Tomlinson–Harashima Precoding With Partial Channel Knowledge

Athanasios P. Liavas, *Member, IEEE*

***Abstract*—We consider minimum mean-square error Tomlinson–Harashima (MMSE-TH) precoding for time-varying frequency-selective channels. We assume that the receiver estimates the channel and sends the channel state information (CSI) estimate to the transmitter through a lossless feedback channel that introduces a certain delay. Thus, the CSI mismatch at the receiver is due to estimation errors, while the CSI mismatch at the transmitter is due to both estimation errors and channel time variations. We exploit *a priori* statistical channel knowledge, and we derive an optimal TH precoder, adopting a Bayesian approach. We use simulations to compare the performance of the so-derived TH precoder with that of the same-complexity MMSE decision-feedback equalizer (DFE). We observe that for low signal-to-noise ratios (SNRs) and sufficiently slow channel time variations, the optimal TH precoder outperforms the DFE, while at high SNR, the opposite happens.**

***Index Terms*—Intersymbol interference (ISI), partial channel knowledge, Rayeigh fading, Tomlinson–Harashima (TH) precoder.**

I. Introduction

INTERSYMBOL interference (ISI) is a significant obstacle against reliable high-speed digital communication through bandlimited channels. The finite-length minimum mean-square error decision-feedback equalizer (MMSE-DFE) has proven an effective structure for combatting ISI. The design of the MMSE-DFE filters requires the knowledge of the channel state information (CSI) at the receiver (acquired through training). A phenomenon that might degrade the MMSE-DFE performance is catastrophic error propagation. If the CSI is available at the transmitter (through a feedback channel), then the feedback portion of the MMSE-DFE can be designed and implemented at the transmitter, where error propagation is impossible. This structure is known as the MMSE Tomlinson–Harashima (MMSE-TH) precoder [1], [2].

In practice, the CSI estimate available at the transmitter is noisy. Potential noise sources include estimation and/or quantization errors, feedback channel errors, and channel time variations. When the quality of the CSI estimate at the transmitter is poor, the performance of the TH precoder may degrade significantly [3], [4]. In this letter, we consider the case where the receiver estimates the channel using a training sequence and sends the estimate to the transmitter through a feedback channel that introduces a certain delay, but *no* errors. Thus, CSI mismatch at the receiver is due to estimation errors, while CSI mismatch at the transmitter is due to both estimation errors and channel time variations. Using a statistical model for the channel time variations and estimation errors, we derive new cost functions for the design of the TH precoder, adopting a Bayesian approach (for related work in different precoding scenarios, see [5]–[9] and the references therein). We use simulations to compare the performance of the resulting TH precoder with that of the same-complexity (i.e., same filter lengths) MMSE-DFE that exploits the statistical channel model. We observe that, at low signal-to-noise ratios (SNRs) and slow channel time variations, the resulting TH precoder outperforms the MMSE-DFE (this fact may be attributed to the use of erroneous previous decisions by the MMSE-DFE). On the other hand, at high SNR, the MMSE-DFE outperforms the TH precoder.

The rest of the letter is organized as follows. In Section II, we assume perfect CSI knowledge and we recall known results for the MMSE-TH precoder. In Section III, we introduce a statistical channel model, we develop new cost functions, using a Bayesian approach, and we derive the optimum TH precoder. In Section IV, we use simulations to compare the performance of the derived TH precoder with that of MMSE-DFE. Conclusions are drawn in Section V.

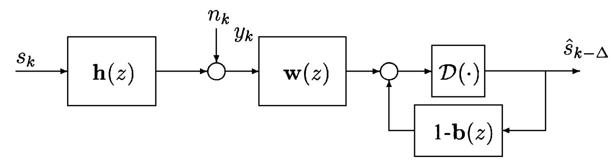

Fig. 1. Channel model and MMSE-DFE.

II. Finite-Length MMSE-TH Precoding With Perfect Channel Knowledge

A. Channel Model

We consider the baseband-equivalent discrete-time noisy communication channel modeled by the νth-order linear time-invariant system depicted in Fig. 1. Its input–output relation is given by the convolution

$$y_k = \sum_{i=0}^{\nu} h_i s_{k-i} + n_k$$

where h_i, $i = 0, \ldots, \nu$ is the discrete-time channel finite impulse response, and s_k, y_k, and n_k are, respectively, samples of the channel input, output, and noise sequences. The channel transfer function is defined as $\mathbf{h}(z) \triangleq \sum_{i=0}^{\nu} h_i z^{-i}$, and the impulse-response vector is defined as $\mathbf{h} \triangleq [h_0 \; \ldots \; h_\nu]^T$, where superscript ^{T} denotes transpose. By stacking N_f successive output samples, we construct the data vector

$$\mathbf{y}_{k:k-N_f+1} \triangleq \left[y_k \; \cdots \; y_{k-N_f+1}\right]^T$$

Paper approved by R. D. Wesel, the Editor for Coding and Communication Theory of the IEEE Communications Society. Manuscript received June 12, 2003; revised March 28, 2004.

The author is with the Department of Electronic and Computer Engineering, Technical University of Crete, Kounoupidiana, 73100 Chania, Greece (e-mail: liavas@telecom.tuc.gr).

Digital Object Identifier 10.1109/TCOMM.2004.840669

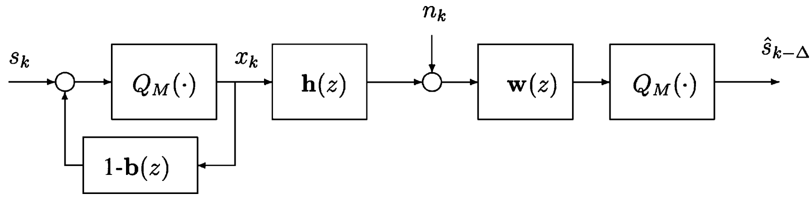

Fig. 2. TH precoding.

which can be expressed as

$$\mathbf{y}_{k:k-N_f+1} = \mathbf{H}\mathbf{s}_{k:k-N_f-\nu+1} + \mathbf{n}_{k:k-N_f+1}$$

where the $N_f \times (\nu + N_f)$ filtering matrix $\mathbf{H}$ is defined as

$$\mathbf{H} \triangleq \begin{bmatrix} h_0 & \cdots & \cdots & h_\nu & & \\ & \ddots & & & \ddots & \\ & & h_0 & \cdots & \cdots & h_\nu \end{bmatrix}$$

and the definitions of $\mathbf{s}_{k:k-N_f-\nu+1}$ and $\mathbf{n}_{k:k-N_f+1}$ are obvious.

B. Finite-Length MMSE-TH Precoding

Our aim is to recover (a delayed version of) the transmitted sequence $s_{k-\Delta}$ for $0 \le \Delta \le N_f + \nu - 1$. A structure that has been widely used for this purpose is the MMSE-DFE, which is also depicted in Fig. 1. The finite-length MMSE-DFE is composed of the following filters:

1) the length-N_f feedforward filter determined by the vector $\mathbf{w}^H \triangleq [w_0^* \cdots w_{N_f-1}^*]$, where ^{H} denotes Hermitian transpose and * denotes complex conjugate. The transfer function of the feedforward filter is $\mathbf{w}(z) \triangleq \sum_{i=0}^{N_f-1} w_i^* z^{-i}$;
2) the length-N_b feedback filter determined by the vector $\mathbf{b}^H \triangleq [1\ b_1^* \cdots b_{N_b}^*]$, with $\mathbf{b}(z) \triangleq 1 + \sum_{i=1}^{N_b} b_i^* z^{-i}$.

The block labeled $\mathcal{D}(\cdot)$ in Fig. 1 represents a symbol decision device.

A problem that might be encountered when using the MMSE-DFE is catastrophic error propagation. If the channel $\mathbf{h}$ is known at the transmitter, then the feedback section may be designed and implemented at the transmitter, resulting in the TH precoder, depicted in Fig. 2. The modulo operation $Q_M(\cdot)$ is defined as $Q_M(v) \triangleq v + a$, where a is the unique integer multiple of M for which $Q_M(v) \in (-M/2, M/2]$. If the input to the modulo operator is complex-valued, then the modulo operation is applied separately to the real and the imaginary parts of the input.

An equivalent structure is shown in Fig. 3, where $1/\mathbf{b}(z)$ denotes the inverse filter of $\mathbf{b}(z)$, $\widetilde{\mathbf{b}}(z) \triangleq z^{-\Delta}\mathbf{b}(z)$, z_k is the ISI term, and $\hat{n}_k$ is noise n_k filtered by the feedforward filter $\mathbf{w}(z)$ [10]. We observe that $\widetilde{\mathbf{b}}(z)$ incorporates a delay of Δ time units. This may be convenient when the channel $\mathbf{h}$ possesses "small" leading terms (a case that is commonly encountered, due to pulse shaping, in bandwidth-efficient systems).

The objective of the MMSE-TH precoder is the minimization of the power of the ISI and the filtered noise terms, $\mathcal{E}[|z_k + \hat{n}_k|^2]$ [10], where $\mathcal{E}[\cdot]$ denotes expectation with respect to the input and the noise. The ISI term is given by

$$z_k = (\mathbf{w}^H\mathbf{H} - \widetilde{\mathbf{b}}^H)\mathbf{x}_{k:k-N_f-\nu+1} \tag{1}$$

where (for notation compactness) we defined $\widetilde{\mathbf{b}} \triangleq [\mathbf{0}_{1\times\Delta}\ \mathbf{b}^T\ \mathbf{0}_{1\times s}]^T$, with $\mathbf{0}_{i\times j}$ denoting the $i \times j$ zero

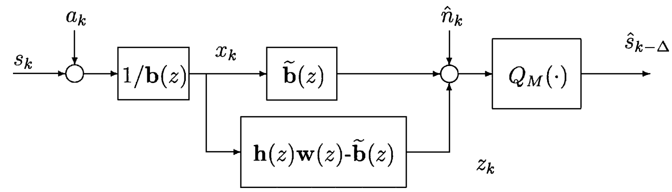

Fig. 3. TH precoding (equivalent structure).

matrix, and $s \triangleq N_f + \nu - \Delta - N_b - 1$. The filtered-noise term is given by

$$\hat{n}_k = \mathbf{w}^H\mathbf{n}_{k:k-N_f+1}.$$

In order to simplify notation in the following, we shall omit the subscripts from $\mathbf{x}_{k:k-N_f-\nu+1}$ and $\mathbf{n}_{k:k-N_f+1}$.

Expanding terms and using the independence of the zero-mean sequences z_k and $\hat{n}_k$, we obtain

$$\begin{aligned}\mathcal{M}(\widetilde{\mathbf{b}}, \mathbf{w}) &\triangleq \mathcal{E}\left[|z_k + \hat{n}_k|^2\right] \\ &= (\mathbf{w}^H\mathbf{H} - \widetilde{\mathbf{b}}^H)\mathbf{R}_{xx}(\mathbf{H}^H\mathbf{w} - \widetilde{\mathbf{b}}) + \mathbf{w}^H\mathbf{R}_{nn}\mathbf{w} \\ &= \widetilde{\mathbf{b}}^H\mathbf{R}_{xx}\widetilde{\mathbf{b}} - \widetilde{\mathbf{b}}^H\mathbf{R}_{xx}\mathbf{H}^H\mathbf{w} - \mathbf{w}^H\mathbf{H}\mathbf{R}_{xx}\widetilde{\mathbf{b}} \\ &\quad + \mathbf{w}^H(\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H + \mathbf{R}_{nn})\mathbf{w}\end{aligned} \tag{2}$$

where $\mathbf{R}_{xx} \triangleq \mathcal{E}[\mathbf{x}\mathbf{x}^H]$ and $\mathbf{R}_{nn} \triangleq \mathcal{E}[\mathbf{n}\mathbf{n}^H]$. This cost function coincides with the cost function for the MMSE-DFE (with $\mathbf{R}_{xx}$ replacing $\mathbf{R}_{ss}$) [11, eq. (11)]. If the input s_k is a sequence of independent, identically distributed (i.i.d.) samples, then a common assumption in the TH precoding literature is that the output of the modulo operator $Q_M(\cdot)$, x_k, is a sequence of independent random variables uniformly distributed in $(-M/2, M/2]$. Assuming that the real and imaginary parts of x_k are independent, we obtain $\mathbf{R}_{xx} = \sigma_x^2\mathbf{I}_{N_f+\nu}$, where $\sigma_x^2 = (M^2/6)$ and $\mathbf{I}_i$ denotes the $i \times i$ identity matrix.

The optimal finite-length MMSE-TH filters, $\mathbf{b}_\text{o}$ and $\mathbf{w}_\text{o}$, can be computed by following steps analogous to those of [11]. More specifically, if we define

$$\mathbf{R} \triangleq \mathbf{R}_{xx} - \mathbf{R}_{xx}\mathbf{H}^H(\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H + \mathbf{R}_{nn})^{-1}\mathbf{H}\mathbf{R}_{xx}$$

and

$$\mathbf{R}_\Delta \triangleq \mathbf{I}_{N_b+1}^{\Delta,s}\mathbf{R}\left(\mathbf{I}_{N_b+1}^{\Delta,s}\right)^T$$

where

$$\mathbf{I}_{N_b+1}^{\Delta,s} \triangleq \left[\mathbf{0}_{(N_b+1)\times\Delta}\ \mathbf{I}_{N_b+1}\ \mathbf{0}_{(N_b+1)\times s}\right]$$

then the optimal filters are given by [11]

$$\mathbf{b}_\text{o} = \frac{\mathbf{R}_\Delta^{-1}\mathbf{e}_1}{\mathbf{e}_1^H\mathbf{R}_\Delta^{-1}\mathbf{e}_1}, \quad \mathbf{w}_\text{o} = (\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H + \mathbf{R}_{nn})^{-1}\mathbf{H}\mathbf{R}_{xx}\widetilde{\mathbf{b}}_\text{o}$$

where $\mathbf{e}_1$ is the vector with 1 at its first position, and 0 elsewhere.

III. MMSE-TH Precoding With Partial Channel Knowledge

In the previous section, we assumed exact CSI at both the receiver and the transmitter. However, this setting seems unrealistic, especially in time-varying environments. In this section, we assume that the receiver possesses an estimate $\hat{\mathbf{h}}$ (acquired through training) of the true (unknown) current channel $\mathbf{h}$, while the transmitter possesses an estimate $\hat{\mathbf{h}}_\tau$ of the true (unknown) channel τ seconds before $\mathbf{h}$, $\mathbf{h}_\tau$ (τ is the time needed for the feedback of the channel estimate from the receiver to the transmitter). Furthermore, we assume that the feedback channel does *not* introduce errors, due to powerful error-control coding (the lossless feedback channel assumption is essential for the satisfactory performance of the derived structures, and is common in all works [5]–[9] that exploit partial channel information at the transmitter). Consequently, the receiver knows, and can use, both $\hat{\mathbf{h}}$ and $\hat{\mathbf{h}}_\tau$.

Thus, due to estimation errors, the receiver possesses a (hopefully, slightly) erroneous CSI. On the other hand, due to estimation errors, channel time variations, and feedback delay, the transmitter possesses a (hopefully, slightly) erroneous estimate of a (hopefully, slightly) outdated CSI.

A. Statistical Channel Model

In this subsection, we provide statistical models for the true and outdated channel and their estimates. More specifically, we assume the following.

1) The true channels $\mathbf{h}$ and $\mathbf{h}_\tau$ are frequency-selective Rayleigh fading, drawn from the same statistical distribution. Their taps are modeled as independent zero-mean complex Gaussian random variables, i.e.,

$$\mathbf{h}, \mathbf{h}_\tau \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma_h}), \quad \mathbf{\Sigma_h} \triangleq \mathrm{diag}\left(\sigma^2_{h_0}, \ldots, \sigma^2_{h_\nu}\right)$$

with the $\sigma^2_{h_i}$ determined by the channel power-delay profile. The channel taps are time varying, according to Jakes' model, with *common* maximum Doppler frequency f_d. Since $\mathbf{h}_\tau$ is the channel realization τ seconds before $\mathbf{h}$, $\mathbf{h}$ and $\mathbf{h}_\tau$ can be modeled as jointly Gaussian with their joint statistics described by the cross-correlation matrix

$$\mathcal{E}\left[\mathbf{h}\mathbf{h}_\tau^H\right] = \rho\Sigma_{\mathbf{h}}.$$

ρ is the normalized correlation coefficient specified by the Jakes model, $\rho = J_0(2\pi f_d \tau)$, where $J_0(x)$ is the zeroth-order Bessel function of the first kind. Thus, we may model the channel time variations as follows:

$$\mathbf{h} = \rho\mathbf{h}_\tau + \boldsymbol{\zeta}$$

where $\boldsymbol{\zeta}$ is independent of $\mathbf{h}$ and $\mathbf{h}_\tau$, and it can be easily seen that it obeys the statistical model

$$\boldsymbol{\zeta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{\zeta}}), \quad \mathbf{\Sigma}_{\boldsymbol{\zeta}} \triangleq \left(1 - |\rho|^2\right)\Sigma_{\mathbf{h}}.$$

2) The channel estimates $\hat{\mathbf{h}}$ and $\hat{\mathbf{h}}_\tau$ are maximum-likelihood (ML) estimates of $\mathbf{h}$ and $\mathbf{h}_\tau$, respectively, acquired through training, and can be expressed as follows:

$$\hat{\mathbf{h}} = \mathbf{h} + \boldsymbol{\epsilon}_{\mathbf{h}}, \quad \hat{\mathbf{h}}_\tau = \mathbf{h}_\tau + \boldsymbol{\epsilon}_{\mathbf{h}_\tau}.$$

The estimation errors $\boldsymbol{\epsilon}_{\mathbf{h}}$ and $\boldsymbol{\epsilon}_{\mathbf{h}_\tau}$ are independent of all other stochastic quantities, and are drawn from the same statistical distribution

$$\boldsymbol{\epsilon}_{\mathbf{h}}, \boldsymbol{\epsilon}_{\mathbf{h}_\tau} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{\epsilon}}).$$

It is well known in the ML estimation literature [12, p. 786] that using an ideal unit-power length-N_{tr} training sequence to estimate a channel with order ν, when the channel noise is additive zero-mean white Gaussian with variance σ_n^2, yields $\mathbf{\Sigma}_{\boldsymbol{\epsilon}} = (\sigma_n^2/(N_{\mathrm{tr}} - \nu))\mathbf{I}_{\nu+1}$ ($N_{\mathrm{tr}} - \nu$ is the number of equations we use to solve the resulting least-squares problem). Under the above assumptions, we obtain

$$\hat{\mathbf{h}}, \hat{\mathbf{h}}_\tau \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma_h} + \mathbf{\Sigma}_{\boldsymbol{\epsilon}}), \ \mathcal{E}[\mathbf{h}\hat{\mathbf{h}}^H] = \mathbf{\Sigma_h}$$
$$\mathcal{E}\left[\mathbf{h}\hat{\mathbf{h}}_\tau^H\right] = \rho\mathbf{\Sigma_h}.$$

B. The New TH Structure

The transmitter, possessing $\hat{\mathbf{h}}_\tau$, exploits the above statistical channel description that relates its channel estimate $\hat{\mathbf{h}}_\tau$ with the true channel $\mathbf{h}$, by minimizing the conditional expectation

$$\mathcal{M}_{\mathrm{tr}}(\widetilde{\mathbf{b}}, \mathbf{w}) \triangleq \mathcal{E}\left[\mathcal{M}(\widetilde{\mathbf{b}}, \mathbf{w})|\hat{\mathbf{h}}_\tau\right].$$

In order to compute the above quantity, we need to compute terms $\mathcal{E}[\mathbf{H}|\hat{\mathbf{h}}_\tau]$ and $\mathcal{E}[\mathbf{H}\mathbf{H}^H|\hat{\mathbf{h}}_\tau]$ (recall the definition of $\mathcal{M}(\widetilde{\mathbf{b}}, \mathbf{w})$ in (2), and that $\mathbf{R}_{xx} = \sigma_x^2\mathbf{I}_{N_f+\nu}$). We start by computing the conditional expectation (MMSE estimate) of $\mathbf{h}$, denoted $\mathbf{h}_{\mathrm{tr}}$ [13, p. 324]

$$\mathbf{h}_{\mathrm{tr}} \triangleq \mathcal{E}[\mathbf{h}|\hat{\mathbf{h}}_\tau] = \mathcal{E}\left[\mathbf{h}\hat{\mathbf{h}}_\tau^H\right]\mathcal{E}\left[\hat{\mathbf{h}}_\tau\hat{\mathbf{h}}_\tau^H\right]^{-1}\hat{\mathbf{h}}_\tau$$
$$= \rho\mathbf{\Sigma_h}(\mathbf{\Sigma_h} + \mathbf{\Sigma}_{\boldsymbol{\epsilon}})^{-1}\hat{\mathbf{h}}_\tau.$$

The corresponding (diagonal) covariance matrix is given by

$$\mathbf{\Sigma}_{\mathbf{h}_{\mathrm{tr}}} = \mathcal{E}[\mathbf{h}\mathbf{h}^H] - \mathcal{E}\left[\mathbf{h}\hat{\mathbf{h}}_\tau^H\right]\mathcal{E}\left[\hat{\mathbf{h}}_\tau\hat{\mathbf{h}}_\tau^H\right]^{-1}\mathcal{E}[\hat{\mathbf{h}}_\tau\mathbf{h}^H]$$
$$= \mathbf{\Sigma_h} - |\rho|^2\mathbf{\Sigma_h}(\mathbf{\Sigma_h} + \mathbf{\Sigma}_{\boldsymbol{\epsilon}})^{-1}\mathbf{\Sigma}_{\mathbf{h}}^H.$$

Thus, if $h_{\mathrm{tr},i}$ and $\Sigma_{\mathbf{h}_{\mathrm{tr}},i}$, $i = 0, \ldots, \nu$, denote, respectively, the elements of $\mathbf{h}_{\mathrm{tr}}$ and the diagonal elements of $\mathbf{\Sigma}_{\mathbf{h}_{\mathrm{tr}}}$, then

$$\mathcal{E}\left[h_i h_j^*|\hat{\mathbf{h}}_\tau\right] = \begin{cases} |h_{\mathrm{tr},i}|^2 + \Sigma_{\mathbf{h}_{\mathrm{tr}},i}, & \text{if } i = j \\ h_{\mathrm{tr},i}h_{\mathrm{tr},j}^*, & \text{otherwise.} \end{cases}$$

Using the above relations, we obtain

$$\mathcal{E}[\mathbf{H}|\hat{\mathbf{h}}_\tau] = \mathbf{H}_{\mathrm{tr}}$$

and

$$\mathcal{E}[\mathbf{H}\mathbf{H}^H|\hat{\mathbf{h}}_\tau] = \mathbf{H}_{\mathrm{tr}}\mathbf{H}_{\mathrm{tr}}^H + \mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{h}_{\mathrm{tr}}}\right)\mathbf{I}_{N_f}$$

where $\mathbf{H}_{\mathrm{tr}}$ is the filtering matrix constructed from $\mathbf{h}_{\mathrm{tr}}$, and $\mathrm{tr}(\cdot)$ denotes the trace of the matrix argument.

The new cost function at the transmitter is expressed as

$$\mathcal{M}_{\mathrm{tr}}(\widetilde{\mathbf{b}}, \mathbf{w}) = \sigma_x^2\left(\widetilde{\mathbf{b}}^H\widetilde{\mathbf{b}} - \widetilde{\mathbf{b}}^H\mathbf{H}_{\mathrm{tr}}^H\mathbf{w} - \mathbf{w}^H\mathbf{H}_{\mathrm{tr}}\widetilde{\mathbf{b}}\right)$$
$$+\mathbf{w}^H\left(\sigma_x^2\mathbf{H}_{\mathrm{tr}}\mathbf{H}_{\mathrm{tr}}^H + \sigma_x^2\mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{h}_{\mathrm{tr}}}\right)\mathbf{I}_{N_f} + \mathbf{R}_{nn}\right)\mathbf{w}.$$

We observe that in the new cost function, the channel matrix $\mathbf{H}$ has been replaced by $\mathbf{H}_{\mathrm{tr}}$, and the extra term $\sigma_x^2\mathrm{tr}(\mathbf{\Sigma}_{\mathbf{h}_{\mathrm{tr}}})\mathbf{w}^H\mathbf{w}$ has appeared, due to channel uncertainties. For perfect CSI knowledge, $\mathcal{M}(\widetilde{\mathbf{b}}, \mathbf{w})$ and $\mathcal{M}_{\mathrm{tr}}(\widetilde{\mathbf{b}}, \mathbf{w})$, of course, coincide.

The optimal feedback filter $\mathbf{b}_{\mathrm{tr}}$ is computed at the transmitter by minimizing $\mathcal{M}_{\mathrm{tr}}(\widetilde{\mathbf{b}}, \mathbf{w})$, following steps analogous to the ones of Section II-B.

The receiver knows $\mathbf{b}_{\rm tr}$ and exploits the statistical channel knowledge by optimizing the cost function

$$\mathcal{M}_{\rm rec}(\mathbf{w}) \triangleq \mathcal{E}\left[\mathcal{M}(\widetilde{\mathbf{b}}_{\rm tr}, \mathbf{w})|\hat{\mathbf{h}}\right]$$

[$\widetilde{\mathbf{b}}_{\rm tr}$ is $\mathbf{b}_{\rm tr}$ appropriately zero-padded; recall the definition of $\widetilde{\mathbf{b}}$ in terms of $\mathbf{b}$ after (1)] which is expressed as

$$\mathcal{M}_{\rm rec}(\mathbf{w}) = \sigma_x^2\left(\widetilde{\mathbf{b}}_{\rm tr}^H\widetilde{\mathbf{b}}_{\rm tr} - \widetilde{\mathbf{b}}_{\rm tr}^H\mathbf{H}_{\rm rec}^H\mathbf{w} - \mathbf{w}^H\mathbf{H}_{\rm rec}\widetilde{\mathbf{b}}_{\rm tr}\right) + \mathbf{w}^H\left(\sigma_x^2\mathbf{H}_{\rm rec}\mathbf{H}_{\rm rec}^H + \sigma_x^2{\rm tr}\left(\mathbf{\Sigma}_{\mathbf{h}_{\rm rec}}\right)\mathbf{I}_{N_f} + \mathbf{R}_{nn}\right)\mathbf{w}$$

where $\mathbf{H}_{\rm rec}$ is the filtering matrix constructed from

$$\mathbf{h}_{\rm rec} = \mathcal{E}[\mathbf{h}|\hat{\mathbf{h}}] = \mathbf{\Sigma}_{\mathbf{h}}(\mathbf{\Sigma}_{\mathbf{h}} + \mathbf{\Sigma}_{\boldsymbol{\epsilon}})^{-1}\hat{\mathbf{h}}$$

and

$$\mathbf{\Sigma}_{\mathbf{h}_{\rm rec}} = \mathbf{\Sigma}_{\mathbf{h}} - \mathbf{\Sigma}_{\mathbf{h}}(\mathbf{\Sigma}_{\mathbf{h}} + \mathbf{\Sigma}_{\boldsymbol{\epsilon}})^{-1}\mathbf{\Sigma}_{\mathbf{h}}^H.$$

The optimal feedforward filter is given by

$$\mathbf{w}_{\rm rec} = \sigma_x^2\left(\sigma_x^2\mathbf{H}_{\rm rec}\mathbf{H}_{\rm rec}^H + \sigma_x^2{\rm tr}\left(\mathbf{\Sigma}_{\mathbf{h}_{\rm rec}}\right)\mathbf{I}_{N_f} + \mathbf{R}_{nn}\right)^{-1} \times\mathbf{H}_{\rm rec}\widetilde{\mathbf{b}}_{\rm tr}.$$

IV. Simulations

In this section, we compare the TH precoder derived in the previous section (termed robust TH precoder) with the same-complexity (i.e., same filter lengths) TH precoder that does not exploit *a priori* statistical channel knowledge (termed conventional TH precoder), and the same-complexity MMSE-DFE that exploits the statistical channel model.

For the conventional TH precoder, the transmitter considers $\hat{\mathbf{h}}_\tau$ as a perfect estimate of the current channel and, following steps like those of Section II-B, optimizes the cost function

$$\widehat{\mathcal{M}}_{\rm tr}(\widetilde{\mathbf{b}}, \mathbf{w}) \triangleq \sigma_x^2\left(\widetilde{\mathbf{b}}^H\widetilde{\mathbf{b}} - \widetilde{\mathbf{b}}^H\widehat{\mathbf{H}}_\tau^H\mathbf{w} - \mathbf{w}^H\widehat{\mathbf{H}}_\tau\widetilde{\mathbf{b}}\right) + \mathbf{w}^H\left(\sigma_x^2\widehat{\mathbf{H}}_\tau\widehat{\mathbf{H}}_\tau^H + \mathbf{R}_{nn}\right)\mathbf{w} \quad (3)$$

with respect to $\mathbf{b}$, obtaining $\hat{\mathbf{b}}_{\rm tr}$. The receiver knows $\hat{\mathbf{b}}_{\rm tr}$, and uses its channel estimate $\hat{\mathbf{h}}$ to compute the optimum feedforward filter as

$$\hat{\mathbf{w}}_{\rm rec} = \sigma_x^2\left(\sigma_x^2\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H + \mathbf{R}_{nn}\right)^{-1}\widehat{\mathbf{H}}\widetilde{\hat{\mathbf{b}}}_{\rm tr}.$$

The DFE filters are computed by optimizing, with respect to $\widetilde{\mathbf{b}}$ and $\mathbf{w}$, the conditional mean of the mean-square error (as defined in (2), but with $\mathbf{R}_{ss}$ replacing $\mathbf{R}_{xx}$), given $\hat{\mathbf{h}}$.

In our simulations, we consider a packet-based system with packet length $N = 126$ data samples, containing $N_{\rm tr} = 26$ training samples. With each packet, we associate a channel realization $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{h}})$ with order $\nu = 6$, obeying the exponential power-delay profile

$$\sigma_{h_k}^2 = e^{-\frac{k}{2}}, \quad k = 0, \ldots, \nu$$

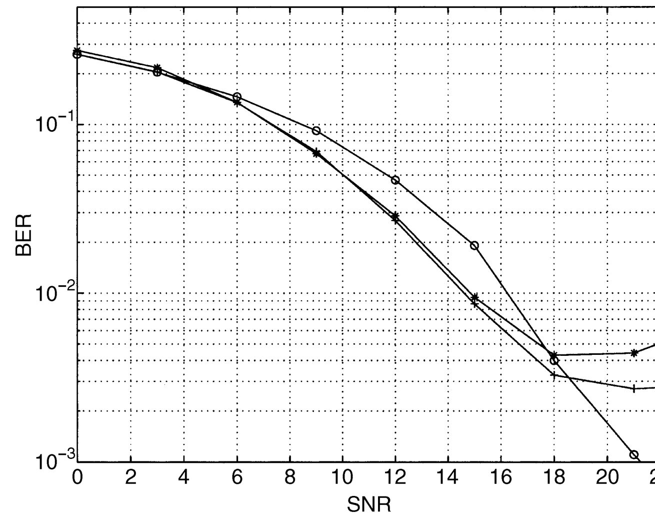

Fig. 4. BER versus SNR for MMSE-DFE (o-), robust TH precoder (+-), and conventional TH precoder (∗-), for $\rho = 0.99$.

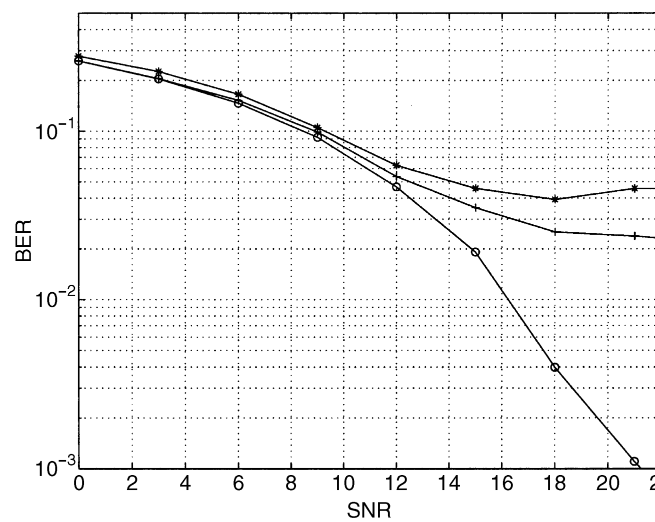

Fig. 5. BER versus SNR for MMSE-DFE (o-), robust TH precoder (+-), and conventional TH precoder (∗-), for $\rho = 0.94$.

and realizations of $\boldsymbol{\zeta} \sim \mathcal{CN}(\mathbf{0}, (1 - |\rho|^2)\mathbf{\Sigma}_{\mathbf{h}})$ and $\boldsymbol{\epsilon}_{\mathbf{h}}$, $\boldsymbol{\epsilon}_{\mathbf{h}_\tau} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{\epsilon}})$. $\mathbf{h}_\tau$, $\hat{\mathbf{h}}$, and $\hat{\mathbf{h}}_\tau$ are constructed from the above quantities, as indicated in Section III-A.

The input s_k is a 4-quadrature amplitude modulation (QAM) i.i.d. sequence, taking with equal probability the values $\pm 1 \pm j$. The SNR is defined as

$${\rm SNR} = 10\log_{10}\left(\frac{\sum_{i=0}^{\nu}|h_i|^2\mathcal{E}\left[|s_k|^2\right]}{\mathcal{E}\left[|n_k|^2\right]}\right)$$

and is the actual signal-power-to-noise-power ratio at the output of the channel for the MMSE-DFE. This does not apply to the TH precoder, because the channel input in this case, x_k, has larger power than s_k (in fact, $\mathcal{E}[|s_k|^2]/\mathcal{E}[|x_k|^2] = 0.75$). We consider delay $\Delta = 2$ and filter lengths $N_f = 5$ and $N_b = 7$.

In Figs. 4–6, we plot the bit-error rate (BER) versus SNR of the robust and the conventional TH precoders and the MMSE-DFE, for channel correlation coefficient $\rho = 0.99$, 0.94, and 0.85, respectively. Our observations are as follows.

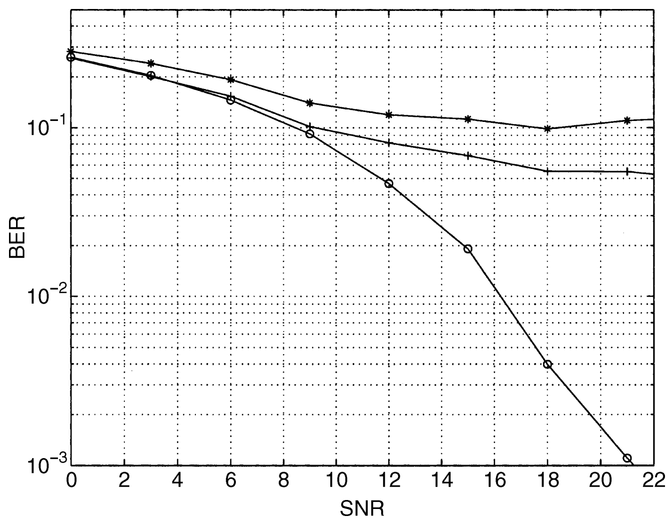

Fig. 6. BER versus SNR for MMSE-DFE (o-), robust TH precoder (+-), and conventional TH precoder ($*$-), for $\rho = 0.85$.

1) The robust TH precoder always outperforms the conventional TH precoder, with the performance difference increasing for increasing the speed of the channel time variations and the SNR. This is an intuitively satisfying observation that supports the exploitation of statistical channel knowledge whenever it is available. We also observe that both TH precoders exhibit an irreducible error floor, directly related to the quality of the CSI at the transmitter.
2) For very slow time variations ($\rho = 0.99$), i.e., very good CSI quality at the transmitter, the robust TH precoder outperforms the MMSE-DFE for low and medium SNRs (in this case, up to 18 dB). The range of SNRs where the robust TH precoder outperforms the MMSE-DFE decreases for increasing the speed of channel time variations, i.e., increasing the degradation of CSI at the transmitter. On the other hand, for high SNR and moderate or fast channel time variations, the MMSE-DFE significantly outperforms both TH precoders.

V. Conclusions

We considered TH precoding with partial channel knowledge. The CSI mismatch at the receiver was due to estimation errors, while the CSI mismatch at the transmitter was due to estimation errors and channel time variations. We designed a robust TH precoder by exploiting a statistical model for the channel time variations and estimation errors. We used simulations to compare the performance of the derived TH precoder with that of the same-complexity MMSE-DFE. We observed that for very slow channel time variations, the robust TH precoder outperforms the MMSE-DFE for low and moderate SNR. On the other hand, for high SNR, the MMSE-DFE significantly outperforms the TH precoder.

References

[1] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, no. 5/6, pp. 138–139, Mar. 1971.
[2] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. COM-20, pp. 774–779, Aug. 1972.
[3] W. Shi and R. Wesel, "The effect of mismatch on desicion-feedback equalization and Tomlinson–Harashima precoding," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, vol. 2, 1998, pp. 1743–1747.
[4] J. E. Smee and S. C. Schwartz, "Adaptive compensation techniques for communications systems with Tomlinson–Harashima precoding," *IEEE Trans. Commun.*, vol. 51, pp. 865–869, Jun. 2003.
[5] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.
[6] E. Visotsky and U. Madhow, "Space–time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2632–2639, Sep. 2001.
[7] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming with orthogonal space–time block coding," *IEEE Trans. Inf. Theory*, vol. 48, pp. 611–627, Mar. 2002.
[8] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space–time block coding based on channel mean feedback," *IEEE Trans. Signal Process.*, vol. 50, pp. 2599–2613, Oct. 2002.
[9] F. Rey, M. Lamarca, and G. Vázquez, "Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI," *IEEE Trans. Signal Process.*, to be published.
[10] R. Wesel and J. Cioffi, "Achievable rates for Tomlinson–Harashima precoding," *IEEE Trans. Inf. Theory*, vol. 44, pp. 824–831, Mar. 1998.
[11] N. Al-Dhahir and J. Cioffi, "MMSE decision-feedback equalizers: Finite-length results," *IEEE Trans. Inf. Theory*, vol. 41, pp. 961–975, Jul. 1995.
[12] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing*. New York: Wiley, 1998.
[13] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.