9

MIMO-THP System with Imperfect CSI

H. Khaleghi Bizaki
Electrical and Electronic Engineering University Complex (EEEUC), Tehran, Iran

1. Introduction

In recent years, it was realized that designing wireless digital communication systems to more efficiently exploit the spatial domain of the transmission medium, allows for a significant increase of spectral efficiency. These systems, in general case, are known as Multiple Input Multiple Output (MIMO) systems and have received considerable attention of researchers and commercial companies due to their potential to dramatically increase the spectral efficiency and simultaneously sending individual information to the corresponding users in wireless systems.

In MIMO channels, the information theoretical results show that the desired throughput can be achieved by using the so called Dirty Paper Coding (DPC) method which employs at the transmitter side. However, due to the computational complexity, this method is not practically used until yet. Tomlinson Harashima Precoding (THP) is a suboptimal method which can achieve the near sum-rate of such channels with much simpler complexity as compared to the optimum DPC approach. In spite of THP's good performance, it is very sensitive to erroneous Channel State Information (CSI). When the CSI at the transmitter is imperfect, the system suffers from performance degradation.

In current chapter, the design of THP in an imperfect CSI scenario is considered for a MIMO-BC (BroadCast) system. At first, the maximum achievable rate of MIMO-THP system in an imperfect CSI is computed by means of information theory concepts. Moreover, a lower bound for capacity loss and optimum as well as suboptimum solutions for power allocation is derived. This bound can be useful in practical system design in an imperfect CSI case.

In order to increase the THP performance in an imperfect CSI, a robust optimization technique is developed for THP based on Minimum Mean Square Error (MMSE) criterion. This robust optimization has more performance than the conventional optimization method. Then, the above optimization is developed for time varying channels and based on this knowledge we design a robust precoder for fast time varying channels. The designed precoder has good performance over correlated MIMO channels in which, the volume of its feed back can be reduced significantly.

Traditionally, channel estimation and pre-equalization are optimized separately and independently. In this chapter, a new robust solution is derived for MIMO THP system, which optimizes jointly the channel estimation and THP filters. The proposed method provides significant improvement with respect to conventional optimization with less increase in complexity.

Notation: Random variables, vectors, and matrices are denoted by lower, lower bold, and upper bold italic letters, respectively. The operators E(.), diag(.), $\perp$, PDF, and CDF stand for expectation, diagonal elements of a vector, statistically independent, Probability Density Function, and Cumulative Distribution Function, respectively.

2. MIMO-BC-THP systems

2.1 Type of MIMO channels

There are three types system can be modeled as MIMO channel [1]:

a. *point-to-point MIMO channel*

This type of MIMO system is a multiple antenna scenario, where both transmitter (TX) and receiver (RX) use several antennas with seperate modulation and demodulation for each antenna. We refer this type of channel as MIMO channel (Central transmitter and receiver).

b. *multipoint-to-point MIMO Channel*

The uplink direction of any multiuser mobile communication system is an example of a MIMO system of this type. The joint receiver at the base station has to recover the individual users' signals. We will refer to this type of channel as the MIMO multiple access channel (Decentralized transmitters and central receiver).

c. *point-to-multipoint MIMO Channel*

The downlink direction of mobile multiuser communication systems is an example of what we call a MIMO broadcast channel (Central transmitter and decentralized receivers).

2.2 Precoding strategy

The main difficulty for transmission over MIMO channels is the separation or equalization of the parallel data streams, i.e., the recovery of the components of the transmitted vector $\boldsymbol{x}$ which interfere at the receiver side. The most obvious strategy for separating the data streams is linear equalization at the receiver side.

It is well-known that linear equalization suffers from noise enhancement and hence has poor power efficiency [2]. This disadvantage can be overcome by spatial decision-feedback equalization (DFE). Unfortunately, in DFE error propagation may occur. Moreover, since immediate decisions are required, the application of channel coding requires some clever interleaving which in turn introduces significant delay [2].

The above methods require CS) only at the receiver side. If CSI is (partly) also available at the transmitter, the users can be separated by means of precoding. Precoding, in general case, stands for all methods applied at the transmitter that facilitate detection at the receiver. If a linear transmitter preprocessing strategy is used, we prefer to denote it as preequalization or linear precoder. In other case we refer it as non-linear precoder.

In MIMO channels a version of DFE by name, matrix DFE is used where is a non-linear spatial equalization strategy at the receiver side. The feedback part of the DFE can be transferred to the transmitter, leading to a scheme known as THP. It is well known that neglecting a very small increase in average transmit power, the performance of DFE and THP is the same, but since THP is a transmitter technique, error propagation at the receiver is avoided [3]. Moreover, channel coding schemes can be applied in the same way as for the ideal additive white Gaussian noise (AWGN) or flat fading channel.

The analogies between temporal equalization methods (in Single Input Single Output (SISO) channels) and their direct counterparts as spatial equalization methods (in MIMO channels) are depicted in Table I [2].

		ISI channel $H(z)$ (temporal Equalization)	MIMO channel $\boldsymbol{H}$ (spatial Equalization)
linear	at Rx	Linear equalization via $1/H(z)$	Linear equalization via $\boldsymbol{H}_t^{-1}$
	at Tx	Linear pre-equalization via $1/H(z)$	Linear pre-equalization via $\boldsymbol{H}_r^{-1}$
	at Tx / Rx	OFDM/DMT, vector precoding	SVD
Non-linear	at Rx	DFE	Matrix DFE
	at Tx / Rx	THP	MIMO-THP

Table 1. Corresponding Equalization Strategies for ISI Channels and MIMO Channels.

2.3 The Principle of THP

The information theory idea behind the THP is based on Costa's "writing on dirty paper result" for interference channels [4], which can be informally summarized as follows:
"*When transmitting over a channel, any interference which is known apriori to the transmitter does not affect the channel capacity. That is, by appropriate coding, transmission at a rate equal to the capacity of the channel without this interference is possible.*"
If we extend the Costa precoding concepts for multiple antenna with Co-Antenna Interference (CAI) then THP structure can be obtained [1, 3]. Consider these subchannels in some arbitrary order. In this case, the encoding for the first subchannel has to be performed accepting full interference from the remaining channels, since at this point the interference is unknown. For the second subchannel, however, if the transmitter is able to calculate the interference from the first subchannel, "Costa precoding" of the data is possible such that the interference from the first subchannel is taken into account. Generally, in the k^{th} subchannel considered, Costa precoding is possible such that interference from subchannels 1 to k-1 is ineffective.
We can apply this result to the MIMO channel [5]: If the precoding operation contains a Costa precoder, no interference can be observed from lower number subchannels into higher number subchannels.
Note that it is possible to transform $\boldsymbol{H}$ into a lower triangular matrix with an orthonormal operation [6]. In this way interference from lower-index subchannels into higher-index subchannels is completely eliminated, and together with Costa precoding adjusted to this modified transmission matrix, effectively only a diagonal matrix remains for the transmission. It turns out that a simple scheme for Costa precoding works analog to the feedbackpart of DFE, now moved to the transmitter side and with the nonlinear decision device replaced by a modulo-operation. This is also known as THP [7, 8], and the link between THP and Costa precoding was first explored in [9].

2.4 MIMO-THP system model

The base station with n_T transmit antenna and n_R user (in which $n_T \leq n_R$) with single antenna can be considered as MIMO broadcast system. A block diagram of this MIMO system together with THP is illustrated in Fig. 1 and is briefly explained here.
The n_T dimensional input symbol vector $\boldsymbol{a}$ passes through feedback filter $\boldsymbol{B}$, which is added to the intended transmit vector to pre-eliminate the interference from previous users.

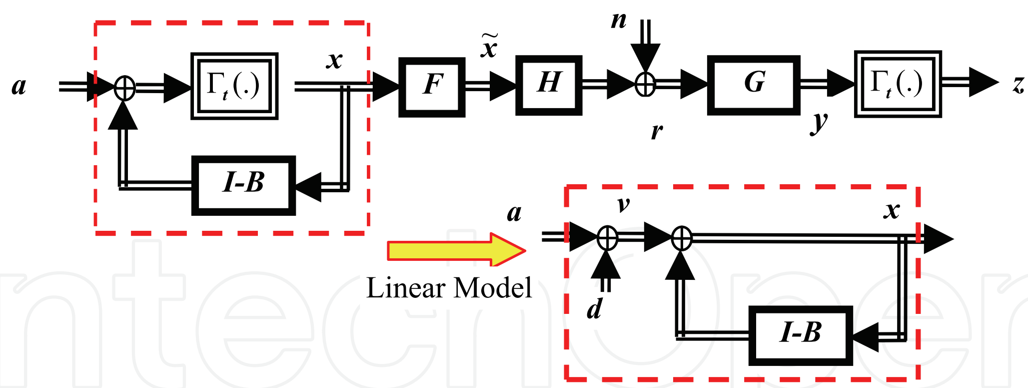


Fig. 1. THP model in a MIMO system

Then the resultant signal is fed to modulo-operator, which serve to limit the transmit power. The output signal of modulo-operator is then passed through a feed forward filter to further remove the interference from future users [10]. Finally, the precoded signal is launched in to the MIMO channel. As all interferences are taken care of at the transmitter side, the receivers at the mobile user side are left with some simple operations including power scaling (diagonal elements of matrix $\boldsymbol{G}$), reverse modulo-operation, and single user detection.
According to Fig. 1, the base band received signal can be modeled as:

$$\boldsymbol{r} = \boldsymbol{H}\tilde{\boldsymbol{x}} + \boldsymbol{n} \tag{1}$$

where $\tilde{\boldsymbol{x}} \in \boldsymbol{C}^{n_T \times 1}$, $\boldsymbol{r} \in \boldsymbol{C}^{n_R \times 1}$, $\boldsymbol{H} \in \boldsymbol{C}^{n_R \times n_T}$ and $\boldsymbol{n} \in \boldsymbol{C}^{n_R \times 1}$ are transmitted, received, channel and noise matrices, respectively (C denotes complex domain). The elements of the noise vector are assumed as independent complex Gaussian random variables with zero mean and variance σ^2, i.e., $\boldsymbol{n} \sim CN(0, \sigma^2 \mathbf{I}_{n_R})$. The elements of matrix $\boldsymbol{H}$ are considered as complex Gaussian random variables (i.e. flat fading case). In other words, the channel tap gain from transmit antenna i to receive antenna j is denoted by h_{ji} which is assumed to be independent zero mean complex Gaussian random variables of equal variance, that is $E[|h_{ji}|^2] = 1$.
The operation of THP is related to the employed signal constellation $\boldsymbol{A}$. Assume that in each of the parallel data streams an M-ary square constellation (M is a squared number) is employed where the coordinates of the signal points are odd integers, i.e., $\boldsymbol{A} = \{a_I + ja_Q \,|\, a_I, a_Q \in \{\pm 1, \pm 3, \ldots, \pm(\sqrt{M} - 1)\}\}$. Then the constellation is bound by the square region of side length $t = 2\sqrt{M}$ which is needed for modular operation [3].
Note: In the rest of the chapter, for means of simplicity, the number of transmit and receive antennas are assumed to be the same (i.e., $n_T = n_R = K$). Also, we consider the flat fading case. Whenever these assumptions are not acceptable we clarify them.
The lower triangular feedback matrix $\boldsymbol{B}$, unitary feed forward matrix $\boldsymbol{F}$ and diagonal scaling matrix $\boldsymbol{G}$ can be found by ZF or MMSE criteria as [11]. The received signal before modulo reduction can be given as:

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{r} = \boldsymbol{G}\boldsymbol{H}\boldsymbol{F}\boldsymbol{B}^{-1}\boldsymbol{v} + \tilde{\boldsymbol{n}} \tag{2}$$

where $\tilde{\boldsymbol{n}} = \boldsymbol{G}\boldsymbol{n}$,and $\boldsymbol{v} = \boldsymbol{a} + \boldsymbol{d}$ is effective input data, and $\boldsymbol{d}$ is the precoding vector used to constrain the value of $\tilde{\boldsymbol{x}}$ [13]. If ZF criterion is used, it requires $\boldsymbol{G}\boldsymbol{H}\boldsymbol{F}\boldsymbol{B}^{-1} = \boldsymbol{I}$. Thus, the processing matrices G, B, and F can be found by performing Cholesky factorization of HH^H as [13]:

$$\begin{aligned} HH^H &= RR^H \\ G &= diag(r_{11}^{-1},...,r_{KK}^{-1}) \\ B &= GR \\ F &= H^{-1}R \end{aligned} \tag{3}$$

where $R=[r_{ij}]$ is a lower triangular matrix. The error covariance matrix can be shown as:

$$\Phi_{\tilde{n}\tilde{n}} = \mathrm{E}[(Gn)(Gn)^H] = \mathrm{diag}[\sigma_n^2 / r_{11}^2,...,\sigma_n^2 / r_{KK}^2] \tag{4}$$

i.e, the noise is white.
If MMSE criterion is used the matrix R can be found through Cholesky factorization of [5]:

$$(H^H H + \zeta I) = R^H R \tag{5}$$

where $\zeta = \sigma_n^2 / \sigma_a^2$. The matrices G, B and F can be found as:

$$\begin{aligned} G &= \mathrm{diag}[r_{11}^{-1},...,r_{KK}^{-1}] \\ B &= GR \\ F &= R^{-H} H^H \end{aligned} \tag{6}$$

The error covariance can be shown as:

$$\Phi_{ee} = \mathrm{E}[ee^H] = \sigma_n^2 G^2 = diag[\sigma_n^2 / r_{11}^2,...,\sigma_n^2 / r_{KK}^2] \tag{7}$$

i.e. error can be considered as white.
In outdated CSI case, the system model, which is considered in Fig. 1, operates in a feedback channel where the CSI is measured in downlink and fed to the transmitter in uplink channel. Time variations of channel lead to a significant outdated (partial) CSI at the transmitter. In fact there will always be a delay between the moment a channel realization is observed and the moment it is actually used by the transmitter. The effect of time variations (or delay) can be considered as: $H = \hat{H} + \Delta H$, where $H, \hat{H}$ and ΔH are true, estimated and channel error due to time variations [13]. We assume that the channel error has Gaussian probability density function with moments $E[\Delta H] = 0$ and $E[\Delta H . \Delta H^H] = C_{\Delta H}$.
According to Fig. 1, the received signal can be considered as:

$$y = Gr = G(\hat{H} + \Delta H) F B^{-1} v + \tilde{n} \tag{8}$$

where $\tilde{n} = Gn$ and v is effective data vector [12]. If ZF criterion is used, it requires:

$$G\hat{H}FB^{-1} = I \tag{9}$$

The processing matrices R, G, B and F can be found by doing Cholesky factorization of $\hat{H}\hat{H}^H$ as [11]:

$$\begin{aligned}
&\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H = \boldsymbol{R}\boldsymbol{R}^H \\
&\boldsymbol{G} = diag(r_{11}^{-1},\dots,r_{KK}^{-1}) \\
&\boldsymbol{B} = \boldsymbol{G}\boldsymbol{R} \\
&\boldsymbol{F} = \hat{\boldsymbol{H}}^{-1}\boldsymbol{R}
\end{aligned} \tag{10}$$

where $\boldsymbol{e} = \boldsymbol{w} + \widetilde{\boldsymbol{n}} = \boldsymbol{G}\Delta\boldsymbol{H}\boldsymbol{F}\boldsymbol{x} + \boldsymbol{G}\boldsymbol{n}$ is considered as an error vector and the term $\boldsymbol{w}$ stands for channel imperfection effect due to outdated CSI. The error covariance matrix can be obtained as:

$$\Phi_{ee} = \boldsymbol{E}[\boldsymbol{e}\boldsymbol{e}^H] = \sigma_x^2\boldsymbol{G}(\boldsymbol{C}_{\Delta\boldsymbol{H}} + \zeta\boldsymbol{I})\boldsymbol{G}^H \tag{11}$$

Note that, with a small channel error assumption (i.e. $\boldsymbol{C}_{\Delta\boldsymbol{H}} \to 0$), the error covariance matrix in an imperfect case tends to the error covariance matrix in a perfect case, i.e.

$$\Phi_{ee} = \sigma_x^2 Diag(1/r_{11}^2, 1/r_{22}^2, \dots, 1/r_{KK}^2) \tag{12}$$

3. MIMO-THP capacity

The first attempt to calculation of achievable rates of THP is done by Wesel and Cioffi in [15]. The authors considered THP for discrete-time SISO consists of Inter-Symbol Interference (ISI) and AWGN. They derived an exact expression for maximum achievable information rate for ZF case and provided information bound for MMSE case. In this section, we develop the achievable rates analysis provided in [15] for MIMO-THP in flat fading channel. We obtain the maximum achievable rate and some upper and lower bounds of it for ZF and MMSE cases with perfect and imperfect CSI.

3.1 Achievable rates of point-to-point MIMO-THP

Consider a point-to-point MIMO system with THP as Fig. 2.

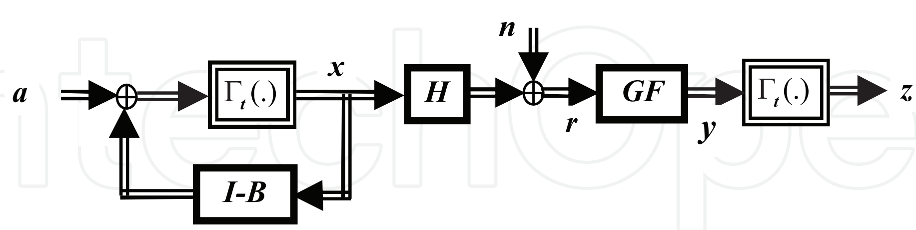


Fig. 2. THP model in a point-to-point MIMO system

The received signal vector can be expressed as:

$$\boldsymbol{z} = \Gamma_t[\boldsymbol{G}\boldsymbol{F}\boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{v} + \boldsymbol{G}\boldsymbol{F}\boldsymbol{n}] = \Gamma_t\left[\boldsymbol{a} + \boldsymbol{w} + \boldsymbol{n}'\right] \tag{13}$$

where $\boldsymbol{w}$ is residual spatial interference after MMSE criterion on THP filters (in the ZF case $\boldsymbol{w} = 0$) and $\Gamma_t(.)$ is modulo $\boldsymbol{t}$ operator so eliminate its output on interval $T = [-t_k/2, t_k/2) \times [-jt_k/2, jt_k/2)$. As $\boldsymbol{n}' = \boldsymbol{G}\boldsymbol{F}\boldsymbol{n}$ is white Gaussian noise and with the assumption that $w_i \perp w_j$ & $a_i \perp w_j$ for $\forall i \neq j$ (so that symbol $\perp$ stand for statistical independence) the received vector $\boldsymbol{r}$ can be decoupled in K parallel streams as[1] [17]:

$$z_k = \Gamma_t \left[a_k + w_k + n'_k\right] \;\; ; \;\; k = 1 \sim K \tag{14}$$

Because of the decoupling of the received information symbols in (14) and assuming independence between elements in $\boldsymbol{a}$ the mutual information between the transmitted symbols and the received signal vector can be expressed as the sum of the mutual information between elements of each vector:

$$I(\boldsymbol{a};\boldsymbol{z}) = \sum_{k=1}^{K} I(a_k ; z_k) \tag{15}$$

where $I(.)$ denote mutual information. Each term in the sum is independently can be considered as:

$$\begin{aligned} I(a_k ; z_k) &= h(z_k) - h(z_k \mid a_k) = h[\Gamma_t(a_k + w_k + n'_k)] - h[\Gamma_t\,((a_k + w_k + n'_k) \mid a_k)] \\ &= h[\Gamma_t(a_k + w_k + n'_k)] - h\left[\Gamma_t\,((w_k + n'_k) \mid a_k)\right] \end{aligned} \tag{16}$$

where $h(.)$ denotes differential entropy. Calculation of the above mutual information seems to be difficult and we try to find an upper and lower bound of (16) by some approximations.
Remark 1: An upper bound on the achievable rate of the channel produced by MMSE-THP of (16) can be found as [17]:

$$I_{Upper}(\boldsymbol{a};\boldsymbol{z}) = \sum_{k=1}^{K}\left(2\log_2(t) - h\left[\Gamma_t\,(n'_k)\right]\right) \tag{17}$$

Also, the upper bound can be obtained essentially by neglecting the spatial interference term w_k in (16) [17]. The lower bound depends largely on the variance of w_k [15]. A lower bound on achievable rate can be found as [17]:

$$I_{Lower}(\boldsymbol{a};\boldsymbol{z}) = \sum_{k=1}^{K} 2\log_2(t) - \log_2(2\pi\sigma^2) - \frac{\gamma^2}{\sigma^2}\log_2 e - 2\log_2(2erf(t/2\sigma)) \tag{18}$$

Thus, a truncated Gaussian [17] with variance of $\gamma^2 = \mathrm{var}(w_k + \Gamma_t(n'_k))$ produces a slightly tighter bound but requires the computation of $\mathrm{var}(\Gamma_t(n'_k))$.
Remark 2: The upper bound attained in (17) can be simplified if some approximations are allowed so that a quasi-optimal (or sub-optimal) closed form solution can be found. This approximations can be done based on the value of t/σ (See [17]).

3.2 General THP in point-to-point MIMO with perfect CSI

Whenever CSI is available at the transmitter in a communication system, since the transmitter has knowledge of the way the transmitted symbols are attenuated and distributed by the channel, it may adjust transmit rate and/or power in an optimized way.

[1] For MMSE case the above assumption for high value of SNR is acceptable and the above results can be true in asymptotic case, so MMSE performance for high SNR values converge to ZF [2].

For instance, in the multi-antenna scenario some of the equivalent parallel channels might have very bad transmission properties or might not be present at all. In this situation, the transmitter might want to adjust to that by either dropping some of the lower diversity order sub channels or by redistributing the data and the available transmission power to improve the average error rate. This can be done by generalization of THP concepts as GTHP by enabling different power transmission for each antenna. GTHP can be done in two main scenarios [17]:
First: If the loading is made according to capacity of system; this structure enables different transmission rate per antenna.
Second: If it is needed to ensure reliable transmission rate for each antenna, the loading should be made according to minimize error rate of system.
Here we consider two different optimization scenarios for loading strategies of THP and extend it's concept in structure that t is not constant, so the modulo interval is different for each sub channel (t_k) [17].

3.2.1 Capacity criterion

In this section, the power adaptation strategy of the second type of GTHP concept is employed. The optimal power allocation is calculated in MIMO-GTHP systems, while regarding the modulation schemes is given. If the loading is made according to capacity of system, this structure enables different transmission rate per antenna. One of the good features of this scenario is that it is scalable architecture, because it allows adding or removing transmitters without losing the precoding structure as explained in [16].
If a assumed as an i.i.d. uniform distribution on T, for such a case, x is also i.i.d. uniform on T (regardless of the choice of matrix B). Thus the transmitted power from k^{th} antenna will be $p_k = E\{|x_k|^2\} = t_k^2/6$ (for real case $p_k = t_k^2/12$). Its corresponding rate will equal the maximum achievable mutual information in (17):

$$I(t_k) = I(a_k; z_k, t_k) = 2\log_2(t_k) - h[\Gamma_{t_k}(n'_k)] \tag{19}$$

Then the maximum achievable rate for a system with THP will be the maximum of the sum of the rates of each stream subject to a maximum total transmitted power constraint, i.e.

$$\begin{cases} C_{General} = \max\limits_{\{t_k\}} \sum\limits_{k=1}^{K} I(t_k) = \max\limits_{\{t_k\}} \sum\limits_{k=1}^{K} (2\log_2(t_k) - h[\Gamma_{t_k}(n'_k)]) \\ s.t. \quad \sum\limits_{k=1}^{K} p_k = \frac{1}{6}\sum\limits_{k=1}^{K} t_k^2 = P_T \end{cases} \tag{20}$$

In order to maximize (20) we assume that all the available streams are classified into two groups[2] (g-Gaussian and u–uniform) based on t/σ values [16]. As shown in [17], the achievable rate of streams belonging to u tends to zero; no power is assigned to these streams, i.e. $t_k = 0 \; \forall k \in u$. Thus the solution of the maximization problem in (20) can be found as assigning the same power to the entire stream in g (and no power to those in u).
The optimal solution can be shown to be [17]:

[2] Based on the value of t/σ for each stream

$$C_{General}^{Adaptive} = \sum_{k \in g} I_k \left(\sqrt{\frac{6 p_T}{|g| \sigma_{n'_k}^2}} \right) \quad , \quad k \in g \tag{21}$$

where $|g|$ denotes the number of active antennas and $\sigma_{n'_k}^2$ is variance of n'_k. Then some kind of adaptive rate algorithm is necessary to achieve the maximum capacity of the GTHP.

3.2.2 Minimum SER criterion

In some application it is needed to ensure reliable transmission rate for each antenna (especially in MIMO broadcast channels). In this section we try to find the optimal sub channel power allocation in MIMO GTHP systems, while regarding the modulation schemes is given. As mentioned, for each sub channel we have:

$$z_k = a_k + n'_k \qquad k = 1,2,...,K \tag{22}$$

where we assumed that $\boldsymbol{w_k}$ tend to zero. For simplicity assume MQAM transmission in all sub channels is used. In this case the approximate average SER for a fixed channel $\boldsymbol{H}$ simply given as [17]:

$$SER \approx \sum_{k=1}^{K} (1 - \frac{1}{\sqrt{M_k}}) Q(\sqrt{\frac{3}{M_k - 1} \frac{|r_{kk}|^2}{\sigma_{n'}^2} E_k}) \tag{23}$$

where $B_k = \frac{3}{M_k - 1} \frac{|r_{kk}|^2}{\sqrt{M_k}}$ and we assumed modulation order (i.e. M_k) can be varied for each sub-channel, so that variable bit allocation is possible (that we didn't consider here). In this case we have [17]:

$$E_k = \frac{1}{B_k} W(\frac{B_k}{A_k} \Lambda) \tag{24}$$

where $A_k = \frac{M_k (M_k - 1)}{(\sqrt{M_k} - 1)^2 |r_{kk}|^2}$ and $W(x)$ is the real valued Lambert's W -function defined as the inverse of the function $f(x) = x.e^x$; $x \geq 0$, i.e., $W(x) = a \Leftrightarrow a.e^a = x$.

Since the $W(x)$ function is real and monotonically increasing for real $x > -1/e$, the value of λ such that $\sum_{k=1}^{K} E_k(\lambda) - K = 0$ holds which can be found by using some classical methods as denoted in [17]. On the other hand, $W(x)$ is a concave and unbounded function with $W(0) = 0$ and $W(x) \leq x$, the unique solution for $\boldsymbol{E} = [E_1, ..., E_K]^T$ can be found by the following simple iterative procedure[14]:

i. Chose a small positive Λ which satisfy

$$\sum_{k=1}^{K} \frac{\Lambda}{A_k} \leq E_T \tag{25}$$

ii. Calculate

$$\hat{E}_T = \sum_{k=1}^{K} \frac{1}{B_k} W(\frac{B_k}{A_k} \Lambda) \tag{26}$$

iii. If $\hat{E}_T$ is not yet sufficiently close to E_T, multiply Λ by $E_T/\hat{E}_T$ and go back to step (ii).

iv. Compute $\boldsymbol{E} = [E_1,...,E_K]^T$ according to (24).

Note that since $W(x)$ for $x > -1/e$ is monotonic function, then according to relation (24) the highest power $(\max E_k)$ assign to the weakest signal so that the SNR value almost stay constant for all sub channels.

3.3 Achievable rate in imperfect CSI

In [17] the scheme proposed in [18] for MIMO THP system was modified by allowing variations of the transmitted power in each antenna. The authors stated the problem of finding the maximum achievable rate for this modified spatial THP scheme and found that Uniform Power Allocation (UPA) with antenna selection is a quasi-optimal transmission scheme with a perfect CSI.

In this sub-section, based on previous researches about SISO and point-to-point MIMO channels, an analytical approach to attain the maximum achievable rate bound in an imperfect CSI case is developed for broadcast channel. It will be shown that this bound depends on the variance of the residual Co-Antenna Interference (CAI) term. Moreover, it will be shown that the power allocation obtained by the UPA in [17] is sub-optimal in an imperfect CSI, too.

3.3.1 Maximum achievable rates

The received signal after modulo operation can be considered as $z = a + \Gamma_{t_k}[G\Delta HFx + \tilde{n}]$. Since x has i.i.d. distribution, $W = G\Delta HFx$ can be considered as an unknown interference with an i.i.d. distribution. Also, for such an a, z is i.i.d. uniform on T. In this case, the received information can be decoupled in K independent parallel data steams and the mutual information between k^{th} element of data vector, a_k, and the corresponding element of the received signal, z_k, is [13]:

$$I(a_k; z_k) = h(z_k) - h(z_k \mid a_k) = h(z_k) - h[\Gamma_{t_k}((w_k + \tilde{n}_k) \mid a_k)] \le \log_2(6p_k) - h\left[\Gamma_{t_k}\left(\sum_{j=1}^{K} \frac{\delta'_{kj}}{r_{kk}} x_j + \tilde{n}_k\right)\right] \tag{27}$$

where $\delta'_{kj} = [\Delta \boldsymbol{HF}]_{kj}$ and $h(.)$ denotes differential entropy. Let us define the random variable e_k as $e_k = \sum_{j=1}^{K} \frac{\delta'_{kj}}{r_{kk}} x_j + \tilde{n}_k$ where its power is $\sigma_{ek}^2 = \frac{\sigma_n^2}{r_{kk}^2} + \sum_{j=1}^{K} \frac{p_j}{r_{kk}^2} \left|\delta_{kj}\right|^2$ and $\delta_{kj} = [\Delta \boldsymbol{H}]_{kj}$. With the assumption of small error, e_k can be approximately modeled as a complex Gaussian random variable. In the case where, the above assumption is true, the mutual information expression (27) can be very well approximated as [13]:

$$I(a_k; z_k) \approx \log_2 [\chi_k]^+ \tag{28}$$

where

$$\begin{cases} \chi_k = 6r_{kk}^2 p_k / \pi e(\sigma_n^2 + \sum_{j=1}^{K} \left|\delta_{kj}\right|^2 p_j) \\ \log[x]^+ = \max[\log(x), 0] \end{cases} \tag{29}$$

The achievable rates for THP in an imperfect CSI case will then be the sum of the mutual information of all K parallel steams as [13]:

$$\begin{cases} C = \max\limits_{\{p_k\}} \sum_{k=1}^{K} I(a_k; z_k) \approx \max\limits_{\{p_k\}} \sum_{k=1}^{K} \log_2 \left[\chi_k\right]^+ \\ s.t. \quad \sum_{k=1}^{K} p_k = P_T \end{cases} \tag{30}$$

Observed that C (or χ_k) depends on three components: $\left|\delta_{kj}\right|^2, r_{kk}^2$ and p_k. In order to maximize χ_k, some kind of spatial ordering is necessary in order to maximize it. For this purpose, it is required to decompose $\boldsymbol{H}$ (in Cholesky factorization) so that the elements of r_{kk}^2 to be maximized (finding the ordering matrix similar to [11]).

On the other hand, it was assumed that by making small error assumption, e_k can be approximately modeled as a complex Gaussian random variable. This is equivalent to assuming $\max\limits_{j} |\delta_{kj}|^2 p_j \le \sigma_n^2 ; \forall j$. Now, we assume that the entries of error matrix are bounded as $\max\limits_{k,j} |\delta_{kj}|^2 \le \alpha_{kj} ; \forall k, j$ [13]. In addition, for the sake of simplicity and without loss of generality, we assume that $\alpha = \max\limits_{k,j} \alpha_{kj}$. Then, the power distribution that will maximize the achievable rates will be the solution of the following maximin problem:

$$\begin{cases} C = \max\limits_{p_i} \min\limits_{\delta_{ij}} \sum_{k=1}^{K} \log_2 \left[\chi_k\right]^+ \\ s.t. \quad \sum_{k=1}^{K} p_k = p_T \ , \ \max\limits_{i,j} \left|\delta_{ij}\right|^2 \le \alpha \quad ; \ \forall i, j \end{cases} \tag{31}$$

In order to solve the above maximin problem the worst-case is assumed, i.e. $\alpha = \max\limits_{i,j} |\delta_{ij}|^2 ; \forall i, j$. With this assumption, the minimum mutual information will be attained for each term in the summation. Then, the resulting maximization problem leads to [13]:

$$\begin{cases} C = \max\limits_{p_i} \sum_{k=1}^{K} \log_2 \left[\dfrac{6 p_k r_{kk}^2}{\pi e(\sigma_n^2 + \alpha\, p_T)} \right]^+ \\ s.t. \quad \sum_{k=1}^{K} p_k = p_T \end{cases} \tag{32}$$

The resulting maximization problem is a standard constrained optimization problem, and can be solved with the use of the Lagrange method in which the solution result is $p_k = const.$ It means that the p_k is independent of k, i.e the distribution of the power, in worst-case, is UPA.

Note that, if we consider different noises with different powers for each user, the distribution of power may not be the UPA.

3.3.2 Capacity loss

In the previous section, it is shown that the capacity of MIMO-THP can be obtained by the UPA. More over, it can be observed from (32) that this capacity, in worst case, depends on the channel error value (i.e. α). We define the capacity loss as difference between the capacity of MIMO-BC-THP in a perfect CSI and in an imperfect CSI, i.e. relation (32), as [19]:

$$\Delta\hat{C} = C_1 - \hat{C}_2 = \sum_{k=1}^{K}\log_2\left[\frac{6p_k r_{kk}^2}{\pi e\sigma_n^2}\right] - \sum_{k=1}^{K}\log_2\left[\frac{6p_k r_{kk}^2}{\pi e(\sigma_n^2 + \alpha\, p_T)}\right] \le K\log_2[1+K] \quad (33)$$

The above bound for capacity loss only is valid for values of α so that the approximation of $\max_j |\delta_{kj}|^2 p_j \le \sigma_n^2 \,;\forall j$ is valid [19]. It means that this bound depends on SNR value and is acceptable for high SNR value, i.e., this capacity loss bound is asymptotic bound for worst-case in which bounds the capacity loss of MIMO-THP . It is desired to bounding the capacity loss of optimal solution of (29). Assume C_2 is the capacity of optimal solution that can be obtained by exactly analysis or by numerical simulation. In this case we can bound $\Delta C = C_1 - C_2$ as [19]:

$$\Delta C = K\log_2[1+K] - K\log_2\left[\left(1+\frac{p_T}{\sigma_n^2}\alpha\right)\right] + \varepsilon \quad (34)$$

where ε is a positive value. The lower bound can be obtained by choosing $\varepsilon = K-2$ [19]:

$$\Delta C \ge \left[K\log_2[1+K] - K\log_2\left[\left(1+\frac{p_T}{\sigma_n^2}\alpha\right)\right] + (K-2)\right]^+ \quad (35)$$

where $[x]^+ = \max[x,0]$. In simulation we refer (35) as theoretic loss.

3.4 Spatial ordering

The VBLAST-Like ordering can be used in order to improve the power loading performance of MIMO-GTHP system in Fig. 1 [1]. To do this, since the loading is based on the SNR values of the equivalent parallel sub-channels, which in turn are proportional to $|r_{kk}|^2$, the distribution of these diagonal entries is an essential parameter in power loading performance. It turns out that by introducing a permutation matrix in the decomposition of ***H***, i.e, allowing different ordering of the sub channels, the distribution of the $|r_{kk}|^2$ values can be modified as [1]:

$$\boldsymbol{P}_{opt} = \arg\min_{\boldsymbol{P}} \,(1/|r_{11}|^2, 1/|r_{22}|^2, \ldots, 1/|r_{KK}|^2) = \arg\min_{\boldsymbol{P}}(|\boldsymbol{G}|^2)$$

It means that in the cholesky factorization of (4), the decomposition should be made so that the square value of diagonal elements of matrix ***R*** minimized. It means that the matrix ***P*** is selected so that the column of ***H*** corresponding to minimum square value of diagonal elements of ***G*** is permuted to the left. Deleting this column from the matrix ***H***, and forming the cholesky factorization of this modified matrix, we can obtain second column of matrix $\boldsymbol{P}$. Continuing this way, constantly updating $\boldsymbol{P}$, the decomposition of $\boldsymbol{H}$ is constructed. The pseudo-code for the algorithm is given in Fig. (3).

$$
\begin{aligned}
&\textit{Initialization :}\\
&\qquad \boldsymbol{P} = [p_{i,j}] = 0\\
&\textit{for } i = 1 \textit{ to } K \textit{ do :}\\
&\qquad [\boldsymbol{Q}\ \ \boldsymbol{R}] = qr(\boldsymbol{H}) \qquad ; \qquad \textit{for ZF}\\
&\qquad \boldsymbol{R} = chol(\boldsymbol{H}\boldsymbol{H}^H + \zeta\boldsymbol{I}) \quad ; \quad \textit{for MMSE}\\
&\qquad \boldsymbol{G} = diag(inv(\boldsymbol{R}))\text{^}2\\
&\qquad l_i = \arg\min_j (g_{i,j})\\
&\qquad P_{l_i,i} = 1\\
&\qquad \boldsymbol{H}_{:,l_i} = zeros(size(\boldsymbol{H}_{:,l_i}))\\
&\textit{end}
\end{aligned}
$$

Fig. 3. The pseudo code for ordering

Note that for ZF or MMSE-THP, the system performance will be dominated by the signal component with the largest noise variance, and we can find the ordering algorithm in the minimax noise variance sense as [1].

3.5 Simulation and results

3.5.1 Perfect CSI

The mutual information for real x_k with i.i.d uniform distribution on the module interval [-t/2, t/2) is plotted in fig. 4, where the average transmitter energy is $t^2/12$. This figure also shows the mutual information curves for the upper (17) and lower (18) bounds for each sub channel. For comparison we also plotted the well-known AWGN channel capacity (with no ISI). Observe that the upper bound lies above the AWGN capacity and lower bound lies below this capacity (especially for high SNR values).

Figs. 5 and 6 give the performance comparison of the MMSE-GTHP with/without power loading (relation (24)) when 4QAM and 16QAM modulations are used, respectively. From these figures, it is clearly seen that the MMSE-GTHP with ordering can achieve better performance than the MMSE-GTHP with or without power allocation (4QAM or 16QAM,). When 4QAM modulation is used, at BER=2×10^{-4} we can observe that the MMSE-GTHP with power allocation achieves about 7dB gain, while this structure with power loading and ordering gives about 11.5dB gain. When 16QAM is used, at BER=5×10^{-4} the MMSE-GTHP with power allocation gives approximately 6.5dB gain, while this structure with power loading and ordering gives about 10dB gain. As can be seen from these figures, the performance of power loading is noticeable, especially when it combined with sub-channel ordering.

3.5.2 Imperfect CSI

For simulation purposes we have considered K=4 users. The entries of $\hat{\boldsymbol{H}}$ and $\Delta\boldsymbol{H}$ have been assumed to be zero mean i.i.d. complex Gaussian random variables, i.e., $\hat{\boldsymbol{H}} \sim CN(0,1)$

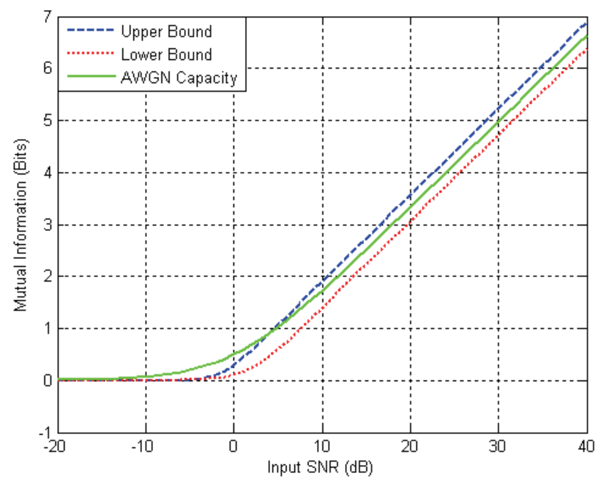


Fig. 4. Upper and lower bound of mutual information

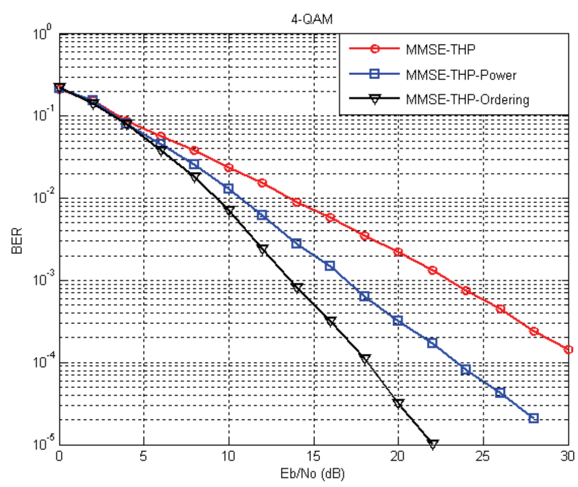


Fig. 5. Performance comparison of MMSE-GTHPwith power loading and ordering for 4QAM.

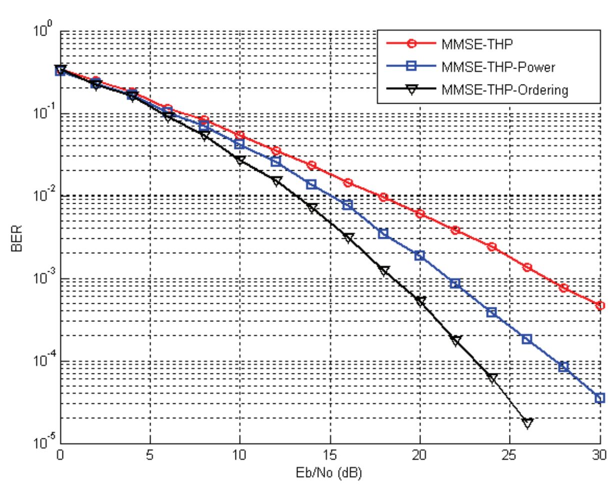


Fig. 6. Performance comparison of MMSE-GTHP. with power loading and ordering for 16QAM

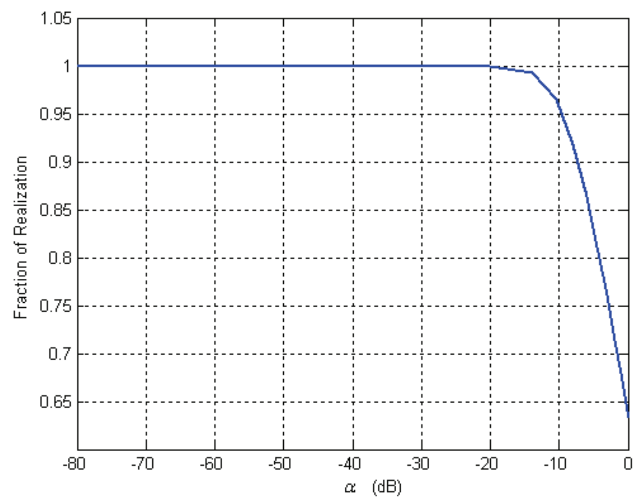


Fig. 7. Validation of approximation of α

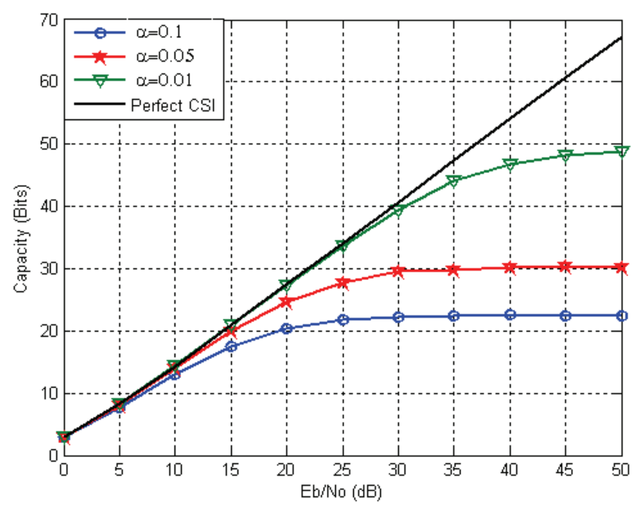


Fig. 8. Capacity with K=4 user and different value of α

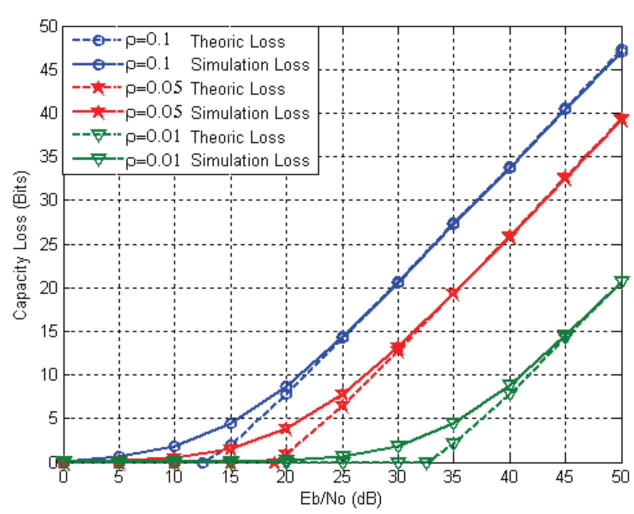


Fig. 9. Capacity loss for K=4 user

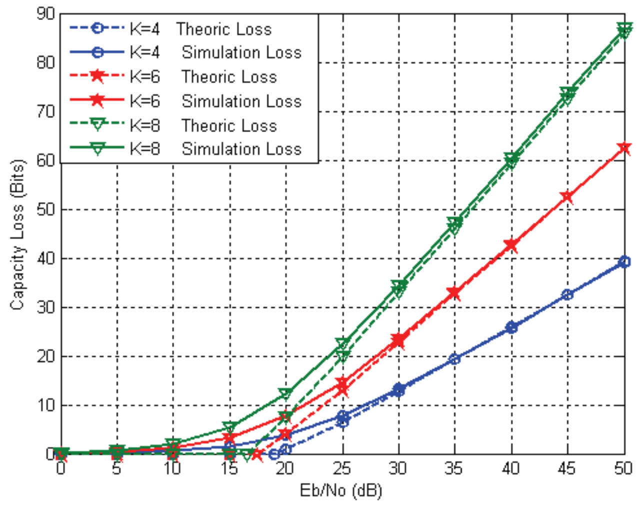


Fig. 10. Capacity loss for different user and $\rho = 0.05$

and $\Delta \boldsymbol{H} \sim CN(0, \rho^2)$. The validity of the approximations of ($\alpha p_j \leq \sigma_n^2 ; \forall j$) for $P_T / \sigma_n^2 = 16\ dB$ is shown in Fig. 7 by plotting the fraction of channel realizations in which the approximations are valid for different values of α. The simulation is done for more than 10^4 channel realization and the validity is calculated as the number of iteration where the inequality is valid. It can be seen that, for the particular values of the simulation parameters taken in this section, the capacity analysis is valid for values of α up to −10 dB. In Fig. 8, we have plotted the mean of the maximum achievable rates for the UPA scheme for different values of α. It can be seen that for values of $\alpha \leq 0.01$ the capacity loss due to the presence of channel errors are negligible and by increasing the SNR value, the capacity increases up to a constant value. It means that by increasing the SNR value the capacity remains almost constant. Figs. 9 and 10 depict asymptotic capacity loss in a time varying channel for different values of ρ and different users, respectively. It can be seen that the attained aproximated bound is valid for variety of rang of the channel errors and number of users, especially for high SNR value (asymptotic case).

4. Robust MIMO-THP

4.1 Design criterion

The error, that is needed to be considered for the system illustrated in Fig. 1, should be the difference between the effective data vector $\boldsymbol{v}$ and the data vector entering the decision module $\boldsymbol{y}$, i.e.:

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{v} = [\boldsymbol{G}(\hat{\boldsymbol{H}} + \Delta \boldsymbol{H})\boldsymbol{F} - \boldsymbol{B}]\boldsymbol{x} + \tilde{\boldsymbol{n}} \tag{36}$$

The MMSE solution should minimize the error signal as:

$$\begin{cases} \underset{\boldsymbol{B},\boldsymbol{F},\boldsymbol{G}}{\arg\min}\ E\left\| [\boldsymbol{G}(\hat{\boldsymbol{H}} + \Delta \boldsymbol{H})\boldsymbol{F} - \boldsymbol{B}]\boldsymbol{x} + \tilde{\boldsymbol{n}} \right\|^2 \\ s.t.\quad E\left\| \tilde{\boldsymbol{x}} \right\|^2 \leq P_T \end{cases} \tag{37}$$

Instead of solving (37), it is easier to use the orthogonallity principle [1]. In this case, the MMSE solution should satisfy:

$$\mathrm{E}[\boldsymbol{e}\boldsymbol{r}^H] = 0 \tag{38}$$

Thus, according to (36) we have:

$$\boldsymbol{G}\Phi_{rr} = \boldsymbol{B}\Phi_{xr} \tag{39}$$

The matrices Φ_{rr} and Φ_{xr} can be computed by using (36) as [13]:

$$\Phi_{rr} = \mathrm{E}[\boldsymbol{r}\boldsymbol{r}^H] = \sigma_x^2(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I} + \boldsymbol{C}_{\Delta H}) \tag{40.a}$$

$$\Phi_{xr} = \mathrm{E}[\boldsymbol{x}\boldsymbol{r}^H] = \sigma_x^2\boldsymbol{F}^H\hat{\boldsymbol{H}}^H \tag{40.b}$$

where $\zeta = \sigma_n^2 / \sigma_x^2$. Substituting relation (40) for (39) and some manipulation, yielding [13]:

$$\boldsymbol{F}^H = \boldsymbol{B}^{-1}\boldsymbol{G}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I} + \boldsymbol{C}_{\Delta H})\hat{\boldsymbol{H}}^{-H} \tag{41}$$

Since $\boldsymbol{F}$ is a unitary matrix [1]:

$$\boldsymbol{R}\boldsymbol{R}^H = (\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I} + \boldsymbol{C}_{\Delta H})\hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I} + \boldsymbol{C}_{\Delta H}) \tag{42}$$

where $\boldsymbol{R} = \boldsymbol{G}^{-1}\boldsymbol{B}$ is assumed. The matrix $\boldsymbol{R}$ can be found through Cholesky factorization of (42) and the matrices $\boldsymbol{G}$, $\boldsymbol{B}$, and $\boldsymbol{F}$ can be obtained as [13]:

$$\begin{aligned} &\boldsymbol{G} = \mathrm{diag}[r_{11}^{-1},\ldots,r_{KK}^{-1}] \\ &\boldsymbol{B} = \boldsymbol{G}\boldsymbol{R} \\ &\boldsymbol{F} = \hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I} + \boldsymbol{C}_{\Delta H})\boldsymbol{R}^{-H} \end{aligned} \tag{43}$$

Using the $\boldsymbol{B}$, $\boldsymbol{F}$ and $\boldsymbol{G}$ obtained in (43), the error covariance matrix can be computed as [13]:

$$\Phi_{ee} = \boldsymbol{E}[\boldsymbol{e}\boldsymbol{e}^H] = \sigma_x^2\boldsymbol{G}(\zeta^2\hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1} + \zeta\boldsymbol{I} + \boldsymbol{C}_{\Delta H})\boldsymbol{G}^H \tag{44}$$

It means that the error covariance matrix in imperfect CSI is the sum of covariance matrix in perfect CSI plus the term $\sigma_x^2\boldsymbol{G}\boldsymbol{C}_{\Delta H}\boldsymbol{G}^H$. This term tends to zero with low channel error assumption (perfect CSI), i.e.:

$$\Phi_{ee} = \boldsymbol{E}[\boldsymbol{e}\boldsymbol{e}^H] = \sigma_n^2\boldsymbol{G}(\zeta\hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1} + \boldsymbol{I})\boldsymbol{G}^H \tag{45}$$

In this case, $\boldsymbol{R}$ can be found through:

$$\boldsymbol{R}\boldsymbol{R}^H = (\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I})\hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I}) \tag{46}$$

and the matrices $\boldsymbol{G}$, $\boldsymbol{B}$ and $\boldsymbol{F}$ can be computed as:

$$\begin{aligned} &\boldsymbol{G} = \mathrm{diag}[r_{11}^{-1},\ldots,r_{KK}^{-1}] \\ &\boldsymbol{B} = \boldsymbol{G}\boldsymbol{R} \\ &\boldsymbol{F}^H = \hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I})\boldsymbol{R}^{-H} \end{aligned} \tag{47}$$

The above results (45-47) are the same as [12], where it is assumed that the perfect CSI is available. In this section, relations 45 to 47 are referred to as conventional optimization and relations (42-44) are referred to as robust optimization.

4.2 Robust optimization with channel estimation consideration

4.2.1 Channel estimation

The received signal at the base station during training period (uplink) , at time stand i, can be modeled as [13]:

$$\boldsymbol{y}(i)=\boldsymbol{H}^{T}\boldsymbol{a}(i)+\boldsymbol{n}(i) \tag{48}$$

During the training period of N symbols in uplink transmission, the received signal can be constructed as [13]:

$$\boldsymbol{y}_s=\boldsymbol{s}\boldsymbol{h}_s+\boldsymbol{n} \tag{49}$$

where $\boldsymbol{y}_s=[\boldsymbol{y}(0),...,\boldsymbol{y}(N-1)]^T$, $\boldsymbol{n}=[\boldsymbol{n}(0),...,\boldsymbol{n}(N-1)]^T$, $\boldsymbol{h}_s=\mathrm{vec}[\boldsymbol{H}^T]$, $\boldsymbol{s}=[\boldsymbol{A}(0),...,\boldsymbol{A}(N-1)]^T$, and $\boldsymbol{A}(i)$ can be constructed as block diagonal matrix with elements of $\boldsymbol{a}(i)^{\mathrm{T}}$. Based on the received signal in (49), the Best Linear Unbiased Estimator (BLUE) channel estimation can be performed as [20]:

$$\hat{\boldsymbol{h}}_s=(\boldsymbol{s}^H\boldsymbol{C}_n^{-1}\boldsymbol{s})^{-1}\boldsymbol{s}^H\boldsymbol{C}_n^{-1}\boldsymbol{y}_s=(\boldsymbol{s}^H\boldsymbol{s})^{-1}\boldsymbol{s}^H\boldsymbol{y}_s=\boldsymbol{W}\boldsymbol{y}_s \tag{50}$$

with the covariance matrix of:

$$\boldsymbol{C}_{\hat{h}_s}=(\boldsymbol{s}^H\boldsymbol{C}_n^{-1}\boldsymbol{s})^{-1}=\sigma_n^2(\boldsymbol{s}^H\boldsymbol{s})^{-1} \tag{51}$$

4.2.2 Improved robust optimization

In the robust optimization, only THP filters were optimized according to the MMSE criterion and the channel estimator was optimized separately from THP. Here, the above solution is extended to optimize THP filters together with the channel estimator conditioned on observed data. In this case, cost function should be optimized with respect to THP filters and the observed data. Thus, the goal is to optimize the precoder directly based on the available observation $\boldsymbol{y}_s$.

Based on the linear model of (49), the conditional PDF $p_{h_s|y_s}(\boldsymbol{h}_s\mid\boldsymbol{y}_s)$ is a complex Gaussian process with moments $\mu_{h|y_s}=\mathrm{E}[\boldsymbol{h}_s\mid\boldsymbol{y}_s]$ and $\boldsymbol{C}_{h|y_s}=\boldsymbol{E}[(\boldsymbol{h}_s-\mu_{h|y_s})(\boldsymbol{h}_s-\mu_{h|y_s})^H\mid\boldsymbol{y}_s]$ [18].

According to the Bayesian Gauss-Markov theorem, the Bayesian estimator can be written as [20]:

$$\mu_{h|y_s}=\boldsymbol{C}_{h_s}\boldsymbol{s}^H(\boldsymbol{s}\boldsymbol{C}_{h_s}\boldsymbol{s}^H+\sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{y}_s=\boldsymbol{W}_s\boldsymbol{y}_s \tag{52}$$

and the covariance matrix of channel estimator is:

$$\boldsymbol{C}_{h|y_s}=\boldsymbol{C}_{h_s}-\boldsymbol{W}_s\boldsymbol{s}\boldsymbol{C}_{h_s} \tag{53}$$

where $\boldsymbol{C}_{h_s}=\boldsymbol{E}[\boldsymbol{h}_s\boldsymbol{h}_s^H]$. In order to optimize the THP filters, the cost function in the previous section should be modified conditional to the observed data, i.e.:

$$\Phi_{ee} = \mathrm{E}_{\mathrm{H}}[\boldsymbol{e}\boldsymbol{e}^{H} \mid \boldsymbol{y}_s] \tag{54}$$

By using the orthogonallity principle, the MMSE solution should be equivalent to:

$$\mathrm{E}_{\mathrm{H}}[\boldsymbol{e}\boldsymbol{r}^{H} \mid \boldsymbol{y}_s] = 0 \tag{55}$$

As relation (39) it is possible to write [13],

$$\boldsymbol{G}\Phi_{rr|y_s} = \boldsymbol{B}\Phi_{xr|y_s} \tag{56}$$

Like the sub-section 4.1, the matrix $\boldsymbol{R}$ can be found through Cholesky factorization of [13]:

$$\boldsymbol{R}\boldsymbol{R}^{H} = (\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{H} + \zeta\boldsymbol{I} + \boldsymbol{C}_{h|y_s})\hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{H} + \zeta\boldsymbol{I} + \boldsymbol{C}_{h|y_s}) \tag{57}$$

and matrices $\boldsymbol{G}$, $\boldsymbol{B}$ and $\boldsymbol{F}$ can be found as:

$$\begin{aligned} &\boldsymbol{G} = \mathrm{diag}[r_{11}^{-1},\ldots,r_{KK}^{-1}] \\ &\boldsymbol{B} = \boldsymbol{G}\boldsymbol{R} \\ &\boldsymbol{F} = \hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{H} + \zeta\boldsymbol{I} + \boldsymbol{C}_{h|y_s})\boldsymbol{R}^{-H} \end{aligned} \tag{58}$$

In this case, the error covariance matrix has the form [13]:

$$\Phi_{ee} = \sigma_x^2\boldsymbol{G}(\zeta^2\hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1} + \zeta\boldsymbol{I} + \boldsymbol{C}_{h|y_s})\boldsymbol{G}^{H} \tag{59}$$

It means that the improved robust optimization can be done by replacing $\boldsymbol{C}_{\Delta H}$ in the robust optimization with its equvalent, i.e. $\boldsymbol{C}_{h|y_s}$.

4.3 Power loading in imperfect CSI

In sub-section 3.4, we discussed about power loading of point-to-point MIMO-THP in perfect CSI. Now we develop this power loading in MIMO-BC-THP for imperfect case.

4.3.1 Optimal solution

It is easy to approximate the SER of each sub-streams for imperfect CSI as [13]:

$$SER \approx \sum_{k=1}^{K}(1-\frac{1}{\sqrt{M_k}})Q(\sqrt{\frac{3}{M_k-1}\frac{p_k}{\sigma_e^2}}) \tag{60}$$

so p_k is the power of transmitted symbols of k^{th} user and from (59) we have,

$$\sigma_e^2 = [\Phi_{ee}]_{kk} = \frac{1}{\left|r_{kk}\right|^2}\left(\sigma_n^2 + \sum_{j=1}^{K}p_j\left|\delta_{kj}\right|^2 + \beta_k\right) \tag{61}$$

where $\delta_{ij} = [\Delta\boldsymbol{H}]_{ij}$, $\beta_k = \sigma_n^4\sum_{j=1}^{K}\left|\widetilde{h}_{kj}\right|^2 / p_j$, and $\widetilde{h}_{ij} = [\hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1}]_{ij}$. Assuming small error in (61), i.e. $\alpha\, p_j \le \sigma_n^2\ ;\forall\, j$, σ_e^2 can be approximated as [13]:

$$\sigma_e^2 \approx \frac{1}{|r_{kk}|^2}\left(\sigma_n^2 + \alpha\, p_T + \beta_k\right) \tag{62}$$

Where, similar to previous section, the worst-case is assumed, i.e., $\alpha = \max_{i,j} |\delta_{ij}|^2 \; ; \forall i, j$. Thus power distribution that will minimize the average SER, when imperfect CSI is presented, can be found with the use of the Lagrange method as [13]:

$$L = \sum_{k=1}^{K} (1 - \frac{1}{\sqrt{M_k}}) Q(\sqrt{p_k B_k}) - \lambda (K - \sum_{k=1}^{K} p_k) \tag{63}$$

where $B_k = \frac{3}{M_k - 1} \frac{|r_{kk}|^2}{\sigma_n^2 + \alpha\, p_T + \beta_k}$. Unfortunately, we did not find any explicit solution to solve (63). Therefore, some numerical or suboptimal solutions are necessary to solve it.

4.3.2 Suboptimal solution

In (63) since β_k is a function of p_k it is difficult to find explicit solution. One simple method to overcome this problem is that the initial power can be approximated as $p_j \approx P_T / K$, so that in this case β_k is not a function of p_k and,

$$\beta_k = \beta \approx \sigma_n^4 (p_T / K) \sum_{j=1}^{K} \left| \tilde{h}_{kj} \right|^2 \tag{64}$$

It means that the initial power distribution is assumed to be uniform, where the power distribution can be attained according to power allocation policies, i.e., relation (63). The simpler solution can be attained by distributing the power according to noise and channel error values, but without the interference term, i.e. neglecting the interference term and assuming $\beta = 0$. (These assumptions are only valid for the case of small error, i.e., $\beta \le \sigma_n^2 + \alpha\, p_T$.) In this case, the initial power is assigned as a uniform distribution with zero value. Solving $\partial L / \partial p_k = 0$ for p_k, yielding [13]:

$$p_k = \frac{1}{B_k} W(\frac{B_k}{A_k} \Lambda) \tag{65}$$

where $A_k = \frac{M_k}{(\sqrt{M_k} - 1)^2} \frac{1}{B_k}$, $B_k = \frac{3}{M_k - 1} \frac{|r_{kk}|^2}{\sigma_n^2 + \alpha\, p_T + \beta}$, $\Lambda = cte$, and $W(x)$ is the real valued Lambert W -function. The unique solution for $\boldsymbol{P} = [p_1, ..., p_K]$ can be found by the simple iterative procedure same as sub-section 3.2.2 [13]:

It should be noted that, the above power loading can be combined with conventional (equal to before section), robust, or improved robust optimization strategies. In the former cases, the power loadings are similar to conventional power loading in imperfect CSI but THP filters should be calculated according to robust/ improved robust optimization.

4.4 Simulation and results

For simulation purposes, K=4 user with 4-QAM signaling are assumed. The entries of $\hat{\boldsymbol{H}}$ and $\Delta \boldsymbol{H}$ have been assumed to be zero mean i.i.d. complex Gaussian random variables,

i.e., $\hat{\boldsymbol{H}} \sim CN(0,1)$ and $\Delta\boldsymbol{H} \sim CN(0,\rho^2)$, respectively. In simulations, in order to compare our results more simply with other contributions, the BER is ploted instead of the SER.
Fig. 11 plots the mean BER versus Eb/No for Robust optimization together with conventional optimization for 4QAM modulation. It is observed that the robust optimization has better performance for all channel imperfection values, especially for high SNR. Fig. 12 compares the performance of the proposed improved robust optimization with conventional optimization. As can be realized, the proposed improved robust optimization algorithm substantially outperforms the conventional optimization, over the whole observation data lengths. In fact, the performance is noticeable for smaller *N*s, where the channel estimator estimates the channel erroneously, especially for high SNR values.
In order to demonstrate the performance of power loading, for simplicity, the suboptimal solution is considered. The validity of the approximation of $\beta \le \sigma_n^2 + \alpha\, p_T$ is shown in Fig. 13 by plotting the fraction of channel realizations in which the approximations are valid for different values of ρ. It can be seen that for the values of SNR>20dB the approximation is

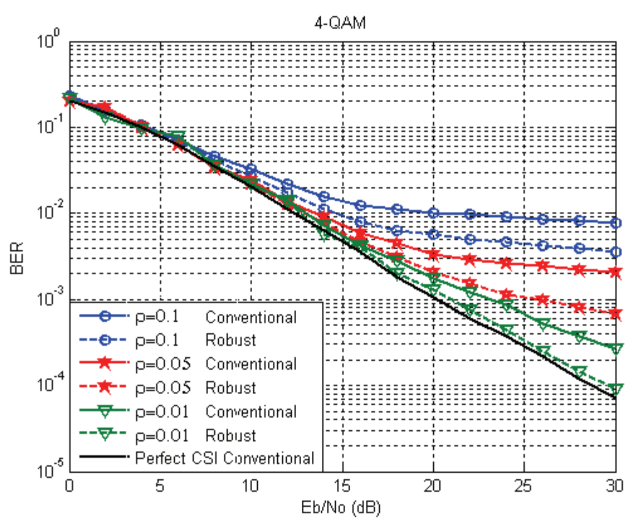


Fig. 11. Robust THP optimization performance

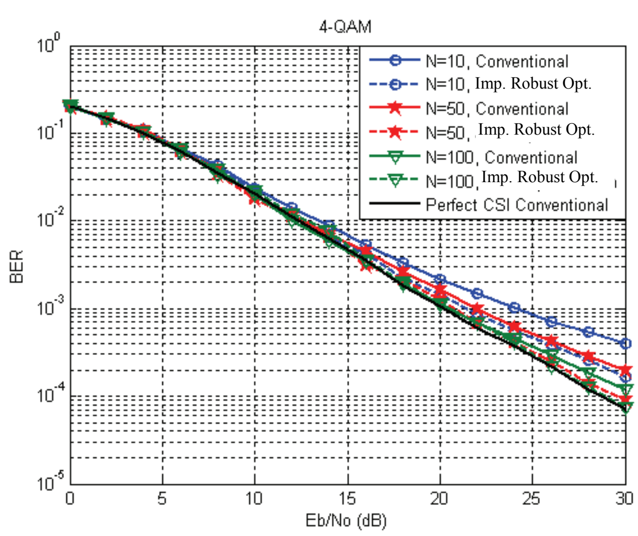


Fig. 12. Improved robust optimization performance with different *N* value

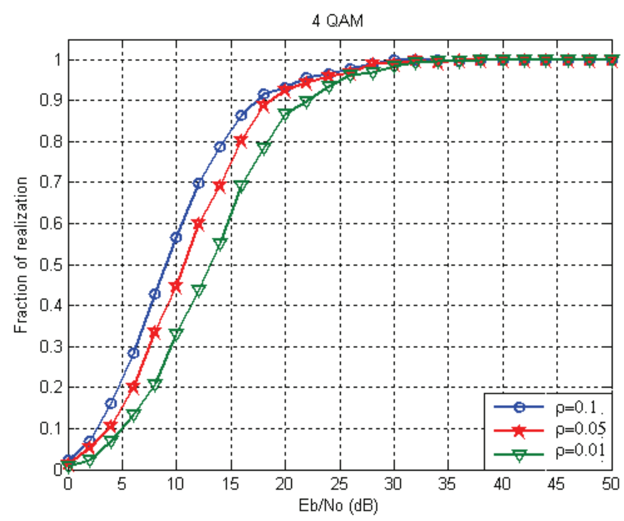


Fig. 13. Validation of approximating $\beta \leq \sigma_n^2 + \alpha\, p_T$

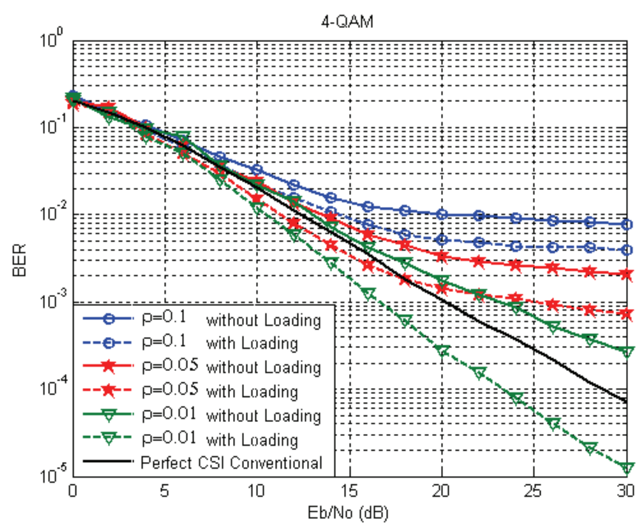


Fig. 14. Suboptimal power loading in conventional optimization

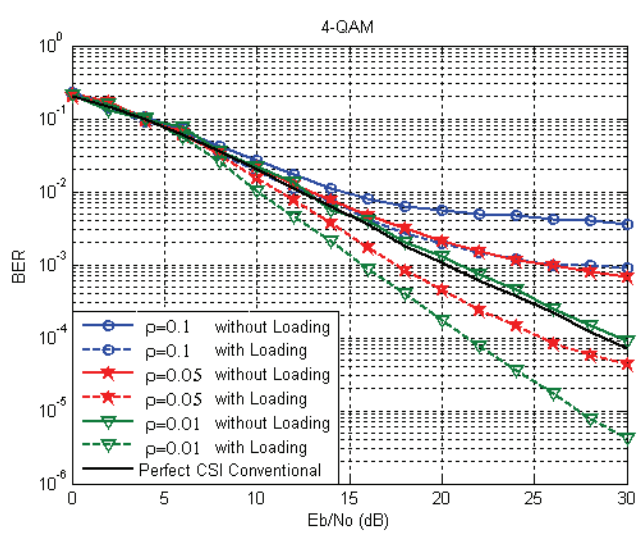


Fig. 15. Suboptimal power loading in robust optimization

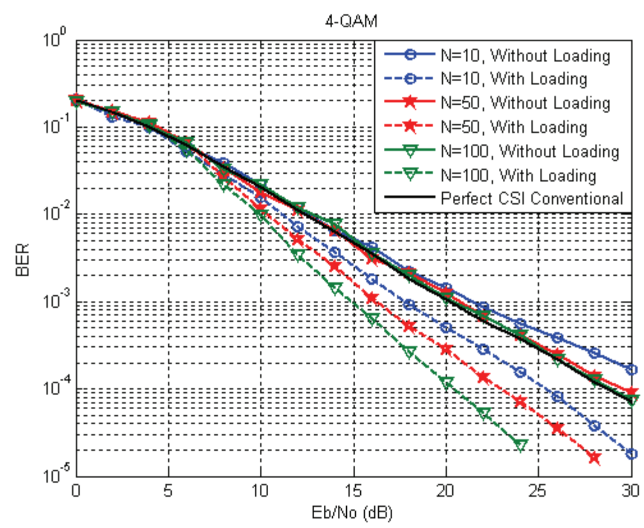


Fig. 16. Suboptimal power loading in improved robust optimization

valid. In figures 14-16 we can observe the performance of power loading for conventional, robust and improved robust optimizations, respectively, with the assumption that $\beta = 0$. As can be seen from these figures, the performance of power loading is noticeable, especially for low errors and high SNR values.

5. Joint channel estimation and spatial pre-equalization

Traditionally, channel estimation and pre-equalization are optimized separately and independently in which results to performance degradation. This loss may be causing some poor performance, especially, in erroneous conditions, as can be seen from previous section. In [21] Dietrich et. al proposed a new method for joint pilot symbol assisted channel estimation and equalization and applied it to the design of the space-time decision feedback equalizer. Their research was developed in [13] for MIMO-BC-THP system with the assumption of the BLUE, as discussed in sub-section 4.2.2. The authors in [13] extracted the explicit solution for THP optimization with a good performance against to separate design of THP and channel estimator.

In this section, their work is extended as joint optimization in which the THP filters are optimized together with the channel estimation conditioned on observation data (with, approximately, the same order of complexity as a separate design). In the other words, in joint optimization, in contrast to separate optimization, the average cost function should be optimized with respect to THP filters and channel estimation, i.e. the expectation is taken with respect to the unknown channel parameters conditioned on the available observation data. It means that in contrast to improved robust optimization in which different channel estimation methods have to be investigated for a given optimized THP to find out the best combination, the best channel estimation can be chosen directly by minimizing the MMSE criterion. As a result, it will be shown that the joint optimization lead to a Linear MMSE (LMMSE) channel estimator and a new structure for THP filters based on the error covariance matrix of the channel estimator.

5.1 Channel estimation

By using the Bayesian Gauss-Markov theorem, the Bayesian LMMSE estimator can be obtained for linear model of (49) [20]:

$$\hat{\boldsymbol{h}} = \mathrm{E}[\boldsymbol{h} \mid \boldsymbol{y}_s] = \boldsymbol{C}_{h_s} \boldsymbol{s}^H (\boldsymbol{s}\boldsymbol{C}_{h_s} \boldsymbol{s}^H + \sigma_n^2 \boldsymbol{I})^{-1} \boldsymbol{y}_s = \boldsymbol{W}_s \boldsymbol{y}_s \tag{66}$$

and

$$\boldsymbol{C}_{h|y_s} = \boldsymbol{C}_{h_s} - \boldsymbol{W}_s \boldsymbol{s} \boldsymbol{C}_{h_s} \tag{67}$$

where $\hat{\boldsymbol{h}}$ indicate the estimation of $\boldsymbol{h}$ and:

$$\begin{aligned} \boldsymbol{C}_{h_s} &= \mathrm{E}[\boldsymbol{h}_s \boldsymbol{h}_s^H] \\ \boldsymbol{W}_s &= \boldsymbol{C}_{h_s} \boldsymbol{s}^H (\boldsymbol{s}\boldsymbol{C}_{h_s} \boldsymbol{s}^H + \sigma_n^2 \boldsymbol{I})^{-1} \end{aligned} \tag{68}$$

5.2 Joint optimization

In conventional THP optimization the error, by using the orthogonality principle of (38), the THP can be optimized in perfect CSI as [11] or in imperfect CSI as [13] in which a specific channel estimator (i.e., BLUE) is assumed and THP is optimized according to this estimator structure (i.e. improved robust optimization). In general aspect, this is not a desired method because it is necessary to select different channel estimators and optimize the THP filters to find the best combination where it is cumbersome work. Nevertheless, in the joint optimization the best channel estimator is determined in which the THP filters and channel estimation can be optimized jointly without any trial method. In this case, since the training sequence and $\boldsymbol{y}_s$ are given, the channel can be modelled as a random variable from point of view of the receiver. Thus, the cost function in (38) is a random variable and should be considered as (55), i.e.:

$$E[\boldsymbol{e}\boldsymbol{r}^H \mid \boldsymbol{y}_s] = 0 \tag{69}$$

where the expectation is taken with respect to the unknown channel parameters. The above equation can be written simpler as:

$$\boldsymbol{G}\Phi_{rr|y_s} = \boldsymbol{B}\Phi_{xr|y_s} \tag{70}$$

where the matrices $\Phi_{rr|y_s}$ and $\Phi_{xr|y_s}$ can be computed by using (1) as [22]:

$$\Phi_{rr|y_s} = \mathrm{E}[\boldsymbol{r}\boldsymbol{r}^H \mid \boldsymbol{y}_s] = \mathrm{E}[(\boldsymbol{H}\boldsymbol{F}\boldsymbol{x} + \boldsymbol{n})(\boldsymbol{H}\boldsymbol{F}\boldsymbol{x} + \boldsymbol{n})^H \mid \boldsymbol{y}_s] = \boldsymbol{V} + \zeta\boldsymbol{I} \tag{71.a}$$

$$\Phi_{xr|y_s} = \mathrm{E}[\boldsymbol{x}\boldsymbol{r}^H \mid \boldsymbol{y}_s] = \mathrm{E}[\boldsymbol{x}(\boldsymbol{H}\boldsymbol{F}\boldsymbol{x} + \boldsymbol{n})^H \mid \boldsymbol{y}_s] = \sigma_x^2 \boldsymbol{F}^H \hat{\boldsymbol{H}}^H \tag{71.b}$$

where $\zeta = \sigma_n^2 / \sigma_x^2$ and $\boldsymbol{V} = \mathrm{E}[\boldsymbol{H}\boldsymbol{H}^H \mid \boldsymbol{y}_s]$. In order to find a closed form solution for THP filters, it is needed to calculate the conditional mean estimate of $\boldsymbol{T} = \boldsymbol{H}\boldsymbol{H}^H$ over observed data, i.e., $\boldsymbol{y}_s$. The matrix $\boldsymbol{T}$ is well known as Gramian matrix where its probability distribution is a Wishart distribution [1]. In order to calculate $\boldsymbol{V}$, consider the cost function as:

$$\boldsymbol{J} = \left\| \hat{\boldsymbol{T}} - \boldsymbol{T} \right\|_F^2 \tag{72}$$

where the lower index stands for Frobenius norm and $\hat{\boldsymbol{T}}$ is a nonlinear function of $\boldsymbol{y}_s$ where should be determined. The minimization of (72) lead to a non-linear conditional mean estimator as [22]:

$$\hat{\boldsymbol{T}} = \mathrm{E}[\boldsymbol{T} \mid \boldsymbol{y}_s] = \mathrm{E}[\boldsymbol{H}\boldsymbol{H}^H \mid \boldsymbol{y}_s] = \sum_{k=1}^{K} \mathrm{E}[\boldsymbol{h}_k \boldsymbol{h}_k^H \mid \boldsymbol{y}_s] \tag{73}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, ..., \boldsymbol{h}_K]$. It is possible to consider each expression in the summation of (73) as [24]:

$$\mathrm{E}[\boldsymbol{h}_k \boldsymbol{h}_k^H \mid \boldsymbol{y}_s] = \mathrm{E}[\boldsymbol{h}_k \mid \boldsymbol{y}_s]\mathrm{E}[\boldsymbol{h}_k \mid \boldsymbol{y}_s]^H + \boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_s} = \hat{\boldsymbol{h}}_k \hat{\boldsymbol{h}}_k^H + \boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_s} \tag{74}$$

where

$$\boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_s} = \mathrm{E}[(\boldsymbol{h}_k - \hat{\boldsymbol{h}}_k)(\boldsymbol{h}_k - \hat{\boldsymbol{h}}_k)^H \mid \boldsymbol{y}_s] \tag{75}$$

Since the error $\boldsymbol{h}_k - \hat{\boldsymbol{h}}_k$ is statically independent from the observation data, we have:

$$\boldsymbol{C}_{\boldsymbol{h}_k|\boldsymbol{y}_s} = \mathrm{E}[(\boldsymbol{h}_k - \hat{\boldsymbol{h}}_k)(\boldsymbol{h}_k - \hat{\boldsymbol{h}}_k)^H] \tag{76}$$

By substituting the relations (74) and (76) in (73) and rearrange the resultant sub-matrices in its original matrix form, we have [22],

$$\hat{\boldsymbol{T}} = \hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \boldsymbol{C}_{\boldsymbol{H}|\boldsymbol{y}_s} \tag{77}$$

where,

$$\boldsymbol{C}_{\boldsymbol{H}|\boldsymbol{y}_s} = \boldsymbol{C}_{\boldsymbol{h}_s} \text{ - } \boldsymbol{W}_s \boldsymbol{S} \boldsymbol{C}_{\boldsymbol{h}_s} \tag{78}$$

The matrices $\boldsymbol{W}_s$ and $\boldsymbol{C}_{\boldsymbol{h}_s}$ are the same as (68) where is used in Bayesian LMMSE channel estimator. On the other hand, it is possible to show that $\hat{\boldsymbol{h}} = \mathrm{E}[\boldsymbol{h} \mid \boldsymbol{y}_s] = \boldsymbol{W}_s \boldsymbol{y}_s$ [20], i.e., this joint optimization lead to a Bayesian LMMSE channel estimator (in the joint optimization, the explicit channel estimation is not needed). In this case, the matrices $\Phi_{rr|\boldsymbol{y}_s}$ in (71.a) can be obtained as [22]:

$$\Phi_{rr|\boldsymbol{y}_s} = \hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \boldsymbol{C}_{\boldsymbol{H}|\boldsymbol{y}_s} + \zeta \boldsymbol{I} \tag{79}$$

Substitution relations (71.b) and (79) in (70) and by some manipulating, the matrix $\boldsymbol{R}$ can be found through the Cholesky factorization of [22]:

$$\boldsymbol{R}\boldsymbol{R}^H = (\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta \boldsymbol{I} + \boldsymbol{C}_{\boldsymbol{H}|\boldsymbol{y}_s})\hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta \boldsymbol{I} + \boldsymbol{C}_{\boldsymbol{H}|\boldsymbol{y}_s}) \tag{80}$$

and matrices $\boldsymbol{G}$, $\boldsymbol{B}$ and $\boldsymbol{F}$ can be found as:

$$\begin{aligned} &\boldsymbol{G} = \mathrm{diag}[r_{11}^{-1}, ..., r_{KK}^{-1}] \\ &\boldsymbol{B} = \boldsymbol{G}\boldsymbol{R} \\ &\boldsymbol{F}^H = \hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta \boldsymbol{I} + \boldsymbol{C}_{\boldsymbol{H}|\boldsymbol{y}_s})\boldsymbol{R}^{-H} \end{aligned} \tag{81}$$

In this case, the error covariance matrix can be computed as [22]:

$$\Phi_{ee} = \sigma_x^2 \boldsymbol{G}(\zeta^2 \hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1} + \zeta \boldsymbol{I} + \boldsymbol{C}_{\boldsymbol{H}|\boldsymbol{y}_s})\boldsymbol{G}^H \tag{82}$$

Note that if the perfect CSI is assumed, i.e. $\boldsymbol{C}_{H|y_s} = 0$, the relations (80-82) are the same as conventional THP optimization where denoted in [12] and here it is referred to as conventional optimization.

5.3 Simulation and results

For simulation purposes, K=4 user with 4-QAM signalling is assumed. The entries of $\boldsymbol{H}$ is assumed to be zero mean i.i.d. complex Gaussian random variables, i.e., $\boldsymbol{H} \sim CN(0,1)$. Fig. 16 compares the performance of the proposed joint optimization with conventional optimization. As can be realized, the proposed joint optimization algorithm substantially outperforms the conventional optimization, over the whole observation data lengths. In fact, the performance is noticeable for smaller Ns, where the channel estimator estimates the channel erroneously, especially for high SNR values. In order to observe the detailed results more precisely, some of the main part of Fig. 17 is reproduced again in Fig. 18.

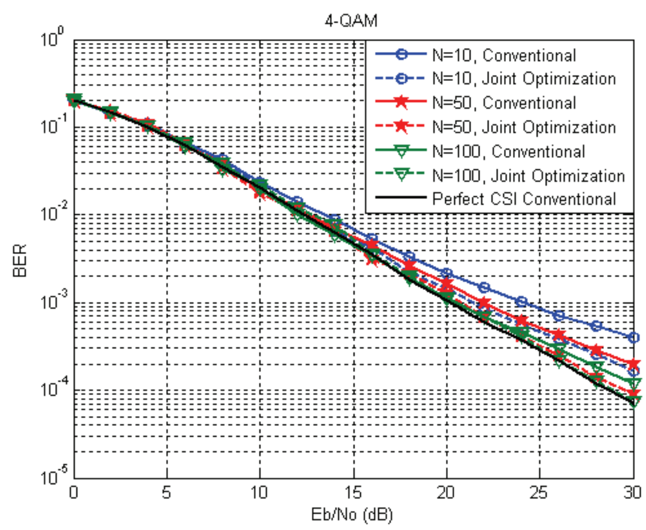


Fig. 17. Joint optimization performance with different observation data length

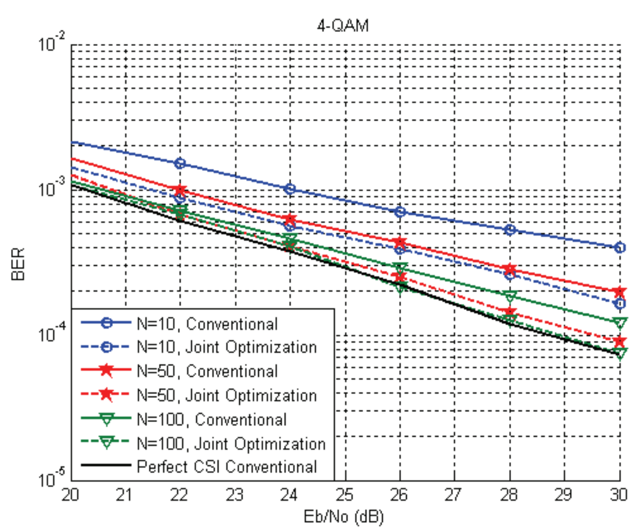


Fig. 18. A detailed reproduction of Fig. 17

6. THP optimization in correlated channel

Mobile channels suffer from multi path fading phenomenon so that it is necessary to adapt the transmitter resources (such as power and rate) with channel characteristics in order to achieve channel capacity. The capacity achieving transmitter adaptation strategy depends on how much CSI is available at the transmitter and the receiver. The use of channel feedback from receiver to transmitter is a standard method in precoding systems, as discussed in previous sections. While knowledge of the channel at the transmitter tends to good performance in precoding systems over slowly time-varying channels, the generation of reliable channel feedback is complicated in fast time-varying channels. This fast time variations impose a significant challenge for precoding design.

In order to increase the THP performance in fast time-varying channels, the THP should be optimized in each (or approximately some) symbol time. On the other hand, the generation of reliable channel feedback in each symbol time is complicated in fast time-varying channel and can be led to high bandwidth overhead. From practical implementation point of view, most of the THPs have been designed assuming that the wireless channel can be regarded as constant over a block of data. In mobile applications where the channel is time-varying, the assumption that the channel is constant over some periods only holds approximately and will affects the THP performance that are designed based on this assumption. Hence, a judicious and innovative THP system design that takes the time-varying nature of the channel into account is the key to solve the above problems.

Based on previous section on precoding and also works denoted in [25], this section extend and unify the THP concepts to time-varying MIMO channels, which will allow us to improve mobile systems performance, as well as to provide guidelines for future precoder designs employing low feedback overhead. A first-order Auto-Regressive (AR) model is used to characterize the channel coefficients that vary from symbol to symbol. Although, traditional THP with a perfect channel estimation in each symbol time has an advantage over MIMO channels, new THP design provides a significant advantage over correlated MIMO channels where the volume of feedback dominate to $1/N$ (N is data block length).

6.1 Correlated channel model

In time-varying channel, it is assumed that the perfect CSI in known at the beginning of each data block, but not during the block. It is desired to optimize THP at time t with the assumption that the outdated CSI, $H_{t-\tau}$, is available. $\boldsymbol{H}_{t-\tau}$ corresponds to the channel state τ seconds earlier (i.e. the beginning of block) where $\tau = lT_s$; $l \le N$. l and T_s denote as number of symbol and symbol time, respectively. Under the assumption of rich scattering, the MIMO channel matrix $\boldsymbol{H}_t$ can be modeled as a complex matrix whose entries are i.i.d. zero-mean complex Gaussian random variables with common variance σ_h^2, i.e. $\boldsymbol{H}_t \sim CN\left(0, \sigma_h^2 \boldsymbol{I}\right)$ [25]. $\boldsymbol{H}_t$ and $\boldsymbol{H}_{t-\tau}$ are correlated realizations of the latter channel distribution. Thus, given the outdated CSI, $\boldsymbol{H}_{t-\tau}$, we can characterize the unknown current CSI, $\boldsymbol{H}_t$, using the conditional CSI model introduced in [25], as follows:

$$\boldsymbol{H}_t \sim CN\left(\rho_t \boldsymbol{H}_{t-\tau}, \sigma_h^2\left(1-\rho_t^{\,2}\right)\boldsymbol{I}\right) \tag{83}$$

where ρ_t is the common time-correlation of the i.i.d. time-varying MIMO channel coefficients, defined as:

$$\rho_t = E\left\{[\boldsymbol{H}_t]_{i,j}[\boldsymbol{H}_{t-\tau}]_{i,j}^{H}\right\}/\sigma_h^2 = R(\tau) \tag{84}$$

where $R(\tau)$ depending on the channel time-variation model.

6.2 THP optimization

The error that is needed to be considered for the system illustrated in Fig. 1, should be the difference between the effective data vector, $\boldsymbol{v}_t$, and the data vector entering the decision module, $\boldsymbol{y}_t$, i.e.:

$$\boldsymbol{e}_t = \boldsymbol{y}_t - \boldsymbol{v}_t = [\boldsymbol{G}_t\boldsymbol{H}_t\boldsymbol{F}_t - \boldsymbol{B}_t]\boldsymbol{x}_t + \tilde{\boldsymbol{n}}_t \tag{85}$$

where $\tilde{\boldsymbol{n}}_t = \boldsymbol{G}_t\boldsymbol{n}_t$. The MMSE solution should minimize the error signal as:

$$\begin{cases} \underset{\boldsymbol{B},\boldsymbol{F},\boldsymbol{G}}{\arg\min}\ E_{H_t|\boldsymbol{H}_{t-\tau}}\left\{E_{a,x}\left\|[\boldsymbol{G}_t\boldsymbol{H}_t\boldsymbol{F}_t - \boldsymbol{B}_t]\boldsymbol{x}_t + \tilde{\boldsymbol{n}}_t\right\|^2\right\} \\ s.t.\quad E\left\|\tilde{\boldsymbol{x}}_t\right\|^2 \le P_T \end{cases} \tag{86}$$

where, P_T is total available power at transmitter. Similar to previous section, instead of solving (86), it is easier to use the orthogonallity principle [13]. In this case, the MMSE solution should satisfy [25]:

$$E_{H_t|\boldsymbol{H}_{t-\tau}}\left\{E_{a,x}[\boldsymbol{e}_t\boldsymbol{r}_t^{\boldsymbol{H}}]\right\} = 0 \tag{87}$$

Thus:

$$\boldsymbol{G}_t\Phi_{rr,t} = \boldsymbol{B}_t\Phi_{xr,t} \tag{88}$$

The matrix $\Phi_{rr,t}$ can be computed by using (1) as:

$$\Phi_{rr,t} = E_{H_t|\boldsymbol{H}_{t-\tau}}\left\{E_{a,x}[\boldsymbol{r}_t\boldsymbol{r}_t^{\boldsymbol{H}}]\right\} = E_{H_t|\boldsymbol{H}_{t-\tau}}[\sigma_x^2\boldsymbol{H}_t\boldsymbol{H}_t^{\boldsymbol{H}} + \sigma_n^2\boldsymbol{I}] \tag{89}$$

The matrix $\boldsymbol{H}_t\boldsymbol{H}_t^{\boldsymbol{H}}$ is well known as Gramian matrix where its probability distribution is a Wishart distribution. Calculation of the conditional expectation, as required in (89), seems to be difficult. We consider an approximate solution to solve it. To do this, by using the channel distribution of (83), we first instantiate the true channel $\boldsymbol{H}_t$ as $\boldsymbol{H}_t = \overline{\boldsymbol{H}}_t + \Delta_t$, where $\overline{\boldsymbol{H}}_t = \rho_t\boldsymbol{H}_{t-\tau}$ and Δ_t is the $CN(0,\sigma_h^2(1-\rho_t{}^2)\boldsymbol{I})$ -distributed uncertainty on the true channel given the outdated CSI [25]. Under the assumption of isotropic scattering and moving terminal, this model describes the time-correlation function as $R(\tau) = J_0(2\pi f_D\tau)$, where J_0 is the zero-th order Bessel function of the first kind and f_D is maximum Doppler frequency [26]. Then, we use the following expression as statistical model for the time variations of the channel [25]:

$$\boldsymbol{H}_t = \rho_t\boldsymbol{H}_{t-\tau} + \sigma_h\sqrt{1-\rho_t{}^2}\,\boldsymbol{E}_t \tag{90}$$

where ρ_t denotes the correlation coefficient between the time instants $t-\tau$ and t, and $\boldsymbol{E}_t$ is a circularly symmetric complex Gaussian matrix with i.i.d. entries, i.e. $\boldsymbol{E}_t \sim CN(0,\boldsymbol{I})$. In this case [25],

$$\Phi_{rr,t} = \boldsymbol{E}_{H_t|\boldsymbol{H}_{t-\tau}}[\sigma_x^2(\overline{\boldsymbol{H}}_t + \Delta_t)(\overline{\boldsymbol{H}}_t + \Delta_t)^H + \sigma_n^2 \boldsymbol{I}] = \sigma_x^2 \overline{\boldsymbol{H}}_t \overline{\boldsymbol{H}}_t^H + \boldsymbol{C}_{\Delta,t} + \sigma_n^2 \boldsymbol{I} \quad (91)$$

where $\boldsymbol{C}_{\Delta,t} = \sigma_h^2(1-\rho_t^2)\boldsymbol{I}$. The matrix $\Phi_{xr,t}$ in (88) can be computed as [25]:

$$\begin{aligned}\Phi_{xr,t} &= \mathrm{E}[\boldsymbol{x}_t \boldsymbol{r}_t^H] = E_{H_t|\boldsymbol{H}_{t-\tau}}\left\{E_{a,x}[\boldsymbol{x}_t \boldsymbol{r}_t^H]\right\} = E_{H_t|\boldsymbol{H}_{t-\tau}}\left\{E_{a,x}[\boldsymbol{x}_t \boldsymbol{x}_t^H \boldsymbol{F}_t^H \boldsymbol{H}_t^H]\right\} \\ &= E_{H_t|\boldsymbol{H}_{t-\tau}}[\sigma_x^2 \boldsymbol{F}_t^H (\overline{\boldsymbol{H}}_t + \Delta_t)^H] = \sigma_x^2 \boldsymbol{F}_t^H \overline{\boldsymbol{H}}_t^H\end{aligned} \quad (92)$$

Substituting relations (89) and (92) in (88) and after some manipulations lead to [25]:

$$\boldsymbol{F}_t^H = \boldsymbol{B}_t^{-1}\boldsymbol{G}_t(\overline{\boldsymbol{H}}_t \overline{\boldsymbol{H}}_t^H + \zeta \boldsymbol{I} + \boldsymbol{C}_{\Delta,t})\overline{\boldsymbol{H}}_t^{-H} \quad (93)$$

where $\zeta = \sigma_n^2 / \sigma_x^2$. Since $\boldsymbol{F}_t$ is unitary matrix, we have:

$$\boldsymbol{R}_t \boldsymbol{R}_t^H = (\overline{\boldsymbol{H}}_t \overline{\boldsymbol{H}}_t^H + \zeta \boldsymbol{I} + \boldsymbol{C}_{\Delta,t})\overline{\boldsymbol{H}}_t^{-H}\overline{\boldsymbol{H}}_t^{-1}(\overline{\boldsymbol{H}}_t \overline{\boldsymbol{H}}_t^H + \zeta \boldsymbol{I} + \boldsymbol{C}_{\Delta,t}) \quad (94)$$

where $\boldsymbol{R}_t = \boldsymbol{G}_t^{-1}\boldsymbol{B}_t$ is assumed. The matrix $\boldsymbol{R}_t$ can be found through Cholesky factorization of (94) and the matrices $\boldsymbol{G}_t$, $\boldsymbol{B}_t$, and $\boldsymbol{F}_t$ can be found as:

$$\begin{aligned}\boldsymbol{G}_t &= \mathrm{diag}[r_{11}^{-1}, \ldots, r_{KK}^{-1}] \\ \boldsymbol{B}_t &= \boldsymbol{G}_t \boldsymbol{R}_t \\ \boldsymbol{F}_t &= (\rho_t \boldsymbol{H}_{t-\tau})^{-1}(\rho_t^2 \boldsymbol{H}_{t-\tau}\boldsymbol{H}_{t-\tau}^H + \sigma_h^2(1-\rho_t^2)\boldsymbol{I} + \zeta \boldsymbol{I})\boldsymbol{R}_t^{-H}\end{aligned} \quad (95)$$

with $\boldsymbol{B}_t$, $\boldsymbol{F}_t$ and $\boldsymbol{G}_t$ that found in (95), the error covariance matrix can be computed as [25]:

$$\Phi_{ee,t} = \boldsymbol{E}[\boldsymbol{e}_t \boldsymbol{e}_t^H] = \sigma_x^2 \boldsymbol{G}_t(\zeta^2 \rho_t^2 \boldsymbol{H}_{t-\tau}^{-H}\boldsymbol{H}_{t-\tau}^{-1} + \zeta \boldsymbol{I} + \sigma_h^2(1-\rho_t^2)\boldsymbol{I})\boldsymbol{G}_t^H \quad (96)$$

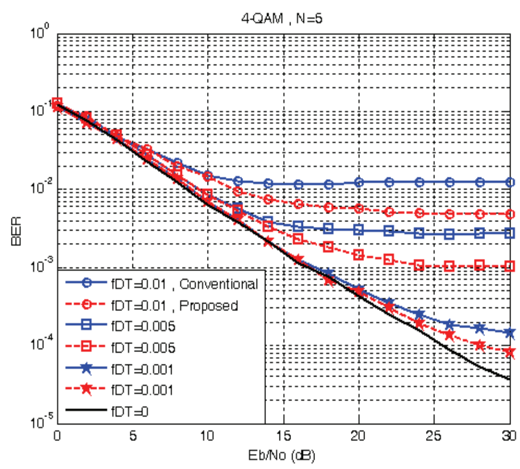
In situation where the channel is assumed to be quasi-static (i.e. $\rho_t \to 1$), the relations (94-96) tends to what is considered as conventional THP in relations (45-47) and also in previous section.

6.3 Simulation and results

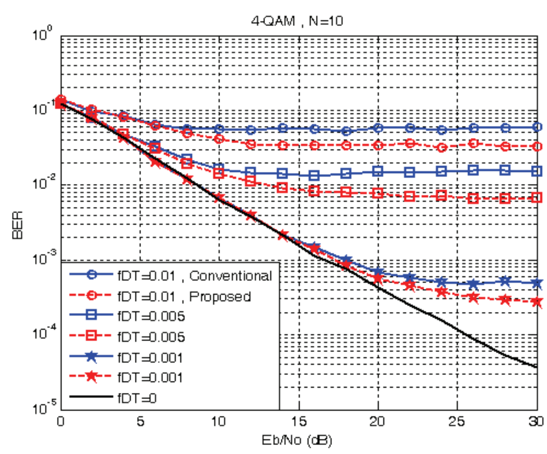
In this section, we illustrate the improvement, in term of average BER, that our proposed MMSE-THP design offers over conventional THP designs that assumes the channel is constant over *N* symbol time. In order to do that, we use the well-known Jakes model [27] to instantiate a realistic outdated CSI model based on (83). In our simulations, we use the normalized Doppler frequency $f_D T = 0.001$, $f_D T = 0.005$, and $f_D T = 0.01$ according to slow, medium, and fast fading, respectively [28]. We further consider the case of a (4, 4) MIMO set-up with 4QAM-modulated data streams.

Figures (19a) to (19c) plot the average BER performances for *N*=5, 10, 20, respectively. Clearly, our robust MMSE-THP design exhibits a lower average BER performance when compared to the conventional design. More specifically, from the figures, it can be observed that the proposed method have more advantage in fast fading over the state-of-the-art design for shorter *N* where the channel correlation is noticeable.

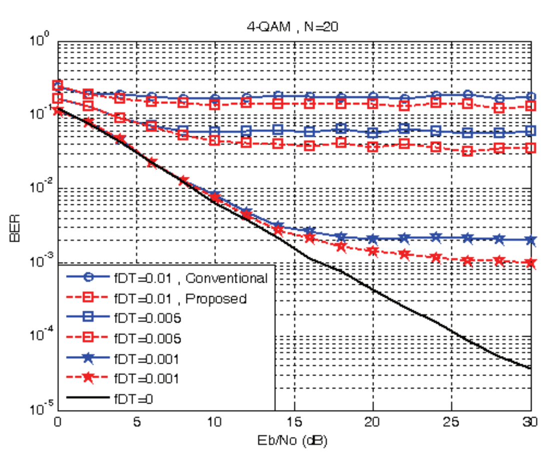
Figures (20a) to (20c) plot the average BER performances for slow, medium, and fast fading, respectively. Observe that for slow fading case, our proposed method has good performance over all values of *N* while in the fast fading case; the performance is noticeable for shorter *N*s.



(a) N=5

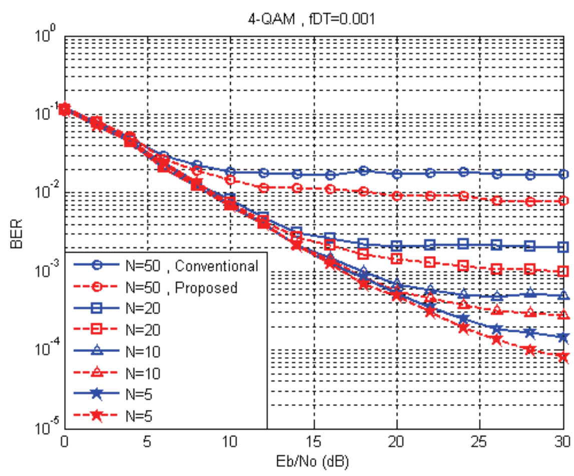


(b) N=10

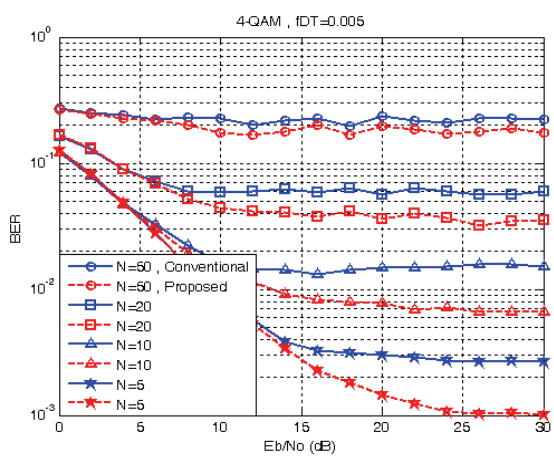


(c) N=20

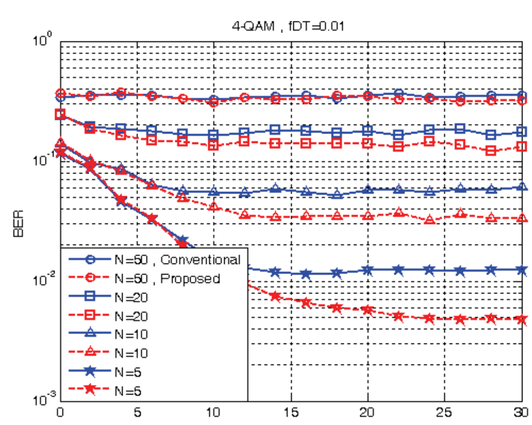
Fig. 19. BER performance for different N values



(a) $f_D T = 0.001$



(b) $f_D T = 0.005$



(c) $f_D T = 0.01$

Fig. 20. BER performance for different Doppler values

7. References

[1] H. Khaleghi Bizaki, "Precoding and Blind/Semi-blind Estimation in MIMO Fading Channels", PhD Thesis, Iran University of Science and Technology (IUSTz), Winter 2008.

[2] R. F. H. Fischer, C. Windpassinger, a. Lampe, and J. B. Huber, "Space-Time Transmission using Tomlinson-Harashima Precoding ", In Proceedings of 4th ITG Conference on Source and Channel Coding, pp. 139-147, Berlin, January 2002.

[3] C. Windpassinger, "Detection and Precoding for Multiple Input Multiple Output Channels", PhD Thesis, Erlangen, 2004.

[4] M. H. M. Costa, "Writing on Dirty Paper", IEEE Transactions on Information Theory, vol. IT-29, No. 3, May 1983.

[5] R. F. H. Fischer, C. Windpassinger, a. Lampe, and J. B. Huber, " MIMO Precoding for Decentralized Receivers"

[6] G. H. Golub and C. F. VanLoan, "Matrix Computations", The Johns Hopkins University Press, Baltimore, MD, USA, 3rd edition, 1996.

[7] M. Tomilson, "New Automatic Equalizer Employing Modulo Arithmetic", Electronic Letters, pp. 138-139, March 1971.

[8] H. Harashima, Miyakawa, "Matched –Transmission Technique for Channels with Intersymbole Interference", IEEE Journal on Communications, pp. 774-780, Aug. 1972.

[9] U. Erez, S. Shamai, and R. Zamir, "Capacity and Lattice Strategies for Cancelling Known Interference", In Proceeding of International Symposium on Information Theory abd Its Applications, Honolulu, Hi, USA, Nov. 2000.

[10] Q. Zhou, H. Dai, and H. Zhang, "Joint Tominson-Harashima Precoding and Scheduling for Multiuser MIMO with Imperfect Feedback", IEEE Wireless Comm. and Networking Conf. (WCNC), Vol. 3, pp: 1233-1238, 2006.

[11] H. Khaleghi Bizaki and A. Falahati, "Power Loading by Minimizing the Average Symbol Error Rate on MIMO-THP Systems", The 9th Int. Conf. on Advanced Comm. Technology (ICACT), Vol. 2, pp: 1323-1326, Feb. 2007

[12] J. Lin, W. A. Krzymein, "Improved Tomlinson Harashima Precoding for the Downlink of Multiple Antenna Multi-User Systems", IEEE Wireless Comm. and Networking Conf. (WCNC), pp: 466-472, March 2006.

[13] H. Khaleghi Bizaki, A. Falahati, "Tomlinson Harashima Precoder with Imperfect Channel State Information ", IET Communication Journal, Volume 2, Issue 1,Page(s):151 – 158, January 2008.

[14] T. Hunziker, D. Dahlhaus, "Optimal Power Adaptation for OFDM Systems with Ideal Bit-Interleaving and Hard-Decision Decoding", IEEE International Conference on Communications (ICC), vol. 5, pp:3392-3397, 2003.

[15] R. D. Wesel, J. Cioffi, "Achievable Rates for Tomlinson-Harashima Precoding", IEEE Transaction on Information Theory, vol. 44, No. 2, March 1998.

[16] M. Payaro, A. P. Neira, M. A. Lagunas, "Achievable Rates for Generalized Spatial Tomlonson-Harashima Precoding in MIMO Systems", IEEE Vehicular Technology Conference (VTC), vol. 4, pp: 2462 – 2466, Fall 2004.

[17] Bizaki, H.K.; Falahati, A., "Achievable Rates and Power Loading in MIMO-THP Systems ", 3rd International Conference on Information and Communication Technologies (ICTTA),Page(s): 1 - 7, 2008

[18] Payaro, M., Neira, A.P., and Lagunas, M.A., "Robustness evaluation of uniform power allocation with antenna selection for spatial Tomlinson-Harashima precoding", IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP 2005), Philadelphia, (USA), 18–23 March 2005.

[19] H. Khaleghi Bizaki, "Channel Imperfection Effects on THP Performance in a Slowly Time Varying MIMO Channels", IEEE WCNIS2010-Wireless Communication, Conference date: 25-27 June 2010

[20] Kay, S.M.: ‘Fundamentals of statistical signal processing: estimation theory’ (Prentice-Hall, 1993)

[21] Dietrich, F.A., Joham, M., and Utschick, W., "Joint optimization of pilot assisted channel estimation and equalization applied to space-time decision feedback equalization", Int. Conf. on Communication (ICC), 2005, Vol. 4, pp. 2162–2167

[22] H. Khaleghi Bizaki and A. Falahati, "Joint Channel Estimation and Spatial Pre-Equalization in MIMO Systems ", IET Electronics Letters, Vol. 43, Issue 24, Nov. 2007

[23] Bizaki, H.Khaleghi, and Falahati, A., "Tomlinson-Harashima precoding with imperfect channel side information", 9th International Conference on Advanced Communication Technology (ICACT), Korea, 2007, Vol. 2, pp. 987–991

[24] Dietrich, F.A., Hoffman, F., and Utschick,W.: ‘Conditional mean estimator for the Gramian matrix of complex gaussian random variables’. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, Pennsylvania, USA, 2005, Vol. 3, pp. 1137–1140

[25] H. Khaleghi Bizaki, "Tomlinson-Harashima Precoding Optimization over Correlated MIMO Channels", IEEE WCNIS2010-Wireless Communication, Conference date: 25-27 June 2010.

[26] N. Khaled, G. Leus, C. Desset and H. De Man, "A Robust Joint Linear Precoder and Decoder MMSE Design for Slowly Time- Varying MIMO Channels", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 4, pp: 485- 488, 2004.

[27] M. Patzold, "Mobile Fading Channels--Modeling, Analysis & Simulation", John Wiley, 2002.

[28] P. M. Castro, L. Castedo and J. Miguez, "Adaptive Precoding in MIMO Wireless Communication Systems Using Blind Channel Precoding Over Frequency Selective Fading Channels", IEEE 13th Workshop on Statistical Signal Processing, pp: 173 – 178, 2005.

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