Joint channel estimation and spatial pre-equalisation in MIMO systems

H. Khaleghi Bizaki and A. Falahati

Most contributions on Tomlinson-Harashima precoding (THP) consider the THP design based on perfect channel state information (CSI) at the transmitter. The perfect CSI is not available in many wireless systems owing to channel estimator error. Traditionally, the channel is estimated by a specified channel estimator and is applied to the THP optimisation function as if it is error-free. This separate optimisation causes some performance degradation, especially in the erroneous estimated channel. Derived is a new robust solution for the THP MIMO system, which optimises jointly the channel estimation and THP filters. The proposed method provides significant improvement with respect to conventional separate optimisation. Simulation results show the performance advantage obtained by the joint optimisation.

Introduction: The Tomlinson-Harashima precoding (THP) solution based on minimum mean square error (MMSE) criterion is one of the most useful pre-equalisation techniques to achieve near multiple input multiple output (MIMO) channel capacity with reasonable complexity. Traditionally, channel estimation and pre-equalisation are optimised separately and independently, which results in performance degradation. This loss may cause poor performance, especially, in erroneous conditions. In [1] Dietrich *et al.* proposed a new method for joint pilot symbol assisted channel estimation and equalisation and applied it to the design of the space-time decision feedback equaliser. Their research was developed in [2] for the THP MIMO system employing the known error covariance matrix of the channel (due to time variation of the channel or imperfect channel estimation). In this Letter, their work is extended as joint optimisation in which the THP filters are optimised together with the channel estimation conditioned on observation data (with approximately the same order of complexity as a separate design). In other words, in joint optimisation, in contrast to separate optimisation, the average cost function should be optimised 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 conventional optimisation in which different channel estimation methods have to be investigated for a given optimised THP to find the best combination, the best channel estimation can be chosen directly by the MMSE criterion. As a result, it will be shown that the joint optimisation leads to a linear MMSE (LMMSE) channel estimator and a new structure for THP filters based on the error covariance matrix of the channel estimator.

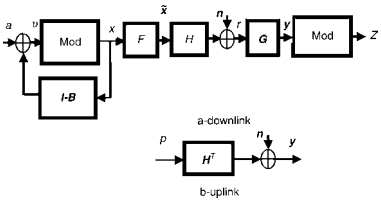

Fig. 1 *THP model in MIMO system*

System overview: The base station (BS) with n_T transmit antennas and n_R users with a single antenna can be considered as a MIMO broadcast system. The system is assumed to be time division duplex (TDD) where the channel state information (CSI) is measured in the uplink channel and used to optimise THP filters in the downlink channel. 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 a passes through lower triangular feedback filter $\boldsymbol{B}$, which is added to the intended transmit vector to pre-eliminate interference from previous users. The resultant signal is then fed to the modulo-operator, which serves to limit the transmit power. The output signal of modulo-operator $\boldsymbol{x}$ is then passed through a unitary feed forward filter $\boldsymbol{F}$ to further remove interference from future users. Finally, the precoded signal $\tilde{\boldsymbol{x}}$ is sent through 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 matrix $\boldsymbol{G}$), reverse modulo-operation, and single user detection.

The matrices $\boldsymbol{B}$, $\boldsymbol{F}$, and $\boldsymbol{G}$ can be found by zero forcing (ZF) or MMSE criteria as in [3]. The received signal before modulo reduction, in a fixed time stand, can be modelled as:

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

where $\tilde{\boldsymbol{x}} = \boldsymbol{F}\boldsymbol{B}^{-1}\boldsymbol{v}$ $[\tilde{x}_1, \ldots, \tilde{x}_{n_T}]^T$, $\boldsymbol{r}$ $[r_1, \ldots, r_{n_R}]^T$ $\boldsymbol{y}$ $[y_1, \ldots, y_{n_R}]^T$, and $\boldsymbol{n}$ $[n_1, \ldots, n_{n_R}]^T$ are transmitted, received before and after the scaling matrix and noise vectors, respectively. $\boldsymbol{H} = [h_{jl}]_{n_R} \times n_T$ is the channel matrix, $\boldsymbol{v} = \boldsymbol{a} + \boldsymbol{d}$ is the effective input data vector, and $\boldsymbol{d}$ is the precoding vector used to constrain the value of $\tilde{\boldsymbol{x}}$ [2, 3]. The elements of the noise vector are assumed independent complex Gaussian random variables with zero mean and variance σ^2, i.e. $\boldsymbol{n} \sim CN(0, \sigma^2 \mathbf{I}_{n_R})$. In addition, the elements of matrix $\boldsymbol{H}$ are considered as complex Gaussian random variables (i.e. flat fading case) with zero mean and unit variance. In the rest of the Letter, for the sake of simplicity and without loss of generality, it is assumed that the number of transmit and receive antennas are the same, i.e. $n_T = n_R = K$.

The received signal at the BS during the training period (uplink), at time stand i, can be modelled as:

$$\boldsymbol{y}(i) = \boldsymbol{H}^T\boldsymbol{p}(i) + \boldsymbol{n}(i) \tag{2}$$

where $\boldsymbol{y}(i) = [y_1(i), \ldots, y_K(i)]^T$, $\boldsymbol{p}(i) = [p_1(i, \ldots, p_K(i))]^T$, $\boldsymbol{n}(i) = [n_1(i), \ldots, n_K(i)]^T$ and $\boldsymbol{H}^T = [\boldsymbol{h}_1^T, \ldots, \boldsymbol{h}_K^T]^T$; $y_k(i)$ is the received signal at the kth receive antenna, p_k (i) is kth user's known symbol (pilot) to train the channel, $\boldsymbol{h}_j = [h_{jl}]$, and h_{jl} is the tap gain from transmit antenna l to receive antenna j.

During the training period (N symbols) in the uplink transmission, the received signal can be considered as [4]:

$$\boldsymbol{y}_s = \boldsymbol{s}\boldsymbol{h}_s + \boldsymbol{n} \tag{3}$$

where $\boldsymbol{y}_s = [y(0), \ldots, \boldsymbol{y}(N-1)]^T$, $\boldsymbol{h}_s = \mathrm{vec}[\boldsymbol{H}^T]$, $\boldsymbol{s} = [\boldsymbol{A}(0), \ldots, \boldsymbol{A}(N-1)]^T$, $\boldsymbol{n} = [\boldsymbol{n}(0), \ldots, \boldsymbol{n}(N-1)]^T$ and $\boldsymbol{A}(i)$ can be constructed as the block diagonal matrix with elements of $\boldsymbol{p}(i)$.

By using the Bayesian Gauss-Markov theorem, the Bayesian LMMSE estimator can be obtained for the linear model of (3) as [5]:

$$\hat{\boldsymbol{h}}_s = \mathrm{E}[\boldsymbol{h}_s|\boldsymbol{y}_s] = \boldsymbol{C}_{h_s}\boldsymbol{s}^H(\boldsymbol{s}\boldsymbol{C}_{h_s}\boldsymbol{s}^H + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{y}_s = \boldsymbol{W}_s\boldsymbol{y}_s \tag{4}$$

and

$$\boldsymbol{C}_{h|\boldsymbol{y}_s} = \boldsymbol{C}_{h_s} - \boldsymbol{W}_s\boldsymbol{s}\boldsymbol{C}_{h_s} \tag{5}$$

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

$$\begin{aligned} \boldsymbol{C}_{h_s} &= \mathrm{E}[\boldsymbol{h}_s\boldsymbol{h}_s^H] \\ \boldsymbol{W}_s &= \boldsymbol{C}_{h_s}\boldsymbol{s}^H(\boldsymbol{s}\boldsymbol{C}_{h_s}\boldsymbol{s}^H + \sigma_n^2\boldsymbol{I})^{-1} \end{aligned} \tag{6}$$

Joint optimisation: In conventional THP optimisation the error that needs 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 $\boldsymbol{y}$ [3], i.e.

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{v} = [\boldsymbol{G}\boldsymbol{H}\boldsymbol{F} - \boldsymbol{B}]\boldsymbol{x} + \tilde{\boldsymbol{n}} \tag{7}$$

where $\tilde{\boldsymbol{n}} = \boldsymbol{G}\boldsymbol{n}$, $\boldsymbol{y} = \boldsymbol{G}\boldsymbol{r}$ and $\boldsymbol{r}$ is the received vector. The MMSE solution should minimise the error signal as:

$$\begin{cases} \underset{B,F,G}{\mathrm{argmin}}\, E\|[\boldsymbol{G}\boldsymbol{H}\boldsymbol{F} - \boldsymbol{B}]\boldsymbol{x} + \tilde{\boldsymbol{n}}\|^2 \\ s.t.\ E\|\tilde{\boldsymbol{x}}\|^2 \le P_T \end{cases} \tag{8}$$

where P_T is the total available power at the transmitter. Instead of solving (8), it is easier to use the orthogonality principle. In this case, the MMSE solution should satisfy [3]:

$$E[\boldsymbol{e}\boldsymbol{r}^{\boldsymbol{H}}] = 0 \tag{9}$$

By solving (9), the THP can be optimised in perfect CSI as [3] or in imperfect CSI as [2] in which a specific channel estimator is assumed and THP is optimised according to error covariance matrix of the channel. From a general aspect, separate optimisation is not a desired method because it is necessary to select different channel estimators and optimise the THP filters to find the best combination (which is cumbersome work). Nevertheless, in the joint optimisation the best channel estimator is determined in a method in which the THP filters and channel estimation can be optimised 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 the point of view of the receiver. Thus, the cost function in (9) is a random variable and should be considered as:

$$E[\boldsymbol{e}\boldsymbol{r}^H|\boldsymbol{y}_s] = 0 \tag{10}$$

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

$$\boldsymbol{G}\Phi_{rr|\boldsymbol{y}_s} = \boldsymbol{B}\Phi_{xr|\boldsymbol{y}_s} \tag{11}$$

where the matrices $\Phi_{rr|\boldsymbol{y}_s}$ and $\Phi_{xr|\boldsymbol{y}_s}$ can be computed using (1) as:

$$\begin{aligned}\Phi_{rr|\boldsymbol{y}_s} &= E[\boldsymbol{r}\boldsymbol{r}^H|\boldsymbol{y}_s] = E[(\boldsymbol{HFx}+\boldsymbol{n})(\boldsymbol{HFx}+\boldsymbol{n})^H|\boldsymbol{y}_s] \\ &= \boldsymbol{V} + \zeta\boldsymbol{I} \end{aligned} \tag{12a}$$

$$\begin{aligned}\Phi_{rr|\boldsymbol{y}_s} &= E[\boldsymbol{x}\boldsymbol{r}^H|\boldsymbol{y}_s] = E[\boldsymbol{x}(\boldsymbol{HFx}+\boldsymbol{n})^H|\boldsymbol{y}_s] \\ &= \sigma_x^2\boldsymbol{F}^H\hat{\boldsymbol{H}}^H \end{aligned} \tag{12b}$$

where $\zeta = \sigma_n^2/\sigma_x^2$ and $\boldsymbol{V} = E[\boldsymbol{H}\boldsymbol{H}^H|\boldsymbol{y}_s]$. To find a closed form solution for THP filters, it is necessary 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 the Gramian matrix where its probability distribution is a Wishart distribution [6]. To calculate $\boldsymbol{V}$, consider the cost function as:

$$\boldsymbol{J} = \|\hat{\boldsymbol{T}} - \boldsymbol{T}\|_F^2 \tag{13}$$

where the lower index stands for the Frobenius norm and $\hat{\boldsymbol{T}}$ is a nonlinear function of $\boldsymbol{y}_s$ which should be determined. The minimisation of (13) leads to a nonlinear conditional mean estimator as [7]:

$$\hat{\boldsymbol{T}} = E[\boldsymbol{T}|\boldsymbol{y}_s] = E[\boldsymbol{H}\boldsymbol{H}^H|\boldsymbol{y}_s] = \sum_{k=1}^{K} E[\boldsymbol{h}_k\boldsymbol{h}_k^H|\boldsymbol{y}_s] \tag{14}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]$. It is possible to consider each expression in the summation as [8]:

$$E[\boldsymbol{h}_k\boldsymbol{h}_k^H|\boldsymbol{y}_s] = E[\boldsymbol{h}_k|\boldsymbol{y}_s]E[\boldsymbol{h}_k|\boldsymbol{y}_s]^H + \boldsymbol{C}_{h_k|\boldsymbol{y}_s} = \hat{\boldsymbol{h}}_k\hat{\boldsymbol{h}}_k^H + \boldsymbol{C}_{h_k|\boldsymbol{y}_s} \tag{15}$$

where

$$\boldsymbol{C}_{h_k|\boldsymbol{y}_s} = E[(\boldsymbol{h}_k - \hat{\boldsymbol{h}}_k)(\boldsymbol{h}_k - \hat{\boldsymbol{h}}_k)^H|\boldsymbol{y}_s] \tag{16}$$

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

$$\boldsymbol{C}_{h_k|\boldsymbol{y}_s} = E[(\boldsymbol{h}_k - \hat{\boldsymbol{h}}_k)(\boldsymbol{h}_k - \hat{\boldsymbol{h}}_k)^H] \tag{17}$$

By substituting the relations (15) and (17) in (14) and rearranging the resultant sub-matrices in original matrix form, we have,

$$\hat{\boldsymbol{T}} = \hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \boldsymbol{C}_{H|\boldsymbol{y}_s} \tag{18}$$

where

$$\boldsymbol{C}_{H|\boldsymbol{y}_s} = \boldsymbol{C}_{h_s} - \boldsymbol{W}_s\boldsymbol{S}\boldsymbol{C}_{h_s} \tag{19}$$

The matrices $\boldsymbol{W}_s$ and $\boldsymbol{C}_{h_s}$ are the same as in (6) which is used for the Bayesian LMMSE channel estimator. On the other hand, it is possible to show that $\hat{\boldsymbol{h}}_s = E[\boldsymbol{h}_s|\boldsymbol{y}_s] = \boldsymbol{W}_s\boldsymbol{y}_s$ [5], i.e. this joint optimisation leads to a Bayesian LMMSE channel estimator. Note that, in the joint optimisation, the explicit channel estimation, i.e. $\hat{\boldsymbol{h}} = \boldsymbol{W}_s\boldsymbol{y}_s$, is not needed. In this case, the matrices $\Phi_{rr|\boldsymbol{y}_s}$ in (12a) can be obtained as:

$$\Phi_{rr|\boldsymbol{y}_s} = \hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \boldsymbol{C}_{H|\boldsymbol{y}_s} + \zeta\boldsymbol{I} \tag{20}$$

Substituting relations (12b) and (20) in (11) and by some manipulating, the lower triangular matrix $\boldsymbol{R}$ (the matrix $\boldsymbol{R}$ should be a lower triangular form in which the feed back matrix $\boldsymbol{B}$ can be attained from $\boldsymbol{R}$ after normalising its diagonal elements by the scaling matrix $\boldsymbol{G}$ as $\boldsymbol{B} = \boldsymbol{B}\boldsymbol{R}$) can be found through the Cholesky factorisation of:

$$\boldsymbol{R}\boldsymbol{R}^H = (\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I} + \boldsymbol{C}_{H|\boldsymbol{y}_s})\hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I} + \boldsymbol{C}_{H|\boldsymbol{y}_s}) \tag{21}$$

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

$$\begin{aligned}\boldsymbol{G} &= \mathrm{diag}[r_{11}^{-1}, \ldots, r_{kk}^{-1}] \\ \boldsymbol{B} &= \boldsymbol{G}\boldsymbol{R} \\ \boldsymbol{F}^H &= \hat{\boldsymbol{H}}^{-1}(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \zeta\boldsymbol{I} + \boldsymbol{C}_{H|\boldsymbol{y}_s})\boldsymbol{R}^{-H}\end{aligned} \tag{22}$$

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

$$\Phi_{ee} = \sigma_x^2\boldsymbol{G}(\zeta^2\hat{\boldsymbol{H}}^{-H}\hat{\boldsymbol{H}}^{-1}\zeta\boldsymbol{I} + \boldsymbol{C}_{H|\boldsymbol{y}_s})\boldsymbol{G}^H \tag{23}$$

Note that if the perfect CSI is assumed, i.e. $\boldsymbol{C}_{H|\boldsymbol{y}_s} = 0$, the relations (21)–(23) are the same as conventional THP optimisation (denoted in [3]) and here it is referred to as conventional optimisation.

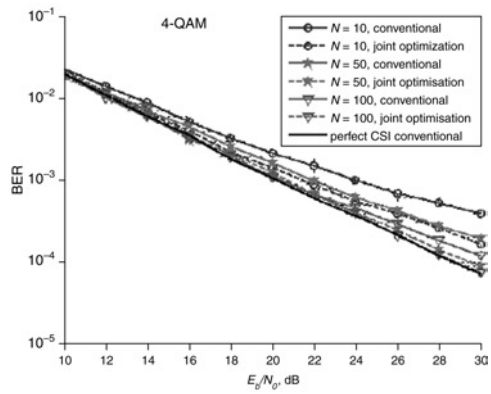

Fig. 2 *Joint optimisation performance with different N*

Simulation results: For simulation purposes, $K = 4$ users with 4-QAM signalling are assumed. The entries of $\boldsymbol{H}$ are assumed to be zero mean IID complex Gaussian random variables, i.e. $\boldsymbol{H} \sim CN(0,1)$. Fig. 2 compares the performance of the proposed joint optimisation with conventional optimisation. As can be realised, the proposed joint optimisation algorithm substantially outperforms the conventional optimisation, over the whole observation data length (N). In fact, the performance is noticeable for smaller Ns, where the channel estimator estimates the channel erroneously, especially for high SNR values.

13 May 2007
Electronics Letters online no: 20071390
doi: 10.1049/el:20071390

H.Khaleghi Bizaki and A. Falahati (*Communication Engineering, Iran University of Science and Technology, University Rd., Tehran Narmak 021, Iran*)

E-mail: h_khaleghi@ee.iust.ac.ir

References

1 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
2 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
3 Liu, J., and Krzymien, W.A.: 'Improved tomlinson harashima precoding for the downlink of multiple antenna multi-user systems'. IEEE Wireless Communications and Networking Conf. (WCNC '05), New Orleans, LA, USA, March 2005, pp. 466–472
4 Bahng, S., Lira, J., Host-Madsen, A., and Wang, X.: 'The effects of channel estimation on Tomlinson-Harashima precoding in TDD MIMO Systems'. IEEE 6th Workshop on Signal Processing Advances in Wireless Communication, New Orleans, LA, USA, 2005, pp. 455–459
5 Kay, S.M.: 'Fundamentals of statistical signal processing: estimation theory' (Prentice Hall, 1993)
6 Mardia, K.V., Kent, J.T., and Bibby, J.M.: 'Multivariate analysis' (Academic Press, Duluth, London, 1979)
7 Papoulis, A., and Pillai, S.U.: 'Probability, random variables and stochastic processes' (McGraw-Hill, 2002, 4th edn.)
8 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