

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

Available online at www.sciencedirect.com

Signal Processing 88 (2008) 1671–1697

**SIGNAL
PROCESSING**www.elsevier.com/locate/sigpro

Adaptive minimum error-rate filtering design: A review

Sheng Chen^{*}, Shuang Tan, Lei Xu, Lajos Hanzo*School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK*

Received 1 May 2007; received in revised form 10 December 2007; accepted 3 January 2008

Available online 20 January 2008

Abstract

Adaptive filtering has been an enabling technology and has found ever-increasing applications in various state-of-the-art communication systems. Traditionally, adaptive filtering has been developed based on the Wiener or minimum mean square error (MMSE) approach, and the famous least mean square algorithm with its low computational complexity readily meets the fast real-time computational constraint of modern high-speed communication systems. For a communication system, however, it is the system's bit error rate (BER), not the mean square error (MSE), that really matters. It has been recognised that minimising the MSE criterion does not necessarily produce the minimum BER (MBER) performance. The introduction of the novel MBER design has opened up a whole new chapter in the optimisation of communication systems, and its design trade-offs have to be documented in contrast to those of the classic but actually still unexhausted MMSE and other often-used optimisation criteria. This contribution continues this theme, and we provide a generic framework for adaptive minimum error-probability filter design suitable for the employment in a variety of communication systems. Advantages and disadvantages of the adaptive minimum error-probability filter design are analysed extensively, in comparison with the classic Wiener filter design.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Adaptive filter; Wiener solution; Minimum mean square error; Minimum bit error rate; Stochastic gradient algorithm; Least mean square algorithm; Least bit error rate algorithm

1. Introduction

Adaptive signal processing has been an enabling technology for the contemporary information society, and adaptive filtering has found wide-ranging applications in modern communication systems. A generic communication system typically includes an inner Modem part and an outer Codec part [1,2]. A

variety of adaptive filters can be found in communication Modems to act as receiver filters or detectors. The state-of-the-art adaptive filtering design for communication applications has traditionally been developed based on the Wiener, also known as the minimum mean square error (MMSE), framework [3,4], and adaptive implementation of the MMSE design can readily be achieved with the low-complexity least mean square (LMS) algorithm which readily meets the real-time computational constraint of modern high-speed communication systems. The Wiener filtering design has its root in other applications of adaptive filtering, such as radar and sonar. For

^{*}Corresponding author. Tel.: +44 23 8059 6660;
fax: +44 23 8059 4508.

E-mail addresses: sqc@ecs.soton.ac.uk (S. Chen),
st104r@ecs.soton.ac.uk (S. Tan), lx04r@ecs.soton.ac.uk (L. Xu),
lh@ecs.soton.ac.uk (L. Hanzo).

communication applications, however, what really matters is the system's bit error rate (BER), not the achievable mean square error (MSE). This has motivated the research for the alternative approach to the MMSE filtering that aims to directly minimise the system's BER.

In the past decade, significant advances have been made in the design of adaptive minimum BER (MBER) filtering for a variety of communication applications, including classical single-user channel equalisation [5–20], multiuser detection in code-division multiple-access (CDMA) systems [21–30], adaptive beamforming assisted receiver for multiple-antenna aided systems [31–39], space-time equalisation assisted multiuser detection for space-division multiple-access (SDMA) induced multiple-input multiple-output (MIMO) systems [40–44], and orthogonal frequency division multiplexing (OFDM) and other multi-carrier systems [45–50]. The MBER filtering design has also been incorporated into turbo iterative detection [51–53]. Other applications include the MBER transmission schemes [54–56], the MBER rake receiver [57,58], the MBER-based optical receiver [59–62], and the power control or allocation based on the MBER criterion [63–66]. It can be seen that the introduction of the novel MBER design has opened up a new research direction and has stimulated wide interests in the communication research community. The purpose of this contribution is to provide a unified framework for the adaptive MBER filtering design and to document its design trade-offs in comparison to the standard yet still often considered MMSE design.

We begin our discussion with an introduction of the generic signal and filter model. Such a filter can be a pure temporal filter as in the classical single-user channel equalisation [1], a pure spatial filter as in the adaptive beamforming assisted receiver for narrow-band channels [67–72], or a combined spatial and temporal filter as in the case of space-time equalisation assisted multiuser detection for SDMA induced MIMO systems [73–75]. This signal and filter model is in fact valid for all the state-of-the-art communication systems. The classical Wiener filter design is then reviewed, and the condition for the Wiener solution to be the optimal MBER solution is discussed. Using the example of the well known matched filter solution for the ideal additive white Gaussian noise (AWGN) channel, it is demonstrated that the MMSE design is also the optimal MBER design only if the conditional

probability density function (PDF) of the filter output for a given transmitted data symbol value is Gaussian. Since this conditional PDF is generally a mixture of Gaussian distributions, and hence non-Gaussian, the MMSE design is inherently sub-optimal with respect to the achievable system's BER.

Based on the generic signal and filter model, the BER expression as the function of the filter's weight vector is derived, and this naturally leads to the MBER design. Adaptive implementation of the optimal MBER filtering design is discussed in full details, and emphasis is placed on the sample-by-sample adaptive algorithm referred to as the least bit error rate (LBER) method [76]. Comparisons are drawn with the adaptive LMS algorithm. Examples are used to highlight the basic concepts and essential properties, as well as to draw insights into how the two designs, the MMSE and MBER, behave differently. Unlike the Wiener filtering whose optimality is linked to the Gaussian assumption, the MBER design, by contrast, can cleverly exploit the non-Gaussian distribution, leading to substantial performance improvements over the sub-optimal MMSE design. Thus, compared with the standard MMSE filtering design which is still often regarded as a state of the art in communication applications, the adaptive MBER filtering design can better combat hostile multipath propagation environments and better suppress multiple access interference, resulting in higher system throughput or user capacity. The penalty to pay for this enhanced system performance is an increase in complexity for the adaptive MBER filter design.

Although the traditional adaptive MMSE filtering design and the novel adaptive MBER filtering design are based on the two very different optimisation criteria, it is interesting to draw some analogy between the two approaches. The MSE is the second-order statistics, while the BER can be viewed as a higher-order statistics, of the underlying filter output's PDF. The second-order statistics required to compute the Wiener solution can be estimated using a block of samples, and by considering a single-sample "estimate" of this second-order statistics, a stochastic gradient adaptive MMSE algorithm, namely the LMS, is derived. The PDF required to determine the BER can be approximated with a Parzen window estimate [77–79] based on a block of samples, and by considering a single-sample

density “estimate”, a stochastic gradient adaptive MBER algorithm is formulated, which is referred to as the LBER algorithm in [76]. For the sake of clearly highlighting the basic concepts, we start our discussion assuming a binary phase shift keying (BPSK) modulation scheme, i.e. data symbols being binary. The approach is then generalised to the minimum symbol error rate (MSER) design for the bandwidth-efficient, high-throughput quadrature amplitude modulation (QAM) scheme. The adaptive filtering model considered in this article is *linear*. Extension of the adaptive MBER design to *nonlinear* filtering is discussed at the end.

2. Signal and filter model

The schematic diagram for the Modem part of a generic communication system is depicted in Fig. 1. The system or channel model is represented by

$$\mathbf{x}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{n}(k), \quad (1)$$

where $\mathbf{s}(k) = [s_1(k) \ s_2(k) \ \dots \ s_M(k)]^T$ is the transmitted data symbol vector at k th symbol index with the uncorrelated BPSK data symbols

$$s_i(k) \in \{\pm 1\}, \quad 1 \leq i \leq M, \quad (2)$$

$\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \dots \ x_L(k)]^T$ denotes the complex-valued received signal vector, \mathbf{H} the $L \times M$ complex-valued channel convolution matrix, and

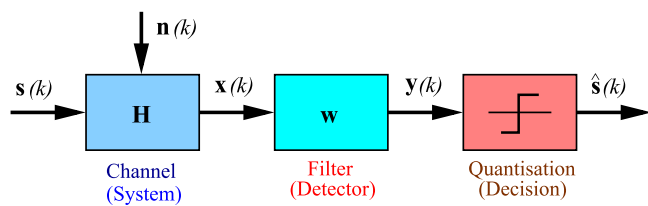


Fig. 1. Schematic diagram of generic communication system.

$\mathbf{n}(k) = [n_1(k) \ n_2(k) \ \dots \ n_L(k)]^T$ is the complex-valued channel AWGN vector with $E[\mathbf{n}(k)\mathbf{n}^H(k)] = 2\sigma_n^2\mathbf{I}_L$ and \mathbf{I}_L denoting the $L \times L$ identity matrix. For the time being, we assume that the transmitted data symbols are binary. Later we will extend the results to the case of QAM data symbols.

The receiver consists of a linear filter characterised by

$$y(k) = \mathbf{w}^H\mathbf{x}(k), \quad (3)$$

where $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_L]^T$ is the complex-valued filter’s coefficient vector. The filter’s output $y(k)$ is passed to the decision device to provide an estimate for the desired user’s data symbol $s_d(k)$, where $1 \leq d \leq M$, and the decision is made according to

$$\hat{s}_d(k) = \text{sgn}(y_R(k)) = \begin{cases} +1 & y_R(k) \geq 0, \\ -1 & y_R(k) < 0, \end{cases} \quad (4)$$

where $y_R(k) = \Re[y(k)]$ is the real-part of $y(k)$. For notational simplification, we avoid using the index d in $y(k)$.

The system and filter models (1) and (3) is very general and is in fact valid for all the state-of-the-art communication systems. We now illustrate the generality of this model using a few examples of the practical communication systems encountered.

2.1. Channel equalisation

Consider the classical single-antenna single-user channel equalisation [1], as depicted in Fig. 2. The multipath distorting channel is characterised by

$$x(k) = \sum_{i=0}^{n_c-1} h_i s(k-i) + n(k), \quad (5)$$

where $n(k)$ is the channel AWGN, n_c is known as the channel order, and h_i are the channel impulse response (CIR) taps. The equaliser $y(k) = \mathbf{w}^H\mathbf{x}(k)$ of

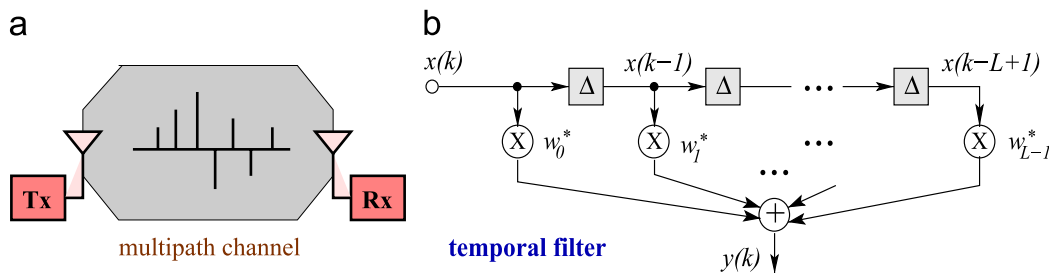


Fig. 2. Multipath channel of single-antenna single-user system (a), and channel equalisation based on a temporal filter (b), where Δ denotes the symbol-spaced delay.

(3) is a pure temporal filter with an order L and a decision delay d . That is, at symbol index k , it detects the transmitted symbol $s(k-d)$. It is straightforward to verify that the signal model for $\mathbf{x}(k)$ is given by (1) with

$$\mathbf{x}(k) = [x(k) \ x(k-1) \ \cdots \ x(k-L+1)]^T,$$

$$\mathbf{s}(k) = [s(k) \ s(k-1) \ \cdots \ s(k-L-n_c+2)]^T,$$

$M = L + n_c - 1$, and the $L \times M$ channel convolution matrix has the following Toeplitz form

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n_c-1} & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_{n_c-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_{n_c-1} \end{bmatrix}. \quad (6)$$

Note that for the decision feedback equaliser (DFE), the decision feedback can be interpreted as a space translation and, on the translated signal space, the DFE takes the same form of a linear equaliser (3) (see, for example, [8,20]).

2.2. Beamforming aided receiver

Consider a communication system that supports M users, where each user transmits on the same carrier frequency of $\omega = 2\pi f$, as illustrated in Fig. 3(a). For such a system, user separation can be achieved in the spatial or angular domain [73–75], and the receiver is equipped with a linear antenna array consisting of L uniformly spaced elements. The geometric structure of the receiver antenna array with respect to the user's direction or angle of arrival (AOA) is depicted in Fig. 3(b). Further assume that the channel is non-dispersive and hence it does not induce intersymbol inter-

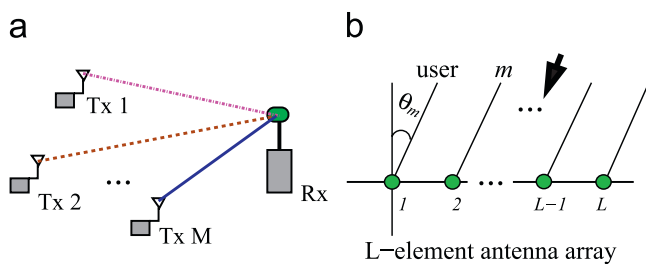


Fig. 3. Beamforming assisted receiver that employs L -element antenna array to support M users: (a) system illustration, and (b) geometric structure of receiver antenna array, where θ_m is the angle of arrival of user m .

ference. Then the received signal vector of the antenna array

$$\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \cdots \ x_L(k)]^T,$$

is modelled as (1), and the $L \times M$ system matrix \mathbf{H} is given by [67,72]

$$\mathbf{H} = [A_1 \mathbf{h}_1 \ A_2 \mathbf{h}_2 \ \cdots \ A_M \mathbf{h}_M], \quad (7)$$

where A_m denotes the m th narrowband complex-valued channel coefficient, and \mathbf{h}_m the steering vector for user m , which is expressed as

$$\mathbf{h}_m = [e^{j\omega t_1(\theta_m)} \ e^{j\omega t_2(\theta_m)} \ \cdots \ e^{j\omega t_L(\theta_m)}]^T \quad (8)$$

with $t_l(\theta_m)$ denoting the relative time delay at array element l for source m , and θ_m being the AOA for source m . The beamformer for detecting the transmitted data symbol of desired user m , $s_m(k)$, is a pure spatial filter expressed as [67,72]

$$y_m(k) = \mathbf{w}_m^H \mathbf{x}(k), \quad (9)$$

where \mathbf{w}_m is the m th beamformer's complex-valued weight vector.

2.3. MIMO space-time equalisation

Consider the SDMA induced MIMO system [73–75], as is depicted in Fig. 4, where each of the Q users is equipped with a single transmit antenna and the receiver is assisted by a P -element antenna array. A bank of the Q space-time equalisers, as shown in Fig. 5, constitutes the multiuser detector. Each space-time equaliser is a combined spatial and temporal filter, with the order of temporal filter being D . It can be shown that the space-time equaliser for user q , where $1 \leq q \leq Q$, is given by the form of $y_q(k) = \mathbf{w}_q^H \mathbf{x}(k)$ with the signal model for $\mathbf{x}(k)$ expressed in the form of (1), see for example [43]. The output of the q th combined spatial and temporal filter, $y_q(k)$, is used to detect the transmitted symbol $s_q(k-\tau)$, where τ is the decision delay of the space-time equaliser. In fact, the signal vector $\mathbf{x}(k)$ is defined as

$$\mathbf{x}(k) = [\mathbf{x}_1^T(k) \ \mathbf{x}_2^T(k) \ \cdots \ \mathbf{x}_p^T(k)]^T \quad (10)$$

with

$$\mathbf{x}_p(k) = [x_p(k) \ x_p(k-1) \ \cdots \ x_p(k-D+1)]^T, \quad 1 \leq p \leq P, \quad (11)$$

and the filter coefficient vector for the q th space-time equaliser is given by

$$\mathbf{w}_q = [\mathbf{w}_{1,q}^T \ \mathbf{w}_{2,q}^T \ \cdots \ \mathbf{w}_{P,q}^T]^T \quad (12)$$

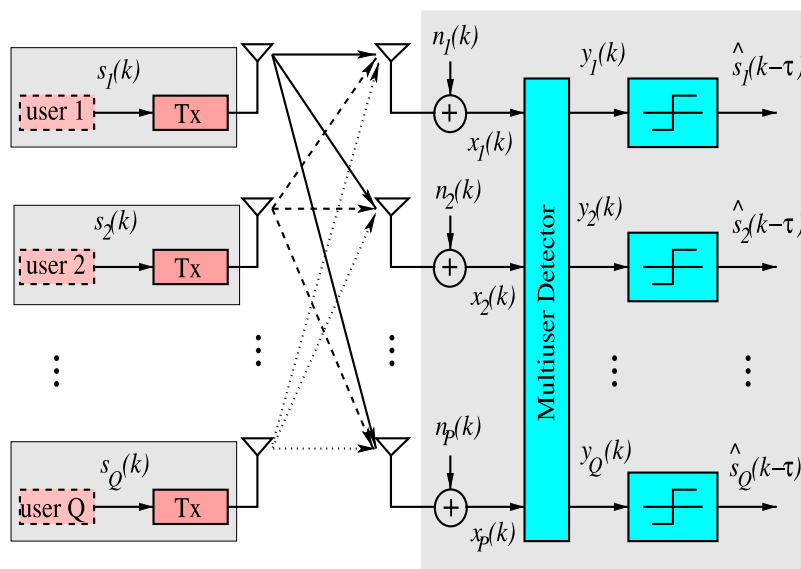


Fig. 4. SDMA induced MIMO system, where each of the Q users is equipped with a single transmit antenna and the receiver is assisted by a P -element antenna array.

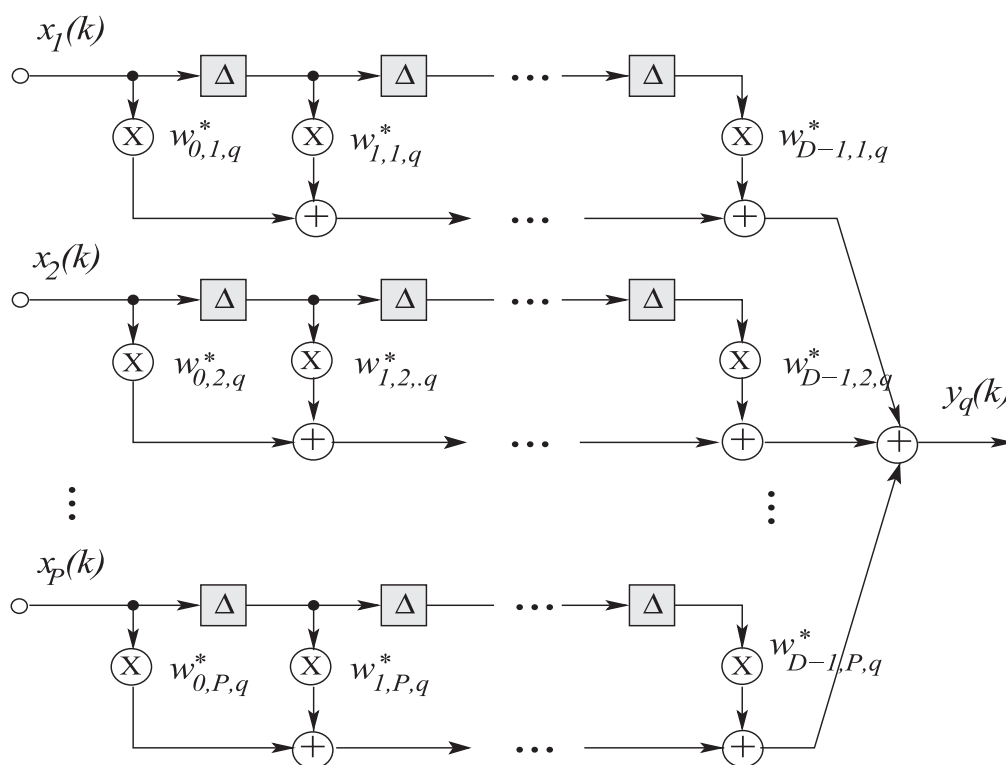


Fig. 5. Space-time equaliser assisted multifuser detector for user q , where Δ denotes the symbol-spaced delay, P is the number of receive antennas, D denotes the length of temporal filter, and $1 \leq q \leq Q$.

with

$$\mathbf{w}_{p,q} = [w_{0,p,q} \ w_{1,p,q} \ \cdots \ w_{D-1,p,q}]^T, \quad 1 \leq p \leq P. \quad (13)$$

Let the CIR connecting the q th user to the p th receive antenna be $h_{0,p,q}, h_{1,p,q}, \dots, h_{n_c-1,p,q}$. Then the $D \times (D + n_c - 1)$ CIR convolution matrix associated with the user q and the receive

antenna p is

$$\mathbf{H}_{p,q} = \begin{bmatrix} h_{0,p,q} & h_{1,p,q} & \cdots & h_{n_c-1,p,q} & 0 & \cdots & 0 \\ 0 & h_{0,p,q} & h_{1,p,q} & \cdots & h_{n_c-1,p,q} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & h_{0,p,q} & h_{1,p,q} & \cdots & h_{n_c-1,p,q} \end{bmatrix} \quad (14)$$

and the overall $L \times M$ system matrix, where $L = P \cdot D$ and $M = Q \cdot (D + n_c - 1)$, is expressed as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \cdots & \mathbf{H}_{1,Q} \\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} & \cdots & \mathbf{H}_{2,Q} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{H}_{P,1} & \mathbf{H}_{P,2} & \cdots & \mathbf{H}_{P,Q} \end{bmatrix}. \quad (15)$$

Similarly, the symbol vector $\mathbf{s}(k)$ is given by

$$\mathbf{s}(k) = [\mathbf{s}_1^T(k) \ \mathbf{s}_2^T(k) \ \cdots \ \mathbf{s}_Q^T(k)]^T \quad (16)$$

with

$$\mathbf{s}_q(k) = [s_q(k) \ s_q(k-1) \ \cdots \ s_q(k-D-n_c+2)]^T, \quad 1 \leq q \leq Q. \quad (17)$$

Again, using the space translation property of decision feedback, the space-time DFE can be translated into the space-time equaliser in the translated observation space [44], and the discussion here is equally applicable to the space-time DFE.

2.4. Conventional filtering design

The above discussion clearly confirms that the signal and filter models (1) and (3) is a generic representation for various communication systems. The classical Wiener filter design is based on minimising the MSE criterion¹

$$J(\mathbf{w}) = E[|s_d(k) - y(k)|^2]. \quad (18)$$

Minimising $J(\mathbf{w})$ with respect to the filter's weight vector \mathbf{w} gives rise to the well-known MMSE

¹It is known that, for the case of the real-valued desired output $s_d(k)$, a better performance in terms of the achievable system's BER can be obtained if the MSE criterion (18) is replaced by $J_{\text{real}}(\mathbf{w}) = E[(s_d(k) - y_R(k))^2]$, see [80–85]. However, the MBER design is still superior over this real-valued MMSE design, see [85]. We will only consider the MSE criterion (18) here, since later we will extend the discussion to the case of complex-valued symbols.

solution [4]

$$\mathbf{w}_{\text{MMSE}} = \left(\mathbf{H}\mathbf{H}^H + \frac{2\sigma_n^2}{\sigma_s^2} \mathbf{I}_L \right)^{-1} \mathbf{h}_d, \quad (19)$$

where $\sigma_s^2 = E[|s_d(k)|^2]$ is the energy of the desired output $s_d(k)$ and \mathbf{h}_d denotes the d th column of \mathbf{H} . The MMSE filtering is attractive, since it is given in a closed-form solution based on the second-order statistics of the underlying system.

When the required second-order statistics are unknown, adaptive implementation of the Wiener filter is particularly simple. Given a block of training samples $\{s_d(k), \mathbf{x}(k)\}_{k=1}^K$, the MSE (18) can be approximated by the following sample-average

$$\hat{J}_K(\mathbf{w}) = \frac{1}{K} \sum_{k=1}^K |s_d(k) - y(k)|^2, \quad (20)$$

where $y(k) = \mathbf{w}^H \mathbf{x}(k)$. Minimising the approximate MSE (20) with respect to \mathbf{w} leads to an approximate MMSE solution. In particular, consider a single-sample “estimate” of the MSE, namely, $|s_d(k) - y(k)|^2$. Minimising this instantaneous squared error leads to the stochastic-gradient adaptive algorithm commonly referred to as the LMS [4]

$$\begin{cases} y(k) = \tilde{\mathbf{w}}^H(k-1)\mathbf{x}(k), \\ \tilde{\mathbf{w}}(k) = \tilde{\mathbf{w}}(k-1) + \mu(s_d(k) - y(k))^* \mathbf{x}(k), \end{cases} \quad (21)$$

where μ is the step size. The LMS algorithm has a very low computational complexity and is particularly suitable for real-time adaptive applications of high-speed communication systems. Convergence properties of the LMS algorithm are also well understood [4].

The Wiener filtering is optimal with respect to the MSE criterion. As pointed out previously, the true performance indicator is the error probability of the decision process (4), i.e. the achievable system's BER. A natural question to ask is under what condition is the MMSE filter also the MBER filter? Let us consider the simplest system, namely, the single-user BPSK communication system over an ideal AWGN channel. The optimal receiver filter for such a system is well known to be the matched filter that maximises the receive signal-to-noise ratio (SNR) [1]. The matched filter's output for such a simple system is expressed as²

$$y_R(k) = A s(k) + n_R(k), \quad (22)$$

²For this ideal AWGN case, the signal model (1) is a scalar $x(k)$ and the filter model (3) constitutes a scalar weighting $y(k) = w_1^* x(k)$.

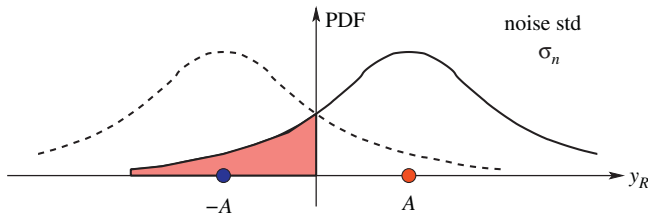


Fig. 6. Error probability of the BPSK communication system over the ideal AWGN channel.

where A is positive and A^2 represents the receive signal power, and $n_R(k)$ denotes the real part of the channel noise at the receiver output and it is Gaussian distributed with a zero mean and a variance of σ_n^2 . The conditional PDFs of $y_R(k)$ given $s(k) = \pm 1$ are

$$p(y_R | s(k) = \pm 1) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-|y_R - (\pm A)|^2 / 2\sigma_n^2}. \quad (23)$$

Assuming $\text{Prob}\{s(k) = 1\} = \text{Prob}\{s(k) = -1\} = \frac{1}{2}$, the error probability of the decision process (4) or BER in this case is well known to be

$$\begin{aligned} P_E &= \int_{-\infty}^0 p(y_R | s(k) = +1) dy_R \\ &= Q\left(\frac{A}{\sigma_n}\right) = Q(\sqrt{\text{SNR}}), \end{aligned} \quad (24)$$

which is also illustrated in Fig. 6, where the Q -function is define by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-u^2/2} du. \quad (25)$$

The match filter that maximises the receive SNR also minimises the error probability P_E . Note that the conditional PDF of $y_R(k)$ given a particular class of symbol value is Gaussian. In this case, clearly the MMSE solution is identical to the MBER solution.³

In general, the MMSE design is also the optimal MBER design only if the conditional PDF of the filter output for a given transmitted data symbol value is Gaussian. This is because if this conditional PDF is Gaussian, i.e. the conditional PDF of $y(k) = \mathbf{w}^H \mathbf{x}(k)$ takes the form of (23), minimising the MSE

³Noting that the signal power $A^2 = \sigma_s^2$ is fixed, minimising the MSE in this ideal case is equivalent to maximising the receive SNR and, therefore, equivalent to minimising the BER. Indeed, the optimisation to achieve maximising the receive SNR is actually carried out by fixing the signal power to constant and minimising the receive noise power, i.e. the MSE in this case (see [1]).

is equivalent to minimising the error probability (24). In certain systems, the conditional PDF of the filter output may be approximated accurately by a Gaussian distribution, and in such cases the MMSE design is preferred over the more complicated MBER design. However, since this conditional PDF is generally a mixture of Gaussian distributions and hence non-Gaussian, as will be shown in the next section, the MMSE design is inherently sub-optimal, in terms of the achievable system's BER.

3. MBER filtering design

Let us revisit the receive signal model (1) and express it as

$$\mathbf{x}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{n}(k) = \bar{\mathbf{x}}(k) + \mathbf{n}(k), \quad (26)$$

where $\bar{\mathbf{x}}(k)$ represents the noise-free part of $\mathbf{x}(k)$. Denote the $N_s = 2^M$ legitimate sequences of $\mathbf{s}(k)$ as \mathbf{s}_i , $1 \leq i \leq N_s$, and further denote the d th element of \mathbf{s}_i , corresponding to the desired symbol $s_d(k)$, as $s_{d,i}$. The noise-free part of the received signal, $\bar{\mathbf{x}}(k)$, takes values from the complex-valued vector set

$$\bar{\mathcal{X}}(k) \in \mathcal{X} = \{\bar{\mathbf{x}}_i = \mathbf{H}\mathbf{s}_i, 1 \leq i \leq N_s\}. \quad (27)$$

\mathcal{X} can be divided into two subsets corresponding to the two values of $s_d(k)$ as follows

$$\mathcal{X}^{(\pm)} = \{\bar{\mathbf{x}}_i^{(\pm)} \in \mathcal{X}, 1 \leq i \leq N_{sb} : s_d(k) = \pm 1\}, \quad (28)$$

where $N_{sb} = N_s/2$. Similarly rewrite the filter model (3) as⁴

$$y(k) = \mathbf{w}^H \mathbf{x}(k) = \bar{y}(k) + e(k), \quad (29)$$

where $e(k) = \mathbf{w}^H \mathbf{n}(k)$ is Gaussian distributed with zero mean and $E[|e(k)|^2] = 2\mathbf{w}^H \mathbf{w} \sigma_n^2$. The noise-free part of the filter output, namely $\bar{y}(k)$, takes values from the complex-valued scalar set

$$\bar{\mathcal{Y}}(k) \in \mathcal{Y} = \{\bar{y}_i = \mathbf{w}^H \bar{\mathbf{x}}_i, \bar{\mathbf{x}}_i \in \mathcal{X}\}. \quad (30)$$

The real part of the filter output $y(k)$ is $y_R(k) = \bar{y}_R(k) + e_R(k)$, and $\bar{y}_R(k)$ takes values from the real-valued scalar set

$$\bar{\mathcal{Y}}_R(k) \in \mathcal{Y}_R = \{\bar{y}_{R_i} = \Re[\bar{y}_i], \bar{y}_i \in \mathcal{Y}\}. \quad (31)$$

The set \mathcal{Y}_R can be divided into the two subsets conditioned on the value of $s_d(k)$, each having a size of the N_{sb} constellation points

$$\mathcal{Y}_R^{(\pm)} = \{\bar{y}_{R_i}^{(\pm)} \in \mathcal{Y}_R : s_d(k) = \pm 1\}. \quad (32)$$

⁴The filter is for detecting the desired symbol $s_d(k)$. As before, we avoid using index d in $y(k)$ for notational simplification.

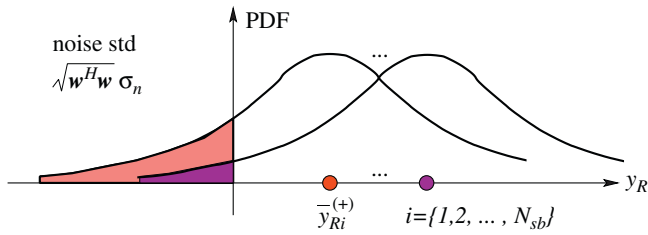


Fig. 7. Derivation of the error probability of the linear filter (3) for the communication system (1) with BPSK symbols.

Definition 1. The system (1) is said to be linearly separable, if there exists a weight vector \mathbf{w} such that the two real-valued scalar subsets, $\mathcal{Y}_R^{(+)}$ and $\mathcal{Y}_R^{(-)}$, can completely be separated by the decision threshold $y_R = 0$.

3.1. BER expression

It is readily seen that the conditional PDF of $y_R(k)$ given $s_d(k) = +1$ is a Gaussian mixture, expressed as

$$\begin{aligned}
 p(y_R | s_d(k) = +1) &= \frac{1}{N_{sb} \sqrt{2\pi \mathbf{w}^H \mathbf{w} \sigma_n^2}} \sum_{y_{Ri}^{(+)} \in \mathcal{Y}_R^{(+)}} e^{-(|y_R - y_{Ri}^{(+)}|^2) / 2\mathbf{w}^H \mathbf{w} \sigma_n^2}.
 \end{aligned} \tag{33}$$

Noting the decision process (4), the error probability or BER of the filter (3) with the weight vector \mathbf{w} is defined by

$$\begin{aligned}
 P_E(\mathbf{w}) &= \int_{-\infty}^0 p(y_R | s_d(k) = +1) dy_R \\
 &= \frac{1}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_i^{(+)}(\mathbf{w})),
 \end{aligned} \tag{34}$$

which is also illustrated in Fig. 7, where

$$g_i^{(+)}(\mathbf{w}) = \frac{\text{sgn}(s_{d,i}) y_{Ri}^{(+)}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}}. \tag{35}$$

The BER can alternatively be calculated using the other subset $\mathcal{Y}_R^{(-)}$, and the length of the filter weight vector, $\mathbf{w}^H \mathbf{w}$, does not affect the BER value (except for zero length).

3.2. MBER solution

The MBER solution for the filter's weight vector is defined as

$$\mathbf{w}_{\text{MBER}} = \arg \min_{\mathbf{w}} P_E(\mathbf{w}). \tag{36}$$

Unlike the MMSE design, there exists no closed-form solution for the MBER design, and the optimisation problem (36) must be solved numerically using for example a gradient-based algorithm. The gradient of $P_E(\mathbf{w})$ with respect to \mathbf{w} can be shown to be

$$\begin{aligned}
 \nabla P_E(\mathbf{w}) &= \frac{1}{2N_{sb} \sqrt{2\pi} \sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \sum_{i=1}^{N_{sb}} e^{-(|y_{Ri}^{(+)}|^2) / 2\sigma_n^2 \mathbf{w}^H \mathbf{w}} \\
 &\quad \times \text{sgn}(s_{d,i}) \left(\frac{y_{Ri}^{(+)} \mathbf{w}}{\mathbf{w}^H \mathbf{w}} - \bar{\mathbf{x}}_i^{(+)} \right),
 \end{aligned} \tag{37}$$

where $\bar{\mathbf{x}}_i^{(+)} \in \mathcal{X}^{(+)}$. The simplified conjugate gradient algorithm [26,86] provides an efficient means of obtaining an MBER solution.

Proposition 1. For linearly separable systems, any local minimiser of the BER cost function (34) is a global minimiser.

Proof. Since the system is linearly separable, there exist weight vectors \mathbf{w} such that

$$\mathbf{w}^H \mathbf{x}_i^{(+)} > 0, \quad \forall \mathbf{x}_i^{(+)} \in \mathcal{X}^{(+)}.$$

The proof then follows similar arguments to those outlined in the proof of Proposition 1 in [87]. \square

Proposition 2. There exist infinitely many global MBER solutions, which form an infinite half line in the filter weight space. There exists a unique unit-length global MBER solution.

Proof. Let \mathbf{w}_{MBER} be a global MBER solution. Since the BER is invariant to the length of \mathbf{w} , weight vectors $\alpha \cdot \mathbf{w}_{\text{MBER}}$, $\alpha > 0$, are all global MBER solutions, which form an infinite half line in the filter weight space. Setting $\alpha = 1 / \sqrt{\mathbf{w}_{\text{MBER}}^H \mathbf{w}_{\text{MBER}}}$ yields the unique unit-length global MBER solution.

Some comments can be made for the MMSE and MBER designs by examining their respective cost functions. The MSE surface (18) is quadratic and has a single global minimum solution. In contrast, the BER surface (34) is much more complex. A comparison of the MSE and BER surfaces is given in Fig. 8, using the simple two-user CDMA system considered in [26,29]. Note that the BER is invariant to a positive scaling of \mathbf{w} . Thus, as can be seen from Fig. 8(b), the BER surface has an infinitely long valley, and any point at the bottom of this valley is a true global MBER solution. It can also be seen from Fig. 8(b) that if we restrict to the unit-length \mathbf{w} , the MBER solution becomes unique. At least for

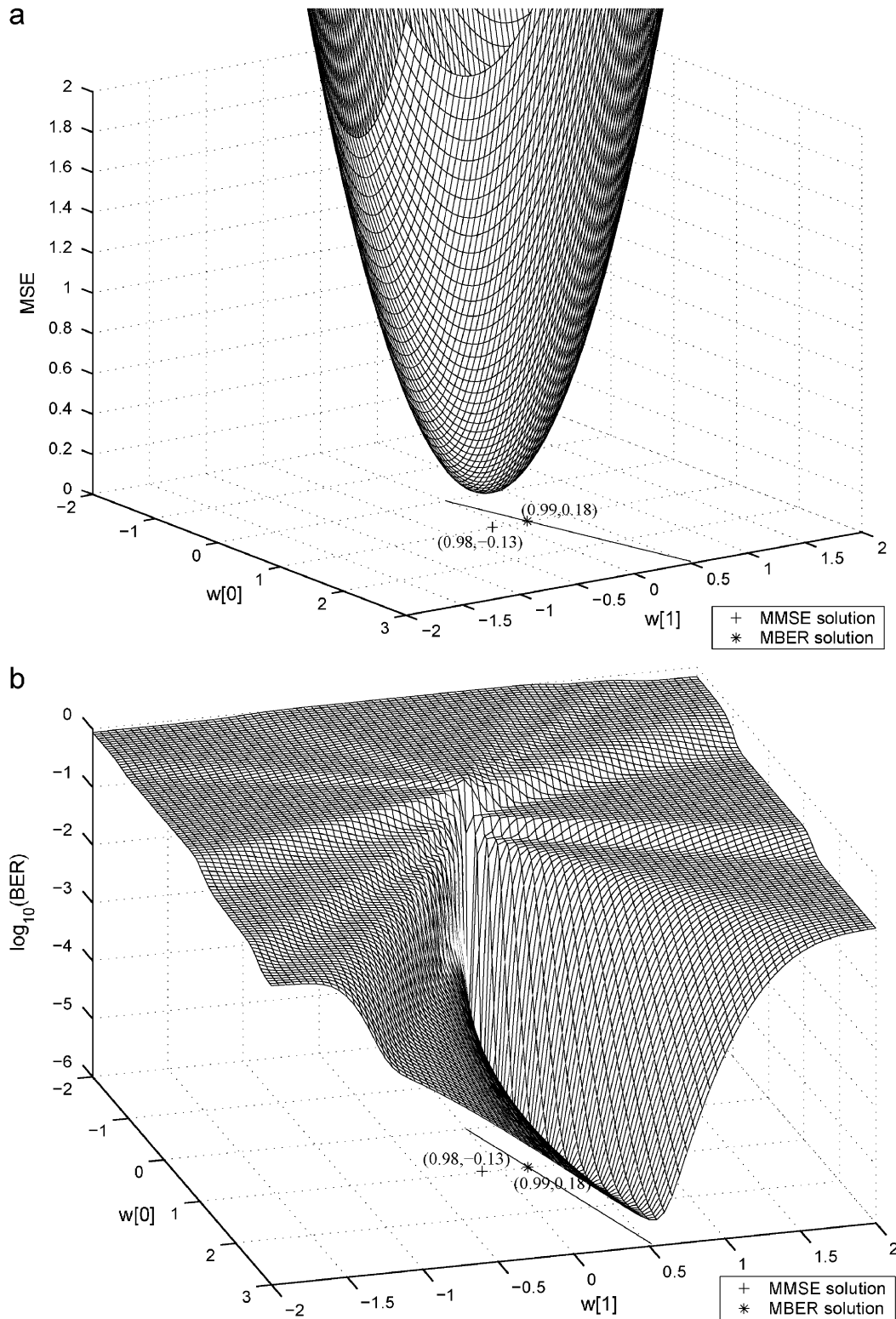


Fig. 8. The MSE surface (a) and the BER surface (b) of a simple two-user CDMA system, taken from [29].

linearly separable systems, there exists no local minimum problem for the optimisation process (36). However, the BER surface may have very small-gradient regions, as can be seen in Fig. 8(b). This indicates that convergence speed of the optimisation

process may depend on the initial choice of the weight vector, particularly when the steepest-descent gradient algorithm is used. By adopting the conjugate gradient-based algorithm, we observe in practice that this problem is alleviated. \square

Before turning to discuss adaptive implementation of the MBER design, we point out that the PDF of $y_R(k)$ is explicitly given by the Gaussian mixture

$$p(y_R) = \frac{1}{N_s \sqrt{2\pi \mathbf{w}^H \mathbf{w} \sigma_n^2}} \sum_{\bar{y}_{R_i} \in \mathcal{Y}_R} e^{-|y_R - \bar{y}_{R_i}|^2 / 2\mathbf{w}^H \mathbf{w} \sigma_n^2}, \quad (38)$$

where the summation is over all the constellation points of \mathcal{Y}_R , and the BER can be computed using

$$P_E(\mathbf{w}) = \frac{1}{N_s} \sum_{i=1}^{N_s} Q(g_i(\mathbf{w})), \quad (39)$$

with

$$g_i(\mathbf{w}) = \frac{\text{sgn}(s_{d,i}) \bar{y}_{R_i}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}}. \quad (40)$$

3.3. Adaptive MBER filtering

The key to adaptive implementation of the MBER filtering design is an effective estimate of the PDF (38). Parzen window estimate [77–79], also known as the kernel density estimate, is a well-known method for estimating a probability distribution. Parzen window method estimates a PDF using a window or block of $y(k)$ by placing a symmetric unimodal kernel function on each $y(k)$ with an equal weighting. Kernel density estimation is capable of producing reliable PDF estimates with relatively short data records and in particular is extremely natural when dealing with Gaussian mixtures, such as the one given in (38).

Given a training data set $\{s_d(k), \mathbf{x}(k)\}_{k=1}^K$, the PDF of the filter's output $y_R(k)$ can be accurately estimated using the following Parzen window estimate

$$\tilde{p}(y_R) = \frac{1}{K \sqrt{2\pi} \rho_n} \sum_{k=1}^K e^{-|y_R - y_R(k)|^2 / 2\rho_n^2}, \quad (41)$$

where ρ_n is called the kernel width. Based on this PDF estimate, a BER estimate is given by

$$\tilde{P}_E(\mathbf{w}) = \frac{1}{K} \sum_{k=1}^K Q(\tilde{g}_k(\mathbf{w})) \quad (42)$$

with

$$\tilde{g}_k(\mathbf{w}) = \frac{\text{sgn}(s_d(k)) y_R(k)}{\rho_n}. \quad (43)$$

Providing that the kernel width ρ_n is chosen appropriately, the kernel density estimate (41) is an accurate estimate of the true PDF $p(y_R)$ and the BER estimate (42) is an accurate estimate of the true BER $P_E(\mathbf{w})$. Given the gradient of $\tilde{P}_E(\mathbf{w})$

$$\nabla \tilde{P}_E(\mathbf{w}) = -\frac{1}{2K \sqrt{2\pi} \rho_n} \times \sum_{k=1}^K e^{-|y_R(k)|^2 / 2\rho_n^2} \text{sgn}(s_d(k)) \mathbf{x}(k), \quad (44)$$

an approximated MBER solution can be obtained iteratively using a gradient-based algorithm, such as the simplified conjugate gradient algorithm [26,86].

In order to derive a sample-by-sample adaptive algorithm, let us adopt a single-sample “estimate” of the PDF $p(y_R)$, namely,

$$\tilde{p}(y_R, k) = \frac{1}{\sqrt{2\pi} \rho_n} e^{-|y_R - y_R(k)|^2 / 2\rho_n^2}. \quad (45)$$

From this one-sample PDF “estimate”, conceptually we have an instantaneous BER “estimate” $\tilde{P}_E(\mathbf{w}, k) = Q(\tilde{g}_k(\mathbf{w}))$. Using the instantaneous stochastic gradient of

$$\nabla \tilde{P}_E(\mathbf{w}, k) = -\frac{\text{sgn}(s_d(k))}{2\sqrt{2\pi} \rho_n} e^{-|y_R(k)|^2 / 2\rho_n^2} \mathbf{x}(k) \quad (46)$$

gives rise to the following LBER algorithm:

$$\begin{cases} y(k) = \tilde{\mathbf{w}}^H(k-1) \mathbf{x}(k), \\ \tilde{\mathbf{w}}(k) = \tilde{\mathbf{w}}(k-1) + \mu \frac{\text{sgn}(s_d(k))}{2\sqrt{2\pi} \rho_n} e^{-|y_R(k)|^2 / 2\rho_n^2} \mathbf{x}(k), \end{cases} \quad (47)$$

where the step size μ and the kernel width ρ_n are the two algorithmic parameters that have to be set appropriately to ensure a fast convergence rate and small steady-state BER misadjustment.

Comment: The factor $1/(2\sqrt{2\pi} \rho_n)$ is a constant that can be absorbed into the step size μ whereas the term $\text{sgn}(s_d(k)) e^{-|y_R(k)|^2 / 2\rho_n^2}$ can be regarded as an “error signal”. In this sense, the LBER algorithm (47) has a similar form to the LMS algorithm (21).

It is interesting to point out that our development of the adaptive MBER filtering design follows a similar path to that of the adaptive MMSE filtering. The BER of the adaptive filter is a higher-order statistics of the underlying filter output's PDF. By estimating this PDF using Parzen window estimate, we are able to estimate the BER using a relatively short block of samples. Further considering a single-sample density “estimate”, we arrive at the

above stochastic-gradient adaptive LBER algorithm. Computational complexity of this LBER algorithm is similar to that of the LMS algorithm. Tuning the LBER algorithm, however, is more complicated than adjusting the LMS algorithm, since the former has two algorithmic parameters, μ and ρ_n , while the latter has only one algorithmic parameter, the step size μ . Convergence properties of the LMS algorithm are well understood and there exist analytical results to predict the steady-state MSE misadjustment. Convergence properties of the LBER algorithm by contrast are less well understood, although extensive simulation experience has suggested that it is not too difficult for tuning the algorithm to achieve fast convergence. Furthermore, at the time of writing there exists no theoretical result for analysing the steady-state BER misadjustment of the LBER algorithm, but in practice we have observed that the steady-state BER misadjustment can often be made very small by carefully tuning the two algorithmic parameters. Convergence behaviour and steady-state BER misadjustment of the LBER algorithm have been extensively investigated in the previous publications [11,14,18,20,25,26,29,31,35,36,38,39,41–44].

3.4. Illustrative examples

Beamforming: The simulation system was illustrated in Fig. 3, where the receiver employed a four-element linear antenna array with a half-wavelength element spacing. The system was first used to support $M = 3$ BPSK users, and the locations of the users in terms of AOA is summarised in Table 1. The simulated channel conditions were $A_i = 1 + j0$ for $1 \leq i \leq 3$, and all the users had an equal power. The user one was the desired user. Fig. 9 compares the BER performance of the two beamformer designs for the desired user one. In order to draw some insights as to how the MBER design behaves differently from the MMSE design, and hence to explain the BER performance difference of the two beamformers as shown in

Table 1
Locations of the users in terms of angle of arrival for the simulated beamforming system that employs a four-element antenna array to support three BPSK users

User i	1	2	3
AOA θ_i ($^\circ$)	0	10	-5

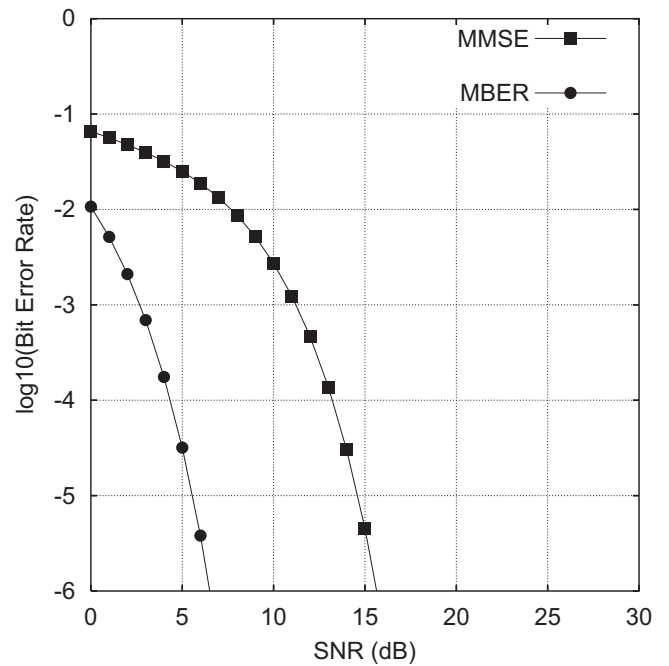


Fig. 9. Desired user's bit error rate comparison of two beamforming designs for the four-element array system supporting three BPSK users as given in Table 1.

Fig. 9, we examined the conditional PDF of the beamformer's output $y(k)$, which is defined by

$$p(y|s_d(k) = +1) = \frac{1}{N_{sb} 2\pi \mathbf{w}^H \mathbf{w} \sigma_n^2} \sum_{\tilde{y}_i^{(+)} \in \mathcal{Y}^{(+)}} e^{-|y - \tilde{y}_i^{(+)}|^2 / 2\mathbf{w}^H \mathbf{w} \sigma_n^2}, \quad (48)$$

where the signal subset $\mathcal{Y}^{(+)}$ is defined as

$$\mathcal{Y}^{(+)} = \{\tilde{y}_i^{(+)} \in \mathcal{Y} : s_d(k) = +1\}, \quad (49)$$

and its marginal conditional PDF, $p(y_R|s_d(k) = +1)$, which is defined in (33). Fig. 10 depicts the conditional PDFs $p(y|s_d(k) = +1)$, the marginal PDFs $p(y_R|s_d(k) = +1)$, the signal subsets $\mathcal{Y}^{(+)}$ and $\mathcal{Y}_R^{(+)}$ for the two beamforming designs given SNR = 6 dB, where the beamformer's weight vector \mathbf{w} was normalised to a unit length. It can be seen that the distribution $p(y|s_d(k) = +1)$ is symmetric with respect to the $\Re[y]$ and $\Im[y]$ axes for the MMSE design, and this was a direct consequence of the minimisation of the MSE criterion (18). By contrast, the MBER design appeared to be more intelligent and was able to shape $p(y|s_d(k) = +1)$ in such a way that the distance between the decision threshold $y_R = 0$ and the signal subset $\mathcal{Y}_R^{(+)}$ was maximised. From the BER expression (34), this is optimal as it ensures the minimisation of the BER. Convergence behaviour and steady-state BER

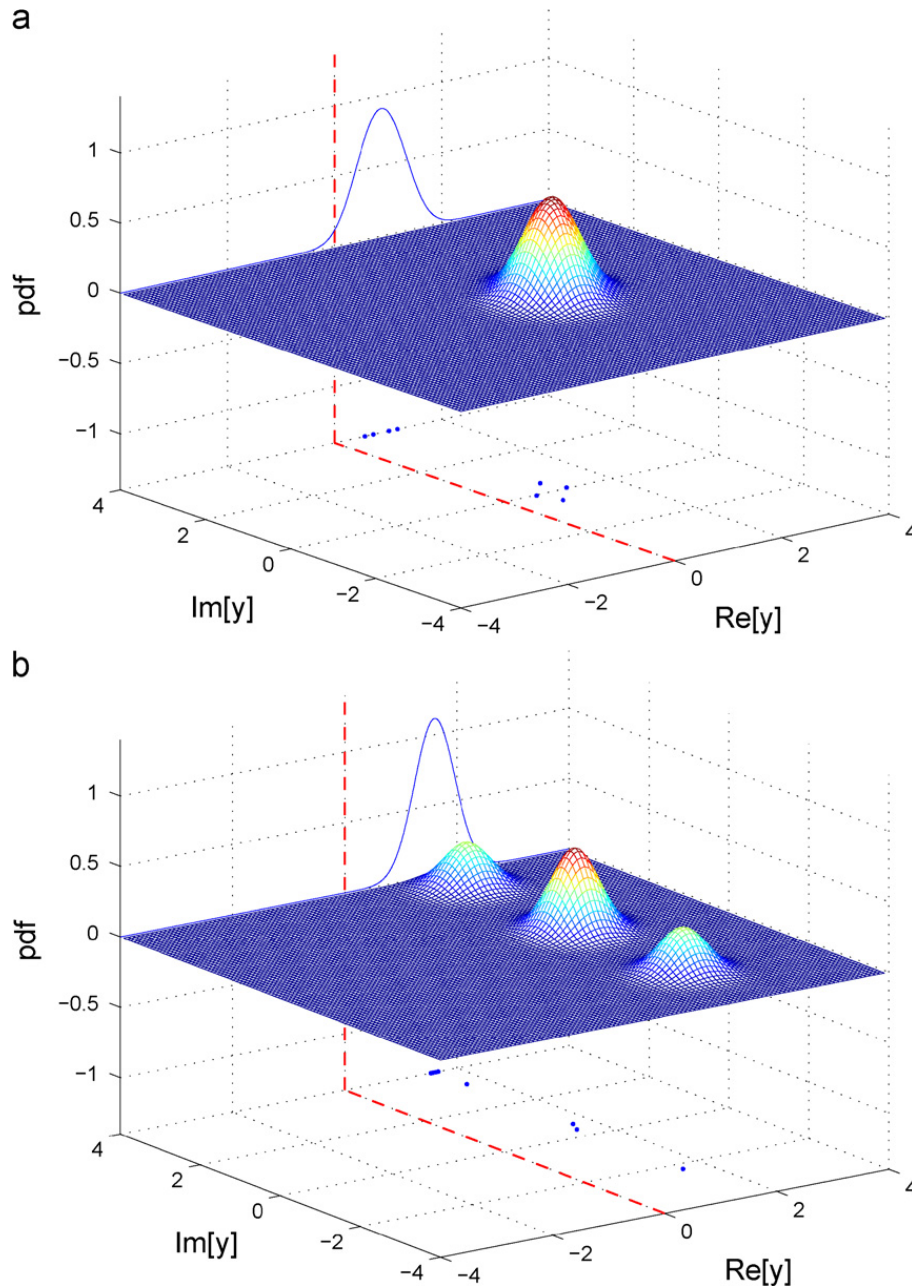


Fig. 10. Conditional PDFs $p(y|s_d(k) = +1)$ (surfaces), marginal conditional PDFs $p(y_R|s_d(k) = +1)$ (curves), signal subsets $\mathcal{Y}^{(+)}$ and $\mathcal{Y}_R^{(+)}$ (points) for the four-element array system supporting three BPSK users of Table 1 with SNR = 6 dB: (a) MMSE design, and (b) MBER design. The beamformer's weight vector is normalised to a unit length.

misadjustment of the two adaptive algorithms, the LMS and LBER, were next investigated. Given SNR = 4 dB, Fig. 11 depicts the learning curves of the two adaptive algorithms, averaged over 100 independent runs. The step size of the LMS was set to $\mu = 0.006$, while for the LBER algorithm we had the step size $\mu = 0.06$ and the kernel variance $\rho_n^2 = 2\sigma_n^2$. Fig. 11(a) shows the training performance, while Fig. 11(b) illustrates the decision-directed (DD) adaptation, started at $k = 30$, using the beamformer's decision $\hat{s}_1(k)$ to substitute for

$s_1(k)$. It can be seen from Fig. 11(b) that the steady-state BER of the LMS algorithm for the given SNR was not sufficiently small for the algorithm to switch to the DD adaptation. By contrast, the LBER algorithm was seen to operate successfully in the DD adaptation. This is a significant advantage of the LBER algorithm, which enables the algorithm to adopt DD updating during data transmission in order to track time-varying channel conditions.

The system was next used to support $M = 8$ BPSK users, and the locations of the users are listed

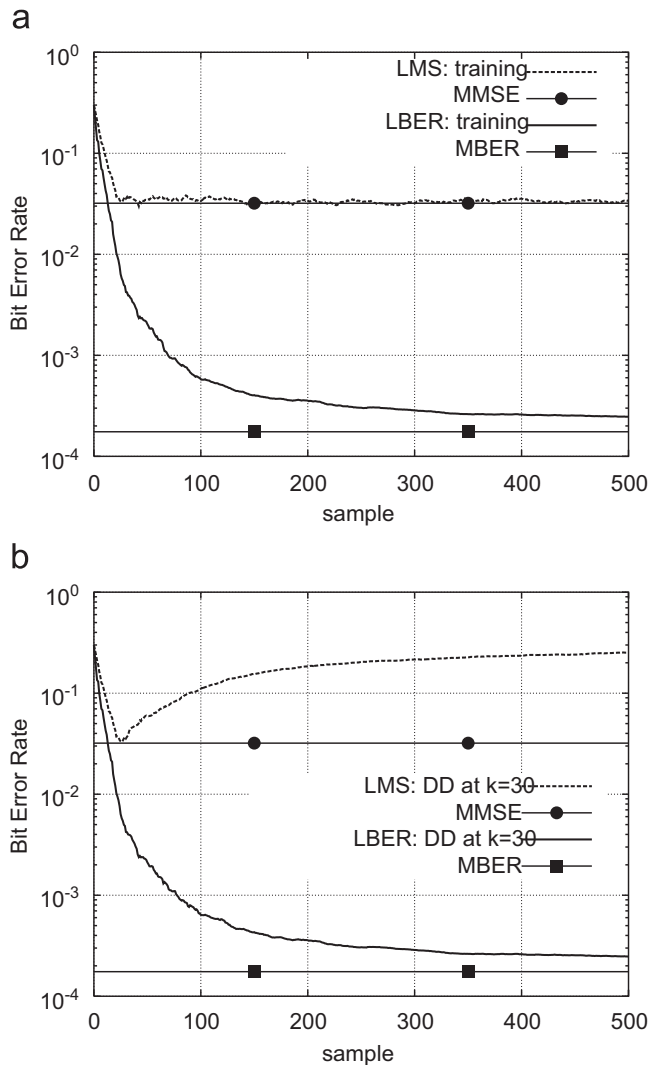


Fig. 11. Learning curves of the two adaptive algorithms averaged over 100 runs for the four-element array system supporting three BPSK users of Table 1 with SNR = 4 dB: (a) training, and (b) decision-directed adaptation starting from $k=30$ with the beamformer's decision $\hat{s}_i(k)$ substituting for $s_i(k)$, where $\tilde{\mathbf{w}}(0) = [0.1 + j0.1 \ 0.0 + j0.0 \ 0.0 + j0.0 \ 0.0 + j0.0]^T$.

Table 2

Locations of the users in terms of angle of arrival for the simulated beamforming system that employs a four-element antenna array to support eight BPSK users

User i	1	2	3	4	5	6	7	8
AOA θ_i (°)	0	10	-15	30	-45	50	60	-55

in Table 2. The user one still remained to be the desired user, all the users had an equal power, and the simulated channel conditions were $A_i = 1 + j0$ for $1 \leq i \leq 8$. Fig. 12 depicts the BER performance of the two beamformer designs for the desired user

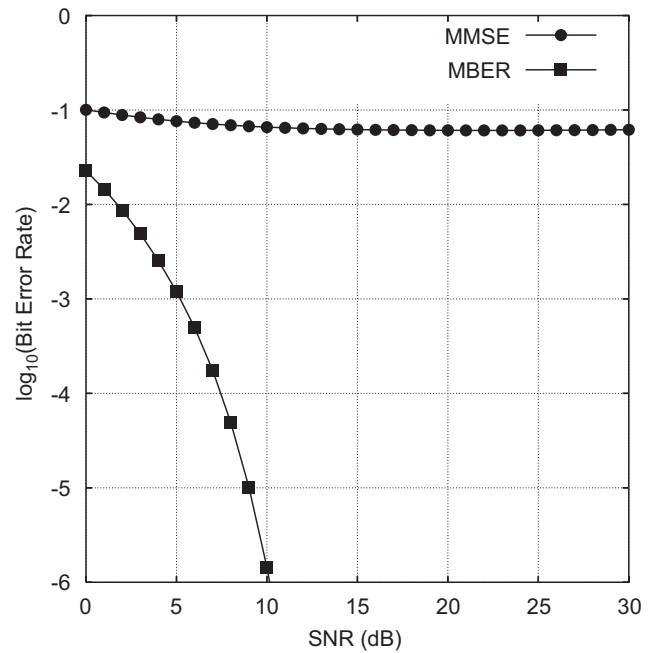


Fig. 12. Desired user's bit error rate comparison of two beamforming designs for the four-element array system supporting eight BPSK users as given in Table 2.

one, while Fig. 13 shows the conditional PDFs $p(y|s_d(k) = 1)$, the marginal PDFs $p(y_R|s_d(k) = +1)$, the signal subsets $\mathcal{Y}^{(+)}$ and $\mathcal{Y}_R^{(+)}$ for the two designs, given SNR = 8 dB. It can be seen from Fig. 13(a) that for the MMSE beamformer several points of $\mathcal{Y}_R^{(+)}$ were in the wrong side of the decision threshold $y_R = 0$, resulting in the high BER floor as shown in Fig. 12. By contrast, the MBER beamformer was capable of ensuring a reasonable distance between $y_R = 0$ and the signal subset $\mathcal{Y}_R^{(+)}$ as can be seen from Fig. 13(b) and, therefore, maintained an adequate BER performance.

In general, it can be demonstrated that the MBER design can better combat channel distortion and interference and offers a larger system user capacity, compared with the suboptimal MMSE design. Extensive investigations [36,38] have also shown that the MBER beamformer is much more robust to the near-far effect. More specifically, when facing strong interfering sources, the MMSE beamforming receiver may exhibit a high BER floor as the underlying signal classes or subsets become linearly inseparable, while the MBER beamforming can often maintain the desired linear separability and hence avoids such a BER floor.

Space-time equalisation: The simulation system was illustrated in Figs. 4 and 5. The MIMO receiver employed $P = 4$ receive antennas to support $Q = 3$ BPSK users. Each of the $Q \cdot P = 12$ dispersive

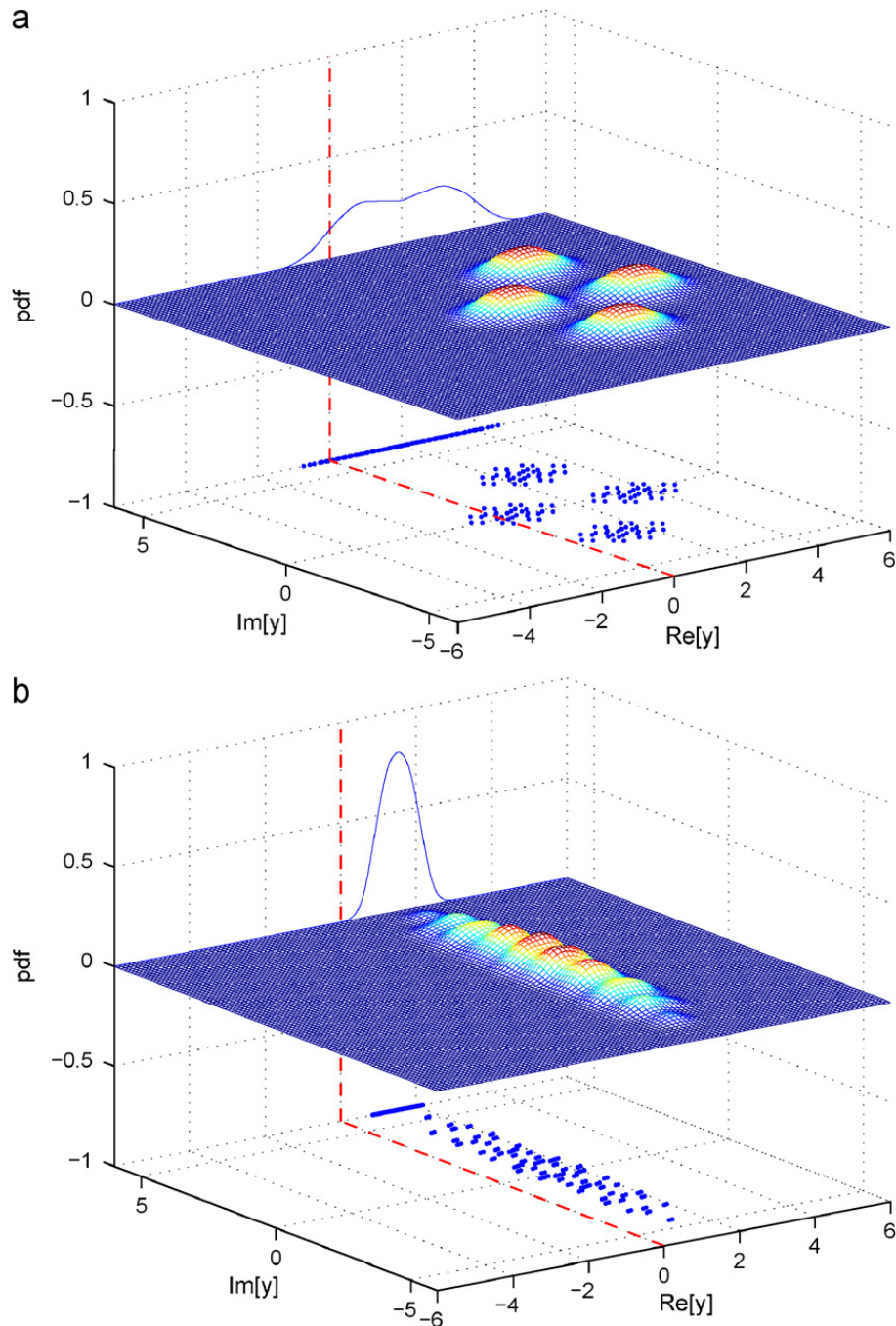


Fig. 13. Conditional PDFs $p(y|s_d(k) = +1)$ (surfaces), marginal conditional PDFs $p(y_R|s_d(k) = +1)$ (curves), signal subsets $\mathcal{Y}^{(+)}$ and $\mathcal{Y}_R^{(+)}$ (points) for the four-element array system supporting eight BPSK users of Table 2 with SNR = 8 dB: (a) MMSE design, and (b) MBER design. The beamformer's weight vector is normalised to a unit length.

channels had a length $n_c = 3$. The magnitudes of the CIR taps were Rayleigh processes, and each CIR tap had the root mean power of $\sqrt{0.5} + j\sqrt{0.5}$. The normalised Doppler frequency for the simulated system was 10^{-5} , which for a carrier of 900 MHz and a symbol rate of 3 Msymbols/s corresponded to a user velocity of 10 m/s (36 km/h). Continuously fluctuating fading was used, which provided a

different fading magnitude and phase of each CIR tap for each transmitted symbol. The transmission frame structure consisted of 50 training symbols followed by 450 data symbols. Each temporal filter of the space-time equaliser had an order $D = 5$, and the equaliser decision delay was set to $\tau = 2$.

The adaptive LMS and LBER based space-time equalisers were investigated. The step size of the

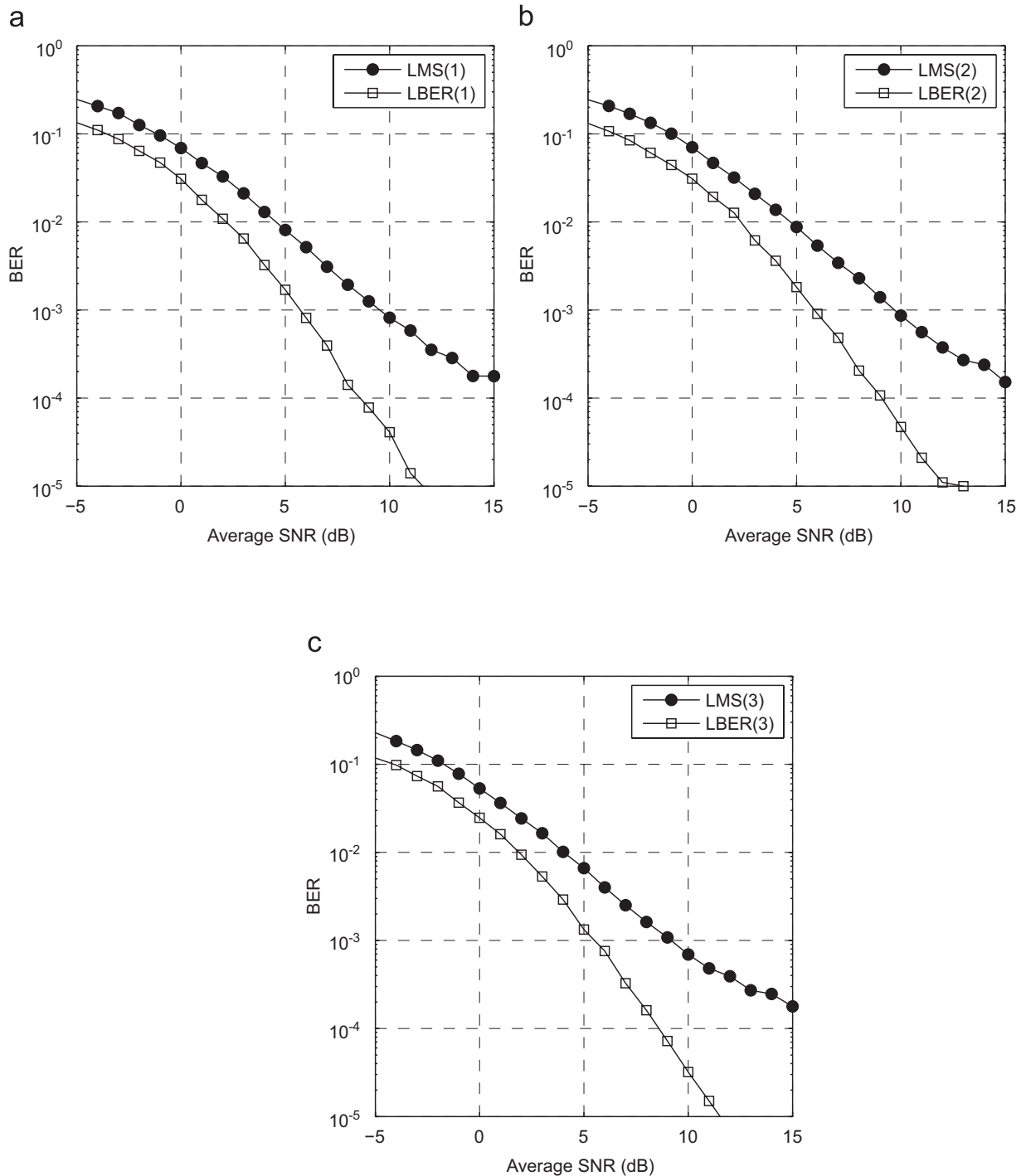


Fig. 14. Bit error rate comparison of the adaptive LMS and LBER space-time equalisation based multiuser detectors for the 3-user 4-antenna slow fading dispersive system: (a) user 1, (b) user 2, and (c) user 3.

LMS algorithm was chosen as $\mu = 0.005$, while for the LBER algorithm the step size $\mu = 0.1$ and the kernel variance $\rho_n^2 = 16\sigma_n^2$ were found empirically to be appropriate. The BER of an adaptive space-time equaliser was calculated on the 450 data symbols of the frame using Monte Carlo simulation averaging

over 10^6 frames. Fig. 14 compares the BERs of the LBER space-time equaliser based multiuser detectors for the three users with those of the LMS based multiuser detectors. It can be seen from Fig. 14 that the LBER space-time equaliser consistently outperformed the LMS space-time equaliser.

4. Extension to QAM

For high-order m -QAM systems [2], it is computationally more attractive to consider the MSER design, rather than the MBER design [39]. For the signal model (1), the data symbols $s_i(k)$ for $1 \leq i \leq M$ now take values from the m -QAM symbol set

$$\mathcal{S} = \{s_{l,q} = u_l + ju_q, 1 \leq l, q \leq \sqrt{m}\}, \quad (50)$$

where the real-part symbol $\Re[s_{l,q}] = u_l = 2l - \sqrt{m} - 1$ and the imaginary-part symbol $\Im[s_{l,q}] = u_q = 2q - \sqrt{m} - 1$. The combined filter and system impulse response is defined by

$$\mathbf{w}^H \mathbf{H} = \mathbf{w}^H [\mathbf{h}_1 \ \mathbf{h}_2 \ \dots \ \mathbf{h}_M] = [c_1 \ c_2 \ \dots \ c_M]. \quad (51)$$

Thus, the filter's output can alternatively be expressed as

$$y(k) = c_d s_d(k) + \sum_{i \neq d} c_i s_i(k) + e(k). \quad (52)$$

Provided that the desired main tap $c_d = c_{R_d} + jc_{I_d}$ satisfies $c_{R_d} > 0$ and $c_{I_d} = 0$, the decision rule for estimating the desired symbol $s_d(k)$ is $\hat{s}_d(k) = \hat{s}_{R_d}(k) + j\hat{s}_{I_d}(k)$, with

$$\hat{s}_{R_d}(k) = \begin{cases} u_1 & \text{if } y_R(k) \leq c_{R_d}(u_1 + 1), \\ u_l & \text{if } c_{R_d}(u_l - 1) < y_R(k) \leq c_{R_d}(u_l + 1) \\ & \text{for } 2 \leq l \leq \sqrt{m} - 1, \\ u_{\sqrt{m}} & \text{if } y_R(k) > c_{R_d}(u_{\sqrt{m}} - 1), \end{cases} \quad (53)$$

and

$$\hat{s}_{I_d}(k) = \begin{cases} u_1 & \text{if } y_I(k) \leq c_{R_d}(u_1 + 1), \\ u_q & \text{if } c_{R_d}(u_q - 1) < y_I(k) \leq c_{R_d}(u_q + 1) \\ & \text{for } 2 \leq q \leq \sqrt{m} - 1, \\ u_{\sqrt{m}} & \text{if } y_I(k) > c_{R_d}(u_{\sqrt{m}} - 1). \end{cases} \quad (54)$$

Note that the main tap c_d must be known in any receiver, i.e. the d th column of the system matrix \mathbf{H} must be known in the receiver. If this fact is overlooked, the decision will be biased [88]. In general, c_d is complex-valued and the rotating operation

$$\mathbf{w}^{(\text{new})} = \frac{c_d^{(\text{old})}}{|c_d^{(\text{old})}|} \mathbf{w}^{(\text{old})} \quad (55)$$

can be used to make c_d real and positive. This rotation is a linear operation and it does not alter the system's symbol error rate (SER).

4.1. Conventional filtering design

The classic Wiener filtering is still given by the closed-form MMSE solution (19), and the LMS based adaptive filtering (21) can be modified to

$$\begin{cases} y(k) = \tilde{\mathbf{w}}^H(k-1)\mathbf{x}(k), \\ \hat{\mathbf{w}}(k) = \tilde{\mathbf{w}}(k-1) + \mu(s_d(k) - y(k))^*\mathbf{x}(k), \\ \hat{c}_d(k) = \hat{\mathbf{w}}^H(k)\tilde{\mathbf{h}}_d, \\ \tilde{\mathbf{w}}(k) = \frac{\hat{c}_d(k)}{|\hat{c}_d(k)|}\hat{\mathbf{w}}(k), \end{cases} \quad (56)$$

where $\tilde{\mathbf{h}}_d$ denotes an estimate of \mathbf{h}_d . Given a block of training samples $\{s_d(k), \mathbf{x}(k)\}_{k=1}^K$, a block-based estimate of \mathbf{h}_d is given by

$$\tilde{\mathbf{h}}_d = \frac{1}{K} \sum_{k=1}^K \frac{\mathbf{x}(k)}{s_d(k)}. \quad (57)$$

Alternatively, the receiver can track \mathbf{h}_d using the simple moving average

$$\tilde{\mathbf{h}}_d(k) = (1 - \alpha)\tilde{\mathbf{h}}_d(k-1) + \alpha \frac{\mathbf{x}(k)}{s_d(k)}, \quad (58)$$

where $0 < \alpha < 1$ is a positive step size. Note that $\tilde{c}_d(k) = \tilde{\mathbf{w}}^H(k)\tilde{\mathbf{h}}_d(k)$ is real-valued and positive.

4.2. SER expression

The complex-valued signal vector set \mathcal{X} now contains $N_s = m^M$ points, and it can be divided into the m subsets, each having $N_{sb} = N_s/m$ points

$$\mathcal{X}_{l,q} = \{\tilde{\mathbf{x}}_i^{(l,q)} \in \mathcal{X}, 1 \leq i \leq N_{sb} : s_d(k) = s_{l,q}\}, \quad 1 \leq l, q \leq \sqrt{m}. \quad (59)$$

Similarly, the complex-valued scalar set \mathcal{Y} can be divided into the m subsets

$$\mathcal{Y}_{l,q} = \{\tilde{y}_i^{(l,q)} \in \mathcal{Y}, 1 \leq i \leq N_{sb} : s_d(k) = s_{l,q}\}, \quad 1 \leq l, q \leq \sqrt{m}. \quad (60)$$

The following two propositions [39] summarise the properties of the signal subsets $\mathcal{Y}_{l,q}$, $1 \leq l, q \leq \sqrt{m}$, which are useful in the derivation of the SER expression for the linear filter (3).

Proposition 3. *The subsets $\mathcal{Y}_{l,q}$, $1 \leq l, q \leq \sqrt{m}$, satisfy the shifting properties*

$$\begin{cases} \mathcal{Y}_{l+1,q} = \mathcal{Y}_{l,q} + 2c_d, & 1 \leq l \leq \sqrt{m} - 1, \\ \mathcal{Y}_{l,q+1} = \mathcal{Y}_{l,q} + j2c_d, & 1 \leq q \leq \sqrt{m} - 1, \\ \mathcal{Y}_{l+1,q+1} = \mathcal{Y}_{l,q} + (2 + j2)c_d, & 1 \leq l, q \leq \sqrt{m} - 1. \end{cases} \quad (61)$$

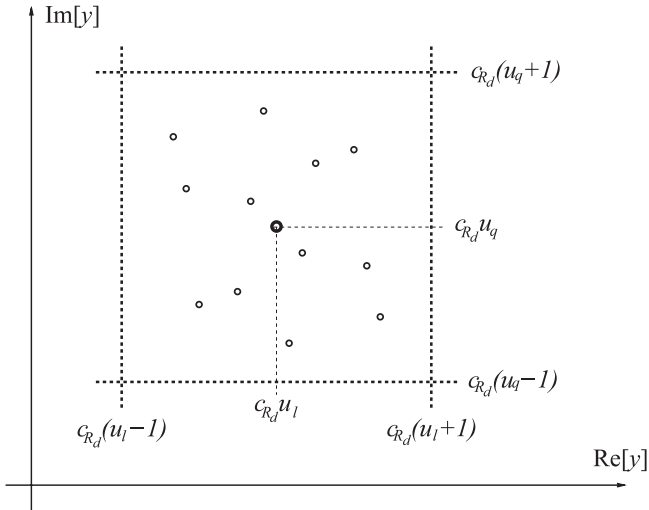


Fig. 15. Decision thresholds associated with symbol point $c_{d s_{l,q}}$ assuming $c_{R_d} > 0$ and $c_{I_d} = 0$, and illustration of symmetric distribution of $\mathcal{Y}_{l,q}$ around $c_{d s_{l,q}}$.

Proposition 4. *The points of $\mathcal{Y}_{l,q}$ are distributed symmetrically around the symbol point $c_{d s_{l,q}}$.*

Fig. 15 depicts the decision thresholds associated with the decision rule (53) and (54) as well as illustrates Proposition 4. For the filter with the weight vector \mathbf{w} , denote

$$\begin{cases} P_E(\mathbf{w}) = \text{Prob}\{\hat{s}_d(k) \neq s_d(k)\}, \\ P_{E_R}(\mathbf{w}) = \text{Prob}\{\hat{s}_{R_d}(k) \neq s_{R_d}(k)\}, \\ P_{E_I}(\mathbf{w}) = \text{Prob}\{\hat{s}_{I_d}(k) \neq s_{I_d}(k)\}. \end{cases} \quad (62)$$

$P_E(\mathbf{w})$ is the total SER, while $P_{E_R}(\mathbf{w})$ and $P_{E_I}(\mathbf{w})$ are the real-part and imaginary-part SERs, respectively. It is then easy to see that the SER is given by

$$P_E(\mathbf{w}) = P_{E_R}(\mathbf{w}) + P_{E_I}(\mathbf{w}) - P_{E_R}(\mathbf{w})P_{E_I}(\mathbf{w}). \quad (63)$$

The conditional PDF of $y(k)$ given $s_d(k) = s_{l,q}$ is a Gaussian mixture defined by

$$p(y|s_d(k) = s_{l,q}) = \frac{1}{N_{sb} 2\pi \sigma_n^2 \mathbf{w}^H \mathbf{w}} \sum_{i=1}^{N_{sb}} e^{-|y - \bar{y}_i^{(l,q)}|^2 / 2\sigma_n^2 \mathbf{w}^H \mathbf{w}}, \quad (64)$$

where $\bar{y}_i^{(l,q)} = \bar{y}_{R_i}^{(l,q)} + j\bar{y}_{I_i}^{(l,q)} \in \mathcal{Y}_{l,q}$. Noting that c_d is real-valued and positive and taking into account the symmetric distribution of $\mathcal{Y}_{l,q}$ (Proposition 4), for $2 \leq l \leq \sqrt{m} - 1$, the conditional error probability of $\hat{s}_{R_d}(k) \neq u_l$ given $s_{R_d}(k) = u_l$ can be shown to be [39]

$$P_{E_{R,d}}(\mathbf{w}) = \frac{2}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{R_i}^{(l,q)}(\mathbf{w})), \quad (65)$$

with

$$g_{R_i}^{(l,q)}(\mathbf{w}) = \frac{\bar{y}_{R_i}^{(l,q)} - c_{R_d}(u_l - 1)}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}}. \quad (66)$$

Further taking into account the shifting property (Proposition 3), it can be shown that

$$P_{E_R}(\mathbf{w}) = \gamma \frac{1}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{R_i}^{(l,q)}(\mathbf{w})), \quad (67)$$

where $\gamma = (2\sqrt{m} - 2)/\sqrt{m}$. It is seen that P_{E_R} can be evaluated using (real part of) any single subset $\mathcal{Y}_{l,q}$. Similarly, P_{E_I} can be evaluated using (imaginary part of) any single subset $\mathcal{Y}_{l,q}$ as

$$P_{E_I}(\mathbf{w}) = \gamma \frac{1}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{I_i}^{(l,q)}(\mathbf{w})) \quad (68)$$

with

$$g_{I_i}^{(l,q)}(\mathbf{w}) = \frac{\bar{y}_{I_i}^{(l,q)} - c_{R_d}(u_q - 1)}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}}. \quad (69)$$

Note that the SER is invariant to a positive scaling of \mathbf{w} .

4.3. MSER filtering

The MSER solution \mathbf{w}_{MSER} is defined as the weight vector that minimises the upper bound of the SER given by

$$P_{E_B}(\mathbf{w}) = P_{E_R}(\mathbf{w}) + P_{E_I}(\mathbf{w}), \quad (70)$$

that is,

$$\mathbf{w}_{\text{MSER}} = \arg \min_{\mathbf{w}} P_{E_B}(\mathbf{w}). \quad (71)$$

The solution obtained by minimising the upper bound (70) is practically equivalent to that of minimising $P_E(\mathbf{w})$, since the bound $P_E(\mathbf{w}) < P_{E_B}(\mathbf{w})$ is very tight, that is, $P_{E_B}(\mathbf{w})$ is very close to the true SER $P_E(\mathbf{w})$. The gradients of $P_{E_R}(\mathbf{w})$ and $P_{E_I}(\mathbf{w})$ with respect to \mathbf{w} can be shown to be, respectively,

$$\begin{aligned} \nabla P_{E_R}(\mathbf{w}) &= \frac{\gamma}{2N_{sb} \sqrt{2\pi} \sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \sum_{i=1}^{N_{sb}} e^{-|\bar{y}_{R_i}^{(l,q)} - c_{R_d}(u_l - 1)|^2 / 2\sigma_n^2 \mathbf{w}^H \mathbf{w}} \\ &\quad \times \left(\frac{\bar{y}_{R_i}^{(l,q)} - c_{R_d}(u_l - 1)}{\mathbf{w}^H \mathbf{w}} \mathbf{w} - \bar{\mathbf{x}}_i^{(l,q)} + (u_l - 1)\mathbf{h}_d \right) \end{aligned} \quad (72)$$

and

$$\begin{aligned} \nabla P_{E_I}(\mathbf{w}) &= \frac{\gamma}{2N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}} \sum_{i=1}^{N_{sb}} e^{-|y_{I_i}^{(l,q)} - c_{R_d}(u_q-1)|^2/2\sigma_n^2\mathbf{w}^H\mathbf{w}} \\ &\times \left(\frac{y_{I_i}^{(l,q)} - c_{R_d}(u_q-1)}{\mathbf{w}^H\mathbf{w}} \mathbf{w} + j\tilde{\mathbf{x}}_i^{(l,q)} + (u_q-1)\tilde{\mathbf{h}}_d \right), \end{aligned} \quad (73)$$

where $\tilde{\mathbf{x}}_i^{(l,q)} \in \mathcal{X}_{l,q}$. With the gradient $\nabla P_{E_B}(\mathbf{w}) = \nabla P_{E_R}(\mathbf{w}) + \nabla P_{E_I}(\mathbf{w})$, the optimisation problem (71) can be solved iteratively using a gradient-based algorithm, such as the simplified conjugate gradient algorithm.

Using the same kernel density estimation approach for deriving the adaptive LBER algorithm, the following adaptive least SER (LSER) algorithm can be obtained [39]

$$\begin{cases} y(k) = \tilde{\mathbf{w}}^H(k-1)\mathbf{x}(k), \\ \tilde{\mathbf{w}}(k) = \tilde{\mathbf{w}}(k-1) + \mu(-\nabla \tilde{P}_{E_B}(\tilde{\mathbf{w}}(k-1), k)), \\ \hat{c}_d(k) = \hat{\mathbf{w}}^H(k)\tilde{\mathbf{h}}_d, \\ \tilde{\mathbf{w}}(k) = \frac{\hat{c}_d(k)}{|\hat{c}_d(k)|} \hat{\mathbf{w}}(k), \end{cases} \quad (74)$$

where the stochastic gradient $\nabla \tilde{P}_{E_B}(\mathbf{w}, k) = \nabla \tilde{P}_{E_R}(\mathbf{w}, k) + \nabla \tilde{P}_{E_I}(\mathbf{w}, k)$ with

$$\begin{aligned} \nabla \tilde{P}_{E_R}(\mathbf{w}, k) &= \frac{\gamma}{2\sqrt{2\pi}\rho_n} e^{-|y_R(k) - \tilde{c}_{R_d}(k-1)(s_{R_d}(k)-1)|^2/2\rho_n^2} \\ &\times (-\mathbf{x}(k) + (s_{R_d}(k)-1)\tilde{\mathbf{h}}_d) \end{aligned} \quad (75)$$

and

$$\begin{aligned} \nabla \tilde{P}_{E_I}(\mathbf{w}, k) &= \frac{\gamma}{2\sqrt{2\pi}\rho_n} e^{-|y_I(k) - \tilde{c}_{R_d}(k-1)(s_{I_d}(k)-1)|^2/2\rho_n^2} \\ &\times (j\mathbf{x}(k) + (s_{I_d}(k)-1)\tilde{\mathbf{h}}_d). \end{aligned} \quad (76)$$

4.4. Illustrative examples

The simulated beamforming system consisted of four m -QAM user sources and a three-element antenna array. Fig. 16 shows the locations of the desired user and the interfering users graphically, where the minimum angular separation between the desired user and the interfering user 4 was $\theta < 65^\circ$. In the simulation study, a perfect $\tilde{\mathbf{h}}_d$ was assumed at the receiver. Hence, our attention was focused on the performance of the adaptive MMSE and MSER beamforming designs, rather than on the adaptive estimator for $\tilde{\mathbf{h}}_d$.

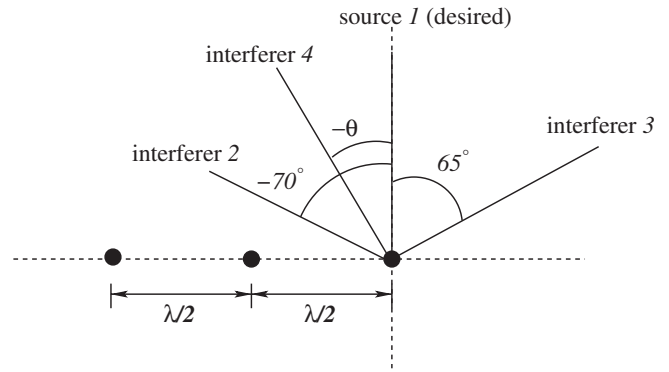


Fig. 16. Locations of the four m -QAM user sources with respect to the three-element array having $\lambda/2$ element spacing, where λ is the wavelength and the minimum angular separation $\theta < 60^\circ$.

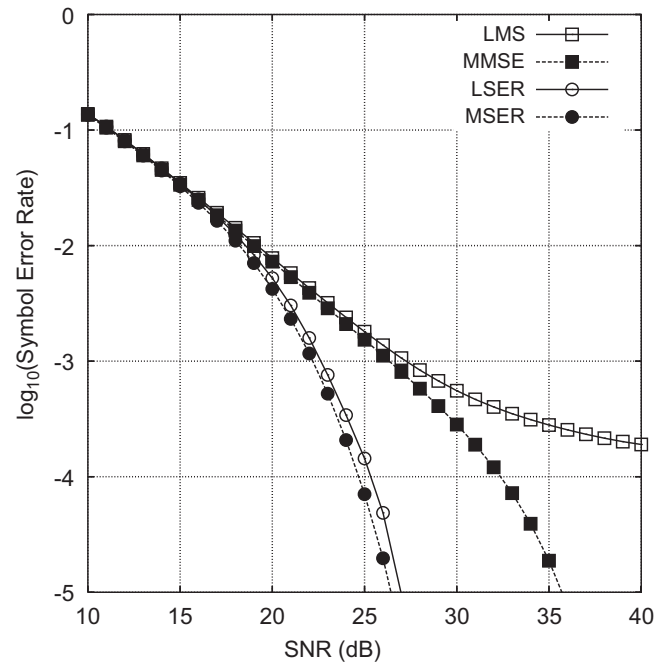


Fig. 17. Desired user's symbol error rate comparison of two beamforming designs and their adaptive implementations for the non-fading system employing the three-element array with a minimum angular separation of $\theta = 30^\circ$ to support four equal-power 16-QAM users.

Stationary system: The modulation scheme was 16-QAM and all the channels $A_i, 1 \leq i \leq 4$, were time-invariant. Fig. 17 compares the SER performance of the MSER beamforming solution to that of the MMSE beamforming solution, under the conditions that the minimum angular separation between the desired user and the interfering user 4 was $\theta = 30^\circ$, and all the four users had an equal signal power. The adaptive performance of the LMS and LSER beamformers are also depicted in Fig. 17, in comparison with their respective theoretic SER

performance. The superiority of the adaptive LSER beamformer over the adaptive LMS beamformer is clearly demonstrated in Fig. 17, where it can be seen that the performance of the LMS beamformer was notably deviated from its theoretic MMSE solution at high SNR values.

The MSER solution is defined as the weight vector that minimises the upper bound SER $P_{E_B}(\mathbf{w})$ of (70). The true SER $P_E(\mathbf{w})$ is given by the sum of the inphase and quadrature components' error rates minus the appropriate correction term used for preventing the “double-counting” error-events, as seen in (63). The probability of simultaneous inphase and quadrature errors, which is represented by the term $P_{E_R}(\mathbf{w})P_{E_I}(\mathbf{w})$ tends to be very small, unless the SNR is extremely low. More explicitly, this term is typically orders of magnitude lower than the first two terms of $P_E(\mathbf{w})$ and, therefore, is negligible. Hence the bound $P_E(\mathbf{w}) < P_{E_B}(\mathbf{w})$ is very tight. In fact, $P_{E_B}(\mathbf{w})$ is almost indistinguishable from $P_E(\mathbf{w})$. This is confirmed by the results of Fig. 18, where both the true SER $P_E(\mathbf{w})$ and its upper bound $P_{E_B}(\mathbf{w})$ are plotted for the MMSE and MSER solutions under the same conditions of Fig. 17.

Fading system: The antenna array structure was as illustrated in Fig. 16, but the modulation scheme

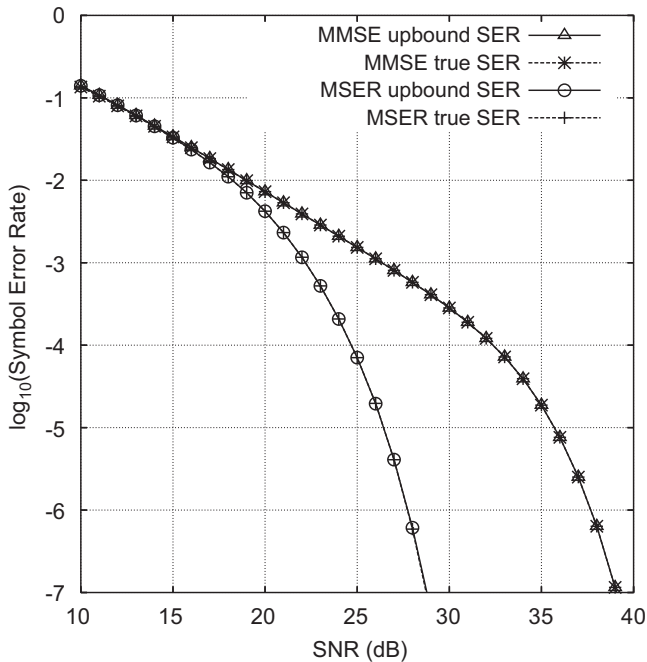


Fig. 18. Comparison of the true symbol error rate and its upper bound for the non-fading system employing the three-element array with a minimum angular separation of $\theta = 30^\circ$ to support four equal-power 16-QAM users.

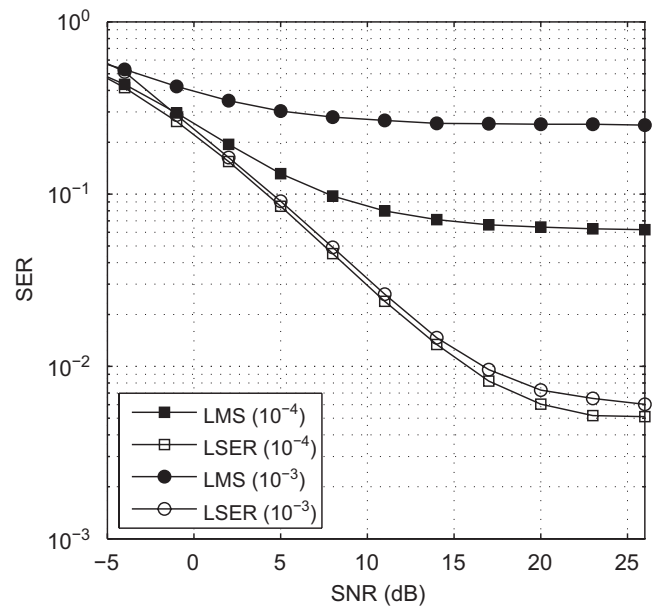


Fig. 19. Desired user's symbol error rate performance comparison for the fading systems of the two normalised Doppler frequencies $\bar{f}_D = 10^{-4}$ and 10^{-3} employing the three-element array with a minimum angular separation of $\theta = 27^\circ$ to support four 64-QAM users. The LMS algorithm has a step size $\mu = 0.0002$, while the LSER algorithm has a step size $\mu = 0.00005$ and a kernel width $\rho_n = 4\sigma_n$.

was 64-QAM. All the four users had an equal power. Fading channels were simulated, where the magnitudes of A_i for $1 \leq i \leq 4$ were Rayleigh processes with the normalised Doppler frequency \bar{f}_D and each A_i had the root mean power of $\sqrt{0.5} + j\sqrt{0.5}$. Continuously fluctuating fading was used, providing a different fading magnitude and phase for each transmitted symbol. The transmission frame structure consisted of 50 training symbols followed by 450 data symbols. Decision-directed adaptation was employed during data transmission, in which the adaptive beamforming detector's decision $\hat{s}_d(k)$ was used to substitute for $s_d(k)$. The SER of an adaptive beamforming detector was calculated using the 450 data symbols of the frame based on Monte Carlo simulation averaging over at least 2×10^5 frames, depending on the value of \bar{f}_D .

Given the minimum angular separation $\theta = 27^\circ$, Fig. 19 compares the SER of the adaptive LSER beamformer with that of the LMS-based one, for the two normalised Doppler frequencies $\bar{f}_D = 10^{-4}$ and 10^{-3} . It can be seen from Fig. 19 that the SER performance of the adaptive LSER beamformer degraded only slightly when the fading rate increased from $\bar{f}_D = 10^{-4}$ to 10^{-3} . This demonstrates

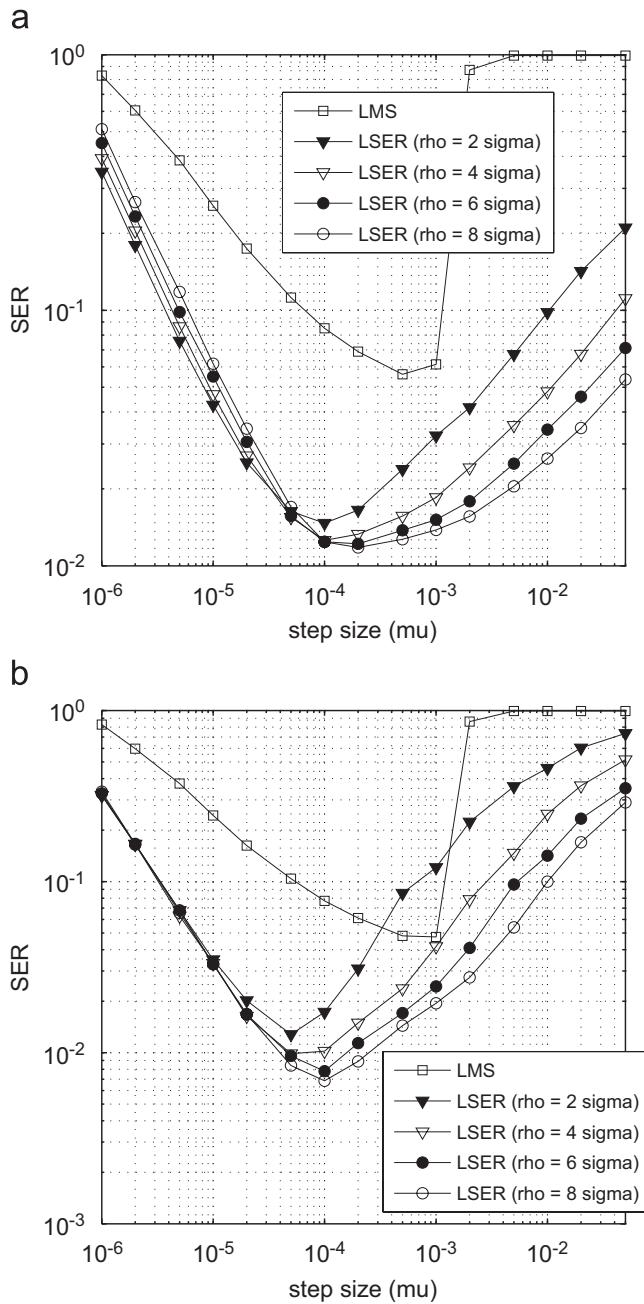


Fig. 20. Influence of the adaptive algorithm's parameters to the SER performance for the fading system of the normalised Doppler frequency $\tilde{f}_D = 10^{-4}$ employing the three-element array with a minimum angular separation of $\theta = 27^\circ$ to support four 64-QAM users: (a) average SNR = 15 dB, and (b) average SNR = 30 dB.

that the LSER algorithm has an excellent tracking ability, capable of operating in fast fading conditions. The influence of the adaptive algorithm's parameters, the step size μ for the LMS algorithm, and the step size μ and kernel width ρ_n for the LSER algorithm, were next investigated. Given $\tilde{f}_D = 10^{-4}$, Fig. 20(a) shows the influence of the adaptive

algorithm's parameters, μ for the LMS algorithm, and μ and ρ_n for the LSER algorithm, on the SER performance for a low average SNR value of 15 dB⁵, while Fig. 20(b) depicts the results for a high average SNR value of 30 dB. These results also explain why $\mu = 0.0002$ for the LMS algorithm and $\mu = 0.00005$ and $\rho_n = 4\sigma_n$ for the LSER algorithm were used in the simulation of Fig. 19.

5. Extension to nonlinear filtering

Our discussion so far restricts to the linear filter model (3), which is most widely used in various communication applications. For notational simplicity in this section we again concentrate on the BPSK modulation scheme (2). For the linear filter (3) to work satisfactorily, an implicit assumption is that the two complex-valued vector subsets, $\mathcal{X}^{(\pm)}$, corresponding to the two values of $s_d(k)$, are linearly separable. That is, there exists a weight vector \mathbf{w} such that the two real-valued scalar subsets, $\mathcal{Y}_R^{(+)}$ and $\mathcal{Y}_R^{(-)}$, can completely be separated by the decision threshold $y_R = 0$. Otherwise, nonlinear filtering is required in order to achieve an adequate performance. Examples of such nonlinear filtering include nonlinear single-user channel equalisation [89–101], nonlinear CDMA multiuser detection [102], nonlinear beamforming assisted detection [103–106], and nonlinear space-time equalisation [107]. Let us consider the generic nonlinear filter of the form

$$y_R(k) = f(\mathbf{x}(k); \mathbf{w}), \quad (77)$$

where $f(\bullet; \bullet)$ is a real-valued nonlinear mapping, and the parameter vector \mathbf{w} contains all the adjustable parameters of the nonlinear filter. Such a nonlinear filter for example may be realised by a neural network. The real-valued filter output $y_R(k)$ is used to estimate the desired data symbol $s_d(k)$ according to the decision rule (4).

5.1. Nonlinear LMS error filtering

Most of the training algorithms for the nonlinear filter (77) adopt the nonlinear MSE criterion

$$J(\mathbf{w}) = E[|s_d(k) - f(\mathbf{x}(k); \mathbf{w})|^2]. \quad (78)$$

In particular, adaptive training of the nonlinear filter can be carried out using an extension of the

⁵Note that this was a 64-QAM system, and a SNR of 15 dB was relatively small.

LMS approach, which we refer to as the nonlinear LMS (NLMS) algorithm. The NLMS algorithm takes the following simple form

$$\begin{cases} y_R(k) = f(\mathbf{x}(k); \tilde{\mathbf{w}}(k-1)), \\ \tilde{\mathbf{w}}(k) = \tilde{\mathbf{w}}(k-1) + \mu(s_d(k) \\ - y_R(k)) \frac{\partial f(\mathbf{x}(k); \tilde{\mathbf{w}}(k-1))}{\partial \mathbf{w}}, \end{cases} \quad (79)$$

where μ is the step size. However, for communication applications, this nonlinear MMSE (NMMSE) approach generally leads to a suboptimal performance, in terms of the achievable system's BER. In fact, it is easy to see that multiplying the output of the nonlinear filter (77) by a positive scalar will change its MSE value $J(\mathbf{w})$ but not its BER.

5.2. Nonlinear least BER filtering

It is highly desired for communication applications to directly minimise the BER of the nonlinear filter (77). Let us define the following signed decision variable

$$y_s(k) = \text{sgn}(s_d(k))y_R(k) \quad (80)$$

and denote the PDF of $y_s(k)$ as $p_s(y_s)$. Then the error probability or BER of the nonlinear filter (77) is given by

$$P_E(\mathbf{w}) = \text{Prob}\{y_s(k) < 0\} = \int_{-\infty}^0 p_s(y_s) dy_s. \quad (81)$$

The nonlinear MBER (NMBER) solution for the filter's parameter vector \mathbf{w} is defined as

$$\mathbf{w}_{\text{NMBER}} = \arg \min_{\mathbf{w}} P_E(\mathbf{w}). \quad (82)$$

The problem associated with this approach is that the PDF of $y_s(k)$ is generally unknown. However, it may be sufficiently accurately estimated using the Parzen window method [77–79]. Given a block of training data $\{s_d(k), \mathbf{x}(k)\}_{k=1}^K$, a Parzen window estimate of $p_s(y_s)$ is readily given as

$$\tilde{p}_s(y_s) = \frac{1}{K\sqrt{2\pi\rho_n}} \sum_{k=1}^K e^{-(y_s - \text{sgn}(s_d(k))y_R(k))^2/2\rho_n^2}, \quad (83)$$

where ρ_n^2 is the chosen kernel variance. With this estimated PDF, the estimated or approximate BER for the nonlinear filter (77) is given by

$$\tilde{P}_E(\mathbf{w}) = \int_{-\infty}^0 \tilde{p}_s(y_s) dy_s = \frac{1}{K} \sum_{k=1}^K Q(\tilde{g}_k(\mathbf{w})), \quad (84)$$

with

$$\tilde{g}_k(\mathbf{w}) = \frac{\text{sgn}(s_d(k))y_R(k)}{\rho_n}. \quad (85)$$

An approximate NMBER solution for \mathbf{w} can be obtained by minimising $\tilde{P}_E(\mathbf{w})$ using a gradient-based optimisation algorithm.

In particular, consider a single-sample PDF “estimate” of $p_s(y_s)$ given by

$$\tilde{p}_s(y_s, k) = \frac{1}{\sqrt{2\pi\rho_n}} e^{-(y_s - \text{sgn}(s_d(k))y_R(k))^2/2\rho_n^2}. \quad (86)$$

With this instantaneous PDF “estimate”, we have a single-sample or instantaneous BER “estimate” $\tilde{P}_E(\mathbf{w}, k) = Q(\tilde{g}_k(\mathbf{w}))$. Using the instantaneous gradient of

$$\nabla \tilde{P}_E(\mathbf{w}, k) = -\frac{1}{\sqrt{2\pi\rho_n}} e^{-y_R^2(k)/2\rho_n^2} \text{sgn}(s_d(k)) \frac{\partial f(\mathbf{x}(k); \mathbf{w})}{\partial \mathbf{w}} \quad (87)$$

gives rise to the following stochastic adaptive algorithm:

$$\begin{cases} y_R(k) = f(\mathbf{x}(k); \tilde{\mathbf{w}}(k-1)), \\ \tilde{\mathbf{w}}(k) = \tilde{\mathbf{w}}(k-1) + \frac{\mu}{\sqrt{2\pi\rho_n}} e^{-y_R^2(k)/2\rho_n^2} \\ \times \text{sgn}(s_d(k)) \frac{\partial f(\mathbf{x}(k); \tilde{\mathbf{w}}(k-1))}{\partial \mathbf{w}}, \end{cases} \quad (88)$$

which we refer to as the nonlinear LBER (NLBER) algorithm. The step size μ and kernel variance ρ_n^2 should be chosen appropriately to achieve a desired convergence performance, both in terms of convergence speed and steady-state BER misadjustment.

5.3. Illustrative example

We will use the beamforming assisted receiver, depicted in Fig. 3, as an example to illustrate the above NLBER filtering. The optimal nonlinear filtering for beamforming detection is known to be the Bayesian detector [103–106], which requires the complete knowledge of the underlying system (1). The Bayesian beamforming detector has an inherently odd symmetry property [105,106]. Thus, the optimal Bayesian solution is specified by the complex-valued vector subset $\mathcal{X}^{(+)}$, which contains N_{sb} states, and the distribution of the noise $\mathbf{n}(k)$.

We consider the following radial basis function (RBF) based nonlinear filter for beamforming

$$y_R(k) = f(\mathbf{x}(k); \mathbf{w}) = \sum_{i=1}^{\hat{N}_{sb}} \alpha_i \phi_i(\mathbf{x}(k)), \quad (89)$$

where α_i is the i th real-valued RBF weight, $\phi_i(\bullet)$ denotes the response of the i th RBF node, \hat{N}_{sb} is the number of RBF nodes used, and \mathbf{w} denotes the vector of all the adjustable parameters of the RBF filter. We adopt the following symmetric RBF node [105,106]

$$\phi_i(\mathbf{x}) = \varphi(\mathbf{x}; \mathbf{c}_i, \sigma_i^2) - \varphi(\mathbf{x}; -\mathbf{c}_i, \sigma_i^2), \quad (90)$$

where $\mathbf{c}_i \in \mathcal{C}^L$ is the i th complex-valued RBF centre, σ_i^2 the i th real-valued and positive RBF variance, and $\varphi(\bullet)$ is the RBF function. The parameter vector \mathbf{w} of this symmetric RBF filter (89) therefore consists of all the RBF weights α_i , RBF centre vectors \mathbf{c}_i as well as RBF variances σ_i^2 . In the following simulation we adopt the Gaussian RBF function of

$$\varphi(\mathbf{x}; \mathbf{c}_i, \sigma_i^2) = e^{-\|\mathbf{x} - \mathbf{c}_i\|^2 / \sigma_i^2}. \quad (91)$$

Note that the RBF filter (89) with the node structure defined in (90) has an inherently odd symmetry, just as the optimal Bayesian solution [105,106]. For the symmetric RBF filter (89) using the Gaussian function (91), the derivatives of the RBF filter's output with respect to the RBF filter's parameters are given by

$$\left. \begin{aligned} \frac{\partial f}{\partial \alpha_i} &= e^{-\|\mathbf{x}(k) - \mathbf{c}_i\|^2 / \sigma_i^2} - e^{-\|\mathbf{x}(k) + \mathbf{c}_i\|^2 / \sigma_i^2}, \\ \frac{\partial f}{\partial \sigma_i^2} &= \alpha_i \left(\frac{e^{-\|\mathbf{x}(k) - \mathbf{c}_i\|^2 / \sigma_i^2} \|\mathbf{x}(k) - \mathbf{c}_i\|^2}{(\sigma_i^2)^2} \right. \\ &\quad \left. - \frac{e^{-\|\mathbf{x}(k) + \mathbf{c}_i\|^2 / \sigma_i^2} \|\mathbf{x}(k) + \mathbf{c}_i\|^2}{(\sigma_i^2)^2} \right), \\ \frac{\partial f}{\partial \mathbf{c}_i} &= \alpha_i \left(\frac{e^{-\|\mathbf{x}(k) - \mathbf{c}_i\|^2 / \sigma_i^2} \mathbf{x}(k) - \mathbf{c}_i}{\sigma_i^2} \right. \\ &\quad \left. + \frac{e^{-\|\mathbf{x}(k) + \mathbf{c}_i\|^2 / \sigma_i^2} \mathbf{x}(k) + \mathbf{c}_i}{\sigma_i^2} \right), \end{aligned} \right\} \quad (92)$$

for $1 \leq i \leq \hat{N}_{sb}$.

A two-element antenna array with half wavelength spacing was designed to support four BPSK signal sources. The users' angular positions are summarised in Table 3. The simulated narrowband channels were $A_i = 1 + j0$, $1 \leq i \leq 4$. The user one was the desired user, and all the users had an equal

Table 3

Locations of the users in terms of angle of arrival for the simulated beamforming system that employs a two-element antenna array to support four BPSK users

User i	1	2	3	4
AOA θ_i ($^\circ$)	0	20	-30	-45

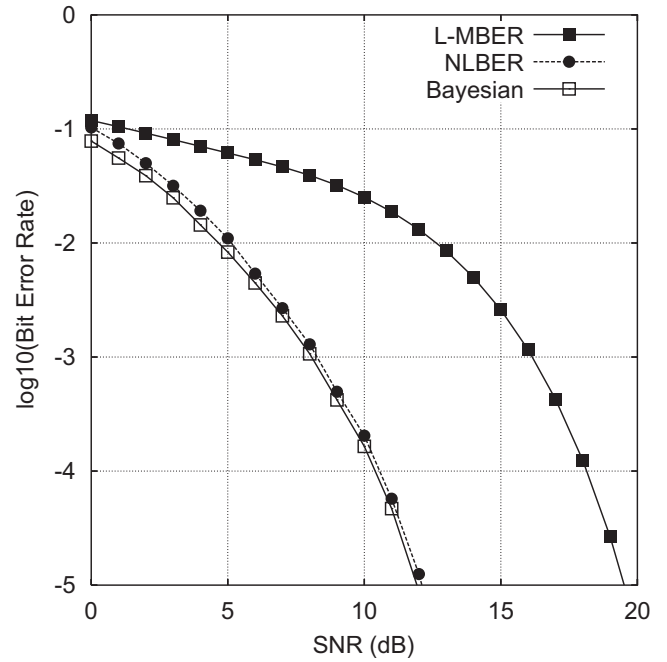


Fig. 21. Desired user's bit error rate comparison of three beamforming designs for the two-element array system supporting four BPSK users as given in Table 3.

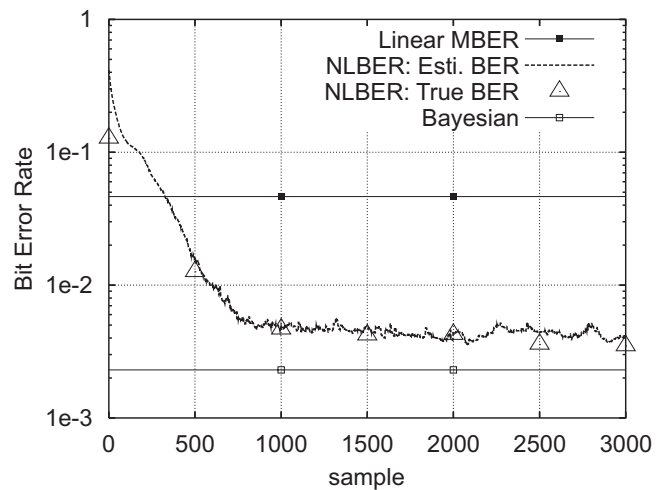


Fig. 22. Learning curve of the NLBER RBF detector averaged over 10 runs for the two-element antenna array supporting four BPSK users at the angular positions of Table 3, where SNR = 7 dB and the RBF detector has $\hat{N}_{sb} = 8$ symmetric RBF nodes.

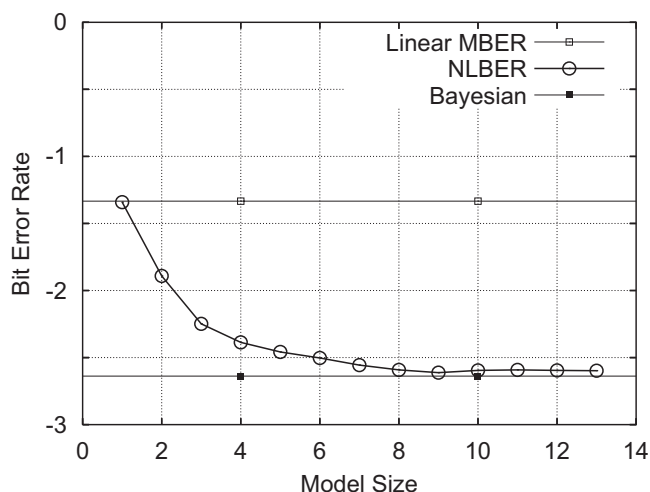


Fig. 23. Influence of the detector's size on the bit error rate performance of the NLBER symmetric RBF detector for the two-element antenna array supporting four BPSK users at the angular positions of Table 3, where SNR = 7 dB.

power. Fig. 21 depicts the BER performance of both the theoretical linear MBER (L-MBER) beamformer and the optimal nonlinear Bayesian detector for the desired user. For this example, the size of the Bayesian detector was specified by the number of symmetric signal states $N_{sb} = 8$.

The convergence performance of the NLBER algorithm was first investigated. Given a SNR value of 7 dB and a RBF filter size of $\hat{N}_{sb} = 8$, Fig. 22 shows the learning curve of the NLBER algorithm averaged over 10 independent simulation runs. The step size and kernel variance of the NLBER algorithm (88) were chosen to be $\mu = 0.4$ and $\rho_n^2 = 10\sigma_n^2$. In fact, μ in the range of 0.3–0.5 and ρ_n^2 in the range of $9\sigma_n^2$ to $11\sigma_n^2$ were found empirically to be appropriate for this example. The learning curve (dashed curve) was the estimated BER $\tilde{P}_E(\tilde{\mathbf{w}}(k))$, calculated using Eq. (84) for each $\tilde{\mathbf{w}}(k)$ in conjunction with a block size of $K = 400$ and a kernel variance of $\tilde{\rho}_n^2 = \sigma_n^2$. Note that $\tilde{\rho}_n^2$ was not the kernel variance of the NLBER algorithm and was only used to approximate the BER. In order to check that the estimated BER $\tilde{P}_E(\tilde{\mathbf{w}}(k))$ gave the correct convergence trend, we also calculated the true BER $P_E(\tilde{\mathbf{w}}(k))$ using Monte Carlo simulation for a number of points, shown in Fig. 22 by the triangles. The results of Fig. 22 confirm that the estimated BER correctly indicated the convergence trend.

The influence of the number of RBF centres \hat{N}_{sb} on the performance of the NLBER-based symmetric RBF detector was studied next. Given

SNR = 7 dB, Fig. 23 illustrates the performance of the NLBER-based detector as a function of the number of RBF centres \hat{N}_{sb} . It can be seen from Fig. 23 that for $\hat{N}_{sb} \geq N_{sb}$ the symmetric RBF detector trained by the stochastic NLBER algorithm becomes capable of closely approaching the optimal Bayesian performance. It is also interesting to observe in Fig. 23 that using a single symmetric RBF node the RBF detector achieves the same performance as the L-MBER solution, since the RBF detector of a single symmetric RBF node is only capable of constructing a linear decision boundary. For each SNR value, the BER performance of the NLBER-based symmetric RBF detector having $\hat{N}_{sb} = 8$ RBF nodes is depicted in Fig. 21, in comparison to the optimal Bayesian performance.

6. Conclusions

A unified framework has been presented for the adaptive linear filtering design based on directly minimising the system's BER. Our motivation has been the well-known fact that the traditional Wiener design is far from optimal for applications in various communication systems. The MMSE filtering corresponds to the MBER solution only if the conditional PDF of the filter's output is Gaussian distributed. Since this conditional PDF is generally a mixture of Gaussian distributions, and hence non-Gaussian, the MMSE solution does not achieve the MBER performance. It has been demonstrated that the MBER design is more intelligent and it better exploits the non-Gaussian nature of the filter's output, leading to significant performance enhancement, in terms of better combating hostile multipath propagation environments and better suppressing multiple access interference as well as achieving higher system throughput or user capacity. Interesting analogy has been drawn between the traditional adaptive filtering approach based on the MMSE criterion and the adaptive MBER filtering approach. In particular, adaptive implementation of the MBER filtering design has been proposed based on a stochastic-gradient algorithm referred to as the LBER method. Extension to adaptive MSER filtering has also been presented, which is suitable for communication systems that employ high throughput QAM modulation schemes. Finally, a novel nonlinear adaptive MBER filtering approach has been proposed for BPSK communication systems.

References

- [1] J.G. Proakis, Digital Communications, third ed., McGraw-Hill, New York, 1995.
- [2] L. Hanzo, W. Webb, T. Keller, Single- and Multi-Carrier Quadrature Amplitude Modulation: Principles and Applications for Personal Communications, WLANs and Broadcasting, Wiley, Chichester, UK, 2000.
- [3] B. Widrow, S.D. Stearns, Adaptive Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1985.
- [4] S. Haykin, Adaptive Filter Theory, third ed., Prentice-Hall, Upper Saddle River, NJ, 1996.
- [5] E. Shamash, K. Yao, On the structure and performance of a linear decision feedback equalizer based on the minimum error probability criterion, in: Proceedings of the ICC'74, 1974, pp. 25F1–25F5.
- [6] S. Chen, E.S. Chng, B. Mulgrew, G. Gibson, Minimum-BER linear-combiner DFE, in: Proceedings of the ICC'96, vol. 2, Dallas, TX, 1996, pp. 1173–1177.
- [7] C.-C. Yeh, J.R. Barry, Approximate minimum bit-error rate equalization for binary signaling, in: Proceedings of the ICC'97, vol. 2, Montreal, Canada, 1997, pp. 1095–1099.
- [8] S. Chen, B. Mulgrew, E.S. Chng, G. Gibson, Space translation properties and the minimum-BER linear-combiner DFE, IEE Proc. Comm. 145 (5) (1998) 316–322.
- [9] S. Chen, B. Mulgrew, The minimum-SER linear-combiner decision feedback equalizer, IEE Proc. Comm. 146 (6) (1999) 347–353.
- [10] S. Chen, C.J. Harris, Design of the optimal separating hyperplane for the decision feedback equalizer using support vector machines, in: Proceedings of the ICASSP'2000, vol. 5, June 5–9, 2000, pp. 2701–2704.
- [11] B. Mulgrew, S. Chen, Stochastic gradient minimum-BER decision feedback equalisers, in: Proceedings of the IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control, Lake Louise, Alberta, Canada, October 1–4, 2000, pp. 93–98.
- [12] C.-C. Yeh, J.R. Barry, Adaptive minimum bit-error rate equalization for binary signaling, IEEE Trans. Comm. 48 (7) (2000) 1226–1235.
- [13] K.A. Phillips, J.H. Reed, W.H. Tranter, Minimum BER adaptive filtering, in: Proceedings of the ICC'2000, vol. 3, 2000, pp. 1675–1680.
- [14] B. Mulgrew, S. Chen, Adaptive minimum-BER decision feedback equalisers for binary signalling, Signal Process. 81 (7) (2001) 1479–1489.
- [15] C.-Y. Chen, C. Heneghan, J.M. Cioffi, A novel decision feedback equalizer design based on generalized space translation, in: Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers, vol. 1, November 7–10, 2004, pp. 727–731.
- [16] J. Riani, S. van Beneden, J.W.M. Bergmans, A. Immink, Near minimum bit-error rate equalizer adaptation for PRML systems, in: Proceedings of GLOBECOM'05, vol. 4, 28 November–2 December 2005, pp. 2123–2128.
- [17] C.-C. Yeh, J.R. Barry, Approximate minimum bit-error rate equalization for pulse-amplitude and quadrature-amplitude modulation, in: Proceedings of ICC'98, vol. 1, June 7–11, 1998, pp. 16–20.
- [18] S. Chen, B. Mulgrew, L. Hanzo, Stochastic least-symbol-error-rate adaptive equalization for pulse-amplitude modulation, in: Proceedings of the ICASSP 2002, vol. 3, Orlando, FL, USA, May 13–17, 2002, pp. 2629–2632.
- [19] C.-C. Yeh, J.R. Barry, Adaptive minimum symbol-error rate equalization for quadrature-amplitude modulation, IEEE Trans. Signal Process. 51 (12) (2003) 3263–3269.
- [20] S. Chen, L. Hanzo, B. Mulgrew, Adaptive minimum symbol-error-rate decision feedback equalization for multi-level pulse-amplitude modulation, IEEE Trans. Signal Process. 52 (7) (2004) 2092–2101.
- [21] N.B. Mandayam, B. Aazhang, Gradient estimation for sensitivity analysis and adaptive multiuser interference rejection in code-division multi-access systems, IEEE Trans. Comm. 45 (7) (1997) 848–858.
- [22] C.-C. Yeh, R.R. Lopes, J.R. Barry, Approximate minimum bit-error rate multiuser detection, in: Proceedings of Globecom'98, Sydney, Australia, 1998, pp. 3590–3595.
- [23] X.F. Wang, W.S. Lu, A. Antoniou, Constrained minimum-BER multiuser detection, in: Proceedings of the ICASSP, vol. 5, Phoenix, AZ, USA, May 14–18, 1999, pp. 2603–2606.
- [24] I.N. Psaromiligkos, S.N. Batalama, D.A. Pados, On adaptive minimum probability of error linear filter receivers for DS-CDMA channels, IEEE Trans. Comm. 47 (7) (1999) 1092–1102.
- [25] S. Chen, A.K. Samingan, B. Mulgrew, L. Hanzo, Adaptive minimum-BER linear multiuser detection, in: Proceedings of the ICASSP, vol. 4, Salt Lake City, UT, USA, May 7–11, 2001, pp. 2253–2256.
- [26] S. Chen, A.K. Samingan, B. Mulgrew, L. Hanzo, Adaptive minimum-BER linear multiuser detection for DS-CDMA signals in multipath channels, IEEE Trans. Signal Process. 49 (6) (2001) 1240–1247.
- [27] R.C. de Lamare, R. Sampaio-Neto, Adaptive multiuser receivers for DS-CDMA using minimum BER gradient-Newton algorithms, in: Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, vol. 3, September 15–18, 2002, pp. 1290–1294.
- [28] R.C. de Lamare, R. Sampaio-Neto, Adaptive MBER decision feedback multiuser receivers in frequency selective fading channels, IEEE Comm. Lett. 7 (2) (2003) 73–75.
- [29] A.K. Samingan, Minimum bit error rate multiuser detection techniques for DS-CDMA, Ph.D. Thesis, School of Electronics and Computer Sciences, University of Southampton, UK, December 2003.
- [30] F.-J. Chen, S.-K. Xiong, G. Wei, A new constrained minimum-BER multiuser detection algorithm, in: Proceedings of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication, vol. 3, 1 May–2 June 2004, pp. 365–368.
- [31] S. Chen, L. Hanzo, N.N. Ahmad, Adaptive minimum bit error rate beamforming assisted receiver for wireless communications, in: Proceedings of the ICASSP 2003, vol. IV, Hong Kong, China, April 6–10, 2003, pp. 640–643.
- [32] A. Wolfgang, N.N. Ahmad, S. Chen, L. Hanzo, Genetic algorithm assisted minimum bit error rate beamforming, in: Proceedings of the VTC 2004-Spring, May 17–19, 2004, pp. 142–146.
- [33] I.D.S. Garcia, J.J.S. Marciano, Jr., R.D. Cajote, Normalized adaptive minimum bit-error-rate beamformers, in: Proceedings of the IEEE Region 10 TENCON Conference, vol. 2, November 21–24, 2004, pp. 625–628.

- [34] Y.-H. Liu, Y.-H. Yang, Adaptive minimum bit error rate multitarget array algorithm, in: Proceedings of the IEEE 6th CAS Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication, vol. 2, 31 May–2 June 2004, pp. 745–748.
- [35] N.N. Ahmad, Minimum bit error ratio beamforming, Ph.D. Thesis, School of Electronics and Computer Sciences, University of Southampton, UK, May 2005.
- [36] S. Chen, N.N. Ahmad, L. Hanzo, Adaptive minimum bit error rate beamforming, *IEEE Trans. Wireless Comm.* 4 (2) (2005) 341–348.
- [37] L.-Y. Fan, H.-B. Zhang, H. Chen, Minimum bit error rate beamforming for pre-FFT OFDM adaptive antenna array, in: Proceedings of the VTC 2005-Fall, vol. 6, September 25–28, 2005, pp. 359–363.
- [38] S. Chen, L. Hanzo, N.N. Ahmad, A. Wolfgang, Adaptive minimum bit error rate beamforming assisted QPSK receiver, in: Proceedings of the ICC 2004, vol. 6, June 22–24, 2004, pp. 3389–3393.
- [39] S. Chen, H.-Q. Du, L. Hanzo, Adaptive minimum symbol error rate beamforming assisted receiver for quadrature amplitude modulation systems, in: Proceedings of VTC 2006-Spring, vol. 5, Melbourne, Australia, May 7–10, 2006, pp. 2236–2240.
- [40] D. Gesber, Robust linear MIMO receivers: a minimum error-rate approach, *IEEE Trans. Signal Process.* 51 (11) (2003) 2863–2871.
- [41] A. Livingstone, S. Chen, Adaptive space-time equalisation for multiple-antenna assisted multiple-input multiple-output systems, in: Proceedings of the 2nd IEE/EURASIP DSP-enabled Radio Conference, September 19–20, 2005, p. 9.
- [42] S. Chen, A. Livingstone, L. Hanzo, Adaptive MBER space-time DFE assisted multiuser detection for SDMA systems, in: Proceedings of the VTC 2006-Spring, vol. 1, Melbourne, Australia, May 7–10, 2006, pp. 12–16.
- [43] S. Chen, A. Livingstone, L. Hanzo, Minimum bite-error rate design for space-time equalization-based multiuser detection, *IEEE Trans. Comm.* 54 (5) (2006) 824–832.
- [44] S. Chen, L. Hanzo, A. Livingstone, MBER Space-time decision feedback equalization assisted multiuser detection for multiple antenna aided SDMA systems, *IEEE Trans. Signal Process.* 54 (8) (2006) 3090–3098.
- [45] M.Y. Alias, A.K. Samangan, S. Chen, L. Hanzo, Multiple antenna aided OFDM employing minimum bit error rate multiuser detection, *Electron. Lett.* 39 (24) (2003) 1769–1770.
- [46] S.J. Yi, C.C. Tsimenidis, O.R. Hinton, B.S. Sharif, Adaptive minimum bit error rate multiuser detection for asynchronous MC-CDMA systems frequency selective Rayleigh fading channels, in: Proceedings of the 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, vol. 2, September 7–10, 2003, pp. 1269–1273.
- [47] M.Y. Alias, S. Chen, L. Hanzo, Genetic algorithm assisted minimum bit error rate multiuser detection in multiple antenna aided OFDM, in: Proceedings of the VTC 2004-Fall, vol. 1, 2004, pp. 548–552.
- [48] M.Y. Alias, Minimum bit error rate multiuser detection for multiple antenna aided uplink OFDM, Ph.D. Thesis, School of Electronics and Computer Sciences, University of Southampton, UK, October 2004.
- [49] Z. Li, M. Latva-Aho, Performance evaluation of interleaved MC-CDMA systems with correlated Nakagami-m fading channels, in: Proceedings of the VTC 2005-Spring, vol. 1, 30 May–1 June 2005, pp. 387–391.
- [50] M.Y. Alias, S. Chen, L. Hanzo, Multiple-antenna-aided OFDM employing genetic-algorithm-assisted minimum bit error rate multiuser detection, *IEEE Trans. Vehic. Technol.* 54 (5) (2005) 1713–1721.
- [51] S. Tan, L. Xu, S. Chen, L. Hanzo, Iterative soft interference cancellation aided minimum bit error rate uplink receiver beamforming, in: Proceedings of the VTC 2006-Spring, vol. 1, Melbourne, Australia, May 7–10, 2006, pp. 17–21.
- [52] L. Xu, S. Tan, S. Chen, L. Hanzo, Iterative minimum bit error rate multiuser detection in multiple antenna aided OFDM, in: Proceedings of the 2006 IEEE Wireless Communications and Networking Conference, vol. 3, Las Vegas, USA, April 3–6, 2006, pp. 1603–1607.
- [53] S. Tan, S. Chen, L. Hanzo, On multi-user EXIT chart analysis aided turbo-detected MBER beamformer design, *IEEE Trans. Wireless Comm.* 7 (1) (2008) 314–323.
- [54] R. Irmer, R. Habendorf, W. Rave, G. Fettweis, Overloaded TDD-CDMA cells with multiuser transmission, in: Proceedings of the 2004 ITG Workshop Smart Antennas, 2004, pp. 235–242.
- [55] A. Hjørungnes, P.S.R. Diniz, M.L.R. de Campos, Jointly minimum BER transmitter and receiver FIR MIMO filters for binary signal vectors, *IEEE Trans. Signal Process.* 52 (4) (2004) 1021–1036.
- [56] A. Hjørungnes, P.S.R. Diniz, Minimum BER prefilter transform for communications systems with binary signaling and known FIR MIMO channel, *IEEE Signal Process. Lett.* 12 (3) (2005) 234–237.
- [57] J. Wehinger, V.R. Anreddy, C.F. Mecklenbräuker, S. Paul, C. Antón-Haro, Adaptive minimum bit error rate space-time rake receiver for the uplink of UMTS frequency division duplex mode, in: Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology, 2003, pp. 294–297.
- [58] J.H. Cho, Q. Zhang, Design of RAKE receivers for ultra-wideband binary block-coded PPM in dense multipath channels, in: Proceedings of the MILCOM 2003, vol. 2, October 13–16, 2003, pp. 874–879.
- [59] Y.R. Zhou, L.R. Watkins, Reducing nonlinear waveform distortion in IM/DD systems by optimized receiver filtering, *IEEE Photon. Technol. Lett.* 6 (9) (1994) 1159–1161.
- [60] S. Arnon, S. Rotman, N.S. Kopeika, Optimum transmitter optics aperture for satellite optical communication, *IEEE Trans. Aerospace Electron. Syst.* 34 (2) (1998) 590–596.
- [61] B.J. Koshy, P.M. Shankar, Spread-spectrum techniques for fiber-fed microcellular networks, *IEEE Trans. Vehic. Technol.* 48 (3) (1999) 847–857.
- [62] G. Katz, D. Sadot, Minimum BER criterion for electrical equalizer in optical communication systems, *J. Lightwave Technol.* 24 (7) (2006) 2844–2850.
- [63] H.-R. Zhuang, L. Dai, S.-D. Zhou, Y. Yao, Low complexity per-antenna rate and power control approach for closed-loop V-BLAST, *IEEE Trans. Comm.* 51 (11) (2003) 1783–1787.
- [64] N. Wang, S.D. Blostein, Power loading for CP-OFDM over frequency-selective fading channels, in: Proceedings of

- the GLOBECOM '03, vol. 4, December 1–5, 2003, pp. 2305–2309.
- [65] N. Wang, S.D. Blostein, Comparison of CP-based single carrier and OFDM with power allocation, *IEEE Trans. Comm.* 53 (3) (2005) 391–394.
- [66] F. Rey, M. Lamarca, G. Vazquez, Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI, *IEEE Trans. Signal Process.* 53 (3) (2005) 1070–1085.
- [67] J. Litva, T.K.Y. Lo, *Digital Beamforming in Wireless Communications*, Artech House, London, 1996.
- [68] L.C. Godara, Applications of antenna arrays to mobile communications, Part I: performance improvement, feasibility, and system considerations, *Proc. IEEE* 85 (7) (1997) 1031–1060.
- [69] R. Kohno, Spatial and temporal communication theory using adaptive antenna array, *IEEE Personal Comm.* 5 (1) (1998) 28–35.
- [70] J.H. Winters, Smart antennas for wireless systems, *IEEE Personal Comm.* 5 (1) (1998) 23–27.
- [71] P. Vandenameele, L. van Der Perre, M. Engels, *Space Division Multiple Access for Wireless Local Area Networks*, Kluwer Academic Publishers, Boston, 2001.
- [72] J.S. Blogh, L. Hanzo, *Third Generation Systems and Intelligent Wireless Networking—Smart Antenna and Adaptive Modulation*, Wiley, Chichester, UK, 2002.
- [73] A. Paulraj, R. Nabar, D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, Cambridge, UK, 2003.
- [74] A.J. Paulraj, D.A. Gore, R.U. Nabar, H. Bölcskei, An overview of MIMO communications—A key to gigabit wireless, *Proc. IEEE* 92 (2) (2004) 198–218.
- [75] D. Tse, P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, Cambridge, UK, 2005.
- [76] S. Chen, Adaptive minimum bit-error-rate filtering, *IEE Proc. Vision, Image Signal Process.* 151 (1) (2004) 76–85.
- [77] E. Parzen, On estimation of a probability density function and mode, *Ann. Math. Statist.* 33 (1962) 1066–1076.
- [78] B.W. Silverman, *Density Estimation*, Chapman Hall, London, 1996.
- [79] A.W. Bowman, A. Azzalini, *Applied Smoothing Techniques for Data Analysis*, Oxford University Press, Oxford, UK, 1997.
- [80] W.M. Brown, R.B. Crane, Conjugate linear filtering, *IEEE Trans. Inform. Theory* IT-15 (4) (1969) 462–465.
- [81] B. Picinbono, P. Chevalier, Widely linear estimation with complex data, *IEEE Trans. Signal Process.* 43 (8) (1995) 2030–2033.
- [82] A.M. Tulino, S. Verdú, Asymptotic analysis of improved linear receivers for BPSK-CDMA subject to fading, *IEEE J. Selected Areas Comm.* 19 (8) (2001) 1544–1555.
- [83] P.J. Schreier, L.L. Scharf, Second-order analysis of improved complex random vectors and processes, *IEEE Trans. Signal Process.* 51 (3) (2003) 714–725.
- [84] R. Schober, W.H. Gerstacker, L. Lampe, On suboptimum receivers for DS-CDMA with BPSK modulation, *Signal Process.* 85 (2005) 1149–1163.
- [85] S. Chen, S. Tan, L. Hanzo, Adaptive beamforming for binary phase shift keying communication systems, *Signal Process.* 87 (2007) 68–78.
- [86] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*, Wiley, New York, 1993.
- [87] X.F. Wang, W.-S. Lu, A. Antoniou, Constrained minimum-BER multiuser detection, *IEEE Trans. Signal Process.* 48 (10) (2000) 2903–2909.
- [88] J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, G.D. Forney Jr., MMSE decision-feedback equalizers and coding—Part I: equalization results, *IEEE Trans. Comm.* 43 (10) (1995) 2582–2594.
- [89] S. Chen, G.J. Gibson, C.F.N. Cowan, P.M. Grant, Adaptive equalization of finite non-linear channels using multilayer perceptrons, *Signal Process.* 20 (2) (1990) 107–119.
- [90] G.J. Gibson, S. Siu, S. Chen, C.F.N. Cowan, P.M. Grant, The application of nonlinear architectures to adaptive channel equalization, in: *Proceedings of the ICC'1990*, Atlanta, USA, 1990, pp. 312.8.1–312.8.5.
- [91] S. Chen, G.J. Gibson, C.F.N. Cowan, P.M. Grant, Reconstruction of binary signals using an adaptive radial-basis-function equalizer, *Signal Process.* 22 (1) (1991) 77–93.
- [92] S. Chen, B. Mulgrew, Overcoming co-channel interference using an adaptive radial basis function equaliser, *Signal Process.* 28 (1) (1992) 91–107.
- [93] S. Chen, B. Mulgrew, P.M. Grant, A clustering technique for digital communications channel equalisation using radial basis function networks, *IEEE Trans. Neural Networks* 4 (4) (1993) 570–579.
- [94] S. Chen, S. McLaughlin, B. Mulgrew, P.M. Grant, Adaptive Bayesian decision feedback equaliser for dispersive mobile radio channels, *IEEE Trans. Comm.* 43 (5) (1995) 1937–1946.
- [95] I. Cha, S.A. Kassam, Channel equalization using adaptive complex radial basis function networks, *IEEE J. Selected Areas Comm.* 13 (1) (1995) 122–131.
- [96] F. Albu, D. Martinez, The application of support vector machines with Gaussian kernels for overcoming co-channel interference, in: *Proceedings of the 9th IEEE International Workshop Neural Networks for Signal Processing*, Madison, WI, August 23–25, 1999, pp. 49–57.
- [97] D.J. Sebald, J.A. Bucklew, Support vector machine techniques for nonlinear equalization, *IEEE Trans. Signal Process.* 48 (11) (2000) 3217–3226.
- [98] S. Chen, S.R. Gunn, C.J. Harris, The relevance vector machine technique for channel equalization application, *IEEE Trans. Neural Networks* 12 (6) (2001) 1529–1532.
- [99] S. Chen, B. Mulgrew, L. Hanzo, Adaptive least error rate algorithm for neural network classifier, in: *Proceedings of the IEEE Workshop Neural Networks for Signal Processing*, Falmouth, MA, USA, September 10–12, 2001, pp. 223–232.
- [100] F. Pérez-Cruz, A. Navia-Vázquez, P.L. Alarcón-Diana, A. Artés-Rodríguez, SVC-based equalizer for burst TDMA transmissions, *Signal Process.* 81 (8) (2001) 1571–1787.
- [101] S. Chen, A. Wolfgang, S. Benedetto, P. Dubamet, L. Hanzo, Symmetric radial basis function network equaliser, in: *Proceedings of the NEWCOM-ACoRN Joint Workshop*, Vienna, Austria, September 20–22, 2006, p. 5.
- [102] S. Chen, A.K. Samangan, L. Hanzo, Support vector machine multiuser receiver for DS-CDMA signals in multipath channels, *IEEE Trans. Neural Networks* 12 (3) (2001) 604–611.
- [103] S. Chen, L. Hanzo, A. Wolfgang, Kernel-based nonlinear beamforming construction using orthogonal forward

- selection with Fisher ratio class separability measure, *IEEE Signal Process. Lett.* 11 (5) (2004) 478–481.
- [104] S. Chen, L. Hanzo, A. Wolfgang, Nonlinear multi-antenna detection methods, *EURASIP J. Appl. Signal Process.* 2004 (9) (2004) 1225–1237.
- [105] S. Chen, K. Labib, R. Kang, L. Hanzo, Adaptive radial basis function detector for beamforming, in: *Proceedings of the ICC'2007, Glasgow, Scotland, June 24–28, 2007*, pp. 2967–2972.
- [106] S. Chen, A. Wolfgang, C.J. Harris, L. Hanzo, Symmetric kernel detector for multiple-antenna aided beamforming systems, in: *Proceedings of the IJCNN, Orlando, Florida, August 12–17, 2007*, pp. 2486–2491.
- [107] A. Wolfgang, S. Chen, L. Hanzo, Radial basis function network assisted space-time equalisation for dispersive fading environments, *Electron. Lett.* 40 (16) (2004) 1006–1007.