

Bayesian decision feedback equaliser for overcoming co-channel interference

S. Chen
S. McLaughlin
B. Mulgrew
P.M. Grant

Indexing terms: Co-channel interference, Bayesian decision feedback equaliser

Abstract: The authors derive a Bayesian decision feedback equaliser which incorporates co-channel interference compensation. By exploiting the structure of co-channel interfering signals, the proposed Bayesian decision feedback equaliser is able to distinguish an interfering signal from white noise and utilises this information to improve performance. Adaptive implementation of this Bayesian decision feedback equaliser includes identifying the channel model using the least mean square algorithm and estimating the co-channel states by means of an unsupervised clustering scheme. Simulation involving a binary signal constellation is used to compare both the theoretical and adaptive performance of this Bayesian decision feedback equaliser with those of the maximum likelihood sequence estimator. The results obtained indicate that, in the presence of severe co-channel interference, the Bayesian decision feedback equaliser employing the proposed simple scheme to compensate co-channel interference can outperform the maximum likelihood sequence estimator that only treats co-channel interference as an additional coloured noise.

1 Introduction

Adaptive equalisers for combating channel intersymbol interference (ISI) and noise can be classified into two categories, namely sequence estimation and symbol decision equalisers. The optimal solution for the class of sequence estimation equalisers is the maximum likelihood sequence estimator (MLSE) [1]. The MLSE provides the lowest error rate attainable for any equaliser when the channel is known but is computationally very expensive. A widely used symbol decision equaliser is the conventional decision feedback equaliser (DFE) [2]

which has a very low computational complexity. The conventional DFE, however, does not achieve the full performance potential of the symbol decision DFE structure, and the optimal symbol decision DFE is known to be the Bayesian DFE [3].

In the previous study [3–5], we have compared the Bayesian DFE with the conventional DFE and the MLSE extensively. In terms of computational requirements, the adaptive Bayesian DFE is more complex than the conventional DFE but is less complex than the adaptive MLSE. The adaptive MLSE requires sophisticated processing capability while the implementation of the Bayesian DFE is relatively straightforward. For stationary channels, the performance of the adaptive Bayesian DFE is much better than the conventional adaptive DFE but is inferior to that of the adaptive MLSE. The adaptive Bayesian DFE however has significant advantages over the adaptive MLSE for rapidly time-varying channels. Extensive simulation results have demonstrated that the adaptive Bayesian DFE actually outperforms the adaptive MLSE in terms of error rate for severely fading channels. It has been suggested that the adaptive MLSE accumulates tracking errors, which causes serious performance degradation [5].

Many communication systems, such as mobile cellular radio and dual polarised microwave radio channels, are impaired not only by channel ISI but also by co-channel interference (CCI). It is well-known that an adaptive equaliser can usually do better against CCI than it can against the same level of noise [6]. However, in doing so, most of the equalisers can only treat CCI as an additional noise source and do not fully exploit the differences between the interfering signals and the noise. For example, a linear equaliser only exploits the spectral characteristics of the interference through its autocorrelations [6, 7]. This is also the case for the conventional DFE studied in [8].

If both the channel and co-channels are known, it is possible to design the MLSE which takes into account both the ISI and CCI. Such a full MLSE, although computationally very complex, will achieve the lowest possible error rate. The difficulty is that there is no practical way of obtaining accurate co-channel models needed. Unlike the case of channel identification, there is generally no training signals available for supervised co-channel identification. Even if a means of identifying the co-channels can be developed, the estimate errors are expected to be large. The MLSE, being a sequence estimation method, is more likely to accumu-

© IEE, 1996

IEE Proceedings online no. 19960612

Paper first received 7th July 1995 and in revised form 25th March 1996

S. Chen is with the Department of Electrical and Electronic Engineering, University of Portsmouth, Anglesea Building, Portsmouth PO1 3DJ, UK

S. McLaughlin, B. Mulgrew and P.M. Grant are with the Department of Electrical Engineering, University of Edinburgh, King's Buildings, Edinburgh EH9 3JL, UK

late the co-channel estimation errors, causing serious performance degradation. In the blind equalisation setting, in theory it is possible to design a joint data detection and channel/co-channels estimation based on the MLSE approach. Such an approach will certainly be computationally too expensive to implement. In practice, interfering signals are often treated as an additional coloured noise in the standard MLSE.

The probability density function (PDF) of an interfering signal is quite different from that of the noise. An ideal equaliser should be capable of distinguishing the interfering signal from the noise. In a previous study [7], a Bayesian transversal equaliser was derived which can effectively exploit the differences between the CCI and the noise and uses this information to improve performance. The adaptive version of this Bayesian equaliser can be implemented easily. The present study extends this result to the DFE structure and incorporates CCI compensation into the Bayesian DFE derived previously for combatting ISI and noise. It is shown that, in the presence of severe CCI, this Bayesian DFE has superior performance over the MLSE which only treats CCI as coloured noise. Adaptive implementation of this Bayesian DFE is then investigated. To effectively compensate for the CCI, the set of co-channel states are required. A simple unsupervised clustering algorithm is used to estimate these co-channel states.

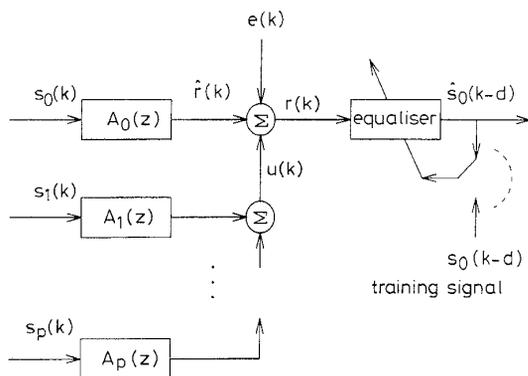


Fig. 1 Discrete-time model of communication system

The system model considered in this study is depicted in Fig. 1. This model [9] is widely used to represent communication systems in the presence of ISI, CCI and noise. The channel $A_0(z)$ and the p interfering co-channels $A_i(z)$, $1 \leq i \leq p$, are modelled by finite impulse response filters

$$A_i(z) = \sum_{j=0}^{n_i-1} a_{i,j} z^{-j}, \quad 0 \leq i \leq p \quad (1)$$

where n_i and $a_{i,j}$ are the length and the tap weights of the i th impulse response, respectively. The transmitted data $s_0(k)$ and the interfering data $s_i(k)$, $1 \leq i \leq p$, are independently identically distributed (iid) and they are mutually independent. The three components of the channel observation,

$$r(k) = \hat{r}(k) + u(k) + e(k) \quad (2)$$

will be referred to as the desired signal, the interfering signal and the noise, respectively. The noise $e(k)$ is assumed to be a Gaussian white noise with variance $E[e^2(k)] = \sigma_e^2$.

The Bayesian DFE [3–5] was derived for complex valued multilevel signals. In the extension to include CCI, for notational simplicity and to highlight the basic concepts, $s_i(k)$, $0 \leq i \leq p$, are assumed to be binary and to take values from the symbol set $\{s^{(1)} = +1, s^{(2)} = -1\}$. The tap weights $a_{i,j}$ are therefore real valued. Application to complex valued $A_i(z)$ and multilevel symbol constellations are straightforward (as in the case of the Bayesian DFE for combatting ISI and noise) but the computational complexity will increase significantly. Let $E[\hat{r}^2(k)] = \sigma_r^2$ and $E[u^2(k)] = \sigma_u^2$. We define the signal to noise ratio (SNR) of the system as $\text{SNR} = \sigma_r^2/\sigma_e^2$, the signal to interference ratio (SIR) of the system as $\text{SIR} = \sigma_r^2/\sigma_u^2$, and the signal to interference and noise ratio (SINR) of the system as $\text{SINR} = \sigma_r^2/(\sigma_u^2 + \sigma_e^2)$, respectively.

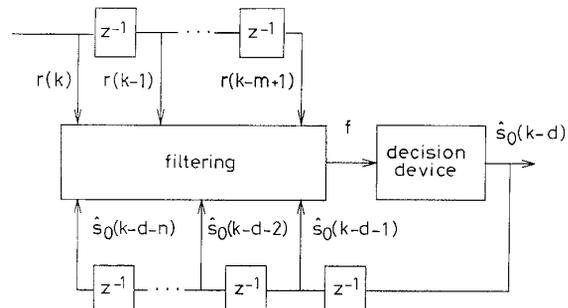


Fig. 2 Schematic of decision feedback equaliser

The structure of the DFE considered in this study is depicted in Fig. 2. The equalisation process defined in Fig. 2 uses the information present in the observed channel output vector,

$$\mathbf{r}(k) = [r(k) \cdots r(k-m+1)]^T \quad (3)$$

and the past detected symbol vector,

$$\hat{\mathbf{s}}_b(k) = [\hat{s}_0(k-d-1) \cdots \hat{s}_0(k-d-n)]^T \quad (4)$$

to produce an estimate $\hat{s}_0(k-d)$ of $s_0(k-d)$. The integers d , m and n are known as the decision delay, the feedforward order and the feedback order, respectively. Without the loss of generality, $d = n_0 - 1$ is chosen to cover the entire channel dispersion $A_0(z)$, m is related to d by $m = d + 1 = n_0$, and n is given by $n = n_0 + m - d - 2 = n_0 - 1$ [3].

2 Bayesian DFE in the absence of CCI

In the presence of the ISI and noise, the optimal solution for the symbol decision structure of Fig. 2 is the Bayesian DFE [3–5]. This Bayesian DFE is first summarised. This will naturally lead to the Bayesian solution in the presence of CCI. A new version of this Bayesian DFE is then presented which has certain practical advantages. Given the channel model $A_0(z)$, the value of the noiseless channel output vector,

$$\hat{\mathbf{r}}(k) = [\hat{r}(k) \cdots \hat{r}(k-m+1)]^T \quad (5)$$

is specified by the symbol sequence $\mathbf{s}(k) = [\mathbf{s}_f^T(k) \mathbf{s}_b^T(k)]^T$, where

$$\left. \begin{aligned} \mathbf{s}_f(k) &= [s_0(k) \cdots s_0(k-d)]^T \\ \mathbf{s}_b(k) &= [s_0(k-d-1) \cdots s_0(k-d-n)]^T \end{aligned} \right\} \quad (6)$$

Under the assumption that the given feedback vector is correct, that is, $\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k)$, the state of $\hat{\mathbf{r}}(k)$ is determined by $\mathbf{s}_f(k)$. Since $\mathbf{s}_f(k)$ has $N_s = 2^{d+1} = 2^m$ combinations, $\hat{\mathbf{r}}(k)$ has N_s states. Let N_s sequences of $\mathbf{s}_f(k)$ be

$$\mathbf{s}_{f,j}(k) = [s_{f,j}(k) \cdots s_{f,j}(k-d)]^T, 1 \leq j \leq N_s \quad (7)$$

The corresponding states of $\hat{\mathbf{r}}(k)$, denoted as \mathbf{r}_j , are given by:

$$\mathbf{r}_j = [\mathbf{F}'' \quad \mathbf{F}'] [s_{f,j}^T(k) \quad \hat{\mathbf{s}}_b^T(k)]^T, 1 \leq j \leq N_s \quad (8)$$

where the $m \times (d+1)$ matrix \mathbf{F}'' has the form:

$$\mathbf{F}'' = \begin{bmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,n_0-1} \\ 0 & a_{0,0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{0,1} \\ 0 & \cdots & 0 & a_{0,0} \end{bmatrix} \quad (9)$$

and the $m \times n$ matrix \mathbf{F}' has the form

$$\mathbf{F}' = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{0,n_0-1} & 0 & \ddots & \vdots \\ a_{0,n_0-2} & a_{0,n_0-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_{0,1} & \cdots & a_{0,n_0-2} & a_{0,n_0-1} \end{bmatrix} \quad (10)$$

The states of $\hat{\mathbf{r}}(k)$ can be grouped into two subsets according to the value of $s_0(k-d)$:

$$R^{(i)} = \{\hat{\mathbf{r}}(k) = \mathbf{r}_j^{(i)} | s_0(k-d) = s^{(i)}\}, 1 \leq i \leq 2 \quad (11)$$

Each $R^{(i)}$ contains $N_s^{(i)} = N_s/2 = 2^d$ states.

The PDF of $\mathbf{r}(k)$ conditioned on $s_0(k-d) = s^{(i)}$ is

$$p_{\mathbf{r}}(\mathbf{r}(k) | s_0(k-d) = s^{(i)}) = \sum_{j=1}^{N_s^{(i)}} \alpha_j^{(i)} p_{\mathbf{e}}(\mathbf{r}(k) - \mathbf{r}_j), 1 \leq i \leq 2 \quad (12)$$

where $\mathbf{r}_j \in R^{(i)}$, $\alpha_j^{(i)}$ are *a priori* probabilities of \mathbf{r}_j , and $p_{\mathbf{e}}(\cdot)$ is the PDF of the noise vector $\mathbf{e}(k) = [e(k) \cdots e(k-m+1)]^T$. Since all the channel states can be assumed to be equiprobable and the noise PDF is Gaussian, eqn. 12 leads to the Bayesian decision variables:

$$\eta_i(k, \mathbf{a}_0) = \sum_{j=1}^{N_s^{(i)}} \exp(-\|\mathbf{r}(k) - \mathbf{r}_j\|^2 / 2\sigma_e^2), 1 \leq i \leq 2 \quad (13)$$

Here $\mathbf{a}_0 = [a_{0,0} a_{0,1} \cdots a_{0,n_0-1}]^T$ is included in the expression to emphasise that the channel states are computed based on the given channel \mathbf{a}_0 . The minimum error probability decision is defined by

$$\hat{s}_0(k-d) = \begin{cases} s^{(1)}, & \text{if } \eta_1(k, \mathbf{a}_0) \geq \eta_2(k, \mathbf{a}_0) \\ s^{(2)}, & \text{otherwise} \end{cases} \quad (14)$$

which provides the optimal solution for the equalisation structure of Fig. 2 in the absence of CCI.

For the above version of the Bayesian DFE originally derived in [3], a different set of the channel states is required at each sample k even when the channel \mathbf{a}_0 is constant because the feedback vector $\hat{\mathbf{s}}_b(k)$ is different at different k . That is, different Bayesian equalisers are used for different decision feedbacks. Analysis and implementation of the Bayesian DFE becomes easier if the following space translation is made. Define:

$$\mathbf{r}'(k) = \mathbf{r}(k) - \mathbf{F}' \hat{\mathbf{s}}_b(k) \quad (15)$$

The elements of $\mathbf{r}'(k)$ can be computed recursively:

$$\left. \begin{aligned} r'(k-i) &= z^{-1} r'(k-i+1) - a_{0,n_0-1} \hat{s}(k-d-1) \\ i &= m-1, \dots, 2, 1 \\ r'(k) &= r(k) \end{aligned} \right\} \quad (16)$$

In the new translated space, the channel states are given by

$$\mathbf{r}'_j = \mathbf{F}'' \mathbf{s}_{f,j}(k), 1 \leq j \leq N_s \quad (17)$$

The Bayesian DFE consists of computing the decision variables:

$$\eta_i(k, \mathbf{a}_0) = \sum_{j=1}^{N_s^{(i)}} \exp(-\|\mathbf{r}'(k) - \mathbf{r}'_j\|^2 / 2\sigma_e^2), 1 \leq i \leq 2 \quad (18)$$

and making the decision according to eqn. 14.

This version of the Bayesian DFE realises the same optimal solution as the original one for the equalisation process defined in Fig. 2. It, however, has certain advantages over the original version. It removes the requirement of different Bayesian equalisers for different decision feedbacks, and has clear advantages in hardware implementation. Using the proposed translation, analysis of the Bayesian DFE can be reduced to one of studying an equivalent Bayesian equaliser 'without decision feedback'. Schematic diagram of this alternative Bayesian DFE is depicted in Fig. 3.

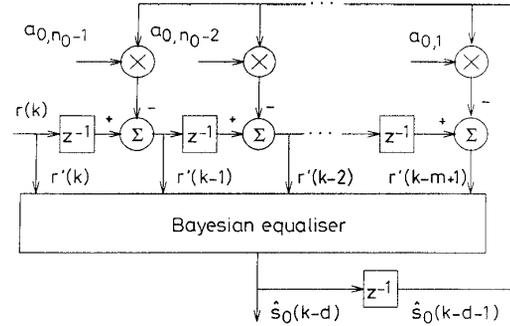


Fig. 3 Alternative implementation of Bayesian DFE

3 Bayesian DFE in the presence of CCI

The Bayesian DFE can now readily be extended to cover CCI. The key to this extension is the fact that similar to the desired signal $\hat{r}(k)$ the interfering signal $u(k)$ can only take some finite number of values. Without loss of generality, we will assume that only one CCI ($p = 1$) is present. The interfering signal $u(k)$ then has $N_{u,s} = 2^m$ scalar states $\{u_j, 1 \leq j \leq N_{u,s}\}$. Therefore, the interfering signal vector,

$$\mathbf{u}(k) = [u(k) \cdots u(k-m+1)]^T \quad (19)$$

has $N_u = 2^{m+n-1}$ states. The set of these co-channel vector states is denoted as $U = \{\mathbf{u}_j, 1 \leq j \leq N_u\}$.

In the presence of this CCI, the PDF of $\mathbf{r}(k)$ conditioned on $s_0(k-d) = s^{(i)}$ is

$$p_{\mathbf{r}}(\mathbf{r}(k) | s_0(k-d) = s^{(i)}) = \sum_{j=1}^{N_s^{(i)}} \sum_{l=1}^{N_u} \alpha_{j,l}^{(i)} p_{\mathbf{e}}(\mathbf{r}(k) - \mathbf{r}_j - \mathbf{u}_l) \quad (20)$$

where $\mathbf{r}_j \in R^{(i)}$, $\mathbf{u}_l \in U$ and $\alpha_{j,l}^{(i)}$ are *a priori* probabilities of $\mathbf{r}_j + \mathbf{u}_l$. Because all the $\mathbf{r}_j + \mathbf{u}_l$ are equiprobable and the noise PDF is Gaussian, the minimum error probability decision is achieved by computing the Bayesian decision variables:

$$\eta_i(k, \mathbf{a}_0) = \sum_{j=1}^{N_s^{(i)}} \sum_{l=1}^{N_u} \exp(-\|\mathbf{r}'(k) - \mathbf{r}'_j - \mathbf{u}_l\|^2 / 2\sigma_e^2) \quad (21)$$

and making the decision according to eqn. 14.

The computational complexity of the Bayesian DFE without CCI compensation is an order of N_s [3, 5]. The complexity of the Bayesian DFE with full CCI compensation is thus an order of $N_s \times N_u$. To reduce the complexity, an approximation of this full Bayesian DFE can be adopted which only approximates co-channel states. The approximation can easily be achieved due to the symmetric structure of co-channel states, and this will be illustrated using an example. Another reason for adopting the approximation is due to practical considerations. The scalar co-channel states u_l can only be estimated based on unsupervised learning. The resolution of unsupervised learning is limited, and it is not always possible to resolve all the co-channel states. In such a situation, it is natural to consider an approximation. Carrying out the approximation to an extreme and approximating the CCI as an additional noise, we obtain the Bayesian DFE with the decision variables:

$$\eta_i(k, \mathbf{a}_0) = \sum_{j=1}^{N_s^{(i)}} \exp(-\|\mathbf{r}'(k) - \mathbf{r}'_j\|^2 / 2\sigma^2), \quad 1 \leq i \leq 2 \quad (22)$$

where $\sigma^2 = \sigma_e^2 + \sigma_u^2$. This has the same form as the Bayesian DFE in the absence of CCI.

Table 1: Scalar co-channel states for $A_1(z) = \lambda(0.50 + 0.81z^{-1} + 0.31z^{-2})$

No.	$s_1(k)$	$s_1(k-1)$	$s_1(k-2)$	u_l
1	1	1	1	1.62 (λ)
2	1	1	-1	1.00 (λ)
3	1	-1	1	0.00 (λ)
4	1	-1	-1	-0.62 (λ)
5	-1	1	1	0.62 (λ)
6	-1	1	-1	0.00 (λ)
7	-1	-1	1	-1.00 (λ)
8	-1	-1	-1	-1.62 (λ)

We now use an example to illustrate the above discussion and to compare the theoretical performance of the Bayesian DFE with that of the MLSE which only treats the CCI as noise. The channel and the interfering co-channel are given by

$$\left. \begin{aligned} A_0(z) &= 0.34 + 0.88z^{-1} + 0.34z^{-2} \\ A_1(z) &= \lambda(0.50 + 0.81z^{-1} + 0.31z^{-2}) \end{aligned} \right\} \quad (23)$$

where the value of the parameter λ dictates the SIR requirement. For example, $\lambda = 0.32$ gives rise to a SIR = 10dB. The set of the scalar co-channel states are listed in Table 1. The symmetric structure of the co-channel states is apparent in Table 1. In general, this symmetric structure is expressed by the relationship:

$$u_{N_{u,s}-l+1} = -u_l, \quad 1 \leq l \leq N_{u,s}/2 \quad (24)$$

The set of the vector co-channel states U is obtained by expanding the scalar states. In this example, U contains 32 vector states as listed in Table 2. The rule to expand the set of the scalar co-channel states into the set of the vector co-channel states can be seen from Table 2. In general, in the table of the vector co-channel states, the last column (corresponding to $u(k-m+1)$) is repeatedly filled with

$$\underbrace{u_1}_{2^0} \underbrace{u_2}_{2^0} \cdots \underbrace{u_{N_{u,s}}}_{2^0}$$

the column corresponding to $u(k-m+2)$ is repeatedly filled with

$$\underbrace{u_1 u_1}_{2^1} \underbrace{u_2 u_2}_{2^1} \cdots \underbrace{u_{N_{u,s}} u_{N_{u,s}}}_{2^1}$$

..., and the first column (corresponding to $u(k)$) is filled with

$$\underbrace{u_1 \cdots u_1}_{2^{m-1}} \underbrace{u_2 \cdots u_2}_{2^{m-1}} \cdots \underbrace{u_{N_{u,s}} \cdots u_{N_{u,s}}}_{2^{m-1}}$$

Table 2: Vector co-channel states for $A_1(z) = \lambda(0.50 + 0.81z^{-1} + 0.31z^{-2})$

No.	$s_1(k)$	$s_1(k-1)$	$s_1(k-2)$	$s_1(k-3)$	$s_1(k-4)$	u_l
1	1	1	1	1	1	$u_1 \ u_1 \ u_1$
2	1	1	1	1	-1	$u_1 \ u_1 \ u_2$
3	1	1	1	-1	1	$u_1 \ u_2 \ u_3$
4	1	1	1	-1	-1	$u_1 \ u_2 \ u_4$
5	1	1	-1	1	1	$u_2 \ u_3 \ u_5$
6	1	1	-1	1	-1	$u_2 \ u_3 \ u_6$
7	1	1	-1	-1	1	$u_2 \ u_4 \ u_7$
8	1	1	-1	-1	-1	$u_2 \ u_4 \ u_8$
9	1	-1	1	1	1	$u_3 \ u_5 \ u_1$
10	1	-1	1	1	-1	$u_3 \ u_5 \ u_2$
11	1	-1	1	-1	1	$u_3 \ u_6 \ u_3$
12	1	-1	1	-1	-1	$u_3 \ u_6 \ u_4$
13	1	-1	-1	1	1	$u_4 \ u_7 \ u_5$
14	1	-1	-1	1	-1	$u_4 \ u_7 \ u_6$
15	1	-1	-1	-1	1	$u_4 \ u_8 \ u_7$
16	1	-1	-1	-1	-1	$u_4 \ u_8 \ u_8$
17	-1	1	1	1	1	$u_5 \ u_1 \ u_1$
18	-1	1	1	1	-1	$u_5 \ u_1 \ u_2$
19	-1	1	1	-1	1	$u_5 \ u_2 \ u_3$
20	-1	1	1	-1	-1	$u_5 \ u_2 \ u_4$
21	-1	1	-1	1	1	$u_6 \ u_3 \ u_5$
22	-1	1	-1	1	-1	$u_6 \ u_3 \ u_6$
23	-1	1	-1	-1	1	$u_6 \ u_4 \ u_7$
24	-1	1	-1	-1	-1	$u_6 \ u_4 \ u_8$
25	-1	-1	1	1	1	$u_7 \ u_5 \ u_1$
26	-1	-1	1	1	-1	$u_7 \ u_5 \ u_2$
27	-1	-1	1	-1	1	$u_7 \ u_6 \ u_3$
28	-1	-1	1	-1	-1	$u_7 \ u_6 \ u_4$
29	-1	-1	-1	1	1	$u_8 \ u_7 \ u_5$
30	-1	-1	-1	1	-1	$u_8 \ u_7 \ u_6$
31	-1	-1	-1	-1	1	$u_8 \ u_8 \ u_7$
32	-1	-1	-1	-1	-1	$u_8 \ u_8 \ u_8$

In this expansion, to obtain correctly the set of the vector co-channel states as shown in Table 2, we need to know the correct order of the scalar co-channel states as indicated in Table 1. The clustering algorithm described in the next Section can only identify the values of the scalar co-channel states and does not provide the information regarding their order. The order of the scalar co-channel states can be sorted out with the help of the state transition diagram. For the case of the eight scalar states, Fig. 4 depicts the state transition diagram. After the set of the eight scalar states has been obtained, by observing a sequence of states through time, their order can easily be arranged

according to the state transition diagram. The symmetric structure of the state transition diagram and the relationship of eqn. 24 helps to speed up this ordering process.

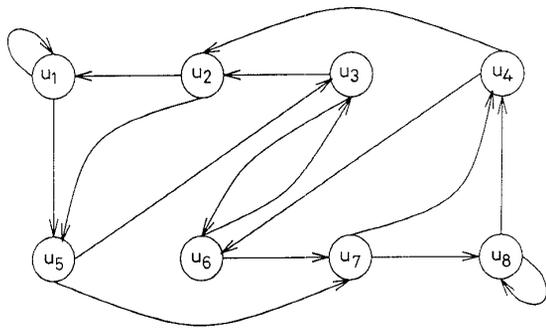


Fig. 4 State transition diagram for case of eight scalar co-channel states

Rearrange the eight scalar co-channel states of Table 1 into

$$(1.62\lambda, 1.00\lambda, 0.62\lambda, 0.00\lambda, -0.00\lambda, -0.62\lambda, -1.00\lambda, -1.62\lambda) \quad (25)$$

We may approximate $(1.62\lambda, 1.00\lambda)$ by its mean 1.31λ and $(0.62\lambda, 0.00\lambda)$ by 0.31λ . This results in four approximated scalar co-channel states:

$$(1.31\lambda, 0.31\lambda, -0.31\lambda, -1.31\lambda) \quad (26)$$

The number of resulting approximated vector co-channel states is 16. This approximation may also be viewed from a different angle. The order of the co-channel is $n_1 = 3$. Suppose that we only have an approximated co-channel order $\hat{n}_1 = 2$. This will give us four scalar co-channel states, and each of these approximated states is the mean of a pair of the true states. These four approximated scalar co-channel states are listed in Table 3 in the correct ordering, and the state transition diagram for the case of the four states is shown in Fig. 5. The Bayesian DFE with decision variables described by eqn. 22 may be viewed as the result of choosing $\hat{n}_1 = 0$.

Table 3: Approximated scalar co-channel states assuming $\hat{n}_1 = 2$ for $A_1(z) = \lambda(0.50 + 0.81z^{-1} + 0.31z^{-2})$

No.	$s_1(k)$	$s_1(k-1)$	u_i
1	1	1	$1.31(\lambda)$
2	1	-1	$-0.31(\lambda)$
3	-1	1	$0.31(\lambda)$
4	-1	-1	$-1.31(\lambda)$

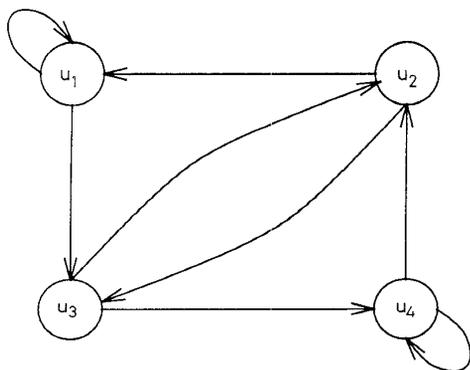


Fig. 5 State transition diagram for case of four scalar co-channel states

Table 4: SIR, SNR and SINR values used to obtain the results of Figs. 6–8

Fig.	SIR	SNR	SINR
6	5 dB	2 – 28 dB	0.3 – 5.0 dB
7	10 dB	2 – 25 dB	1.4 – 9.8 dB
8	15 dB	2 – 21 dB	1.8 – 14.0 dB

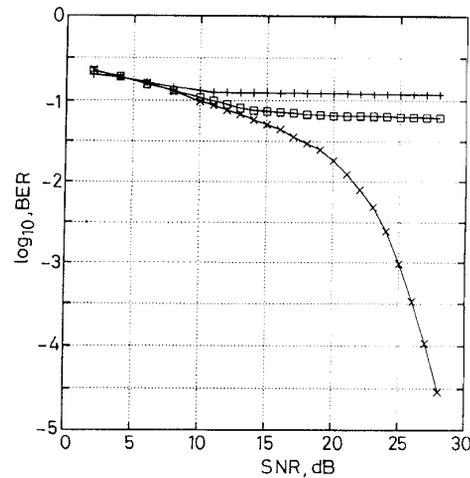


Fig. 6 Theoretical performance for $SIR = 5\text{ dB}$
 —+— MLSE treating CCI as noise
 —□— Bayesian DFE with approximated CCI compensation ($\hat{n}_1 = 2$)
 —x— Bayesian DFE with full CCI compensation ($\hat{n}_1 = n_1 = 3$)
 Bayesian DFE without CCI compensation ($\hat{n}_1 = 0$) is similar to —+—

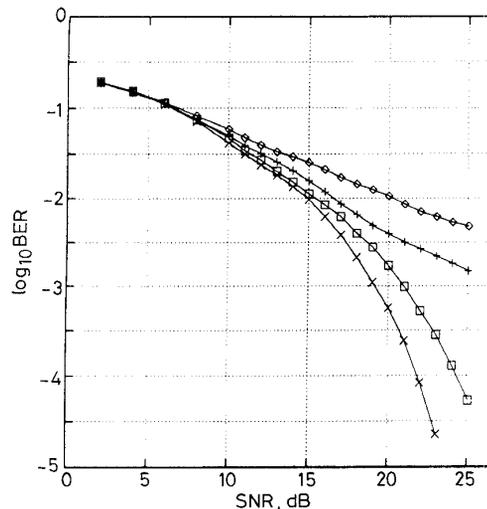


Fig. 7 Theoretical performance for $SIR = 10\text{ dB}$
 —◇— Bayesian DFE without CCI compensation ($\hat{n}_1 = 0$)
 —+— MLSE treating CCI as noise
 —□— Bayesian DFE with approximated CCI compensation ($\hat{n}_1 = 2$)
 —x— Bayesian DFE with full CCI compensation ($\hat{n}_1 = n_1 = 3$)

Figs. 6 to 8 plot the performance curves of the Bayesian DFE without CCI compensation ($\hat{n}_1 = 0$), the MLSE which treats CCI as noise, the Bayesian DFE with an approximated CCI compensation ($\hat{n}_1 = 2$) and the Bayesian DFE with the full CCI compensation ($\hat{n}_1 = n_1 = 3$) for three different SIR conditions respectively. Table 4 summarises the SIR, SNR and SINR values used to obtain the results shown in Figs. 6 to 8. The performance of the Bayesian DFEs were obtained with detected symbols being fed back. When the CCI is negligible, the MLSE has superior performance over the Bayesian DFE, as can be seen from the results of Fig. 8. However, in the presence of severe CCI, the

Bayesian DFE with an effective compensation of the CCI can outperform the MLSE that only treats the CCI as noise, as clearly shown in Figs. 6 and 7.

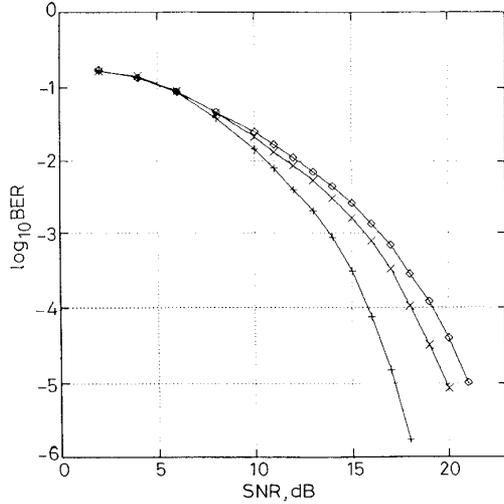


Fig. 8 Theoretical performance for $SIR = 15$ dB
 —◇— Bayesian DFE without CCI compensation ($\hat{n}_1 = 0$)
 —+— MLSE treating CCI as noise
 —×— Bayesian DFE with full CCI compensation ($\hat{n}_1 = n_1 = 3$)

It should be emphasised that the performance of the MLSE which fully takes into account the ISI and CCI will always be superior over the Bayesian DFE with full CCI compensation. As discussed previously, such a full MLSE is practically very difficult to implement. In contrast, the Bayesian DFE with CCI compensation can easily be implemented, as shown in the next Section. Providing the Bayesian DFE with ‘incorrect’ or approximate co-channel states, performance can improve compared with no CCI compensation. Given ‘incorrect’ co-channels to the MLSE, performance is likely to degrade because of the accumulation of estimate errors. Evidence from the fading channel equalisation [5] supports this view. In terms of computational complexity, the full MLSE is more complex than the full Bayesian DFE. In practice, the MLSE is usually implemented by considering CCI as additional noise, which is less complex than the Bayesian DFE with full CCI compensation. The Bayesian DFE with approximate CCI compensation however can be designed to be less complex than the MLSE that treats CCI as noise, depending on approximation level.

4 Adaptive implementation

Since the equaliser has access to the transmitted data $\{s_0(k)\}$ during the training and the estimated data $\{\hat{s}_0(k)\}$ during the data communication, identifying the channel states \mathbf{r}'_j presents no difficulty. The channel model \mathbf{a}_0 can be identified, for example, using the normalised least square (NLMS) algorithm:

$$\left. \begin{aligned} \epsilon(k) &= r(k) - \sum_{j=0}^{n_0-1} \hat{a}_{0,j}(k-1)s_0(k-j) \\ \rho(k) &= \sum_{j=0}^{n_0-1} s_0^2(k-j) \\ \hat{a}_{0,j}(k) &= \hat{a}_{0,j}(k-1) + \frac{g_a}{\rho(k)}\epsilon(k)s_0(k-j) \\ 0 &\leq j \leq n_0 - 1 \end{aligned} \right\} \quad (27)$$

where g_a is an adaptive gain. Given the channel estimate $\hat{\mathbf{a}}_0$, it is straightforward to calculate the set of the channel states \mathbf{r}'_j from eqn. 17.

The equaliser does not have access to the interfering data $\{s_1(k)\}$ or its estimate, and supervised learning such as the NLMS algorithm is not applicable for identifying the co-channel states. In the previous study [7], the unsupervised κ -means clustering algorithm [10] is used to estimate the co-channel states. The κ -means clustering algorithm is known to be sensitive to the initial positions of the cluster centres. Recently, an enhanced κ -means clustering algorithm has been proposed [11], which overcomes this drawback. This enhanced κ -means clustering algorithm is optimal in the sense that the variances of every cluster are equal after convergence. This property is particularly relevant for the application to estimate co-channel states since all the cluster variances in this case should be equal. Using this enhanced κ -means clustering algorithm, we propose the following procedure to estimate the co-channel states:

(i) Compute the channel residual

$$\epsilon(k) = r(k) - \sum_{j=0}^{n_0-1} \hat{a}_{0,j}s_0(k-j) \quad (28)$$

where $\hat{\mathbf{a}}_0 = [\hat{a}_{0,0} \dots \hat{a}_{0,n_0-1}]^T$ is the current channel estimate.

(ii) Compute the cluster variance weighted squared distances between the residual $\epsilon(k)$ and the scalar co-channel states $u_l(k-1)$, $1 \leq l \leq \hat{N}_{u,s}$

$$\begin{aligned} \bar{\zeta}_l(k) &= v_l(k-1)\zeta_l(k) \\ &= v_l(k-1)(\epsilon(k) - u_l(k-1))^2, \quad 1 \leq l \leq \hat{N}_{u,s} \end{aligned} \quad (29)$$

where $\hat{N}_{u,s} = 2^{\hat{n}_1}$, \hat{n}_1 is an estimate of the co-channel order, $v_l(k-1)$ is the current variance of the l th cluster and $\zeta_l(k)$ is the squared distance between $\epsilon(k)$ and $u_l(k-1)$. Find the minimum weighted distance:

$$\bar{\zeta}_l^*(k) = \min\{\bar{\zeta}_l(k), 1 \leq l \leq \hat{N}_{u,s}\} \quad (30)$$

(iii) Update the l^* th and $(\hat{N}_{u,s} - l^* + 1)$ th co-channel states:

$$\left. \begin{aligned} u_{l^*}(k) &= u_{l^*}(k-1) + g_u(\epsilon(k) - u_{l^*}(k-1)) \\ u_{\hat{N}_{u,s}-l^*+1}(k) &= -u_{l^*}(k) \end{aligned} \right\} \quad (31)$$

where g_u is an adaptive gain. The cluster variances are then updated according to the rule:

$$\left. \begin{aligned} v_l(k) &= \alpha v_l(k-1), \quad 1 \leq l \leq \hat{N}_{u,s} \text{ and } l \neq l^*, \hat{N}_{u,s} - l^* + 1 \\ v_{\hat{N}_{u,s}-l^*+1}(k) &= v_{l^*}(k) = \alpha v_{l^*}(k-1) + (1.0 - \alpha)\bar{\zeta}_{l^*}^*(k) \end{aligned} \right\} \quad (32)$$

where α is a positive constant slightly less than 1.0. The initial $v_l(0)$, $1 \leq l \leq \hat{N}_{u,s}$ can be set to a same small value. Setting $v_{\hat{N}_{u,s}-l^*+1}(k) = v_{l^*}(k)$ together with the rule $v_{\hat{N}_{u,s}-l^*+1} = -u_{l^*}(k)$ exploits the symmetric structure of the co-channel states, and this accelerates convergence.

The scalar co-channel states are then arranged in the correct order and expanded to obtain the set of the vector co-channel states, U . Alternatively, the vector co-channel states can be estimated directly using the same κ -means clustering algorithm. This however requires longer learning, since the number of the vector co-channel states is much larger. An advantage of the latter is that there is no need to order the vector states. The resolution of the above unsupervised clustering

algorithm obviously depends on the noise and the co-channel itself. A common-sense rule based on the amplitude of the channel residual can be adopted to choose the number of scalar co-channel states needed. For example, if the channel residual lies in the range $(-0.1, 0.1)$, it is unnecessary to choose $\hat{N}_{u,s} > 8$. In this case, it may be sufficient to choose $\hat{N}_{u,s} = 4$, regardless of the true number of the scalar co-channel states.

For the system defined in eqn. 23 with SIR = 10dB and SNR = 15dB, the combined NLMS and clustering learning was used to estimate the channel model and the scalar co-channel states. The channel order was assumed to be known and only an estimated co-channel order $\hat{n}_1 = 2$ was assumed to be available. This gave rise to the four scalar co-channel states. The gain of the NLMS algorithm was chosen to be $g_a = 0.08$. The parameters of the clustering procedure were set to: $\alpha = 0.999$, $g_u = 0.05$ and $v_l(0) = 0.000001$ for all l . Fig. 9 depicts a typical set of the scalar co-channel state trajectories obtained, where the lines indicate the expected values.

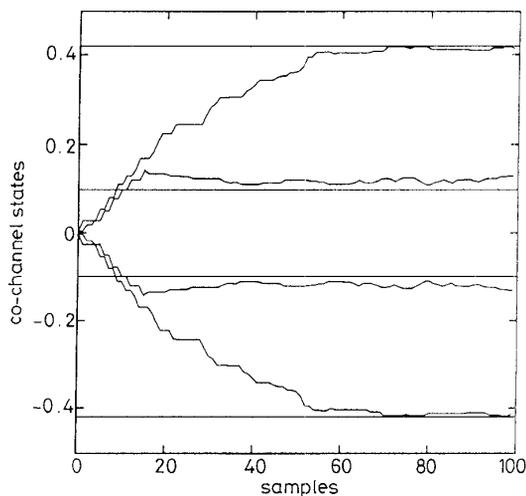


Fig. 9 Trajectories of scalar co-channel states obtained using clustering algorithm
Lines indicate expected values

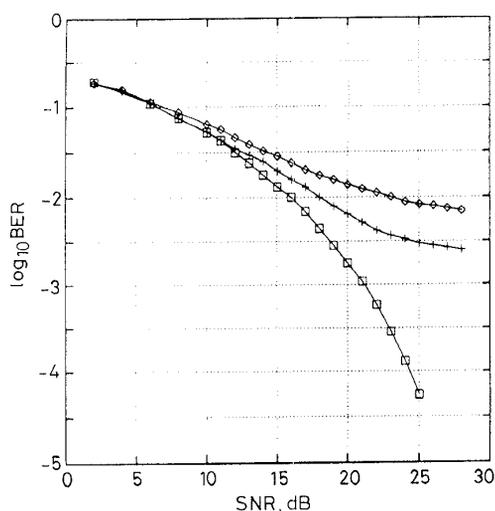


Fig. 10 Adaptive performance for SIR = 10dB
 —◇— Bayesian DFE without CCI compensation ($\hat{n}_1 = 0$)
 —+— MLSE treating CCI as noise
 —□— Bayesian DFE with approximated CCI compensation ($\hat{n}_1 = 2$)

For the same system (eqn. 23) with SIR = 10dB, Fig. 10 compares the adaptive performance of the Bayesian DFE without CCI compensation ($\hat{n}_1 = 0$), the MLSE which only treats CCI as noise and the Bayesian DFE with an approximate CCI compensation ($\hat{n}_1 = 2$). In the first two cases, the NLMS algorithm used 100 training pairs (channel observations and transmitted symbols) to identify the channel model. For the last case, in addition, the clustering algorithm used 100 channel observation samples to estimate the four scalar co-channel states. The adaptive performance of the Bayesian DFE with an approximate CCI compensation is very close to its theoretical performance, and is significantly better than that of the MLSE without CCI compensation.

5 Conclusions

Adaptive equalisation in the presence of ISI, additive Gaussian white noise and CCI has been investigated. It has been shown that, by exploiting the nature of interfering signals, the Bayesian DFE is capable of distinguishing an interfering signal from the noise. Simulation results have demonstrated that, in the presence of severe CCI, the Bayesian DFE which incorporates CCI compensation can outperform the MLSE without CCI compensation. In theory, if an accurate knowledge of the channel and co-channels is known, the MLSE can be designed to take into account both the ISI and CCI and hence outperforms the Bayesian DFE. In practice, however, adaptive implementation of such a MLSE is very difficult. Adaptive implementation of the Bayesian DFE has been studied, and a simple unsupervised clustering algorithm has been suggested to learn the co-channel states. This adaptive Bayesian DFE is particularly effective in compensating one or a few dominant interferences. A drawback of this adaptive scheme is that its computational complexity increases quickly as the size of the symbol constellation increases.

6 References

- FORNEY, G.D.: 'Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference', *IEEE Trans.*, 1972, **IT-18**, (3), pp. 363-378
- QURESHI, S.U.H.: 'Adaptive equalization', *Proc. IEEE*, 1985, **73**, (9), pp. 1349-1387
- CHEN, S., MULGREW, B., and MCLAUGHLIN, S.: 'Adaptive Bayesian equaliser with decision feedback', *IEEE Trans.*, 1993, **SP-41**, (9), pp. 2918-2927
- CHEN, S., MCLAUGHLIN, S., and MULGREW, B.: 'Complex-valued radial basis function network. Part II: application to digital communications channel equalisation', *EURASIP Signal Proc. J.*, 1994, **36**, pp. 175-188
- CHEN, S., MCLAUGHLIN, S., MULGREW, B., and GRANT, P.M.: 'Adaptive Bayesian decision feedback equaliser for dispersive mobile radio channels', *IEEE Trans.*, 1995, **COM-43**, (5), pp. 1937-1946
- PETERSEN, B.R., and FALCONER, D.D.: 'Exploiting cyclostationary subscriber-loop interference by equalization'. Proceedings of GLOBECOM'90, San Diego, 1990, pp. 1156-1160
- CHEN, S., and MULGREW, B.: 'Overcoming co-channel interference using an adaptive radial basis function equaliser', *EURASIP Signal Proc. J.*, 1992, **28**, (1), pp. 91-107
- LO, N.W.K., FALCONER, D.D., and SHEIKH, A.U.H.: 'Adaptive equalization for co-channel interference in a multipath fading environment', *IEEE Trans.*, 1995, **COM-43**, (5), pp. 1441-1453
- CAMPBELL, J.C., GIBBS, A.J., and SMITH, B.M.: 'The cyclostationary nature of crosstalk interference from digital signals in multipair cable. Part I: fundamentals; Part II: applications and further results', *IEEE Trans.*, 1983, **COM-31**, (5), pp. 629-649
- DUDA, R.O., and HART, P.E.: 'Pattern classification and scene analysis' (John Wiley, New York, 1973)
- CHINRUNGRUENG, C., and SEQUIN, C.H.: 'Optimal adaptive κ -means algorithm with dynamic adjustment of learning rate', *IEEE Trans.*, 1995, **NN-6**, (1), pp. 157-169