterion. This chapter will consider the above two criteria based THP schemes' analysis and design, respectively.

For the ZF-THP scheme, initial research mainly focuses on the scenarios that each receiver is equipped with a single antenna (Windpassinger et al., 2004), where there exists only the transmit diversity, but without any receive diversity. Presently, the receive diversity due to multiple antennas at each receiver is taken into account (Stankovic & Haardt, 2005; Wang et al., 2006).In these literatures, it is commonly assumed that the total number of receive antennas is less than or equal to that of transmit antennas. Under this assumption, firstly the layers are divided into groups which correspond to different users, and then the dominant eigenmode transmission is performed for each group. Hereby, this kind of schemes is regarded as *per-user* processing. Actually, it is more common in the cellular multiuser downlink systems that the number of users is not less than that of transmit antennas at the base station (BS), which is investigated as the generalized case with THP in this chapter. In order to avoid complicated user selection and concentrate on the essential of transceiver filters design, our consideration is limited into a unique case that the number of users equals the number of transmit antennas, denoted as M. Besides, it is assumed that the channels of these M users have the same large-scale power attenuation.[1] In this case a so-called *per-layer* processing can be applied by the regulation that each user be provided with only one

[1] In practice, when the number of users is large enough, we can find M users whose large-scale power attenuations are approximately equal by scheduling.

subchannel and all the M users be served simultaneously. Based on the criterion of maximum system sum-capacity, two per-layer joint transmit and receive filters design schemes can be employed which apply receive antenna beamforming (RAB) and receive antenna selection (RAS), respectively. Through a theorem and two corollaries, the differences of the equivalent channel gains and capacities between these two schemes are developed. Theoretical analysis and simulation results indicate that compared with linear-ZF and per-user processing, these per-layer schemes can achieve better rate region and sum-capacity performance.

For the MMSE-THP scheme, we address the problem of joint transceiver design under both perfect and imperfect channel state information (CSI). The authors in (Joham et al., 2004) designed THP based on the MMSE criterion for the MISO system where the users are restricted to use a common scalar receiving weight. This restriction was relaxed in (Schubert & Shi, 2005), i.e., the users may use different scalar receiving weights, where the authors used the MSE duality between the uplink and downlink and an exhaustive search method to tackle the problem. The problem of joint THP transceiver design for multiuser MIMO systems has been studied in (Doostnejad et al., 2005) based on the MMSE criterion. However, a per-user power constraint is imposed, which may not be reasonable in the downlink. Morevoer, only the inter-user interference is pre-canceled nonlinearly, whereas the data streams of the same user are linearly precoded. The work of (Doostnejad et al., 2005) has been improved in (Mezghani et al., 2006) under a total transmit power constraint, where the users apply the MSE dualtiy between the uplink and downlink and the *projected gradient* algorithm to calculate the solution iteratively. Again, only the inter-user interference is pre-subtracted.

The above schemes have a common assumption that the BS, has perfect CSI. In a realistic scenario, however, the CSI is generally imperfect due to limited number of training symbols or channel time-variations. Therefore, the robust transceiver design which takes into account the uncertainties of CSI at the transmitter (CSIT) is required. Several robust schemes have been proposed for THP in the multiuser MISO downlink, which can be classified into the worst-case approach (Payaro et al., 2007; Shenouda & Davidson, 2007) and the stochastic approach (Dietrich et al., 2007; Shenouda & Davidson, 2007). The worst-case approach optimizes the worst system performance for any channel error in a predefined uncertainty region. In (Payaro et al., 2007) a robust power allocation scheme for THP was proposed, which maximizes the achievable rates for the worst-case errors in the CSI in the small errors regime. The authors of (Shenouda & Davidson, 2007) designed the THP transmitter to minimize the worst-case MSE over all admissible channel uncertainties subject to power constraints on each antenna, or a total power constraint. On the other hand, the stochastic approach optimizes a statistical measure of the system performance assuming that the statistics of the uncertainty is known. A robust nonlinear transmit zero-forcing filter with THP was presented in (Hunger et al., 2004) using a conditional-expectation approach, and has been extended lately in (Shenouda & Davidson, 2007) by relaxing the zero-forcing constraint and using the MMSE criterion. The problem of combined optimization of channel estimation and THP was considered in (Dietrich et al., 2007) and a conditional-expectation approach is adopted to solve the problem. All the above robust schemes are designed for the MISO downlink where each user has only one single antenna.

In this chapter for the MMSE scheme, we propose novel joint THP transceiver designs for the multiuser MIMO downlink with both perfect and imperfect CSIT. The transmitter performs nonlinear stream-wise (both inter-user and intra-user) interference pre-cancelation. We first consider the transceiver optimization problem under perfect CSIT and formulate it as minimizing the total mean square error (T-MSE) of the downlink (Zhang et al., 2005)

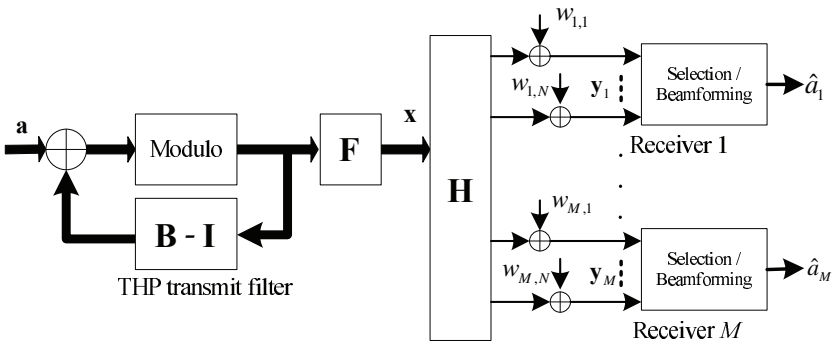

Fig. 1. Block diagram of ZF-THP in multiuser MIMO downlink system.

under a total transmit power constraint. Under the optimization criterion of minimizing the T-MSE, the stream-wise interference pre-cancelation structure is superior to the structure of inter-user only interference pre-cancelation combined with intra-user linear precoding adopted in (Doostnejad et al., 2005) and (Mezghani et al., 2006), which has already been proven to be true in a particular case, i.e., the single-user MIMO case (Shenouda & Davidson, 2008). By some convex analysis of the optimization problem, we find the necessary conditions for the optimal solution, by which the optimal transmitter and receivers are inter-dependent. We extend the iterative algorithm developed in (Zhang et al., 2005) to handle our problem. Although the iterative algorithm does not assure to converge to the globally optimal solution, it is guaranteed to converge to a locally optimal solution. Then, we make an extension of our design under perfect CSIT to the imperfect CSIT case which leads to a robust transceiver design against the channel uncertainty. The robust optimization problem is mathematically formulated as minimizing the expectation of the T-MSE conditioned on the channel estimates at the BS under a total transmit power constraint. An iterative optimization algorithm similar to its perfect CSIT counterpart can also be applied. Extensive simulation results are presented to illustrate the efficacy of our proposed schemes and their superiority over existing MMSE-based THP schemes.

The organization of the rest of this chapter is as follows. In Section 2, the system models for the multiuser MIMO downlink with THP established. In Section 3, two per-layer ZF-THP schemes are proposed and the analysis of the equivalent channel gains is given. In Section 4, the problem of the MMSE-THP design and analysis under both perfect and imperfect CSI is addressed. Simulation results are presented in Section 5. Section 6 concludes the chapter.

2. System models of multiuser MIMO downlink with THP

In this section, we will consider two system models for ZF-THP and MMSE-THP schemes, respectively.

2.1 System model for ZF-THP scheme

As mentioned in Section 1, for ZF-THP scheme, we consider the unique case that the number of users equals the number of transmit antennas at BS, denoted as M. Therein, each user is equipped with N receive antennas, as shown in Fig. 1. Perfect CSI is assumed at the transmitter (Windpassinger et al., 2004).

THP transmit filter group consists of a forward filter $\mathbf{F}$, a backward filter $\mathbf{B}$, and a modulo operator (Windpassinger et al., 2004). The transmit data symbol is denoted by the $M \times 1$ vector $\mathbf{a}$. After $\mathbf{a}$ passes through the THP transmit filter, the precoded symbol, which is denoted by the $M \times 1$ vector $\mathbf{x}$, is generated. It is assumed that the channel is flat fading. Denote the MIMO channel of user k by an $N \times M$ matrix $\mathbf{H}_k$. Each entry in $\mathbf{H}_k$ satisfies zero-mean unit-variance complex-Gaussian distribution, denoted by $CN(0,1)$. Through the channels, each user's $N \times 1$ received signal vector is

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{w}_k, \quad k = 1, 2, \cdots, M. \tag{1}$$

Therein the noise $\mathbf{w}_k$ is an $N \times 1$ vector, whose entries are independent and identically distributed (i.i.d.) random variables with the distribution $CN(0, \sigma_n^2)$.
Under the regulation that only one sub-channel be allocated to a user and all the M users be served simultaneously, every user's receive filter is a $1 \times N$ row vector, denoted by $\mathbf{r}_k$. For normalization, we assume $\|\mathbf{r}_k\|_2^2 = 1$, where $\|\cdot\|_2$ stands for the Euclidean norm of a vector. Thus, the detected signal can be expressed as

$$\hat{a}_k = \mathbf{r}_k \mathbf{H}_k \mathbf{x} + \mathbf{r}_k \mathbf{w}_k = \widetilde{\mathbf{h}}_k \mathbf{x} + \mathbf{r}_k \mathbf{w}_k, \quad k = 1, 2, \cdots, M, \tag{2}$$

where $\widetilde{\mathbf{h}}_k \triangleq \mathbf{r}_k \mathbf{H}_k$ is the equivalent channel row vector of user k. Construct the entire equivalent channel as $\widetilde{\mathbf{H}} \triangleq \left[\widetilde{\mathbf{h}}_1^H \quad \widetilde{\mathbf{h}}_2^H \quad \cdots \quad \widetilde{\mathbf{h}}_M^H\right]^H$.

2.2 System model for MMSE-THP scheme

Different from the above system model for ZF-THP scheme, we consider a more generalized model for MMSE-THP scheme, in which the number of users is not necessarily equal to that of transmit antennas. Therein, the BS is with M transmit antennas and K users are with N_k receive antennas at the kth user, $k = 1, \ldots, K$ (see Fig. 2). Let $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$ denote the channel between the BS and the kth user. The vector $\mathbf{d}_k \in \mathbb{C}^{L_k \times 1}$ represents the transmitted data vector for user k, where each entry belongs to the interval $[-\tau/2, \tau/2) + j \cdot [-\tau/2, \tau/2)$ (τ is the *modulo* base of THP as introduced later) and L_k is the number of data streams transmitted for user k. The data vectors are stacked into $\mathbf{d} \triangleq [\mathbf{d}_1^T \ \mathbf{d}_2^T \ \ldots \ \mathbf{d}_K^T]^T$, which is first reordered by a permutation matrix $\mathbf{\Pi} \in \mathbb{C}^{L \times L}$ ($\mathbf{\Pi}\mathbf{\Pi}^T = \mathbf{\Pi}^T\mathbf{\Pi} = \mathbf{I}_L$, $L \triangleq \sum_{k=1}^{K} L_k$) and then successively precoded using THP (see Fig. 2). The feedback matrix $\mathbf{F} \in \mathbb{C}^{L \times L}$ is a lower triangular matrix with zero diagonal. The structure of $\mathbf{F}$ enables inter-stream interference pre-cancelation and is different from the one used in (Mezghani et al., 2006) which only enables inter-user interference pre-cancelation. The *modulo* device performs a mod τ operation to avoid transmit power enhancement. Each entry of the output $\mathbf{w}$ of the *modulo* device is constrained in the interval $[-\tau/2, \tau/2) + j \cdot [-\tau/2, \tau/2)$. A common assumption in the literature is that the entries of $\mathbf{w}$ are uniformly distributed with unit variance (i.e., $\tau = \sqrt{6}$) and are mutually uncorrelated. Then $\mathbf{w}$ is linearly precoded by a feedforward matrix $\mathbf{P} \in \mathbb{C}^{M \times L}$ and transmitted over the downlink channel to the K users.
At the kth receiver, a decoding matrix $\mathbf{G}_k \in \mathbb{C}^{L_k \times N_k}$ and a *modulo* device are employed to estimate the data vector $\mathbf{d}_k$. Denote the estimate of $\mathbf{d}_k$ by $\tilde{\mathbf{d}}_k$, then it is given by

$$\tilde{\mathbf{d}}_k = (\mathbf{G}_k \mathbf{H}_k \mathbf{P} \mathbf{w} + \mathbf{G}_k \mathbf{n}_k) \bmod \tau, \tag{3}$$

in which $\mathbf{n}_k \in \mathbb{C}^{N_k \times 1}$ is the additive Gaussian noise vector at user k with zero mean and covariance matrix $\mathcal{E}\{\mathbf{n}_k \mathbf{n}_k^H\} = \sigma_{n,k}^2 \mathbf{I}_{N_k}$. We assume that there is a total power constraint P_T at the BS so that $\mathrm{tr}\left(\mathbf{P}^H \mathbf{P}\right) = P_T$.

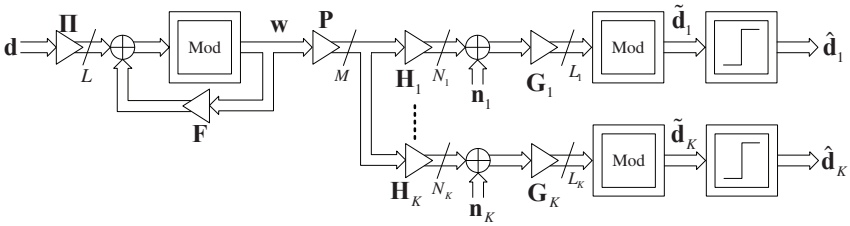

Fig. 2. Block diagram of MMSE-THP in multiuser MIMO downlink system.

3. Per-layer ZF-THP design and analysis

In this section, we will firstly propose two *per-layer* ZF-THP schemes for multiuser MIMO downlinks based on the system model in Fig. 1.

3.1 Capacity analysis for ZF-THP

Perform QR factorization to the conjugate transpose of the equivalent channel matrix $\widetilde{\mathbf{H}}$ in (2) in Section 2.1. This generates

$$\widetilde{\mathbf{H}} = \mathbf{S}\mathbf{F}^H, \tag{4}$$

where $\mathbf{F}$ is a unitary matrix and $\mathbf{S}$ is a lower-triangular matrix. In (Windpassinger et al., 2004), it is given that without account of the precoding loss (Yu et al., 2005), the sum-capacity of all layers is equivalent to

$$C^{sum} = \sum_{l=1}^{M} C_l = \sum_{l=1}^{M} \log\left(1 + \frac{\sigma_{x,l}^2}{\sigma_n^2}|s_{ll}|^2\right), \tag{5}$$

where $\sigma_{x,l}^2$ is the signal power in layer l. $|s_{ll}|^2$ can be interpreted as the equivalent channel gain of the lth layer.

If all the entries in $\widetilde{\mathbf{H}}$ have the distribution $CN(0,1)$, $|s_{ll}|^2$ is a random variable with the chi-square distribution of $2(M-l+1)$ degrees of freedom (Windpassinger et al., 2004). Nevertheless, when the total number of receive antennas is more than that of transmit antennas, the distribution of entries in $\widetilde{\mathbf{H}}$ will change with different precessing methods for receive antennas.

With the assumption that the channels of all the users have the same power attenuation, serving all the M users simultaneously means that the obtained receive diversity gain, which is defined by the negative slope of the outage probability versus signal-to-noise ratio (SNR) curve on a log-log scale (Tse & Viswanath, 2005), can be scaled by MN. In comparison, in per-user processing only $\lceil \frac{M}{N} \rceil$ users are served at one time, so the obtained receive diversity gain is scaled by M. Therefore, the strategy of serving all the M users simultaneously leads to N times larger receive diversity gain, which implies that each user should be provided with only one subchannel.

3.2 Per-layer transmit and receive filters design

From (2), the equivalent channel matrix $\widetilde{\mathbf{H}}$ is derived from the receive filters $\{\mathbf{r}_l, l = 1, \cdots, M\}$. Due to the channel matrix trianglization in (4), the higher layers will interfere with, but not be influenced by the lower ones. Denote the mapping $f_l : \bar{\mathbf{h}}_l = f_l\big(\{\mathbf{r}_p, p = 1, \cdots, l\}\big)$, then $|s_{ll}|^2 = \|\bar{\mathbf{h}}_l\|_2^2 = \|f_l\big(\{\mathbf{r}_p\}\big)\|_2^2$.

So the optimal $\{\mathbf{r}_l, l = 1, \cdots, M\}$ that maximizes the sum-capacity can be expressed as

$$\{\mathbf{r}_l^{opt}\} = \arg\max_{\{\mathbf{r}_l\}} \sum_{l=1}^{M} \log\left(1 + \frac{\sigma_{x,l}^2}{\sigma_n^2}\|f_l(\{\mathbf{r}_p\})\|_2^2\right), \quad \text{s. t.} \quad \|\mathbf{r}_l\|_2^2 = 1, \quad l = 1, \cdots, M. \tag{6}$$

According to (6), the design of one layer should take into account of its impact upon all the lower layers, and for each layer except the last one, there are multiple candidate users. So the solution of this optimization problem is very complicated. To make it practical, we employ a suboptimal approach, which conducts a per-layer optimization from high to low and converts the global optimization (6) into a series of greedy optimization as follows.

$$\mathbf{r}_l^{opt} = \arg\max_{\mathbf{r}_l} \log\left(1 + \frac{\sigma_{x,l}^2}{\sigma_n^2}\|f_l(\{\mathbf{r}_p\})\|_2^2\right), \quad \text{s. t.} \quad \|\mathbf{r}_l\|_2^2 = 1, \quad l = 1, \cdots, M. \tag{7}$$

When processing one layer, say layer l, we disregard its impact upon other layers and just maximize the power in $\bar{\mathbf{h}}_l$. Specifically, we suppose all the rest users as candidates, and generate their own receive filters according to some proper criterion. Thus, for layer l and user k, the equivalent channel row vector, denoted by $\widetilde{\mathbf{h}}_{l,k}^{equi}$, can be obtained. Here, $\bar{\mathbf{H}}_{l-1}^{null}$ represents the subspace orthogonal to that spanned by $\{\widetilde{\mathbf{h}}_p^H, p = 1, \cdots, l-1\}$, and the projection power of $\widetilde{\mathbf{h}}_{l,k}^{equi,H}$ onto $\bar{\mathbf{H}}_{l-1}^{null}$ is interpreted as user k's *residual* channel gain in layer l. Then, the user with the largest residual channel gain is selected and placed into layer l. In this way, all the users can be arranged into the sequence of layers and $\{\bar{\mathbf{h}}_l, l = 1, \cdots, M\}$ can be obtained sequentially. Within this per-layer approach, the key is how to design the receive filters. For layer l and user k, we denote $\bar{\mathbf{H}}_k^{(l)}$ as the projection of $\mathbf{H}_k^H$ onto $\bar{\mathbf{H}}_{l-1}^{null}$, then the optimal receive filter $\mathbf{r}_{k,l}$ can be obtained by

$$\mathbf{r}_{l,k}^{opt} = \arg\max_{\mathbf{r}} \|\mathbf{r}\left(\bar{\mathbf{H}}_k^{(l)}\right)^H\|_2^2, \quad \text{s. t.} \quad \|\mathbf{r}\|_2^2 = 1. \tag{8}$$

The solution of this maximization problem can be given by the theory of Rayleigh quotient (Horn & Johnson, 1985). That is, $\mathbf{r}_{l,k}^{opt}$ is the conjugate transpose of the eigenvector corresponding to the maximum eigenvalue of the matrix $\left(\bar{\mathbf{H}}_k^{(l)}\right)^H \bar{\mathbf{H}}_k^{(l)}$. In essential, this processing method aims to maximize power gain and diversity gain of each layer through the design of receive antenna beamforming (RAB). The per-layer RAB scheme is summarized in Table 1-a. Therein EVD$(\cdot)$ returns the set of eigenvalues and eigenvectors, and Householder$(\cdot)$ returns the Householder matrix. $\mathbf{I}_N$ stands for an $N \times N$ identity matrix.

By this scheme, the user ordering $\{\pi_l\}$, the receive filters $\{\hat{\mathbf{r}}_l\}$, and the transmit filter $\hat{\mathbf{F}} = \mathbf{F}^{(M)} \cdots \mathbf{F}^{(1)}$ are all generated. However, the operations of eigenvalue decomposition (EVD) still consume a certain complexity. To further reduce the complexity and employ less analog chains at the receivers (Gorokhov et al., 2003), RAB can be replaced by receive antenna selection (RAS). Specifically, for a layer and a candidate user, instead of computing the the eigenvector, we just select the receive antenna whose equivalent channel vector has the maximum Euclidean norm, as shown in Table 1-b.

Remarks:

- For each layer, the aim of the receive filter design is to adjust the weights of receive antennas to maximize the power in the equivalent channel vector's component orthogonal to the higher layers' dimensions (i.e., $\|\bar{\mathbf{h}}_l\|_2^2$), but not the power in the equivalent channel vector itself (i.e., $\|\widetilde{\mathbf{h}}_l\|_2^2$).

(a) The scheme of per-layer RAB	(b) The scheme of per-layer RAS
Given all user's $N \times M$ *channel matrices* $\mathbf{H}_k$, $k = 1, 2, \cdots, M$.	*Given all user's* $N \times M$ *channel matrices* $\mathbf{H}_k$, $k = 1, 2, \cdots, M$.
Initialization: The candidate user set $\Phi = \{1, 2, \cdots, M\}$, $\mathbf{F}^{(0)} = \mathbf{I}_M$, $\bar{\mathbf{H}}_k^{(0)} = \mathbf{H}_k^H$, $k = 1, 2, \cdots, M$.	*Initialization: The candidate user set* $\Phi = \{1, 2, \cdots, M\}$, $\mathbf{F}^{(0)} = \mathbf{I}_M$, $\bar{\mathbf{H}}_k^{(0)} = \mathbf{H}_k^H$, $k = 1, 2, \cdots, M$.
For the layer index $l : 1 \rightarrow M$	*For the layer index* $l : 1 \rightarrow M$
For the user index $k \in \Phi$	*For the user index* $k \in \Phi$
$\bar{\mathbf{H}}_k^{(l)} = \mathbf{F}^{(l-1)} \bar{\mathbf{H}}_k^{(l-1)}$	$\bar{\mathbf{H}}_k^{(l)} = \mathbf{F}^{(l-1)} \bar{\mathbf{H}}_k^{(l-1)}$
$\bar{\mathbf{H}}_k^{(l),proj}$ *is comprised by the* l*th to* M*th rows of* $\bar{\mathbf{H}}_k^{(l)}$	$\bar{\mathbf{H}}_k^{(l),proj}$ *is comprised by the* l*th to* M*th rows of* $\bar{\mathbf{H}}_k^{(l)}$
$\{[\lambda_n \quad \mathbf{u}_n], n = 1, \cdots, N\} = \mathrm{EVD}\left((\bar{\mathbf{H}}_k^{(l),proj})^H \bar{\mathbf{H}}_k^{(l),proj}\right)$	p_n *is the Euclidean norm of the* n*th column of* $\bar{\mathbf{H}}_k^{(l),proj}$
$n_{max} = \arg\max_n \{\lambda_n\}$	$n_{max} = \arg\max_n \{p_n\}$
$\mathbf{r}_k = \mathbf{u}_{n_{max}}^H$	$\mathbf{r}_k$ *is a* $1 \times N$ *vector,* $[0, \cdots, 0, \star]$
$\lambda^{(k)} = \lambda_{n_{max}}$	$p^{(k)} = p_{n_{max}}$
end	*end*
$\widetilde{k} = \arg\max_{k \in \Phi} \{\lambda^{(k)}\}$	$\widetilde{k} = \arg\max_{k \in \Phi} \{p^{(k)}\}$
$\Phi = \Phi \backslash \{\widetilde{k}\}$	$\Phi = \Phi \backslash \{\widetilde{k}\}$
$\pi_l = \widetilde{k}$	$\pi_l = \widetilde{k}$
$\hat{\mathbf{r}}_l = \mathbf{r}_{\widetilde{k}}$	$\hat{\mathbf{r}}_l = \mathbf{r}_{\widetilde{k}}$
$\bar{\mathbf{h}}_l$ *is comprised by the* l*th to* M*th rows of* $\bar{\mathbf{H}}_{\widetilde{k}}^{(l)} \mathbf{r}_{\widetilde{k}}^H$	$\bar{\mathbf{h}}_l$ *is comprised by* l*th to* M*th rows of* $\bar{\mathbf{H}}_{\widetilde{k}}^{(l)} \mathbf{r}_{\widetilde{k}}^H$
$\mathbf{F}^{(l)} = \left[\begin{array}{c\|c} \mathbf{I}_{l-1} & \mathbf{0}_{(l-1)\times(M-l+1)} \\ \hline \mathbf{0}_{(M-l+1)\times(l-1)} & \mathrm{Householder}(\bar{\mathbf{h}}_l) \end{array}\right]$	$\mathbf{F}^{(l)} = \left[\begin{array}{c\|c} \mathbf{I}_{l-1} & \mathbf{0}_{(l-1)\times(M-l+1)} \\ \hline \mathbf{0}_{(M-l+1)\times(l-1)} & \mathrm{Householder}(\bar{\mathbf{h}}_l) \end{array}\right]$
end	*end*

Table 1. The schemes of per-layer RAB and per-layer RAS.

- In the successive mechanism of THP, the higher a layer, the less it costs for the interference suppression. In the per-layer schemes, the users with large residual channel gains are placed into the high layers. In this way, the power wasted in the interference suppression can be decreased, but the power contributing to the sum-capacity can be increased.
- As a suboptimal solution of (8), per-layer RAS is inferior to per-layer RAB. However, for the sake of practice, in per-layer RAS only the indexes of the selected antennas should be informed to the receivers, but in per-layer RAB, the counterparts are the designed receive filter weights.

3.3 Comparison between per-layer RAB and RAS

Here, we do not order the users and consider the lth layer's projected channel matrix $\left(\bar{\mathbf{H}}_k^{(l),proj}\right)^H, \forall k$, whose entries have i.i.d. $CN(0,1)$ distribution. Using per-layer RAB, the equivalent channel gain is the square of its maximum singular value, while using per-layer RAS, the equivalent channel gain is the square of its maximum row vector's Euclidean norm. We denote these two kinds of channel gains by $\delta_{RAB}^2(l)$ and $\delta_{RAS}^2(l)$, respectively. With the decrease of l, the relative difference between $\delta_{RAB}^2(l)$ and $\delta_{RAS}^2(l)$ tends to decrease, which can be deduced below.

Theorem 1. *Given a* $2 \times n$ *matrix* $\boldsymbol{A}$ *such that all its entries have i.i.d.* $CN(0,1)$ *distribution. Denote the eigenvalues of* $\boldsymbol{A}\boldsymbol{A}^H$ *as* λ_i', $i = 1, 2$. *Let* $\lambda_1 \triangleq \max_i\{\lambda_i'\}$, $\lambda_2 \triangleq \min_i\{\lambda_i'\}$, *and denote* $\Delta\lambda \triangleq \lambda_1 - \lambda_2$. *Then with* $n \to \infty$, *the ratio* $E(\Delta\lambda)/E(\lambda_1) \to 0$.

Proof: By the bidiagonalization (Th. 3.4 in (Wang et al., 2006)), $\mathbf{A}$ is unitarily similar to a lower-triangular matrix Λ, where

$$\Lambda \triangleq \begin{bmatrix} x_{2n} & 0 & \cdots 0 \\ y_2 & x_{2(n-1)} & \cdots 0 \end{bmatrix}, \tag{9}$$

x_{2n}^2, $x_{2(n-1)}^2$ and y_2^2 are independent chi-square distributed random variables with the degrees of freedom $2n$, $2(n-1)$ and 2, respectively. Then,

$$\mathbf{A}\mathbf{A}^H = \Lambda\Lambda^H = \begin{bmatrix} x_{2n}^2 & x_{2n}y_2 \\ x_{2n}y_2 & x_{2(n-1)}^2 + y_2^2 \end{bmatrix}. \tag{10}$$

Therein we denote a new chi-square random variable $y_{2n}^2 \triangleq x_{2(n-1)}^2 + y_2^2$. Further, the eigenvalues of $\mathbf{A}\mathbf{A}^H$ can be obtained as

$$\lambda_{1,2} = \frac{x_{2n}^2 + y_{2n}^2}{2} \pm \sqrt{x_{2n}^2 y_2^2 + \left(\frac{x_{2n}^2 + y_{2n}^2}{2}\right)^2 - x_{2n}^2 y_{2n}^2}. \tag{11}$$

Substitute $x_{2n}^2 + y_{2n}^2$ by x_{4n}^2, then

$$\Delta\lambda = 2\sqrt{x_{2n}^2 y_2^2 + \left(\frac{x_{4n}^2}{2}\right)^2 - x_{2n}^2 y_{2n}^2}. \tag{12}$$

$$E(\Delta\lambda) \le \sqrt{E\big((\Delta\lambda)^2\big)} = 2\sqrt{E\big(x_{2n}^2 y_2^2\big) + E\Big(\big(\frac{x_{4n}^2}{2}\big)^2\Big) - E\big(x_{2n}^2 y_{2n}^2\big)}. \tag{13}$$

According to the character of chi-square distribution (Horn & Johnson, 1985), we have

$$E\big(x_{2n}^2 y_2^2\big) = E\big(x_{2n}^2\big)E\big(y_2^2\big) = 2n \cdot 2 = 4n, \tag{14}$$

$$E\Big(\big(\frac{x_{4n}^2}{2}\big)^2\Big) = \frac{4n(4n+2)}{4} = 4n^2 + 2n, \tag{15}$$

$$E\big(x_{2n}^2 y_{2n}^2\big) = E\big(x_{2n}^2\big)E\big(y_{2n}^2\big) = 2n \cdot 2n = 4n^2. \tag{16}$$

Plunging to (13),

$$E(\Delta\lambda) \le 2\sqrt{6n}. \tag{17}$$

On the other hand,

$$E(\lambda_1) \ge E\left(\frac{x_{2n}^2 + y_{2n}^2}{2} + \sqrt{\left(\frac{x_{2n}^2 + y_{2n}^2}{2}\right)^2 - x_{2n}^2 y_{2n}^2}\right) = E\big(x_{2n}^2\big) = 2n. \tag{18}$$

So $E(\Delta\lambda)/E(\lambda_1) \leq \sqrt{6}n^{-1/2}$.
With $n \to \infty$, $\sqrt{6}n^{-1/2} \to 0$. Considering $E(\Delta\lambda)/E(\lambda_1) \geq 0$, it can be concluded that with $n \to \infty$, $E(\Delta\lambda)/E(\lambda_1) \to 0$.

■

Corollary 1. *Given the same conditions as Theorem 1. Additionally denote the row vectors of* $\boldsymbol{A}$ *as* $\boldsymbol{a}_i$, $i = 1, 2$. *Then with* $n \to \infty$, *the ratio* $E(\lambda_1 - \max_i \|\boldsymbol{a}_i\|_2^2)/E(\lambda_1) \to 0$.

Proof: By the character of Rayleigh quotient (Horn & Johnson, 1985), $\forall r \in \mathbb{C}^{1\times 2}$, $rr^H = 1$, the maximum and minimum values of $r\mathbf{A}\mathbf{A}^H r^H$ are λ_1 and λ_2, respectively. Let $r' = [0 \quad 1]$ or $r' = [1 \quad 0]$, then the value of $r'\mathbf{A}\mathbf{A}^H r'^H$ is surely between λ_1 and λ_2. Obviously, $r'\mathbf{A}\mathbf{A}^H r'^H$ is equivalent to $\|\mathbf{a}_i\|_2^2$, $i = 1, 2$. Thus,

$$0 \leq \lambda_1 - \max_i \|\mathbf{a}_i\|_2^2 \leq \Delta\lambda. \tag{19}$$

According to Theorem 1, with $n \to \infty$, the ratio $E\big(\lambda_1 - \max_i \|\mathbf{a}_i\|_2^2\big)/E(\lambda_1) \to 0$. ■

According to Corollary 1, when the number of transmit antennas K tends to be infinite, from the bottom layer ($l = K$) to the top layer ($l = 1$), the relative difference between $\delta_{RAB}(l)$ and $\delta_{RAS}(l)$, denoted by $\Delta G_l \triangleq \big(\delta_{RAB}(l) - \delta_{RAS}(l)\big)/\delta_{RAB}(l)$, tends to decrease until zero.
Next, consider the capacity of each layer in both per-layer RAB and per-layer RAS. Denote the capacities of layer l in these two schemes as C_l^{RAB} and C_l^{RAS}, respectively, and denote their difference as $\Delta C_l \triangleq C_l^{RAB} - C_l^{RAS}$. Let $n = K - l + 1$, which means the degree of freedom in layer l, then ΔC_l can be rewritten as $\Delta C(n)$. Denote $\gamma \triangleq \sigma_x^2/\sigma_n^2$. In the medium and high (SNR) scenarios, the characteristic of $\Delta C(n)$ is described in the following corollary.

Corollary 2. *Given the same conditions as in Corollary 1,*

$$E\big(\Delta C(n)\big) = E\Big(\log\big(1 + \gamma\lambda_1(n)\big) - \log\big(1 + \gamma \max_i \|\boldsymbol{a}_i(n)\|_2^2\big)\Big). \tag{20}$$

In the medium and high SNR scenarios, with $n \to \infty$, *it holds that* $E\big(\Delta C(n)\big) \to 0$.

The details of the proof of Corollary 2 are omitted due to page limit[2]. Then, we consider the low SNR scenarios, where $\gamma \to 0$,

$$E\big(\Delta C(n)\big) \approx E\big(\gamma\lambda_1(n) - \gamma \max_i \|\mathbf{a}_i(n)\|_2^2\big) \leq E\big(\gamma\Delta\lambda(n)\big). \tag{21}$$

From (17), it can be inferred that in the low SNR scenarios $E\big(\Delta C(n)\big)$ increases with n. This trend is opposite to that in the medium and high SNR scenarios.
Theorem 1 and its corollaries indicate the case of a two-row matrix, which corresponds to the scenarios with two receive antennas at each receiver. Thus, a conclusion can be drawn that when the number of transmit antennas increases infinitely, both ΔG_l and ΔC_l (at medium and high SNR) in those high layers will asymptotically tend to zero. This implies that in the case of a large number of transmit antennas, for those higher layers, whether applying RAB or RAS, the differences of channel gains could be approximately negligible, but RAS consumes much less complexity.

[2] Please refer to (Huang et al., 2010).

Scheme	Layer 1	Layer 2
per-layer RAB	$\widetilde{\lambda}^{max,2}_{M-2l+2}$	$\widetilde{\lambda}^{max,2}_{M-2l+1}$
per-layer RAS	$\widetilde{p}^{max,2}_{M-2l+2}$	$\widetilde{p}^{max,2}_{M-2l+1}$
per-user	$\widetilde{\lambda}^{max,2}_{M-2l+2}$	$\widetilde{\lambda}^{min,2}_{M-2l+2}$

Table 2. The equivalent channel gains of a unit in per-layer RAB, per-layer RAS and per-user processing.

3.4 Comparison with per-user processing

We still consider the case of two receive antennas for each receiver. In the per-user processing, each receiver owns a group of two adjacent layers, represented by a $2 \times M$ channel matrix. The channel matrices of lower groups should be orthogonally projected onto those of higher groups. Hence the equivalent channel matrix for group l, which includes layers $2l-1$ and $2l$, is a $2 \times (M-2l+2)$ matrix, and all its entries can be assumed to be i.i.d. $CN(0,1)$ random variables. Here, each group of two adjacent layers is interpreted as a basic unit. For the lth unit, the two equivalent channel gains are the squares of singular values of a $2 \times (M-2l+2)$ matrix, denoted by $\widetilde{\lambda}^{max,2}_{M-2l+2}$ and $\widetilde{\lambda}^{min,2}_{M-2l+2}$, respectively.

Accordingly, we also bind every two adjacent layers as a unit in both per-layer RAB and per-layer RAS schemes. In this way, for the lth unit, in per-layer RAB, one equivalent channel gain equals the square of the maximum singular value of a $2 \times (M-2l+2)$ matrix, and the other equals the square of the maximum singular value of a $2 \times (M-2l+1)$ matrix, denoted by $\widetilde{\lambda}^{max,2}_{M-2l+2}$ and $\widetilde{\lambda}^{max,2}_{M-2l+1}$, respectively; while in per-layer RAS, they are the squares of the maximum row-norm of a $2 \times (M-2l+2)$ matrix and the maximum row-norm of a $2 \times (M-2l+1)$ matrix, denoted by $\widetilde{p}^{max,2}_{M-2l+2}$ and $\widetilde{p}^{max,2}_{M-2l+1}$, respectively. The equivalent channel gains of a unit in these three schemes are summarized in Table 2.

Remarks:

- In layer 1, per-layer RAB and per-user processing have the same equivalent channel gain, which is larger than that of per-layer RAS.
- In layer 2, the chances of exploiting the maximum singular value or the maximum row-norm still exist in per-layer RAB and per-layer RAS; while in per-user processing only the minimum singular value is used, and hence, the diversity gain is lost.

Based on the above observations, per-layer RAB outperforms the other two schemes evidently. But the relation between per-layer RAS and per-user processing is indistinct. We analyze two extreme cases: with very low SNR, where the maximum sum-capacity is approximately achieved by allocating all the signal power into the best layer, or with very high SNR, where by allocating the power into all the layers averagely (Tse & Viswanath, 2005).

It can be derived from (19) that $E\left(\widetilde{p}^{max}_{K-2l+2}\right) < E\left(\widetilde{\lambda}^{max,2}_{K-2l+2}\right)$, thus, at low SNR, per-layer RAS has smaller sum-capacity than per-user processing.

At very high SNR, the capacity depends on the product of the channel gains of two layers. Let $n = K-2l+2$, then the lower bounds can be developed in **Appendix A** that for per-user processing $E\left(\widetilde{\lambda}^{max,2}_{n}\widetilde{\lambda}^{min,2}_{n}\right) \geq 4n^2 - 6n$ and for per-layer RAS $E\left(\widetilde{p}^{max,2}_{n}\widetilde{p}^{max,2}_{n-1}\right) \geq 4n^2 - 4n$. Though the tightness of these two lower bounds are not proved, the advantage of per-layer RAS over per-user processing at high SNR can be additionally validated by the simulation results in Subsection 5.1.

4. Stream-wise MMSE-THP design and analysis

In this section, we firstly propose our joint THP transceiver design under perfect CSIT using a minimum total mean square error (MT-MSE) criterion. Using convex analysis for the optimization problem we derive the necessary conditions for the optimal transceiver in Subsection 4.1.1. Then the iterative algorithm proposed in (Zhang et al., 2005) is extended in Subsection 4.1.2 to obtain a locally optimal transceiver. Furthermore, we introduce a robust THP transceiver design for the multiuser MIMO downlink in Subsection 4.2, which is more effective against the uncertainty in the CSIT than the above simple solution. The robust optimization problem is mathematically formulated as minimizing the expectation of the T-MSE conditioned on the channel estimates at the BS under a total transmit power constraint. Then the iterative algorithm proposed in Subsection 4.1 is applied to solve the problem.

4.1 Transceiver optimization under perfect CSIT

4.1.1 Problem reformulation

Our design is based on the linear representation (Joham et al., 2004) (see Fig. 3) of the system in Fig. 2, where the *modulo* devices at the transmitter and receivers are replaced by the additive vector $\mathbf{a} \triangleq [\mathbf{a}_1^T\ \mathbf{a}_2^T\ \dots\ \mathbf{a}_K^T]^T$ and $-\tilde{\mathbf{a}}_k$, $k = 1, ..., K$, where $\mathbf{a} \in \tau\left(\mathbb{Z}^{L\times 1} + j\cdot\mathbb{Z}^{L\times 1}\right)$ and $\tilde{\mathbf{a}}_k \in \tau\left(\mathbb{Z}^{L_k\times 1} + j\cdot\mathbb{Z}^{L_k\times 1}\right)$. The vectors $\mathbf{a}$ and $\tilde{\mathbf{a}}_k$ are chosen to make the same $\mathbf{w}$ and $\tilde{\mathbf{d}}_k$ as the *modulo* devices at the transmitter and receivers output respectively.

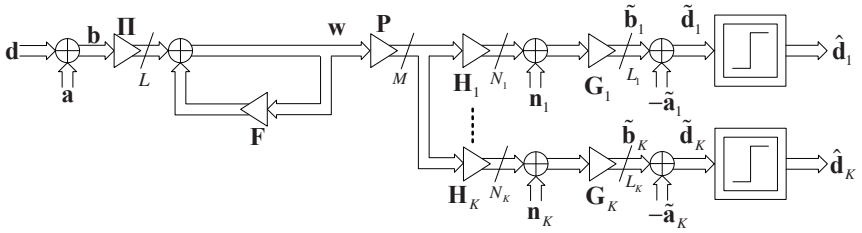

Fig. 3. Equivalent linear representation of THP in Fig. 2.

Define $\mathbf{b}_k \triangleq \mathbf{d}_k + \mathbf{a}_k$ and $\tilde{\mathbf{b}}_k \triangleq \tilde{\mathbf{d}}_k + \tilde{\mathbf{a}}_k$ and stack them into $\mathbf{b} \triangleq [\mathbf{b}_1^T\ \dots\ \mathbf{b}_K^T]^T$ and $\tilde{\mathbf{b}} \triangleq [\tilde{\mathbf{b}}_1^T\ \dots\ \tilde{\mathbf{b}}_K^T]^T$. Let $\mathbf{H} \triangleq \left[\mathbf{H}_1^T \dots \mathbf{H}_K^T\right]^T$, $\mathbf{n} \triangleq \left[\mathbf{n}_1^T \dots \mathbf{n}_K^T\right]^T$ and $\mathbf{G} \triangleq \text{blockdiag}\left(\mathbf{G}_1, ..., \mathbf{G}_K\right)$, then from Fig. 3 we have

$$\mathbf{\Pi b} + \mathbf{Fw} = \mathbf{w} \ \Rightarrow\ \mathbf{b} = \mathbf{\Pi}^T(\mathbf{I}_L - \mathbf{F})\mathbf{w} \tag{22}$$

and

$$\tilde{\mathbf{b}} = \mathbf{GHPw} + \mathbf{Gn}. \tag{23}$$

We consider the MSE between $\mathbf{b}$ and $\tilde{\mathbf{b}}$ rather than $\mathbf{d}$ and $\tilde{\mathbf{d}}$ in order to bypass the impact of the *modulo* operations and define it as the total MSE (T-MSE) of the downlink, which is written as follows:

$$\begin{aligned}\text{T-MSE} &= \mathcal{E}_{\mathbf{w},\mathbf{n}}\left\{\left\|\tilde{\mathbf{b}} - \mathbf{b}\right\|_2^2\right\} = \mathcal{E}_{\mathbf{w},\mathbf{n}}\left\{\left\|\left(\mathbf{GHP} - \mathbf{\Pi}^T(\mathbf{I}_L - \mathbf{F})\right)\mathbf{w} + \mathbf{Gn}\right\|_2^2\right\} \\ &= \left\|\mathbf{GHP} - \mathbf{\Pi}^T(\mathbf{I}_L - \mathbf{F})\right\|_F^2 + \text{tr}(\mathbf{G}^H\mathbf{G}\mathbf{\Sigma}_\mathbf{n}),\end{aligned} \tag{24}$$

where $\mathbf{\Sigma_n} \triangleq \mathcal{E}\{\mathbf{n}\mathbf{n}^H\} = \text{blockdiag}\left(\sigma_{n,1}^2\mathbf{I}_{N_1}, ..., \sigma_{n,K}^2\mathbf{I}_{N_K}\right)$.

So our transceiver design problem is to find a set $\left\{\mathbf{\Pi},\ \mathbf{F},\ \mathbf{P},\ \{\mathbf{G}_k\}_{k=1}^K\right\}$ that minimizes the T-MSE defined in (24) under a total transmit power constraint. Mathematically it can be formulated as follows:

$$\begin{aligned} &\min_{\mathbf{\Pi},\ \mathbf{F},\ \mathbf{P},\ \{\mathbf{G}_k\}_{k=1}^K} \text{T-MSE} \\ &\text{s.t.} \quad \text{tr}\left(\mathbf{P}^H\mathbf{P}\right) = P_T, \\ &\qquad [\mathbf{F}]_{m,n} = 0,\ \forall\, 1 \leqslant m, n \leqslant L \text{ and } m \leqslant n. \end{aligned} \tag{25}$$

4.1.2 Iterative algorithm

In this subsection, through some analysis, we find the necessary conditions for the optimal $\mathbf{\Pi}$, $\mathbf{F}$, $\mathbf{P}$ and $\{\mathbf{G}_k\}_{k=1}^K$, which form an inter-dependence among them. This kind of inter-dependence leads to an iterative algorithm similar to the one proposed in (Zhang et al., 2005). In each iteration, we first determine the suboptimal reordering matrix $\mathbf{\Pi}$ and update $\mathbf{P}$ and $\mathbf{F}$ using the updated $\{\mathbf{G}_k\}_{k=1}^K$ in the last iteration, then update $\{\mathbf{G}_k\}_{k=1}^K$ using the above updated $\mathbf{\Pi}$, $\mathbf{P}$ and $\mathbf{F}$.

For ease of derivation, we introduce two new matrix variables $\mathbf{T} \triangleq \beta^{-1}\mathbf{P}$ and $\mathbf{R} \triangleq \beta\mathbf{G}$ to replace $\mathbf{P}$ and $\mathbf{G}$, where β is a positive real number. Then (24) is rewritten as

$$\text{T-MSE} = \left\|\mathbf{RHT} - \mathbf{\Pi}^T(\mathbf{I}_L - \mathbf{F})\right\|_F^2 + \beta^{-2} \cdot \text{tr}(\mathbf{R}^H\mathbf{R}\mathbf{\Sigma_n}). \tag{26}$$

Moreover, using the total power constraint in (25) we obtain

$$\beta = P_T^{\frac{1}{2}}\left(\text{tr}(\mathbf{T}\mathbf{T}^H)\right)^{-\frac{1}{2}}. \tag{27}$$

Note that $\mathbf{F}$ only appears in the first term of (26). We expand the first term of (26) as follows:

$$\begin{aligned} &\left\|\mathbf{RHT} - \mathbf{\Pi}^T(\mathbf{I}_L - \mathbf{F})\right\|_F^2 \\ =\ &\|\mathbf{\Pi}\mathbf{RHT} - (\mathbf{I}_L - \mathbf{F})\|_F^2 \quad (\text{for } \mathbf{\Pi}^T\mathbf{\Pi} = \mathbf{\Pi}\mathbf{\Pi}^T = \mathbf{I}_L) \qquad (28) \\ =\ &\sum_{i=1}^{L}\left\|\begin{bmatrix}\mathbf{A}_i \\ \mathbf{B}_i\end{bmatrix}\mathbf{t}_i - \mathbf{e}_i + \mathbf{f}_i\right\|_2^2, \qquad (29) \end{aligned}$$

where

$$\begin{bmatrix}\mathbf{A}_i \\ \mathbf{B}_i\end{bmatrix} = \mathbf{\Pi}\mathbf{RH},\ \mathbf{A}_i \in \mathbb{C}^{i\times M},\ \mathbf{B}_i \in \mathbb{C}^{(L-i)\times M}, \tag{30}$$

$\mathbf{t}_i$, $\mathbf{e}_i$ and $\mathbf{f}_i$ are the ith columns of $\mathbf{T}$, $\mathbf{I}_L$ and $\mathbf{F}$ respectively. The equality in (28) follows from the fact that the Frobenius norm of a matrix remains constant after the multiplication of a unitary matrix (Horn & Johnson, 1985). For fixed $\mathbf{\Pi}$, $\mathbf{T}$ and $\mathbf{R}$, each term in the summation in (29) can be minimized separately. With the lower triangular and zero diagonal structure of $\mathbf{F}$, the optimal $\mathbf{f}_i$ that minimizes the ith term of (29) is easily computed as:

$$\mathbf{f}_i = -\begin{bmatrix}\mathbf{0}_{i\times M} \\ \mathbf{B}_i\end{bmatrix}\mathbf{t}_i,\ i = 1, \ldots, L. \tag{31}$$

By substituting (27) and (31) into (26), we rewrite the T-MSE as:

$$\text{T-MSE} = \sum_{i=1}^{L} \left\| \begin{bmatrix} \mathbf{A}_i \\ \mathbf{0} \end{bmatrix} \mathbf{t}_i - \mathbf{e}_i \right\|_2^2 + \xi \sum_{i=1}^{L} \text{tr}(\mathbf{t}_i^H \mathbf{t}_i), \tag{32}$$

where $\xi \triangleq P_T^{-1} \sum_{k=1}^{K} \sigma_{n,k}^2 \text{tr}\left(\mathbf{R}_k^H \mathbf{R}_k\right)$ is a nonnegative real number, i.e., $\xi \geq 0$.
For fixed $\mathbf{\Pi}$ and $\mathbf{R}$, the optimization problem in (25) can be reformulated as:

$$\min_{\mathbf{T}} g(\mathbf{T}) \triangleq \sum_{i=1}^{L} \left\| \begin{bmatrix} \mathbf{A}_i \\ \mathbf{0} \end{bmatrix} \mathbf{t}_i - \mathbf{e}_i \right\|_2^2 + \xi \sum_{i=1}^{L} \text{tr}(\mathbf{t}_i^H \mathbf{t}_i). \tag{33}$$

Notice that by the introduction of β the power constraint in the original optimization problem has been absorbed into the objective function, so (33) is an unconstraint optimization problem. The Hessian matrix of (32) with respect to $\mathbf{t}_i$ is calculated and shown below:

$$\nabla_{\mathbf{t}_i^T} \left[\nabla_{\mathbf{t}_i^*} g(\mathbf{T}) \right] = \mathbf{A}_i^H \mathbf{A}_i + \xi \mathbf{I}_M \succeq \mathbf{0}. \tag{34}$$

The Hessian matrix in (34) being positive semidefinite indicates that $g(\mathbf{T})$ is convex respect to $\mathbf{t}_i$. Then the optimal $\mathbf{t}_i$ is derived by calculating the first order derivative with respect to $\mathbf{t}_i^*$ and setting it to zero, i.e.,

$$\frac{\partial g(\mathbf{T})}{\partial \mathbf{t}_i^*} = \mathbf{A}_i^H \mathbf{A}_i \mathbf{t}_i - \begin{bmatrix} \mathbf{A}_i \\ \mathbf{0} \end{bmatrix}^H \mathbf{e}_i + \xi \mathbf{t}_i = \mathbf{0} \Rightarrow \mathbf{t}_i = \left(\mathbf{A}_i^H \mathbf{A}_i + \xi \mathbf{I}_M \right)^{-1} \begin{bmatrix} \mathbf{A}_i^H & \mathbf{0} \end{bmatrix} \mathbf{e}_i. \tag{35}$$

Now we consider the problem of the optimal ordering, i.e., the optimal $\mathbf{\Pi}$. (32) can be rewritten as:

$$\text{T-MSE} = \sum_{i=1}^{L} \left(\mathbf{t}_i^H \left(\mathbf{A}_i^H \mathbf{A}_i + \xi \mathbf{I}_M \right) \mathbf{t}_i - \mathbf{t}_i^H \begin{bmatrix} \mathbf{A}_i^H & \mathbf{0} \end{bmatrix} \mathbf{e}_i - \mathbf{e}_i^H \begin{bmatrix} \mathbf{A}_i \\ \mathbf{0} \end{bmatrix} \mathbf{t}_i + 1 \right). \tag{36}$$

Substituting (35) into (36) and after some algebraic manipulations, we rewrite the T-MSE as

$$\text{T-MSE} = L - \sum_{i=1}^{L} \text{tr} \left(\mathbf{e}_i^H \begin{bmatrix} \mathbf{A}_i \\ \mathbf{0} \end{bmatrix} \left(\mathbf{A}_i^H \mathbf{A}_i + \xi \mathbf{I}_M \right)^{-1} \begin{bmatrix} \mathbf{A}_i^H & \mathbf{0} \end{bmatrix} \mathbf{e}_i \right). \tag{37}$$

The T-MSE in (37) is a function of $\mathbf{\Pi}$ for fixed $\mathbf{R}$. An exhaustive search is needed to find the optimal reordering matrix that minimizes (37). To avoid the high complexity of this global optimal approach, we adopt a suboptimal successive reordering algorithm that only maximizes one term of the summation in (37) and starts from the L-th term till the 1st term. The maximization of the ith term determines the ith row of $\mathbf{\Pi}$. The procedure of the reordering algorithm is listed in Table 3.
Till now we have found the suboptimal $\mathbf{\Pi}$, the optimal $\mathbf{F}$, $\mathbf{T}$ and β for fixed $\mathbf{R}$. Next we calculate the optimal $\mathbf{R}$ under fixed $\mathbf{\Pi}$, $\mathbf{F}$ and $\mathbf{T}$.
The T-MSE in (26) can be expanded as the summation of the K users' MSEs, and the MSE for the kth user is written as follows:

$$\text{MSE}_k = \mathcal{E}_{\mathbf{w},\mathbf{n}} \left\{ \left\| \tilde{\mathbf{b}}_k - \mathbf{b}_k \right\|_2^2 \right\} = \left\| \mathbf{R}_k \mathbf{H}_k \mathbf{T} - \mathbf{E}_k^T \mathbf{\Pi}^T (\mathbf{I}_L - \mathbf{F}) \right\|_F^2 + \zeta_k \cdot \text{tr}(\mathbf{R}_k^H \mathbf{R}_k), \tag{38}$$

in which $\mathbf{E}_k \triangleq \left[\mathbf{e}_{\sum_{i=1}^{k-1} L_i + 1}, \ \ldots, \ \mathbf{e}_{\sum_{i=1}^{k} L_i} \right]$ and $\zeta_k \triangleq P_T^{-1} \sigma_{n,k}^2 \text{tr}(\mathbf{T}^H \mathbf{T}) \geq 0$.

Initialization:
$\mathbf{A} = \mathbf{RH}$, $\mathbf{\Pi} = \mathbf{0}_{L\times L}$, $\xi = P_T^{-1}\sum_{k=1}^{K}\sigma_{n,k}^2 \mathrm{tr}\left(\mathbf{R}_k^H\mathbf{R}_k\right)$.
For $i = L:-1:1$,
$\mathbf{M} = \mathbf{A}\left(\mathbf{A}^H\mathbf{A} + \xi\mathbf{I}_M\right)^{-1}\mathbf{A}^H$.
$l^* = \max\limits_{1\le l\le L}[\mathbf{M}]_{l,l}$.
The ith row $\mathbf{ß}_i$ of $\mathbf{\Pi}$ is obtained as: $\mathbf{ß}_i = \mathbf{e}_{l^*}^T$.
Then set the entries of the l^*th row of $\mathbf{A}$ to zeros.
end

Table 3. The suboptimal ordering algorithm for THP.

Since $\mathbf{R}_k$ is only related to MSE_k, the Hessian matrix of T-MSE with respect to $\mathbf{R}_k$ is equal to that of MSE_k, which is calculated as

$$\frac{\partial}{\partial(\mathrm{vec}(\mathbf{R}_k))^T}\left(\frac{\partial \text{T-MSE}}{\partial(\mathrm{vec}(\mathbf{R}_k))^*}\right) = \frac{\partial}{\partial(\mathrm{vec}(\mathbf{R}_k))^T}\left(\frac{\partial \mathrm{MSE}_k}{\partial(\mathrm{vec}(\mathbf{R}_k))^*}\right)$$
$$= \left(\mathbf{H}_k\mathbf{T}\mathbf{T}^H\mathbf{H}_k^H + \zeta_k\mathbf{I}_{N_k}\right)^T \otimes \mathbf{I}_{L_k} \succeq \mathbf{0}. \tag{39}$$

The Hessian matrix in (39) being positive semidefinite indicates that the T-MSE is also convex with respect to $\mathbf{R}_k$. Then the optimal $\mathbf{R}_k$ is calculated in the same way as (35) :

$$\frac{\partial \text{T-MSE}}{\partial \mathbf{R}_k^*} = \frac{\partial \mathrm{MSE}_k}{\partial \mathbf{R}_k^*} = \mathbf{R}_k\mathbf{H}_k\mathbf{T}\mathbf{T}^H\mathbf{H}_k^H - \mathbf{E}_k^T\mathbf{\Pi}^T(\mathbf{I}_L - \mathbf{F})\mathbf{T}^H\mathbf{H}_k^H + \zeta_k\mathbf{R}_k = \mathbf{0}.$$
$$\Rightarrow \mathbf{R}_k = \mathbf{E}_k^T\mathbf{\Pi}^T(\mathbf{I}_L - \mathbf{F})\mathbf{T}^H\mathbf{H}_k^H\left(\mathbf{H}_k\mathbf{T}\mathbf{T}^H\mathbf{H}_k^H + \zeta_k\mathbf{I}_{N_k}\right)^{-1}. \tag{40}$$

As the inter-dependence among the optimal $\mathbf{\Pi}$, $\mathbf{F}$, $\mathbf{T}$, β and $\{\mathbf{R}_k\}_{k=1}^K$ has been found, we now summarize our iterative algorithm in Table 4, where the notations with the superscript $(\cdot)^{(n)}$ denote the related variables in the nth iteration.
The convergence of our proposed iterative algorithm can be guaranteed. The proof of convergence is in **Appendix B**.

4.2 Robust optimization of transceivers under imperfect CSIT

4.2.1 Channel uncertainty model

We consider a TDD system where the BS estimates the CSI using the training sequences in the uplink. The maximum-likelihood estimate of the actual channel matrix $\mathbf{H}_k$ can be modeled as (Hassibi & Hochwald, 2003) $\widehat{\mathbf{H}}_k = \mathbf{H}_k + \Delta\mathbf{H}_k$, where $\Delta\mathbf{H}_k$ denotes the error matrix whose entries are i.i.d. complex Gaussian distributed with zero mean and variance $\sigma_{e,k}^2$. $\Delta\mathbf{H}_k$ is statistically independent of $\mathbf{H}_k$. According to (Kay, 1993), the distribution of $\mathbf{H}_k$ conditioned on $\widehat{\mathbf{H}}_k$ is Gaussian and can be expressed as

$$\mathbf{H}_k \mid \widehat{\mathbf{H}}_k = \rho_k\widehat{\mathbf{H}}_k + \Delta\widetilde{\mathbf{H}}_k, \tag{41}$$

where $\rho_k = \sigma_{h,k}^2/(\sigma_{h,k}^2 + \sigma_{e,k}^2)$ and the entries of $\Delta\widetilde{\mathbf{H}}_k$ are i.i.d. complex Gaussian distributed with zero mean and variance $\tilde{\sigma}_k^2 = \sigma_{e,k}^2\sigma_{h,k}^2/(\sigma_{h,k}^2 + \sigma_{e,k}^2)$. We assume that the information of $\widehat{\mathbf{H}}_k$, $\sigma_{h,k}^2$ and $\sigma_{e,k}^2$, $k = 1,\ldots,K$ is known at the BS.
Note that the channel uncertainty caused by the slow time-variations of the channel can also be modeled in the same manner as (41) except that ρ_k has a different relationship with $\tilde{\sigma}_k^2$ (Khaled et al., 2004).

Step (1) Set the iteration number $n = 0$ and initialize $\mathbf{\Pi}^{(0)} = \mathbf{I}_L$ and $\mathbf{R}_k^{(0)} = \mathbf{U}_k^H$, where $\mathbf{U}_k$ comprises the L_k dominant left singular vectors of $\mathbf{H}_k$.

Step (2) Set $n = n + 1$. Calculate the reordering matrix $\widetilde{\mathbf{\Pi}}$ using the algorithm described in Table 3 and $\mathbf{R}^{(n-1)}$.
Calculate (37) using $\widetilde{\mathbf{\Pi}}$ and $\mathbf{R}^{(n-1)}$ and denote the result as $\widetilde{C}$.

Calculate (37) using $\mathbf{\Pi}^{(n-1)}$ and $\mathbf{R}^{(n-1)}$ and denote the result as C.
If $\widetilde{C} \leqslant C$
$\mathbf{\Pi}^{(n)} = \widetilde{\mathbf{\Pi}}$,
else
$\mathbf{\Pi}^{(n)} = \mathbf{\Pi}^{(n-1)}$.
end
Update transmit processing:
$\mathbf{t}_i^{(n)} = \left(\mathbf{A}_i^{(n),H}\mathbf{A}_i^{(n)} + \xi\mathbf{I}_M\right)^{-1}\left[\mathbf{A}_i^{(n),H}\ \mathbf{0}\right]\mathbf{e}_i$,
$\mathbf{f}_i^{(n)} = -\begin{bmatrix}\mathbf{0}\\ \mathbf{B}_i^{(n)}\end{bmatrix}\mathbf{t}_i^{(n)}, i = 1,\ldots,L$,
and $\beta^{(n)} = P_T^{\frac{1}{2}}\left(\mathrm{tr}(\mathbf{T}^{(n)}\mathbf{T}^{(n),H})\right)^{-\frac{1}{2}}$,
where
$\begin{bmatrix}\mathbf{A}_i^{(n)}\\ \mathbf{B}_i^{(n)}\end{bmatrix} = \mathbf{\Pi}^{(n)}\mathbf{R}^{(n-1)}\mathbf{H}$, $\mathbf{A}_i^{(n)} \in \mathbb{C}^{i\times M}$, $\mathbf{B}_i^{(n)} \in \mathbb{C}^{(L-i)\times M}$,
and $\xi = P_T^{-1}\sum_{k=1}^{K}\sigma_{n,k}^2\mathrm{tr}\left(\mathbf{R}_k^{(n-1),H}\mathbf{R}_k^{(n-1)}\right)$.
Update receiver processing:
$\mathbf{R}_k^{(n)} = \mathbf{E}_k^T\mathbf{\Pi}^{(n),T}(\mathbf{I}_L - \mathbf{F}^{(n)})\mathbf{T}^{(n),H}\mathbf{H}_k^H \cdot \left(\mathbf{H}_k\mathbf{T}^{(n)}\mathbf{T}^{(n),H}\mathbf{H}_k^H + \zeta_k\mathbf{I}_{N_k}\right)^{-1}, k = 1,\ldots,K$,
where $\zeta_k = P_T^{-1}\sigma_{n,k}^2\mathrm{tr}(\mathbf{T}^{(n),H}\mathbf{T}^{(n)})$.

Step (3) If $\| \mathbf{R}_k^{(n)} - \mathbf{R}_k^{(n-1)} \|_F^2 \geq \epsilon$, $\exists\, k \in \{1,\ldots,K\}$, then go to **Step (2)**. Otherwise, stop the iteration and the solution is given by $\mathbf{\Pi} = \mathbf{\Pi}^{(n)}$, $\mathbf{P} = \beta^{(n)}\mathbf{T}^{(n)}$, $\mathbf{F} = \mathbf{F}^{(n)}$, $\beta = \beta^{(n)}$ and $\mathbf{G}_k = \left(\beta^{(n)}\right)^{-1}\mathbf{R}_k^{(n)}$.

Table 4. The iterative algorithm for joint THP transceiver design.

4.2.2 Robust optimization problem formulation and iterative algorithm

When only the channel estimates $\widehat{\mathbf{H}}_k$, $k = 1,\ldots,K$ are available at the BS, the definition of T-MSE in (26) and MSE_k in (38) cannot be directly applied to the transceiver design. Instead, the expectation of MSE conditioned on $\widehat{\mathbf{H}}_k$ is an applicable performance measure and provides the robustness against the channel uncertainties in an average manner (Dietrich et al., 2007; Shenouda & Davidson, 2007). By (38), the conditional expectation of MSE of user k is expressed

as

$$\begin{aligned}
&\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\{\mathrm{MSE}_k\} = \mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\left\{\left\|\mathbf{R}_k\mathbf{H}_k\mathbf{T}-\mathbf{E}_k^T\mathbf{\Pi}^T(\mathbf{I}_L-\mathbf{F})\right\|_F^2+\zeta_k\cdot\mathrm{tr}(\mathbf{R}_k^H\mathbf{R}_k)\right\}\\
&=\mathrm{tr}\Bigg(\mathbf{R}_k\cdot\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\left\{\mathbf{H}_k\mathbf{T}\mathbf{T}^H\mathbf{H}_k^H\right\}\cdot\mathbf{R}_k^H-\mathbf{R}_k\cdot\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\{\mathbf{H}_k\}\cdot\mathbf{T}\left(\mathbf{E}_k^T\mathbf{\Pi}^T(\mathbf{I}_L-\mathbf{F})\right)^H\\
&-\mathbf{E}_k^T\mathbf{\Pi}^T(\mathbf{I}_L-\mathbf{F})\mathbf{T}^H\cdot\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\left\{\mathbf{H}_k^H\right\}\cdot\mathbf{R}_k^H+\mathbf{E}_k^T\mathbf{\Pi}^T(\mathbf{I}_L-\mathbf{F})\left(\mathbf{E}_k^T\mathbf{\Pi}^T(\mathbf{I}_L-\mathbf{F})\right)^H\Bigg)\\
&+\zeta_k\cdot\mathrm{tr}(\mathbf{R}_k^H\mathbf{R}_k).
\end{aligned}\tag{42}$$

Using (41) it can be easily verified that

$$\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\{\mathbf{H}_k\}=\rho_k\widehat{\mathbf{H}}_k,\tag{43}$$

$$\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\left\{\mathbf{H}_k^H\right\}=\rho_k\widehat{\mathbf{H}}_k^H.\tag{44}$$

Next we calculate the quadratic term $\mathbf{Q}_k \triangleq \mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\{\mathbf{H}_k\mathbf{T}\mathbf{T}^H\mathbf{H}_k^H\}$. Let $\mathbf{H}_k = \left[\mathbf{h}_{k,1}^T\,\mathbf{h}_{k,2}^T\ldots\mathbf{h}_{k,N_k}^T\right]^T$, where $\mathbf{h}_{k,m}$ is the mth row of $\mathbf{H}_k$. Then the element at the mth row and nth column of $\mathbf{Q}_k$ can be written as

$$\begin{aligned}
[\mathbf{Q}_k]_{m,n} &= \mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\left\{\mathbf{h}_{k,m}\mathbf{T}\mathbf{T}^H\mathbf{h}_{k,n}^H\right\}=\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\left\{\mathrm{tr}\left(\mathbf{h}_{k,m}\mathbf{T}\mathbf{T}^H\mathbf{h}_{k,n}^H\right)\right\}\\
&= \mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\left\{\mathrm{tr}\left(\mathbf{h}_{k,n}^H\mathbf{h}_{k,m}\mathbf{T}\mathbf{T}^H\right)\right\}=\mathrm{tr}\left(\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\left\{\mathbf{h}_{k,n}^H\mathbf{h}_{k,m}\right\}\mathbf{T}\mathbf{T}^H\right).
\end{aligned}\tag{45}$$

According to (41), we have

$$\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\left\{\mathbf{h}_{k,n}^H\mathbf{h}_{k,m}\right\}=\rho_k^2\hat{\mathbf{h}}_{k,n}^H\hat{\mathbf{h}}_{k,m}+\mathcal{E}\left\{\Delta\tilde{\mathbf{h}}_{k,n}^H\Delta\tilde{\mathbf{h}}_{k,m}\right\},\tag{46}$$

where

$$\mathcal{E}\left\{\Delta\tilde{\mathbf{h}}_{k,n}^H\Delta\tilde{\mathbf{h}}_{k,m}\right\}=\begin{cases}\tilde{\sigma}_k^2\mathbf{I}_M, & m=n\\ 0. & m\neq n\end{cases}$$

Therefore,

$$[\mathbf{Q}_k]_{m,n}=\begin{cases}\rho_k^2\hat{\mathbf{h}}_{k,m}\mathbf{T}\mathbf{T}^H\hat{\mathbf{h}}_{k,n}^H+\tilde{\sigma}_k^2\mathrm{tr}\left(\mathbf{T}\mathbf{T}^H\right), & m=n\\ \rho_k^2\hat{\mathbf{h}}_{k,m}\mathbf{T}\mathbf{T}^H\hat{\mathbf{h}}_{k,n}^H. & m\neq n\end{cases}$$

Finally, $\mathbf{Q}_k$ can be expressed as

$$\mathbf{Q}_k=\rho_k^2\widehat{\mathbf{H}}_k\mathbf{T}\mathbf{T}^H\widehat{\mathbf{H}}_k^H+\tilde{\sigma}_k^2\mathrm{tr}\left(\mathbf{T}\mathbf{T}^H\right)\mathbf{I}_{N_k}.\tag{47}$$

By substituting (43), (44) and (47) into (42), we can obtain the explicit expression of $\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\{\mathrm{MSE}_k\}$ as follows:

$$\mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\{\mathrm{MSE}_k\}=\left\|\rho_k\mathbf{R}_k\widehat{\mathbf{H}}_k\mathbf{T}-\mathbf{E}_k^T\mathbf{\Pi}^T(\mathbf{I}_L-\mathbf{F})\right\|_F^2+\tilde{\zeta}_k\cdot\mathrm{tr}(\mathbf{R}_k^H\mathbf{R}_k),\tag{48}$$

in which $\tilde{\zeta}_k \triangleq (P_T^{-1}\sigma_{n,k}^2 + \tilde{\sigma}_k^2) \cdot \mathrm{tr}(\mathbf{T}^H\mathbf{T})$. Then the conditional expectation of the T-MSE can be expressed as:

$$\mathcal{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\{\text{T-MSE}\} = \sum_{k=1}^{K} \mathcal{E}_{\mathbf{H}_k|\widehat{\mathbf{H}}_k}\{\text{MSE}_k\} = \left\|\mathbf{R}\widehat{\mathbf{H}}\mathbf{T} - \mathbf{\Pi}^T(\mathbf{I}_L - \mathbf{F})\right\|_F^2 + \tilde{\xi} \cdot \mathrm{tr}(\mathbf{T}^H\mathbf{T}), \tag{49}$$

where $\widehat{\mathbf{H}} \triangleq \left[\rho_1\widehat{\mathbf{H}}_1^T \ldots \rho_K\widehat{\mathbf{H}}_K^T\right]^T$ and $\tilde{\xi} \triangleq \sum_{k=1}^{K}\left(P_T^{-1}\sigma_{n,k}^2 + \tilde{\sigma}_k^2\right)\mathrm{tr}\left(\mathbf{R}_k^H\mathbf{R}_k\right)$.
Finally, the robust transceiver optimization problem is mathematically formulated as below:

$$\begin{aligned} &\min_{\mathbf{\Pi},\,\mathbf{F},\,\mathbf{T},\,\{\mathbf{R}_k\}_{k=1}^K} \mathcal{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\{\text{T-MSE}\} \\ &\text{s.t.} \quad [\mathbf{F}]_{m,n} = 0,\ \forall\, 1 \leqslant m,n \leqslant L \text{ and } m \leqslant n. \end{aligned} \tag{50}$$

Following the derivation of transceiver design under perfect CSIT, we can easily obtain the necessary conditions for the optimal robust transceiver under imperfect CSIT as follows. For fixed $\mathbf{\Pi}$ and $\mathbf{R}$, the optimal $\mathbf{f}_i$ and $\mathbf{t}_i$ are listed below:

$$\mathbf{f}_i = -\begin{bmatrix}\mathbf{0}_{i\times M}\\ \widehat{\mathbf{B}}_i\end{bmatrix}\mathbf{t}_i, \tag{51}$$

$$\mathbf{t}_i = \left(\widehat{\mathbf{A}}_i^H\widehat{\mathbf{A}}_i + \tilde{\xi}\mathbf{I}_M\right)^{-1}\begin{bmatrix}\widehat{\mathbf{A}}_i^H & \mathbf{0}\end{bmatrix}\mathbf{e}_i,\ i = 1,\ldots,L, \tag{52}$$

where

$$\begin{bmatrix}\widehat{\mathbf{A}}_i\\ \widehat{\mathbf{B}}_i\end{bmatrix} = \mathbf{\Pi}\mathbf{R}\widehat{\mathbf{H}},\ \widehat{\mathbf{A}}_i \in \mathbb{C}^{i\times M},\ \widehat{\mathbf{B}}_i \in \mathbb{C}^{(L-i)\times M}. \tag{53}$$

By substituting (51) and (52) into (49), we reformulate $\mathcal{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\{\text{T-MSE}\}$ as:

$$\mathcal{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\{\text{T-MSE}\} = L - \sum_{i=1}^{L}\mathrm{tr}\left(\mathbf{e}_i^H\begin{bmatrix}\widehat{\mathbf{A}}_i\\ \mathbf{0}\end{bmatrix}\left(\widehat{\mathbf{A}}_i^H\widehat{\mathbf{A}}_i + \tilde{\xi}\mathbf{I}_M\right)^{-1}\begin{bmatrix}\widehat{\mathbf{A}}_i^H & \mathbf{0}\end{bmatrix}\mathbf{e}_i\right). \tag{54}$$

Then the suboptimal successive ordering algorithm presented in Table 3 can be applied with $\mathbf{H}$ replaced by $\widehat{\mathbf{H}}$ and the expression of ξ replaced by $\tilde{\xi} = \sum_{k=1}^{K}\left(P_T^{-1}\sigma_{n,k}^2 + \tilde{\sigma}_k^2\right)\mathrm{tr}\left(\mathbf{R}_k^H\mathbf{R}_k\right)$.
For fixed $\mathbf{\Pi}$, $\mathbf{F}$ and $\mathbf{T}$, the optimal $\mathbf{R}_k$ is given by

$$\mathbf{R}_k = \rho_k\mathbf{E}_k^T\mathbf{\Pi}^T(\mathbf{I}_L - \mathbf{F})\mathbf{T}^H\mathbf{H}_k^H\left(\rho_k^2\mathbf{H}_k\mathbf{T}\mathbf{T}^H\mathbf{H}_k^H + \tilde{\zeta}_k\mathbf{I}_{N_k}\right)^{-1}. \tag{55}$$

Now that we have found the inter-dependence among $\mathbf{\Pi}$, $\mathbf{F}$, $\mathbf{T}$ and $\{\mathbf{R}_k\}_{k=1}^K$ as shown in (51)-(55), the thread of the iterative algorithm proposed in Subsection 4.1 can be again adopted here to compute the robust transceiver. In each iteration, we first determine the suboptimal reordering matrix $\mathbf{\Pi}$ and update $\mathbf{T}$ and $\mathbf{F}$ using $\{\mathbf{R}_k\}_{k=1}^K$ updated in the last iteration, then update $\{\mathbf{R}_k\}_{k=1}^K$ using the above updated $\mathbf{\Pi}$, $\mathbf{T}$ and $\mathbf{F}$ until convergence. The formulation of the algorithm is similar to that described in Table 4 except for some notation and expression changes so we do not list the details of the modified algorithm.
The convergence of the iterative algorithm is also guaranteed. The proof of convergence is also similar to that in **Appendix B**.

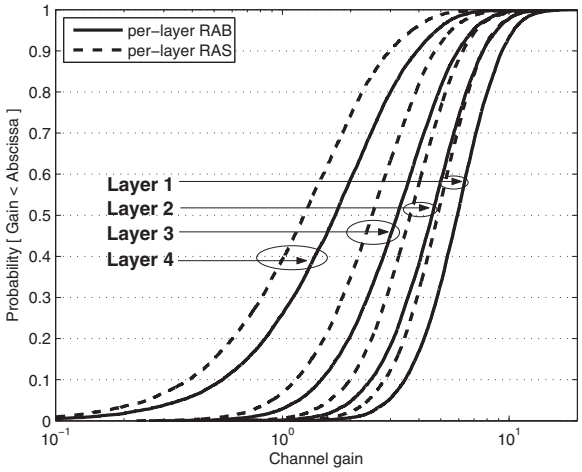

Fig. 4. The cumulative distribution function of equivalent channel gains of four layers applying per-layer RAB and RAS in $(2 \times 4) \times 4$ systems.

5. Simulation results

In this section, we will present some simulation results to verify the effectiveness of our proposed per-layer ZF-THP schemes and stream-wise MMSE-THP schemes, respectively.

5.1 Results of per-layer ZF-THP

We consider the system settings of one transmitter with M antennas and M receivers each with N antennas, denoted by $(N \times M) \times M$. For comparison, we also simulate the performance of the parallel linear-ZF precoding where each user combines its receive antennas according to the eigenmode, and the performance of the per-user processing which selects $\lceil \frac{M}{N} \rceil$ users with the largest channel gains and allocates each of them with N adjacent layers.

Fig. 4 and Fig. 5 compare the equivalent channel gains and capacities respectively of the four layers by per-layer RAB and RAS with the equal power allocation among layers. Here, we place users into layers according to their indexes, i.e., $\pi_l = l, l = 1, \cdots, 4$. It can be found that for the layers from 4 to 1, the relative differences of channel gains between these two schemes gradually decrease, though the rates of decreasing get slower. At the same time, the order of layers according to the differences of channel capacities between per-layer RAB and per-layer RAS exhibits two results. When SNR<3dB, the capacity difference is the largest in layer 1, and the smallest in layer 4. On the contrary, when SNR>7dB, the order from large to small is layer 4 to layer 1. These results are consistent with Corollary 1 and 2.

In Fig. 6, the rate regions of the $(2 \times 2) \times 2$ systems are considered, whose boundaries are generated by averaging channel realizations. For THP, the rate regions are asymmetric, which means the higher layer has the larger capacity than the lower one. With the ideal power allocation among layers, the order of the maximum sum-capacities from large to small is RAB, RAS, per-user, and linear-ZF.

In Fig. 7, the ergodic sum-capacity of the $(2 \times 4) \times 4$ system is evaluated. Here, in order to achieve the potential sum-capacity, the water-filling power allocation (Tse & Viswanath, 2005) among layers is applied. Among these four curves, the advantage of THP over linear-ZF, and

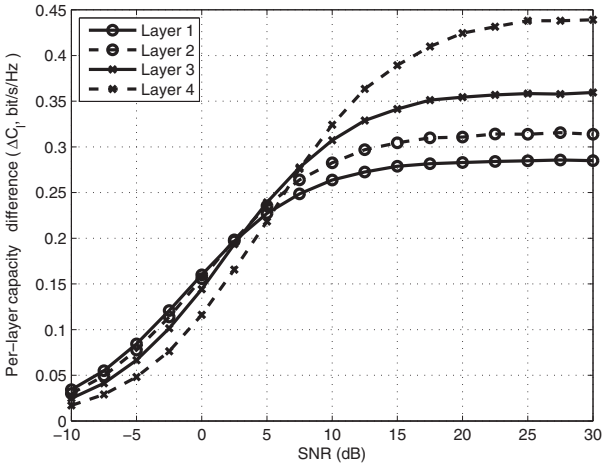

Fig. 5. The channel capacity differences of four layers between per-layer RAB and per-layer RAS in $(2 \times 4) \times 4$ systems.

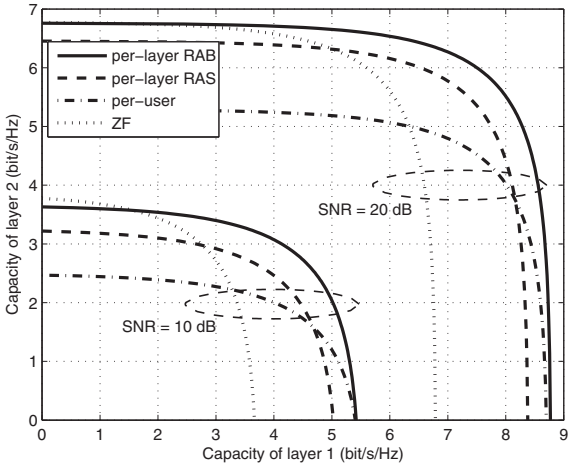

Fig. 6. The rate regions in $(2 \times 2) \times 2$ systems.

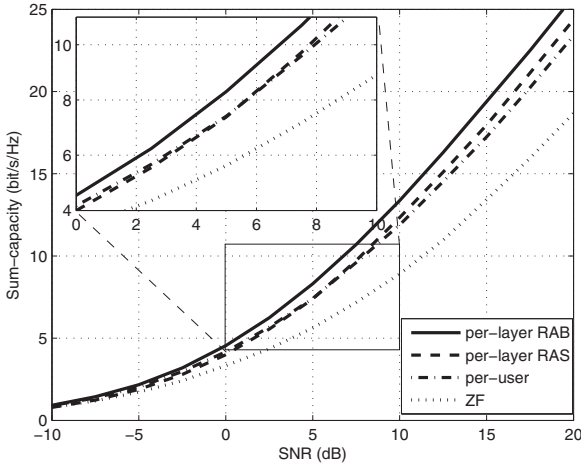

Fig. 7. The ergodic sum-capacities in $(2 \times 4) \times 4$ systems. The water-filling power allocation is applied.

the advantage of RAB over RAS and per-user are obvious; the curves of RAS and per-user are intersected at SNR≈5dB, i.e., when SNR<5dB, per-user outperforms RAS, but when SNR>5dB, it is the opposite.
To sum up, the essential of these comparison results among THP schemes is as follows.

- With the larger power gain and diversity gain, per-layer beamforming outperforms per-user processing.
- Per-layer RAS and per-user processing actually exploit the same number of receive antennas, though for the former, these antennas belong to K users, while for the latter, they belong to $\lceil \frac{K}{N} \rceil$ users. In the low SNR scenarios, per-user method can obtain larger power gains due to its dominant intra-user processing and less inter-user interference suppression. However, in the high SNR scenarios, this effect of power gains makes trivial contribution to the system sum-capacity, but the larger multiuser and multi-antenna selection diversity gain in per-layer RAS turns to the dominant factor.

5.2 Results of stream-wise MMSE-THP

In this subsection, some results are presented to show the performance superiority of our proposed joint stream-wise THP transceiver designs in comparison with some existing THP schemes. The illustration is divided into two parts. The first part illustrates the performance under perfect CSIT and the second part focuses on the performance under imperfect CSIT. We assume quasi-static i.i.d. Rayleigh flat fading channel with unit channel variance between each transmit antenna at the BS and each receive antenna at each user. We also assume $\sigma_{n,k}^2 = 1$, $k = 1, \ldots, K$. The signal-to-noise ratio (SNR) in the following figures is defined as SNR $\triangleq 10 \cdot \log_{10} P_T$. QPSK or 16-QAM modulations are used in the simulations. We set the convergence threshold in the iterative algorithms $\epsilon = 10^{-5}$.

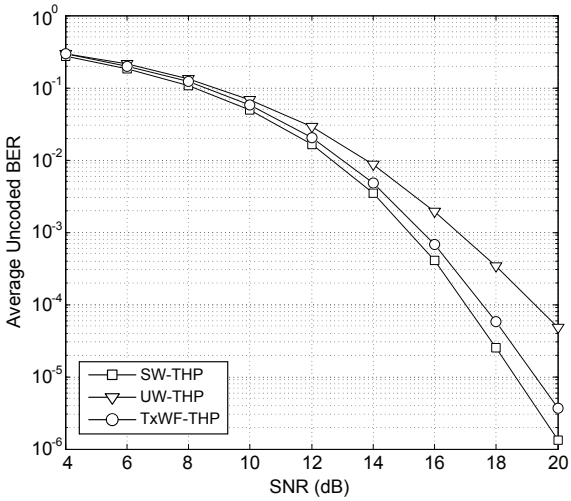

Fig. 8. Performance comparison of different schemes with $M = 6,\ K = 3,\ N_k = 2,\ L_k = 2,\ \forall k$ and QPSK.

5.2.1 Performance under perfect CSIT

We examine the performance of our proposed joint THP transceiver design under perfect CSIT in comparison with some existing *MMSE-based* THP schemes.

Figs. 8-9 compare our proposed *stream-wise* THP transceiver (denote as "SW-THP") with the *user-wise* THP transceiver in (Mezghani et al., 2006) (denoted as "UW-THP") and "TxWF-THP" in (Joham et al., 2004) in terms of average uncoded bit error rate (BER). In Fig. 8, we set $M = 6,\ K = 3,\ N_k = N = 2,\ L_k = L = 2,\ \forall k$ and use QPSK. For "TxWF-THP" we assume that the 2 antennas at the same receiver are decentralized, i.e., there are 6 *virtual* users in the system. From the simulation results we can see that our scheme clearly outperforms the other two schemes. The superiority over "UW-THP" comes from that our scheme performs stream-wise interference pre-cancelation while "UW-THP" only enables inter-user interference pre-cancelation and multiple streams of the same user are linearly precoded. Moreover, our scheme outperforms "TxWF-THP" because in our scheme the received signals from multiple antennas of one user can be jointly processed while in "TxWF-THP", the receive antennas of the same user are assumed to be decentralized and a common scaling factor is imposed for each single receive antenna, which is apparently suboptimal.

In Fig. 9, we increase N by 1 and keep the other parameters unchanged. Since $N > L$, "TxWF-THP" cannot be directly applied. Here we simply select $\mathbf{U}_k^H$ as the receiver for user k, where $\mathbf{U}_k$ comprises the L_k dominant left singular vectors of $\mathbf{H}_k$, then apply $\mathbf{U}_k^H\mathbf{H}_k$ as the equivalent $L_k \times M$ channel matrix, which is suitable for implementation of "TxWF-THP". It is shown in this figure that our scheme still performs best.

It is an interesting phenomenon that the comparison results of "UW-THP" and "TxWF-THP" in Fig. 8 and 9 are just opposite, so are those for 16-QAM[3]. This can be explained that for Fig. 8, more interferences are pre-canceled in "TxWF-THP" than in "UW-THP". For Fig. 9, however,

[3] Due to page limit, we don't show the results of 16-QAM here. Please refer to (Miao et al., 2009).

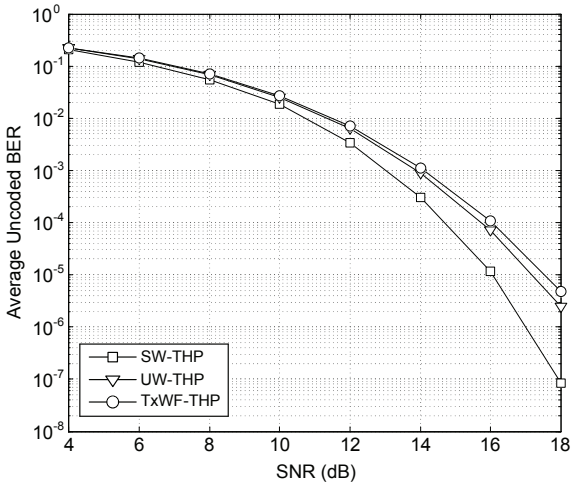

Fig. 9. Performance comparison of different schemes with $M = 6,\ K = 3,\ N_k = 3,\ L_k = 2,\ \forall k$ and QPSK.

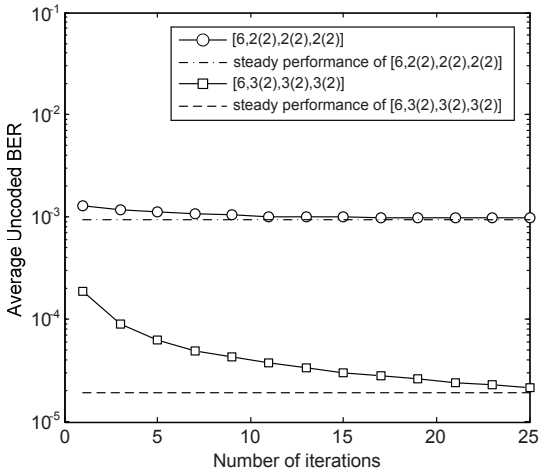

Fig. 10. verage uncoded 16-QAM BER performance of SW-THP under different number of iterations at SNR = 22dB.

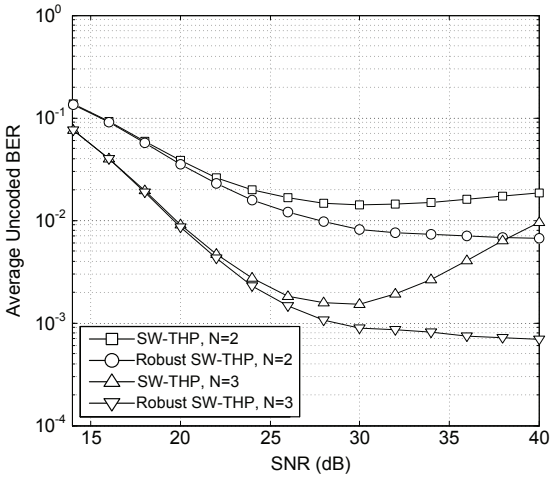

Fig. 11. Performance comparison of non-robust and robust SW-THP with $M = 6,\ K = 3,\ L = 2,\ \sigma_e^2 = 0.01$ and 16-QAM.

the additional receive antenna at each user provides more diversity for each data stream in "UW-THP" through joint transceiver design while the simple and suboptimal receiver structure we have adopted for "TxWF-THP" restricts its performance improvement.

We also present some simulation results to verify the convergence of our proposed iterative algorithm. We use the notation $[M, N_1(L_1), ..., N_K(L_K)]$ to denote a K-user MIMO system that has M antennas at the BS and N_k antennas at the kth receiver to support L_k data streams for user k, $k = 1, ..., K$. Fig. 10 shows the average uncoded BER of 16-QAM versus the number of iterations for our scheme under two system configurations at a fixed SNR = 22dB. The dashed lines denote the steady state performance. It can be seen that after about 10 iterations (for $[6, 2(2), 2(2), 2(2)]$) or 25 iterations (for $[6, 3(2), 3(2), 3(2)]$) the BER performance is quite close to the steady state performance. Similar results hold for QPSK, thus are not shown here for brevity.

5.2.2 Performance under imperfect CSIT

In this subsection, we examine the performance of SW-THP introduced in Subsection 4.1 and the robust version of SW-THP introduced in Subsection 4.2 under imperfect CSIT.

Fig. 11 compare "SW-THP" and "Robust SW-THP" for $M = 6$, $K = 3$, $L_k = 2$, $N_k = N = 2$ and 3 (the parameters are consistent with those in Subsection 5.1). We assume $\sigma_{e,k}^2 = \sigma_e^2 = 0.01$, $k = 1, \ldots, K$ for 16-QAM. It can be seen that the non-robust and robust schemes have close performance at low SNR since the noise is dominant in this regime. However, the gap between the two schemes enlarges as the SNR increases because the channel uncertainty gradually dominates and the efficacy of the robust scheme shows up.

In Fig. 12 we test the performance of "SW-THP" and "Robust SW-THP" under different channel errors. The antenna configurations are the same as above, i.e., $M = 6$, $K = 3$, $L_k = 2$, $N_k = N = 2$ and 3. We fix SNR = 40 dB, and change σ_e^2 from 0.01 to 0.1 for 16-QAM in Fig. 12. The simulated curves reveal that "Robust SW-THP" always holds a considerable superiority over the non-robust "SW-THP" as the channel error increases.

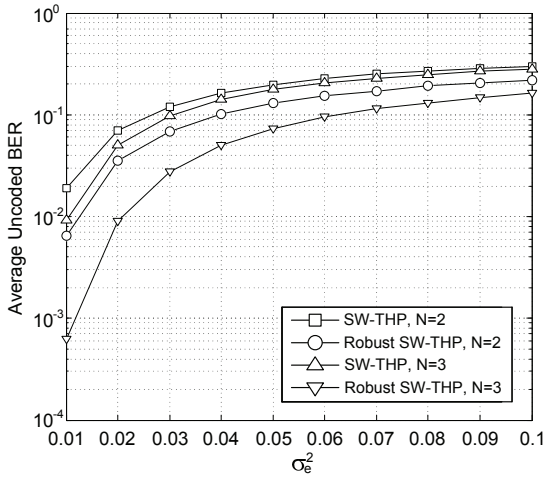

Fig. 12. Performance comparison of non-robust and robust SW-THP with $M = 6,\ K = 3,\ L = 2$, SNR = 40dB and 16-QAM for different channel errors.

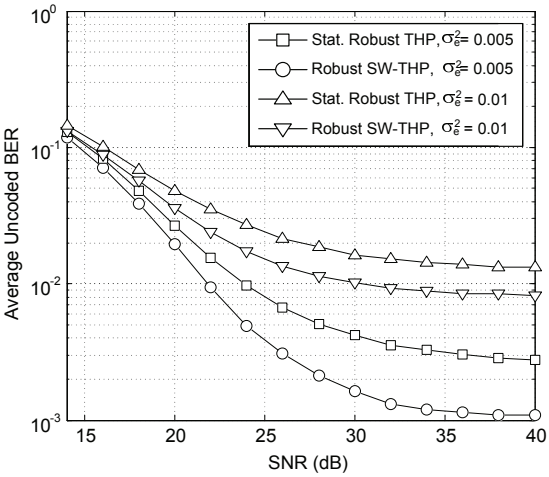

Fig. 13. Performance comparison of robust SW-THP and Stat. Robust THP in (Shenouda & Davidson, 2007) with $M = 4,\ K = 2,\ N = 2,\ L = 2$ and 16-QAM.

We also compare our robust THP scheme with "Statistically Robust Tomlinson-Harashima precoding" (denoted as "Stat. Robust THP") reported lately in (Shenouda & Davidson, 2007). The results are shown in Fig. 13. We choose $M = 4$, $K = 2$, $L_k = L = 2$, $N_k = N = 2$. For 16-QAM we set $\sigma_{e,k}^2 = \sigma_e^2 = 0.005$ and 0.01. Since "Stat. Robust THP" is designed only for the MISO downlink, it cannot be directly applied here. So for "Stat. Robust THP", we assume that the 2 antennas at the same receiver are decentralized, i.e., there are 4 *virtual* users in the system. From Fig. 13 we can find out that "Robust SW-THP" outperforms "Stat. Robust THP" in the entire SNR regime and also has a much lower BER floor at high SNR. The advantage of our scheme comes from two aspects. First, we jointly optimize the transmitter, the receivers and the precoding order with consideration of the channel errors. In contrast, only the transmitter of "Stat. Robust THP" is robustly designed and the receivers and the precoding order are determined as if the estimated CSI was the true one. Second, the received signals from the two antennas of the same user can be jointly processed using "Robust SW-THP" but can only be processed independently by "Stat. Robust THP".

6. Conclusions

In this chapter, the multiuser MIMO downlink systems with THP based on ZF and MMSE criteria have been investigated, respectively.

For the ZF-THP schemes, based on the criterion of maximum system sum-capacity, two per-layer joint transmit and receive filters design schemes have been studied, which apply RAB and RAS, respectively. Furthermore, the comparison between equivalent channel gains and capacities within these two per-layer schemes have been proved, including the trend with the layers from low to high and the trend with the increase of the number of transmit antennas to infinite.

For the MMSE-THP schemes, we have jointly designed the THP transceiver under both perfect and imperfect CSIT. First, we deal with the perfect CSIT case. Based on the equivalent linear representation of the THP downlink, we formulate a T-MSE minimization problem under the total power constraint. After some analysis, an iterative algorithm is proposed to obtain a locally optimal solution. Next, we consider the imperfect CSIT case. By minimizing the conditional-expectation of the T-MSE, we extend the algorithm derived under perfect CSIT to the robust version that is more effective against the channel uncertainty. The two characteristics of our schemes, i.e., the stream-wise interference pre-cancelation and the joint design of transceiver, ensure its performance superiority over the existing MMSE-based THP schmes, which has been verified by the simulation results.[4]

7. Appendix

7.1 Appendix A: Lower bounds for per-user processing and per-layer RAS

Derived from the proof of Theorem 1, $E\left(\widetilde{\lambda}_n^{max,2} - \widetilde{\lambda}_n^{min,2}\right)^2 = 24n$. Besides,

$$E\left(\widetilde{\lambda}_n^{max,2} + \widetilde{\lambda}_n^{min,2}\right)^2 \geq \left[E\left(\widetilde{\lambda}_n^{max,2} + \widetilde{\lambda}_n^{min,2}\right)\right]^2 = (4n)^2. \tag{56}$$

[4] This work is partially supported by National Basic Research Program (2007CB310608), National Science and Technology Pillar Program (2008BAH30B09), National Natural Science Foundation of China (60832008), National Major Project (2008ZX03O03-004), China 863 Project (2009AA011501) and PCSIRT.

For per-user processing,

$$E\left(\widetilde{\lambda}_n^{max,2}\widetilde{\lambda}_n^{min,2}\right) = \frac{1}{4}\left[E\left(\widetilde{\lambda}_n^{max,2} + \widetilde{\lambda}_n^{min,2}\right)^2 - E\left(\widetilde{\lambda}_n^{max,2} - \widetilde{\lambda}_n^{min,2}\right)^2\right] \geq 4n^2 - 6n. \tag{57}$$

For per-layer RAS, due to the operations of maximization,

$$E\left(\widetilde{p}_n^{max,2}\widetilde{p}_{n-1}^{max,2}\right) \geq E\left(x_{2n}^2 x_{2n-2}^2\right) = 2n \cdot (2n-2) = 4n^2 - 4n, \tag{58}$$

where x_{2n}^2 and x_{2n-2}^2 stand for two chi-square distributed random variables with $2n$ and $2n-2$ degrees of freedom, respectively.

7.2 Appendix B: Proof of convergence of the proposed iterative algorithm

We use the notation $f\left(\mathbf{\Pi}^{(n)}, \mathbf{F}^{(n)}, \mathbf{T}^{(n)}, \mathbf{R}^{(n)}\right)$ to represent the T-MSE calculated after the nth iteration of our iterative algorithm. If we fix $\mathbf{\Pi}^{(n)}$ and $\mathbf{R}^{(n)}$, and update $\mathbf{F}^{(n)}$ and $\mathbf{T}^{(n)}$ using the optimal conditions derived in (31) and (35), then due to the convexity of T-MSE with respect to $\mathbf{t}_i$, we have the following inequality:

$$f\left(\mathbf{\Pi}^{(n)}, \mathbf{F}^{(n)}, \mathbf{T}^{(n)}, \mathbf{R}^{(n)}\right) \geqslant f\left(\mathbf{\Pi}^{(n)}, \mathbf{F}', \mathbf{T}', \mathbf{R}^{(n)}\right), \tag{59}$$

where $\mathbf{F}'$ and $\mathbf{T}'$ are the updated matrices using $\mathbf{\Pi}^{(n)}$ and $\mathbf{R}^{(n)}$.
We define $h\left(\mathbf{\Pi}, \mathbf{R}\right) \triangleq \min_{\mathbf{F},\mathbf{T}} f\left(\mathbf{\Pi}, \mathbf{F}, \mathbf{T}, \mathbf{R}\right)$. Obviously $h\left(\mathbf{\Pi}, \mathbf{R}\right)$ is equal to (37). From the definition of h we know that

$$h\left(\mathbf{\Pi}^{(n)}, \mathbf{R}^{(n)}\right) = f\left(\mathbf{\Pi}^{(n)}, \mathbf{F}', \mathbf{T}', \mathbf{R}^{(n)}\right). \tag{60}$$

We denote the permutation matrix calculated using $\mathbf{R}^{(n)}$ and the suboptimal ordering algorithm in Table 3 as $\mathbf{\Pi}'$. As stated in Table 4, if $\mathbf{\Pi}'$ leads to a smaller value of (37) than $\mathbf{\Pi}^{(n)}$, then $\mathbf{\Pi}^{(n+1)} = \mathbf{\Pi}'$. Otherwise, $\mathbf{\Pi}^{(n+1)} = \mathbf{\Pi}^{(n)}$. Then the following inequality holds:

$$h\left(\mathbf{\Pi}^{(n)}, \mathbf{R}^{(n)}\right) \geqslant h\left(\mathbf{\Pi}^{(n+1)}, \mathbf{R}^{(n)}\right). \tag{61}$$

From the definition of h and iteration procedure, we have:

$$h\left(\mathbf{\Pi}^{(n+1)}, \mathbf{R}^{(n)}\right) = f\left(\mathbf{\Pi}^{(n+1)}, \mathbf{F}^{(n+1)}, \mathbf{T}^{(n+1)}, \mathbf{R}^{(n)}\right). \tag{62}$$

Since f is also convex with respect to $\mathbf{R}_k$, the following inequality holds:

$$f\left(\mathbf{\Pi}^{(n+1)}, \mathbf{F}^{(n+1)}, \mathbf{T}^{(n+1)}, \mathbf{R}^{(n)}\right) \geqslant f\left(\mathbf{\Pi}^{(n+1)}, \mathbf{F}^{(n+1)}, \mathbf{T}^{(n+1)}, \mathbf{R}^{(n+1)}\right)$$

Combing (59)-(63) we finally get:

$$f\left(\mathbf{\Pi}^{(n+1)}, \mathbf{F}^{(n+1)}, \mathbf{T}^{(n+1)}, \mathbf{R}^{(n+1)}\right) \leqslant f\left(\mathbf{\Pi}^{(n)}, \mathbf{F}^{(n)}, \mathbf{T}^{(n)}, \mathbf{R}^{(n)}\right). \tag{63}$$

The inequality in (63) indicates that during the iterations of our proposed iterative algorithm, the T-MSE is decreasing. Moreover, T-MSE is obviously lower bounded by 0. Therefore, the iterative algorithm guarantees the convergence.
Thus, we have completed the proof. ■

8. References

Caire, G. & Shamai, S. (2003). On the achievable throughput of a multi-antenna gaussian broadcast channel, *IEEE Trans. Inform. Theory* Vol.49(No.7): 1691–1706.

Choi, R. & Murch, R. (2004). A transmit pre-processing technique for multiuser mimo systems: a decomposition approach, *IEEE Trans. Wireless Commun.* Vol.3(No.1): 20–24.

Dietrich, F., Breun, P. & Utschick, W. (2007). Robust tomlinson-harashima precoding for the wireless broadcast channel, *IEEE Trans. Sig. Process.* Vol.55(No.2): 631–644.

Doostnejad, R., Lim, T. & Sousa, E. (2005). Joint precoding and beamforming design for the downlink in a multiuser mimo system, *Proc. Conf. on Wireless and Mobile Computing, Networking and Communications 2005*, Montreal, Canada, pp. 153–159.

Gorokhov, A., Gore, D. & Paulraj, A. (2003). Receive antenna selection for mimo flat-fading channels: theory and algorithms, *IEEE Trans. Information Theory* Vol.49(No.10): 2687–2696.

Hassibi, B. & Hochwald, B. (2003). How much training is needed in multiple-antenna wireless links?, *IEEE Trans. Inform. Theory* Vol.49(No.4): 951–963.

Horn, R. & Johnson, C. (1985). *Matrix Analysis*, Cambridge University Press, Cambridge, UK.

Huang, M., Chen, X., Zhou, S. & Wang, J. (2010). Per-layer transmit and receive filters design for tomlinson−harashima precoding in multiuser mimo systems, *submitted to Science in China Series−F* .

Hunger, R., Dietrich, F., Joham, M. & Utschick, W. (2004). Robust transmit zero-forcing filters, *Proc. ITG Workshop Smart Antennas*, Munich, Germany, pp. 130–137.

Joham, M., Brehmer, J. & Utschick, W. (2004). Mmse approaches to multiuser spatio-temporal tomlinson-harashima precoding, *Proc. 5th Int. ITG Conf. on Source and Channel Coding*, Erlangen, Germany, pp. 1117–1120.

Kay, S. (1993). *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall PTR, London, UK.

Khaled, N., Leus, G., Desset, C. & Man, H. (2004). A robust joint linear precoder and decoder mmse design for slowly time-varying mimo channels, *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Montreal, Canada, pp. 485–488.

Mezghani, A., Hunger, R., Joham, M. & Utschick, W. (2006). Iterative thp transceiver optimization for multi-user mimo systems based on weighted sum-mse minimization, *Proc. IEEE 7th Workshop on Signal Processing Advances in Wireless Communications*, Cannes, France, pp. 1–5.

Miao, W., Chen, X., Zhou, S. & Wang, J. (2009). Joint stream-wise thp transceiver design for the multiuser mimo downlink, *IEICE Transactions on Communications* Vol.E92-B(No.1): 209–218.

Payaro, M., Pascual-Iserte, A., Perez-Neira, A. & Lagunas, M. (2007). Robust design of spatial tomlinson-harashima precoding in the presence of errors in the csi, *IEEE Trans. Wireless Commun.* Vol.6(No.7): 2396–2401.

Schubert, M. & Shi, S. (2005). Mmse transmit optimization with interference pre-compensation, *Proc. Vehicular Technology Conf. 2005-Spring*, Stockholm, Sweden, pp. 845–849.

Shenouda, M. & Davidson, T. (2007). Tomlinson-harashima precoding for broadcast channels with uncertainty, *IEEE J. Sel. Area. Comm.* Vol.25(No.7): 1380–1389.

Shenouda, M. & Davidson, T. (2008). A framework for designing mimo systems with decision feedback equalization or tomlinson-harashima precoding, *IEEE J. Sel. Area. Comm.* Vol.26(No.2): 401–411.

Stankovic, V. & Haardt, M. (2005). Successive optimization tomlinson-harashima precoding *sothp* for multi-user mimo systems, *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing 2005*, Philadelphia, USA, pp. 1117–1120.

Tse, D. & Viswanath, P. (2005). *Fundamentals of wireless communication*, Cambridge University Press, Cambridge, UK.

Vishwanath, S., Jindal, N. & Goldsmith, A. (2003). Duality, achievable rates and sum capacity of gaussian mimo broadcast channels, *IEEE Trans. Inform. Theory* Vol.49(No.10): 2658–2668.

Viswanath, P. & Tse, D. (2003). Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality, *IEEE Trans. Inform. Theory* Vol.49(No.8): 1912–1921.

Wang, D., Jorswieck, E., Sezgin, A. & Costa, E. (2006). Joint tomlinson-harashima precoding with diversity techniques for multiuser mimo system, *Proc. IEEE VTC 2005Spring*, Stockholm, Sweden, pp. 1017–1021.

Weingarten, H., Steinberg, Y. & Shamai, S. (2006). The capacity region of the gaussian multiple-input multiple-output broadcast hannel, *IEEE Trans. Inform. Theory* Vol.52(No.9): 3936–3964.

Windpassinger, C., Fischer, R., Vencel, T. & Huber, J. (2004). Precoding in multiantenna and multiuser communications, *IEEE Trans. Wireless Commun.* Vol.3(No.4): 1305–1316.

Yu, W., Varodayan, D. & Cioffi, J. (2005). Trellis and convolutional precoding for transmitter-based interference presubtraction, *IEEE Trans. Commun.* Vol.53(No.7): 1220–1230.

Zhang, J., Wu, Y., Zhou, S. & Wang, J. (2005). Joint linear transmitter and receiver design for the downlink of multiuser mimo systems, *IEEE Commun. Letters* Vol.9(No.11): 991–993.

MIMO Systems, Theory and Applications

Edited by Dr. Hossein Khaleghi Bizaki

ISBN 978-953-307-245-6
Hard cover, 488 pages
Publisher InTech
Published online 04, April, 2011
Published in print edition April, 2011

In recent years, it was realized that the MIMO communication systems seems to be inevitable in accelerated evolution of high data rates applications due to their potential to dramatically increase the spectral efficiency and simultaneously sending individual information to the corresponding users in wireless systems. This book, intends to provide highlights of the current research topics in the field of MIMO system, to offer a snapshot of the recent advances and major issues faced today by the researchers in the MIMO related areas. The book is written by specialists working in universities and research centers all over the world to cover the fundamental principles and main advanced topics on high data rates wireless communications systems over MIMO channels. Moreover, the book has the advantage of providing a collection of applications that are completely independent and self-contained; thus, the interested reader can choose any chapter and skip to another without losing continuity.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821