Adaptive channel aided decision feedback equalisation for SISO and MIMO systems

Y. Lee and W.-R. Wu

Abstract: Error propagation can seriously affect the performance of an adaptive decision feedback equaliser (DFE), especially when operated in time-varying channel environments. Error propagation not only affects DFE decisions, but also disturbs the DFE adaptation. The paper focuses on improving the robustness against error propagation for the least-mean-square (LMS) based minimum mean-squared-error DFE (MMSE-DFE). A specifically designed channel estimator is introduced to help the DFE adaptation in the decision-directed (DD) mode. Unlike the conventional DFE, the proposed adaptive channel-aided DFE (ACA-DFE) only adapts the feedforward filter with the LMS algorithm. The feedback filter, however, is obtained from the postcursors of the estimated channel convolved with the feedforward filter. As a result, the proposed ACA-DFE can reduce the error propagation effect and perform better than the conventional adaptive DFE. We also demonstrate that the ACA-DFE can be extended to multiple-input multiple-out (MIMO) systems improving the performance of the conventional MIMO DFE.

1 Introduction

Decision feedback equaliser (DFE) is a well-known channel equaliser in single-input single-output (SISO) systems [1–3]. It has been widely used in digital communications to suppress inter-symbol interference (ISI) for over several decades. When the channel spectrum exhibits spectral nulls due to multipath propagation, the DFE performs significantly better than the linear equaliser (LE). Though the maximum likelihood sequence estimator (MLSE) [4] can have better performance than the DFE, the computational complexity is much more higher.

A DFE incorporates a feedforward filter (FFF) operating on the received signal to suppress precursor ISI, and a feedback filter (FBF) operating on previously detected symbols to suppress postcursor ISI. A DFE uses a nonlinear decision device at the output, and the output represents a noise-free replica of the transmitted symbol assuming that the probability of decision error is small. However, if a symbol is detected incorrectly, the next input to the FBF will be in error. As this error advances through the feedback loop, the probability of making an error in the detection of subsequent symbols will be increased. It can result in error propagation (EP) that causes bursts of incorrect decisions and a corresponding increase in the decision-error rate [5]. A number of schemes were proposed to reduce EP for DFE. A technique combining DFE with partial response precoding and detection was presented in [6]. In [7–9], soft decisions and specifically designed constraints were suggested to prevent questionable decisions from being used in the feedback loop and thereby the probability of error burst was reduced. Besides, a periodic transmission of a short resetting sequence calculated based on a certain steady-state error probability was used to control the error behaviour of DFE [10].

Recently, much attention is paid in the development of multiple-input multiple-out (MIMO) systems. With the use of multiple antennas at both transmitter and receiver, the spectral efficiency of a communication system can be increased dramatically [11]. For high data-rate transmission, frequency selective fading is present between pairs of transmit and receive antennas. This brings a great design challenge at MIMO receivers. One solution for this problem is to use an MIMO DFE, where both the FFF and the FBF are extended to have multiple inputs and multiple outputs, i.e., multi-dimensional FFF and multi-dimensional FBF performing multi-dimensional channel equalisation [12–14]. For the MIMO DFE, the problem of EP is even more severe than that of its SISO companion owing to the complicated channel configuration and the need to detect signal buried in ISI in addition to co-channel interference (CCI), in addition to noise.

Since the communication environment may be time-varying, tap weights in the DFE should be updated dynamically for better performance [15]. The least-mean-square (LMS) adaptive algorithm [16] is well-known for its simplicity and robustness, and is often utilised to adapt both the FFF and the FBF in SISO DFE systems. It can be shown that the LMS algorithm is also attractive to the adaptive MIMO DFE for dispersive MIMO channels [17, 18]. As described, the EP effect will have a greater impact in the adaptive implementation of the DFE. A decision error not only affects the DFE future outputs, but also disturbs the reference signal of the adaptive algorithm. As a result, the DFE will be adapted toward an incorrect direction. In the worst case, EP can diverge the DFE adaptation.

The most popular design strategy for channel equalisation by far is the use of the minimum mean-squared-error (MMSE) criterion. Its well-accepted theoretical framework

IEE Proceedings online no. 20050569

doi:10.1049/ip-com:20050569

Paper first received 17th October 2005 and in revised form 17th February 2006

Y. Lee is with the Graduate Institute of Communication Engineering, National Chi Nan University, Nantou Hsien 545, Taiwan, Republic of China

W.-R. Wu is with the Department of Communication Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, Republic of China

E-mail: ymlee@ncnu.edu.tw

and amenability to adaptive implementation make it very attractive for practical usage. Another strategy for equaliser design is to use the minimum bit-error rate/minimum symbol-error rate (MBER/MSER) criterion [19]. Various adaptive realisations for the MBER/MSER equalisation were proposed in [20–23]. Though better results can be obtained in terms of this criterion, there is no guarantee that the global minimum can be reached. In addition, the convergence rate may be slower and the computational complexity may be higher. All these may make the MBER/MSER equaliser less effective in time-varying channel environments. In the following, we only consider the DFE optimised by the MMSE criterion.

In this paper, an LMS-based MMSE-DFE is proposed to reduce the EP effect. A particularly designed channel estimator is introduced to the conventional DFE structure. The resultant adaptive channel-aided (ACA) DFE can perform better than the conventional adaptive DFE and the EP effect can be effectively reduced. This approach is different from those channel-estimation-based DFEs proposed in [24, 25], where both the FFF and the FBF are calculated based on the estimated channel response. Since matrix multiplications and inversions are involved, its computational complexity will be high for time-varying channels. In the proposed ACA-DFE, however, the adaptive structure is remained. Only the FFF is adapted with the LMS algorithm, and the FBF is obtained from the postcursors of the up-to-date estimated channel convolved with the FFF. Generally, this will result in lower computational complexity. We will also show that our SISO ACA-DFE can be extended to an MIMO ACA-DFE.

This paper is organised as follows. In Section 2, the background materials for the MMSE-DFE for both SISO and MIMO systems are described. In Section 3, we propose the new ACA-DFE and explain its operation mechanisms. This result can be extended to use in MIMO channels resulting an MIMO ACA-DFE. Finally, simulation results and conclusions are presented in Section 4 and 5, respectively. Throughout the paper, we utilise the superscripts $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ to denote conjugation, transposition, and Hermitian transposition, respectively, and the operator $\mathrm{E}\{\cdot\}$ to denote mathematical expectation.

2 Background

2.1 Conventional DFE for SISO systems

Let the tap weights of the FFF and the FBF of a DFE be denoted by the column vectors $\boldsymbol{f}$ with length α and $\boldsymbol{b}$ with length β, respectively. The complex dispersive channel is modelled by discrete path h^l with $0 \le l \le L-1$, in which L is the channel order. We assume that the transmitted symbol $a(k)$ is randomly generated and the noise sample sequence $n(k)$ is zero mean, white, and Gaussian distributed. The received discrete-time equivalent baseband signal at the kth time instant can then be modelled as

$$x(k) = \sum_{l=0}^{L-1} h^l a(k-l) + n(k) = \boldsymbol{h}^T \boldsymbol{a}(k) + n(k) \tag{1}$$

with $\boldsymbol{h} = [h^0 \ h^1 \cdots h^{L-1}]^T$ and $\boldsymbol{a}(k) = [a(k) \ a(k-1) \cdots a(k-L+1)]^T$. Let $\boldsymbol{x}(k)$ be the input vector of the FFF with length α, i.e. $\boldsymbol{x}(k) = [x(k) \ x(k-1) \cdots x(k-\alpha+1)]^T$, and $\hat{\boldsymbol{a}}(k)$ be the input vector of the FBF with length β, i.e. $\hat{\boldsymbol{a}}(k) = [\hat{a}(k-\kappa-1) \ \hat{a}(k-\kappa-2) \cdots \hat{a}(k-\kappa-\beta)]^T$, where κ is a suitably chosen decision delay. For the training based MMSE-DFE, the error signal can then be written as

$$e(k) = a(k-\kappa) - (\boldsymbol{f}^H \boldsymbol{x}(k) - \boldsymbol{b}^H \hat{\boldsymbol{a}}(k)) \tag{2}$$

Assuming decisions are correct, i.e., $\hat{a}(k) = a(k)$, we can write the MSE as

$$\begin{aligned}\mathrm{E}\{|e(k)|^2\} &= \boldsymbol{f}^H \boldsymbol{R}_{xx} \boldsymbol{f} - \boldsymbol{f}^H \boldsymbol{R}_{xa} \boldsymbol{b} - \boldsymbol{f}^H \boldsymbol{p}_{xa} - \boldsymbol{b}^H \boldsymbol{R}_{xa}^H \boldsymbol{f} \\ &\quad + \boldsymbol{b}^H \boldsymbol{R}_{aa} \boldsymbol{b} + \boldsymbol{b}^H \boldsymbol{p}_{aa} - \boldsymbol{p}_{xa}^H \boldsymbol{f} + \boldsymbol{p}_{aa}^H \boldsymbol{b} + \sigma_a^2\end{aligned} \tag{3}$$

with $\boldsymbol{R}_{xx} = \mathrm{E}\{\boldsymbol{x}(k)\boldsymbol{x}^H(k)\}$, $\boldsymbol{R}_{aa} = \mathrm{E}\{\boldsymbol{a}(k)\boldsymbol{a}^H(k)\}$, $\boldsymbol{R}_{xa} = \mathrm{E}\{\boldsymbol{x}(k)\boldsymbol{a}^H(k)\}$, $\boldsymbol{p}_{xa} = \mathrm{E}\{\boldsymbol{x}(k)a^*(k-\kappa)\}$, $\boldsymbol{p}_{aa} = \mathrm{E}\{\boldsymbol{a}(k)a^*(k-\kappa)\}$, and $\sigma_a^2 = \mathrm{E}\{a(k-\kappa)a^*(k-\kappa)\}$. To obtain the optimum solution, we set the gradient of $\mathrm{E}\{|e(k)|^2\}$ with respect to $\boldsymbol{f}^*$ and $\boldsymbol{b}^*$ to zero. This results in

$$\boldsymbol{f}_{\text{opt}} = \left(\boldsymbol{R}_{xx} - \frac{1}{\sigma_a^2} \boldsymbol{R}_{xa} \boldsymbol{R}_{xa}^H\right)^{-1} \boldsymbol{p}_{xa} \tag{4}$$

$$\boldsymbol{b}_{\text{opt}} = \frac{1}{\sigma_a^2} \boldsymbol{R}_{xa}^H \boldsymbol{f}_{\text{opt}} \tag{5}$$

As we can see, the optimum solution relies on the correlation matrices which cannot be known in advance, and the matrix inverse operation in (4) requires extensive computation. A simple alternative to find the optimum tap weights is to use an adaptive training method. The LMS algorithm is known to be a simple yet effective choice. The LMS update equations for $\boldsymbol{f}$ and $\boldsymbol{b}$ are expressed as [16]

$$\boldsymbol{f}(k+1) = \boldsymbol{f}(k) + \mu_{\mathrm{f}} \boldsymbol{x}(k) e^*(k) \tag{6}$$

$$\boldsymbol{b}(k+1) = \boldsymbol{b}(k) - \mu_{\mathrm{b}} \hat{\boldsymbol{a}}(k) e^*(k) \tag{7}$$

where $\boldsymbol{f}(k)$ and $\boldsymbol{b}(k)$ are the estimates of $\boldsymbol{f}_{\text{opt}}$ and $\boldsymbol{b}_{\text{opt}}$ at the kth time instant, μ_{f} and μ_{b} are the step sizes controlling the convergence rate, and $e(k)$ is the error signal given in (2). A typical adaptation process consists of a training mode and a decision-directed (DD) mode. Initially, the training mode is launched and sufficient training symbols are transmitted to let both $\boldsymbol{f}(k)$ and $\boldsymbol{b}(k)$ converge around the optimum. Then, the DFE switches to the DD mode in which DFE decisions are used as the reference signal and the DFE is continuously adapted. However, DFE decisions may not be always reliable, especially in time-varying channels. Decision errors not only affect the DFE future output, but also disturb the DFE adaptation. In the worst case, the adaptive DFE can diverge, and another training period needs to be re-initiated.

2.2 Conventional DFE for MIMO systems

The SISO DFE can be extended to an MIMO DFE for the equalisation of MIMO channels. Here, we use $M \times N$ to signify the configuration with M transmit and N receive antennas, and L to indicate the maximum order for the multi-dimensional channel. Generally, $M \le N$ is assumed. A sequence of data symbols $a_m(k)$ $(1 \le m \le M)$ is transmitted from the mth antenna. We define $\boldsymbol{a}_m(k) = [a_m(k) \ a_m(k-1) \cdots a_m(k-L+1)]^T$ as a collection of L successive data symbols from the mth antenna. These data symbols are randomly generated (both in time and space domain) and drawn from the same signal constellation with a variance of σ_a^2. All M data sequences are transmitted over the MIMO channel. The sampled channel response from the mth transmit antenna to the nth receive antenna is given by

$$\boldsymbol{h}_{nm} = [h_{nm}^0 \quad h_{nm}^1 \cdots h_{nm}^{L-1}]^T \tag{8}$$

for $m = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$. We can assemble the vectors $\boldsymbol{h}_{nm}$ into a matrix of size $L \times N$ as

$$\boldsymbol{H}_m = [\boldsymbol{h}_{1m} \quad \boldsymbol{h}_{2m} \cdots \boldsymbol{h}_{Nm}] \tag{9}$$

for $m = 1, 2, \ldots, M$. We also let $\boldsymbol{n}(k) = [n_1(k) \; n_2(k) \; \cdots \; n_N(k)]^T$ be an N-dimensional noise vector with zero mean, white, and Gaussian distributed elements. With the formulation, M different symbols are simultaneously transmitted through M antennas and received by N antennas to yield the N-dimensional signal vector $\boldsymbol{x}(k) = [x_1(k) \; x_2(k) \; \cdots \; x_N(k)]^T$. With this premise, the received discrete-time equivalent baseband signal vector can be written as

$$\boldsymbol{x}(k) = \sum_{m=1}^{M} \boldsymbol{H}_m^T \boldsymbol{a}_m(k) + \boldsymbol{n}(k) \tag{10}$$

The formulation of the MIMO DFE is similar to that of the SISO DFE. Nevertheless, M decision devices are employed for M different data sequences. For simplicity, we only consider the most basic form of the MIMO DFE which does not include any successive interference cancellation (SIC) action [2]. To be consistent with the previous derivation for the SISO DFE, we first arrange the structure of the FFF into M matrices $\boldsymbol{F}_m$, for $m = 1, 2, \ldots, M$, with dimension $\alpha \times N$, and the FBF into M matrices $\boldsymbol{B}_m$, for $m = 1, 2, \ldots, M$, with dimension $\beta \times M$. Both α and β are selected to be long enough to cover the ISI effect in the multi-dimensional channel. The matrices $\boldsymbol{F}_m$ and $\boldsymbol{B}_m$ have the forms as

$$\boldsymbol{F}_m = \begin{bmatrix} f_{m1}^0 & f_{m2}^0 & \cdots & f_{mN}^0 \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1}^{\alpha-1} & f_{m2}^{\alpha-1} & \cdots & f_{mN}^{\alpha-1} \end{bmatrix} = [\boldsymbol{f}_{m1} \quad \boldsymbol{f}_{m2} \; \ldots \; \boldsymbol{f}_{mN}] \tag{11}$$

$$\boldsymbol{B}_m = \begin{bmatrix} b_{m1}^0 & b_{m2}^0 & \cdots & b_{mM}^0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1}^{\beta-1} & b_{m2}^{\beta-1} & \cdots & b_{mM}^{\beta-1} \end{bmatrix} = [\boldsymbol{b}_{m1} \quad \boldsymbol{b}_{m2} \; \ldots \; \boldsymbol{b}_{mM}] \tag{12}$$

To be more compact, we stack the components in the above matrices to form the following vectors

$$\bar{\boldsymbol{f}}_m = [\boldsymbol{f}_{m1}^T \quad \boldsymbol{f}_{m2}^T \quad \ldots \quad \boldsymbol{f}_{mN}^T]^T \tag{13}$$

$$\bar{\boldsymbol{b}}_m = [\boldsymbol{b}_{m1}^T \quad \boldsymbol{b}_{m2}^T \quad \ldots \quad \boldsymbol{b}_{mM}^T]^T \tag{14}$$

for the mth FFF and the mth FBF, respectively. Similarly, the successive received signal of (10) for the nth antenna can be first grouped as $\boldsymbol{x}_n(k) = [x_n(k) \; x_n(k-1) \cdots x_n(k-\alpha+1)]^T$, for $n = 1, 2, \ldots, N$, and then the total received signal vector is described as $\bar{\boldsymbol{x}}(k) = [\boldsymbol{x}_1^T(k) \; \boldsymbol{x}_2^T(k) \cdots \boldsymbol{x}_N^T(k)]^T$, which serves as the input to the FFF. The most recent β decisions from the output of the mth decision device after delay κ_m are labelled as $\hat{\boldsymbol{a}}_m(k) = [\hat{a}_m(k-\kappa_m-1) \; \hat{a}_m(k-\kappa_m-2) \cdots \hat{a}_m(k-\kappa_m-\beta)]^T$, for $m = 1, 2, \ldots, M$. Here, we assume that all decision delays κ_m are known at the receiver. Then, we can write the overall decision vector as $\bar{\boldsymbol{a}}(k) = [\boldsymbol{a}_1^T(k) \; \boldsymbol{a}_2^T(k) \cdots \boldsymbol{a}_M^T(k)]^T$, which constitutes the input to the FBF. After that, we may express the estimate in the output of the MIMO DFE prior to the mth decision device at the kth time instant as

$$y_m(k) = \bar{\boldsymbol{f}}_m^H \bar{\boldsymbol{x}}(k) - \bar{\boldsymbol{b}}_m^H \bar{\boldsymbol{a}}(k) \tag{15}$$

and the estimation error for it as

$$e_m(k) = a_m(k-\kappa_m) - y_m(k) \tag{16}$$

for $m = 1, 2, \ldots, M$. We see that the error signal given in (16) is similar to that of the SISO case except dimension expansion. Architecture-wise, the MIMO DFE can be treated as a generalisation of the SISO DFE, where the scalar delay line, the taps, and the decision are replaced by the vector delay line, the matrix taps, and the decision vector, respectively. With reference to (4) and (5), for each data sequence and the corresponding decision device, we may have the optimum FFF and FBF expressed as

$$\bar{\boldsymbol{f}}_{m,\mathrm{opt}} = \left(\boldsymbol{R}_{\bar{x}\bar{x}} - \frac{1}{\sigma_a^2} \boldsymbol{R}_{\bar{x}\bar{a}} \boldsymbol{R}_{\bar{x}\bar{a}}^H \right)^{-1} \boldsymbol{p}_{\bar{x}a_m} \tag{17}$$

$$\bar{\boldsymbol{b}}_{m,\mathrm{opt}} = \frac{1}{\sigma_a^2} \boldsymbol{R}_{\bar{x}\bar{a}}^H \bar{\boldsymbol{f}}_{m,\mathrm{opt}} \tag{18}$$

with the matrix and vector elements defined similar to those for the SISO DFE. Again, to avoid the matrix inverse operation in the FFF calculation, we may adopt the LMS algorithm to find the optimum tap weights recursively. It is not difficult to obtain the update equations for the MIMO DFE as

$$\bar{\boldsymbol{f}}_m(k+1) = \bar{\boldsymbol{f}}_m(k) + \mu_{\mathrm{f}} \bar{\boldsymbol{x}}(k) e_m^*(k) \tag{19}$$

$$\bar{\boldsymbol{b}}_m(k+1) = \bar{\boldsymbol{b}}_m(k) - \mu_{\mathrm{b}} \bar{\boldsymbol{a}}(k) e_m^*(k) \tag{20}$$

for $m = 1, 2, \ldots, M$. Since the received signal is also corrupted by CCI in MIMO channel environments, it tends to make the error signal in the above update equations more noisy. For similar ISI conditions, the performance of the adaptive MIMO DFE is worse than that of the adaptive SISO DFE.

3 Proposed adaptive channel-aided DFE (ACA-DFE)

3.1 *ACA-DFE for SISO systems*

Figure 1 is the block diagram of the proposed ACA-DFE for SISO systems. To obtain the channel response, we first introduce a channel estimator in the DFE structure. Let the coefficients of the included channel estimator be denoted as $\boldsymbol{q}$ and its dimension is $\gamma \times 1$. The value of γ is chosen to be larger than or equal to that of the channel order L. For convenience, we choose $\gamma = L$. According to Fig. 1, the channel estimator $\boldsymbol{q}$ is tuned by a new error signal $e_{\mathrm{q}}(k)$, and the cost function for the optimisation of $\boldsymbol{q}$ can be written as

$$\min_{\boldsymbol{q}} \mathrm{E}\{|e_{\mathrm{q}}(k)|^2\} = \min_{\boldsymbol{q}} \mathrm{E}\{|x(k) - \boldsymbol{q}^H \tilde{\boldsymbol{a}}(k)|^2\} \tag{21}$$

where $x(k)$ is the received signal and $\tilde{\boldsymbol{a}}(k) = [\hat{a}(k) \; \hat{a}(k-1) \cdots \hat{a}(k-\gamma+1)]^T$ is the input vector to the channel estimator. Assume that decisions are correct and input data symbols are white. We can then calculate the input correlation matrix for $\boldsymbol{q}$ as $\boldsymbol{R}_{\tilde{a}\tilde{a}} = E\{\tilde{\boldsymbol{a}}(k)\tilde{\boldsymbol{a}}^H(k)\} = \sigma_a^2 \boldsymbol{I}_\gamma$, where $\boldsymbol{I}_\gamma$ is a $\gamma \times \gamma$ identity matrix, and the cross-correlation vector for $\boldsymbol{a}(k)$ and $x(k)$ as $\boldsymbol{p}_{\tilde{a}x} = E\{\tilde{\boldsymbol{a}}(k)x^*(k)\} = \sigma_a^2 \boldsymbol{h}^*$. From (21), the optimum $\boldsymbol{q}$ solved

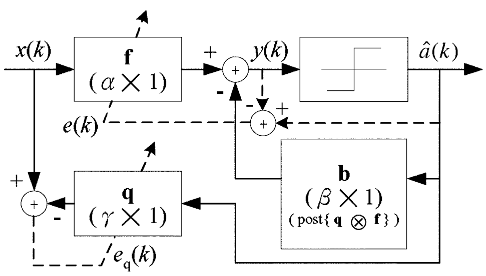

Fig. 1 *ACA-DFE in DD mode for SISO systems*

by the classical Wiener solution is

$$\boldsymbol{q}_{\text{opt}} = \boldsymbol{R}_{\tilde{a}\tilde{a}}^{-1}\boldsymbol{p}_{ax} = \boldsymbol{h}^* \tag{22}$$

We see that the optimum $\boldsymbol{q}$ equals the complex conjugate of the channel response. As previous, we can use the LMS algorithm to approach $\boldsymbol{q}_{\text{opt}}$ recursively. The update equation is stated as

$$\boldsymbol{q}(k+1) = \boldsymbol{q}(k) + \mu_{\text{q}}\tilde{\boldsymbol{a}}(k)e_{\text{q}}^*(k) \tag{23}$$

where $\boldsymbol{q}(k)$ is the estimate of $\boldsymbol{q}_{\text{opt}}$ at the kth time instant and μ_{q} is the step size for the adaptation. We observe that the channel estimation problem is essentially a system identification problem. For uncorrelated input data symbols, the eigenvalues of the input correlation matrix are all identical, and thus the eigenvalue spread equals unity, which is the minimum possible value. It is well-known that the convergence rate of the LMS algorithm is inversely proportional to the eigenvalue spread [16]. Thus, the convergence of $\boldsymbol{q}(k)$ is expected to be fast and stable.

Here, we make use of this channel estimator and propose a new DFE structure, i.e. the ACA-DFE. Our approach uses a basic property of the DFE, i.e., the postcursors of the channel response convolved with the FFF is cancelled by the FBF. For completeness, we now show the property formally. It is simple to see that the convolution of the channel and the FFF results in a response of length $\alpha + \gamma - 1$. Thus, for perfect postcursor cancellation, we must have $\beta \geq \alpha + \gamma - 2 - \kappa$. Without loss of generality, we let $\beta = \alpha + \gamma - 2 - \kappa$. Represent the convolution of $\boldsymbol{q}_{\text{opt}}$ and $\boldsymbol{f}_{\text{opt}}$ as $\boldsymbol{\Pi}\boldsymbol{f}_{\text{opt}}$, where $\boldsymbol{\Pi}$ is an $(\alpha + \gamma - 1) \times \alpha$ matrix as

$$\boldsymbol{\Pi} = \begin{bmatrix} h^0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ h^1 & h^0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & \ddots & & & & & \vdots \\ h^{\gamma-1} & h^{\gamma-2} & \cdots & h^0 & 0 & \cdots & \cdots & 0 \\ 0 & h^{\gamma-1} & h^{\gamma-2} & \cdots & h^0 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \ddots & & \vdots \\ 0 & \cdots & 0 & h^{\gamma-1} & h^{\gamma-2} & \cdots & h^0 & 0 \\ 0 & \cdots & \cdots & 0 & h^{\gamma-1} & h^{\gamma-2} & \cdots & h^0 \\ \vdots & & & & & \ddots & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & h^{\gamma-1} & h^{\gamma-2} \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & h^{\gamma-1} \end{bmatrix}^* \tag{24}$$

We can further partition $\boldsymbol{\Pi}$ as $\boldsymbol{\Pi} = [\boldsymbol{\Pi}_{\text{r}}^T \ \boldsymbol{\Pi}_{\text{p}}^T]^T$, where $\boldsymbol{\Pi}_r$ is of dimension $(\kappa + 1) \times \alpha$ and $\boldsymbol{\Pi}_{\text{p}}$ is of dimension $(\alpha + \gamma - 2 - \kappa) \times \alpha$. It is not difficult to observe that $\boldsymbol{\Pi}_{\text{r}}\boldsymbol{f}_{\text{opt}}$ corresponds to the precursor response of $\boldsymbol{\Pi} f_{\text{opt}}$ while $\boldsymbol{\Pi}_{\text{p}}\boldsymbol{f}_{\text{opt}}$ the postcursor response. Recall that the optimum FFF and FBF for SISO systems is calculated using (4) and (5), respectively. With some manipulations, we can derive

$$\frac{1}{\sigma_a^2}\boldsymbol{R}_{\boldsymbol{xa}}^H = \boldsymbol{\Pi}_{\text{p}} \tag{25}$$

From (5), we then obtain $\boldsymbol{b}_{\text{opt}} = \boldsymbol{\Pi}_{\text{p}}\boldsymbol{f}_{\text{opt}}$. This result can be re-stated as

$$\boldsymbol{b}_{\text{opt}} = \text{post}\{\boldsymbol{q}_{\text{opt}} \otimes \boldsymbol{f}_{\text{opt}}\} \tag{26}$$

where $\otimes$ denotes the convolution operation and $\text{post}\{\cdot\}$ denotes the postcursor-taking operation. This result suggests an adaptation approach for the training-based MMSE-DFE. Let $\boldsymbol{f}(k)$ and $\boldsymbol{b}(k)$ be the FFF and FBF at the kth time instant. With reference to (26), we can let

$$\boldsymbol{b}(k) = \text{post}\{\boldsymbol{q}(k) \otimes \boldsymbol{f}(k)\} \tag{27}$$

in which $\boldsymbol{q}(k)$ is the channel estimate at the kth time instant. If $\boldsymbol{q}(k)$ converges to $\boldsymbol{q}_{\text{opt}}$, $\boldsymbol{b}(k)$ will converge to $\boldsymbol{b}_{\text{opt}}$ too. The difference between this approach and the conventional method lies in that only $\boldsymbol{f}(k)$ is adapted (not both $\boldsymbol{f}(k)$ and $\boldsymbol{b}(k)$). For the conventional adaptive DFE in the DD mode, the scenario is that both $\boldsymbol{f}(k)$ and $\boldsymbol{b}(k)$ are adapted based on the LMS update equations as given in (6) and (7). We observe that if there is a decision error, the error will immediately reflect to $\hat{\boldsymbol{a}}(k)$ and then $e(k)$. Note that the adaptation of $\boldsymbol{f}(k)$ involves erroneous $e(k)$ only while that of $\boldsymbol{b}(k)$ involves both erroneous $\hat{\boldsymbol{a}}(k)$ and erroneous $e(k)$. The two error sources in (7) will make $\boldsymbol{b}(k)$ quite sensitive to decision errors. Alternatively, in the proposed method, only $\boldsymbol{f}(k)$ is adapted as given in (6). By using (27) to calculate $\boldsymbol{b}(k)$, the overall DFE will perform much better. Although the effect of decision error will also pass to $e_{\text{q}}(k)$ which will perturb the adaptation of $\boldsymbol{q}(k)$, the influence is smaller. This is because the convergence of $\boldsymbol{q}(k)$ for channel estimate is much faster and more stable than that of the DFE. In one word, with the proposed operation, the resultant ACA-DFE will be less sensitive to decision error and the EP effect will be reduced.

3.2 ACA-DFE for MIMO systems

In Section 2.2, we have already shown that for the mth decision device in MIMO DFE, the optimum formulation is similar to that for SISO DFE except dimension expansion. This motivates us to extend the idea of the SISO ACA-DFE to MIMO ACA-DFE. The block diagram of the MIMO ACA-DFE is described in Fig. 2. First, we define the channel estimators $\boldsymbol{q}_{nm}$ with dimension $\gamma \times 1$, for $m = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$, to estimate $\boldsymbol{h}_{nm}$ given in (8). Following the development presented previously, we can express this task as a system identification problem as

$$\min_{\boldsymbol{q}_{nm}} \ \text{E}\{|e_{\text{q},nm}(k)|^2\} = \min_{\boldsymbol{q}_{nm}} \ \text{E}\{|x_n(k) - \boldsymbol{q}_{nm}^H\tilde{\boldsymbol{a}}_m(k)|^2\} \tag{28}$$

where $x_n(k)$ is the received signal from the nth antenna and $\tilde{\boldsymbol{a}}_m(k) = [\hat{a}_m(k) \ \hat{a}_m(k-1) \ \cdots \ \hat{a}_m(k-\gamma+1)]^T$ is the decision vector from the mth decision device as the input to the corresponding channel estimators. Similar to (22), the solution is in the form as

$$\boldsymbol{q}_{nm,\text{opt}} = \boldsymbol{h}_{nm}^* \tag{29}$$

for $m = 1, 2, \ldots, M$ and $n = 1, 2, \cdots, N$. Referring to (23), we can then use the LMS algorithm to approach $\boldsymbol{q}_{nm,\text{opt}}$

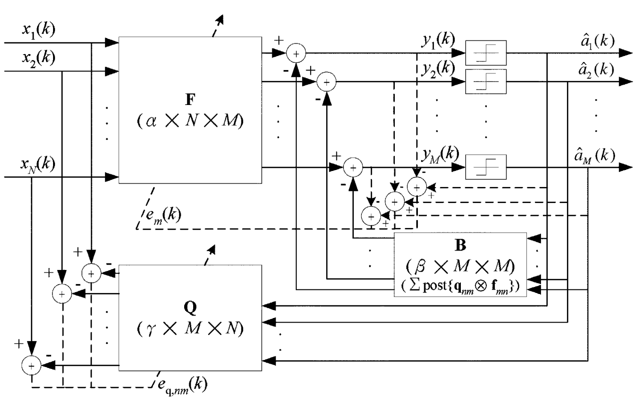

Fig. 2 *ACA-DFE in DD mode for MIMO systems*

too. Likewise, we define matrices $\boldsymbol{\Pi}_{mn}$ with dimension $(\alpha+\gamma-1)\times\alpha$, for $m=1,2,\ldots,M$ and $n=1,2,\ldots,N$, as

$$\boldsymbol{\Pi}_{mn}=\begin{bmatrix} h_{nm}^{0} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ h_{nm}^{1} & h_{nm}^{0} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & \ddots & & & & & \vdots \\ h_{nm}^{\gamma-1} & h_{nm}^{\gamma-2} & \cdots & h_{nm}^{0} & 0 & \cdots & \cdots & 0 \\ 0 & h_{nm}^{\gamma-1} & h_{nm}^{\gamma-2} & \cdots & h_{nm}^{0} & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \ddots & & \vdots \\ 0 & \cdots & 0 & h_{nm}^{\gamma-1} & h_{nm}^{\gamma-2} & \cdots & h_{nm}^{0} & 0 \\ 0 & \cdots & \cdots & 0 & h_{nm}^{\gamma-1} & h_{nm}^{\gamma-2} & \cdots & h_{nm}^{0} \\ \vdots & & & & & \ddots & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & h_{nm}^{\gamma-1} & h_{nm}^{\gamma-2} \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & h_{nm}^{\gamma-1} \end{bmatrix}^{*} \tag{30}$$

We then partition $\boldsymbol{\Pi}_{mn}$ as $\boldsymbol{\Pi}_{mn}=[\boldsymbol{\Pi}_{\mathrm{r},mn}^{T}\ \boldsymbol{\Pi}_{\mathrm{p},mn}^{T}]^{T}$, where $\boldsymbol{\Pi}_{\mathrm{r},mn}$ is of dimension $(\kappa_m+1)\times\alpha$ and $\boldsymbol{\Pi}_{\mathrm{p},mn}$ is of dimension $(\alpha+\gamma-2-\kappa_m)\times\alpha$. By the definition that $\boldsymbol{R}_{\bar{x}\bar{a}}=\mathrm{E}\{\bar{\boldsymbol{x}}(k)\bar{\boldsymbol{a}}^{H}(k)\}$, we can have

$$\frac{1}{\sigma_a^2}\boldsymbol{R}_{\bar{x}\bar{a}}^{H}=\begin{bmatrix} \boldsymbol{\Pi}_{\mathrm{p},11} & \boldsymbol{\Pi}_{\mathrm{p},12} & \cdots & \boldsymbol{\Pi}_{\mathrm{p},1N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Pi}_{\mathrm{p},M1} & \boldsymbol{\Pi}_{\mathrm{p},M2} & \cdots & \boldsymbol{\Pi}_{\mathrm{p},MN} \end{bmatrix} \tag{31}$$

and the mth optimum FBF given in (18) can be specified as

$$\bar{\boldsymbol{b}}_{m,\mathrm{opt}}=\begin{bmatrix} \sum_{n=1}^{N}\mathrm{post}\{\boldsymbol{q}_{n1,\mathrm{opt}}\otimes\boldsymbol{f}_{mn,\mathrm{opt}}\} \\ \sum_{n=1}^{N}\mathrm{post}\{\boldsymbol{q}_{n2,\mathrm{opt}}\otimes\boldsymbol{f}_{mn,\mathrm{opt}}\} \\ \vdots \\ \sum_{n=1}^{N}\mathrm{post}\{\boldsymbol{q}_{nM,\mathrm{opt}}\otimes\boldsymbol{f}_{mn,\mathrm{opt}}\} \end{bmatrix} \tag{32}$$

in which $\boldsymbol{f}_{mn,\mathrm{opt}}$ is the nth sub-vector in $\bar{\boldsymbol{f}}_{m,\mathrm{opt}}$, as represented in (13). While both $\boldsymbol{q}_{nm,\mathrm{opt}}$ and $\bar{\boldsymbol{f}}_{m,\mathrm{opt}}$ are estimated using the LMS algorithm, the estimation of $\bar{\boldsymbol{b}}_{m,\mathrm{opt}}$ at the kth time instant can then be calculated as

$$\bar{\boldsymbol{b}}_{m}(k)=\begin{bmatrix} \sum_{n=1}^{N}\mathrm{post}\{\boldsymbol{q}_{n1}(k)\otimes\boldsymbol{f}_{mn}(k)\} \\ \sum_{n=1}^{N}\mathrm{post}\{\boldsymbol{q}_{n2}(k)\otimes\boldsymbol{f}_{mn}(k)\} \\ \vdots \\ \sum_{n=1}^{N}\mathrm{post}\{\boldsymbol{q}_{nM}(k)\otimes\boldsymbol{f}_{mn}(k)\} \end{bmatrix} \tag{33}$$

Since $\boldsymbol{q}_{nm}$ can estimate the corresponding channel response $\boldsymbol{h}_{nm}$, for the same reason described in the SISO case, the proposed operation in (33) can enhance the adaptation of the FBF. The resultant MIMO ACA-DFE can then improve the robustness against EP for MIMO channel equalisation.

4 Simulations

Computer simulations are conducted to demonstrate the effectiveness of the proposed ACA-DFE and MIMO ACA-DFE. In the first part, we consider SISO channels. In the second part, we consider MIMO channels. All transmitted symbols are randomly generated and then modulated by quadrature phase-shift keying (QPSK). All decision delays are chosen to optimise the performance. In all figures, at least 500 simulation runs are averaged to obtain each simulated result.

4.1 Experiment 1: ACA-DFE

In this set of simulations, we demonstrate that the proposed ACA-DFE can provide more robust and stable performance than the conventional adaptive DFE against EP under severe ISI environments. We first consider a static channel chosen from [4, p. 616], which is $[0.227\ \ 0.460\ \ 0.688\ \ 0.460\ \ 0.227]^{T}$ (Proakis C channel). The parameters α, β and γ for those filters are set to be 9, 9 and 5, respectively. For comparison, we also show the case of ACA-DFE with perfect channel state information (CSI). The signal-to-noise ratio (SNR) is set as 25 dB. Here, $\mu_{\mathrm{f}}=\mu_{\mathrm{b}}=0.005$, $\mu_{\mathrm{q}}=0.002$, and the number of training symbols $T_{\mathrm{t}}=2000$. The DD mode follows immediately after the training mode. Figure 3 gives the learning curves for various equalisation schemes. We see that there is no big difference in performance between the proposed ACA-DFE with and without perfect CSI. It implies that the channel estimator works fairly well. The ACA-DFE performs better than the conventional adaptive DFE in the DD mode in this severe ISI scenario. To demonstrate the merits of the proposed ACA-DFE further, we give the relation between the average SER and the step size used in the FFF (the same step size is used in the FBF of the conventional adaptive DFE as well) in Fig. 4. This figure reveals that the ACA-DFE always has lower SER than the conventional adaptive DFE with the same step size. There are a couple of things that we can observe from the figure. First, there is an optimum step size for a DFE. As known, for the LMS algorithm, the smaller the step size, the smaller the output MSE in the steady state (possibly the lower the average SER). However, a smaller step size will also make the convergence slower. As a result, there exists an optimum step size balancing these two effects. The optimum μ_{f} giving the lowest SER is around 0.005 for both schemes, and the

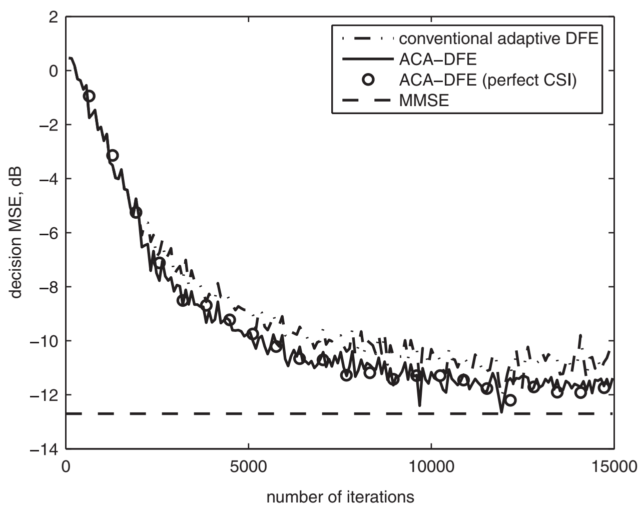

Fig. 3 *MSE learning curves for static Proakis C channel*

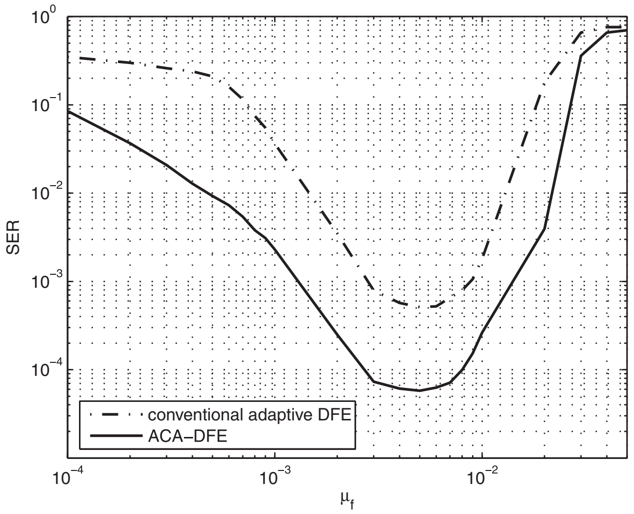

Fig. 4 *Average SER for static Proakis C channel with different step sizes*

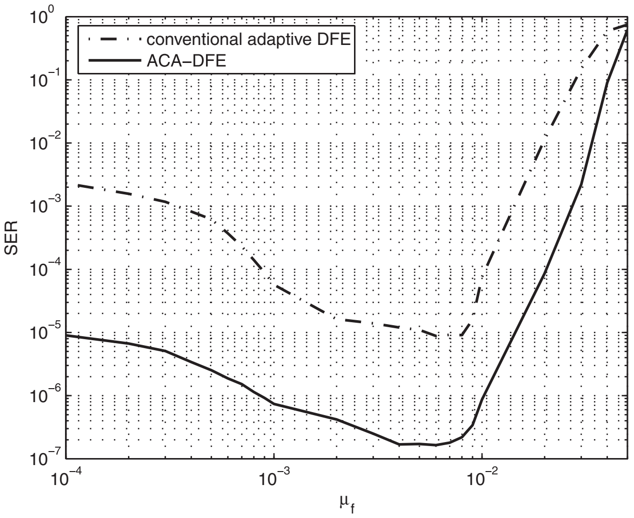

Fig. 5 *Average SER for time-varying Proakis C channel ($f_d T_s = 5 \times 10^{-4}$) with different step sizes*

SER improvement with the ACA-DFE is almost an order of magnitude. Second, the ACA-DFE outperforms the conventional adaptive DFE for any step size. Last, given a target SER, the applicable step-size range of the ACA-DFE is wider than that of the conventional adaptive DFE. We then consider a time-varying channel constructed from Proakis C channel used previously. The first, second, forth and fifth paths in the channel now undergo fading according to Jake's model [26], and are upper bounded by their corresponding path magnitudes. The normalised Doppler frequency $f_\mathrm{d} T_\mathrm{s}$ equals 5×10^{-4}, in which f_d is the Doppler frequency and T_s is the symbol duration. This time-varying channel is normalised to keep the SNR at a constant level. Figure 5 shows the average SER against the step size used in the DFE. We observe that the SER improvement for this time-varying case is quite significant, and is more than an order of magnitude most of the time. In practice, it is difficult to know the exact channel variation pattern and the optimum step size. For the ACA-DFE, since the applicable step-size range is wider and the resultant SER is always lower, making the choice of the step size becomes much easier. This enables the ACA-DFE to work adequately in general time-varying environments.

4.2 Experiment 2: MIMO ACA-DFE

In this part, we consider the MIMO ACA-DFE for dispersive MIMO channels. First, we use the static 2×2 MIMO channel given in [2]: $\boldsymbol{h}_{11} = [0.781\ \ 0.625]^T$, $\boldsymbol{h}_{12} = [0.781\ \ -0.625]^T$, $\boldsymbol{h}_{21} = [0.895\ \ -0.447]^T$, and $\boldsymbol{h}_{22} = [0.958\ \ 0.287]^T$. The parameters used are $T_\mathrm{t} = 200$, SNR = 20 dB, $\mu_\mathrm{f} = \mu_\mathrm{b} = 0.005$, and $\mu_\mathrm{q} = 0.002$. The learning curves for the MIMO ACA-DFE and the conventional adaptive MIMO DFE are shown in Fig. 6. We see that the MIMO ACA-DFE can achieve an MSE lower than that of the conventional adaptive MIMO DFE in the DD mode. Similarly, Fig. 7 presents the relationship between the average SER and the step size used in the MIMO DFEs. The MIMO ACA-DFE generally achieves lower SER than the conventional adaptive MIMO DFE. Next, we conduct the experiment under time-varying channel environments, in which we let the second channel tap in $\boldsymbol{h}_{nm}$ ($n = 1, 2$ and $m = 1, 2$) be varied with Jake's model and upper bounded by its corresponding path magnitude. The normalised Doppler frequency is now changed to 2×10^{-4}. Figure 8 shows the average SER versus the step size used for the time-varying channel. We can see that the MIMO ACA-DFE still provides better performance. In this scenario, the

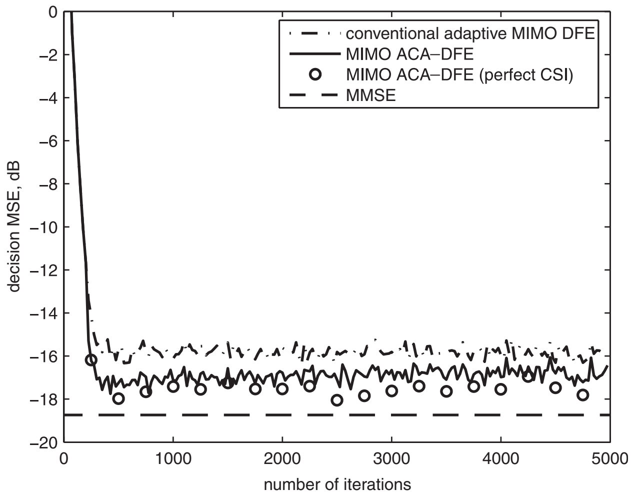

Fig. 6 *MSE learning curves for static MIMO channel*

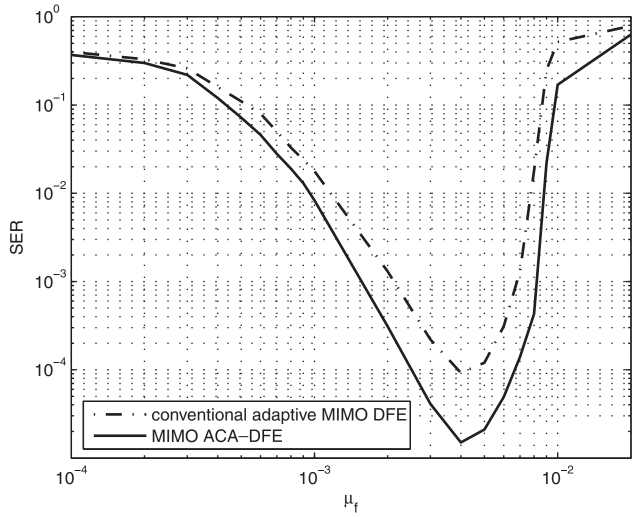

Fig. 7 *Average SER for static MIMO channel with different step sizes*

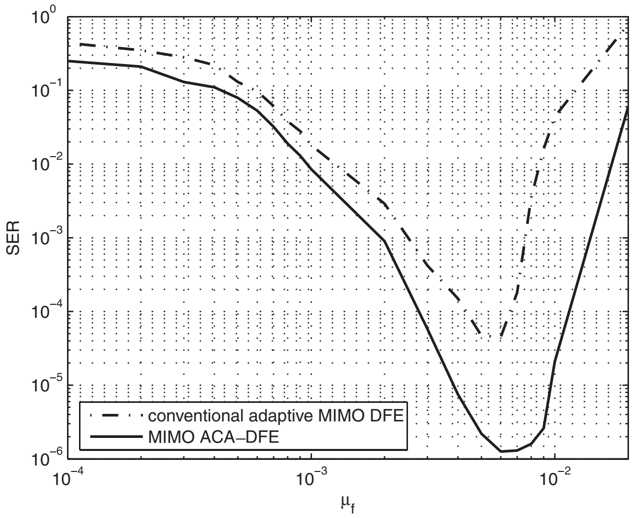

Fig. 8 *Average SER for time-varying MIMO channel ($f_dT_s = 2 \times 10^{-4}$) with different step sizes*

optimum step size for both DFE schemes is around 0.006, and the SER improvement with the MIMO ACA-DFE is more than an order of magnitude. However, note that the performance obtained with the proposed method is not as good as that in the SISO scenario. It is because the MIMO environment induces CCI for each transmitted sequence and this lowers the input SNR. Also, the channel becomes multi-dimensional and is more difficult to estimate.

5 Conclusions

In this paper, we have developed the ACA-DFE for SISO systems and MIMO systems. With the additional channel estimator(s) and the special operation for the FBF, the stability and robustness against EP are improved. Simulation results confirm the usefulness of these proposed schemes. Note that the DFE considered here for MIMO systems is generally referred to as the parallel interference cancellation (PIC) scheme in which decisions are made for all recovered bit streams simultaneously. To further enhance the performance in MIMO systems, we can apply the SIC technique where the decision for each bit stream is made sequentially. The decision sequence is determined according to the decision-error probabilities in the recovered bit streams. Combining SIC and the method proposed in this paper, we can reduce the EP effect even more effectively. Research on this subject is now underway.

6 References

1 Belfiore, C., and Park, Jr., J.: 'Decision feedback equalization', *Proc. IEEE*, 1979, **67**, (8), pp. 1143–1156

2 Al-Dhahir, N., and Cioffi, J.M.: 'MMSE decision-feedback equalizers: Finite-length results', *IEEE Trans. Inf. Theory*, 1995, **41**, (7), pp. 961–975

3 Cioffi, J.M., Dudevoir, G.P., Eyuboglu, M.V., and Forney, G.D.: 'MMSE decision-feedback equalizers and coding. I. Equalization results', *IEEE Trans. Commun.*, 1995, **43**, (10), pp. 2582–2594

4 Proakis, J.G.: 'Digital communications' (McGraw-Hill, New York, 1995, 3rd edn.)

5 Smee, J.E., and Beaulieu, N.C.: 'Error-rate evaluation of linear equalization and decision feedback equalization with error propagation', *IEEE Trans. Commun.*, 1998, **46**, (5), pp. 656–665

6 Russell, M., and Bergmans, W.M.: 'A technique to reduce error propagation in M-ary decision feedback equalization', *IEEE Trans. Commun.*, 1995, **43**, (12), pp. 2878–2881

7 Chiani, M.: 'Introducing erasures in decision-feedback equalization to reduce error propagation', *IEEE Trans. Commun.*, 1997, **45**, (7), pp. 757–760

8 Fertner, A.: 'Improvement of bit-error-rate in decision feedback equalizer by preventing decision error propagation', *IEEE Trans. Signal Process.*, 1998, **46**, (7), pp. 1872–1877

9 Reuter, M., Allen, J.C., Zeidler, J.R., and North, R.C.: 'Mitigating error propagation effects in a decision feedback equalizer', *IEEE Trans. Commun.*, 2001, **49**, (11), pp. 2028–2041

10 Wulich, D., and Geva, A.: 'On a periodic training sequence in DFE to reduce the steady-state error probability', *IEEE Trans. Commun.*, 1999, **47**, (9), pp. 1288–1292

11 Foschini, G.J., and Gans, M.J.: 'On the limits of wireless communications in a fading environment when using multiple antennas', *Wirel. Pers. Commun.*, 1998, **6**, (3), pp. 315–335

12 Al-Dhahir, N., and Sayed, A.H.: 'The finite-length multi-input multi-output MMSE-DFE', *IEEE Trans. Signal Process.*, 2000, **48**, (10), pp. 2921–2936

13 Tidestav, C., Ahlén, A., and Sternad, M.: 'Realizable MIMO decision feedback equalizers: Structure and design', *IEEE Trans. Signal Process.*, 2001, **49**, (1), pp. 121–133

14 Fischer, R., Huber, J., and Windpassinger, C.: 'Signal processing in decision-feedback equalization of intersymbol-interference and multiple-input/multiple-output channels: A unified view', *Signal Process.*, 2003, **83**, (8), pp. 1633–1642

15 Qureshi, S.U.H.: 'Adaptive equalization', *Proc. IEEE*, 1985, **73**, (9), pp. 1349–1387

16 Haykin, S.: 'Adaptive filter theory' (Prentice-Hall Inc, 1996, 3rd edn.)

17 Frigon, J.F., and Daneshrad, B.: 'A multiple input-multiple output (MIMO) adaptive decision feedback equalizer (DFE) with cancellation for wideband space-time communications', *Int. J. Wirel. Inf. Netw.*, 2002, **9**, (1), pp. 13–23

18 Komninakis, C., Fragouli, C., Sayed, A.H., and Wesel, R.D.: 'Multi-input multiple-output fading channel tracking and equalization using Kalman estimation', *IEEE Trans. Signal Process.*, 2002, **50**, (5), pp. 1065–1076

19 Aaron, M.R., and Tufts, D.W.: 'Intersymbol interference and error probability', *IEEE Trans. Inf. Theory*, 1966, **12**, (1), pp. 26–34

20 Chen, S., Mulgrew, B., Chng, E.S., and Gibson, G.J.: 'Space translation properties and the minimum-BER linear-combiner DFE', *IEE Proc. Commun.*, 1998, **145**, (10), pp. 316–322

21 Chen, S., Hanzo, L., and Mulgrew, B.: 'Adaptive minimum symbol-error-rate decision feedback equalization for multilevel pulse-amplitude modulation', *IEEE Trans. Signal Process.*, 2004, **52**, (7), pp. 2092–2101

22 Yeh, C.-C., and Barry, J.R.: 'Adaptive minimum bit-error rate equalization for binary signaling', *IEEE Trans. Commun.*, 2000, **48**, (7), pp. 1226–1235

23 Yeh, C.-C., and Barry, J.R.: 'Adaptive minimum symbol-error rate equalization for quadrature-amplitude modulation', *IEEE Trans. Signal Process.*, 2003, **51**, (12), pp. 3263–3269

24 Shukla, P.K., and Turner, L.F.: 'Channel-estimation-based adaptive DFE for fading multipath radio channels', *IEE Proc. I, Commun. Speech Vis.*, 1991, **138**, (12), pp. 525–543

25 Rontogiannis, A.A., and Berberidis, K.: 'Bandwidth efficient transmission through sparse channels using a parametric channel-estimation-based DFE', *IEE Proc. Commun.*, 2005, **152**, (4), pp. 251–256

26 Dent, P., Bottomley, G.E., and Croft, T.: 'Jakes fading model revisited', *Electron. Lett.*, 1993, **29**, (6), pp. 1162–1163