Adaptive Nonlinear Equalizer Using a Mixture of Gaussian-Based Online Density Estimator

Hao Chen, *Student Member, IEEE*, Yu Gong, *Member, IEEE*, Xia Hong, *Senior Member, IEEE*, and Sheng Chen, *Fellow, IEEE*

***Abstract*—This paper introduces a new adaptive nonlinear equalizer relying on a radial basis function (RBF) model, which is designed based on the minimum bit error rate (MBER) criterion, in the system setting of the intersymbol interference channel plus cochannel interference (CCI). Our proposed algorithm is referred to as the online mixture of Gaussian-estimator-aided MBER (OMG-MBER) equalizer. Specifically, a mixture of Gaussian-based probability density function (pdf) estimator is used to model the pdf of the decision variable, for which a novel online pdf update algorithm is derived to track the incoming data. With the aid of this novel online mixture of Gaussian-based sample-by-sample updated pdf estimator, our adaptive nonlinear equalizer is capable of updating its equalizer's parameters sample by sample to aim directly at minimizing the RBF nonlinear equalizer's achievable bit error rate (BER). The proposed OMG-MBER equalizer significantly outperforms the existing online nonlinear MBER equalizer, known as the least bit error rate equalizer, in terms of both the convergence speed and the achievable BER, as is confirmed in our simulation study.**

***Index Terms*—Adaptive nonlinear equalizer, minimum bit error rate (MBER), mixture of Gaussians, probability density function (pdf), radial basis function (RBF).**

I. Introduction

CHANNEL equalization as a standard approach in communications to combat the dispersive effects of a channel has been very well studied [1]–[4]. Both linear and nonlinear equalizers have been proposed [5]–[9] and the majority of them are based on the MMSE criterion. The MMSE equalizer has the advantage of easy implementation with good performance. The MMSE linear equalizer also has a natural link to adaptive filters [10] and admits a very simple yet powerful online adaptation by the least mean square (LMS) algorithm [11]. However, the MMSE criterion is not equivalent to the minimum bit error rate (MBER) criterion, and the latter is the ultimate performance criterion for communication channel equalization [12]–[32]. Equalizers based on the MBER criterion are thus of great interest and have attracted considerable attention [14], [20]–[23], [33]–[35]. The least bit error rate (LBER) nonlinear equalizer for binary transmission was proposed in [33], which was shown to achieve significant performance gain over both the adaptive linear and nonlinear MMSE equalizers.

The equalization problem can alternatively be regarded as a classification problem, where the task of the equalizer is to determine the decision boundaries for separating the different classes of data. It is shown in [33] that the adaptive MBER nonlinear equalizer achieves the decision boundary that is close to the optimum Bayesian equalizer and so is its equalization performance. While both the batch and sample-by-sample adaptive nonlinear MBER equalizers were proposed in [33], the latter, referred to as the LBER, is of particular interest since it is an online adaptive algorithm with very low computational complexity and is capable of tracking the channel variation well. To update the gradient of the BER with the new coming data, which is required for adapting the MBER nonlinear equalizer's parameters, it is necessary to adaptively estimate the probability density function (pdf) of the signed decision variable constructed based on the desired signal and the equalizer output. The LBER nonlinear equalizer [33] adopts a one-sample or single Gaussian kernel estimate for the pdf of the signed decision variable to realize sample-by-sample adaptation of the nonlinear equalizer's parameters, in a manner similar to the LMS algorithm that updates the linear equalizer's parameters using a one-sample estimate for the MSE [10].

The performance of an adaptive MBER nonlinear equalizer is to a large extent determined by the online pdf estimator employed. Despite of its computational simplicity and its superior performance over the adaptive MMSE equalizer, the LBER nonlinear equalizer in [33] has a drawback, owing to the fact that it adopts a one-sample pdf estimate, which is stochastic by nature and is very sensitive to the noise in the received signal sample. In statistical data learning, there exist a large number of works [36]–[41] using the Gaussian mixture model to estimate pdf. These kernel density estimators based on a mixture of Gaussians however are batch learning algorithms by nature and are unsuitable for online applications. In this paper, we also use a kernel density estimator, consisting of a small number of Gaussian kernels to estimate the pdf of the signed decision variable. Our novel contribution is to propose a new online mixture of Gaussian estimator (OMG) to update the pdf estimate sample by sample. To be specific, a new Gaussian kernel is formed for each new data, and it is then merged with

Manuscript received August 28, 2013; revised January 16, 2014; accepted March 12, 2014. Date of publication March 24, 2014; date of current version November 6, 2014. This work was supported in part by the U.K. Engineering and Physical Sciences Research Council and in part by the Defence Science and Technology Laboratory under Grant EP/H012516/1. The review of this paper was coordinated by Dr. C. Cozzo.

H. Chen and X. Hong are with the School of Systems Engineering, University of Reading, Reading RG6 6AY, U.K. (e-mail: hao.chen@pgr.reading.ac.uk; x.hong@reading.ac.uk).

Y. Gong is with the School of Electronic, Electrical and Systems Engineering, Loughborough University, Loughborough LE11 3TU, U.K. (e-mail: alex.yugong@gmail.com).

S. Chen is with the Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K., and also with King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: sqc@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2014.2313458

0018-9545 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

the "nearest" existing Gaussian kernel in the kernel density estimator. With the aid of this OMG for online estimation of the signed decision variable's pdf, the nonlinear equalizer's parameters can be adapted sample by sample in the way similar to the LBER nonlinear equalizer of [33]. We refer to this new adaptive MBER nonlinear equalizer as the OMG-aided MBER (OMG-MBER) algorithm. Because the proposed OMG-MBER equalizer relies on a much more accurate online pdf estimate, unlike the one-sample pdf estimator of the LBER algorithm, it outperforms the LBER nonlinear equalizer in terms of both the convergence speed and the equalizer's achievable BER. A simulation study is carried out, and the results obtained confirm that the OMG-MBER equalizer significantly improve the equalization performance, compared with the existing LBER nonlinear equalizer.

This paper is organized as follows. Section II introduces the equalization problem in the setting of the intersymbol interference channel plus cochannel interference (CCI) and reviews the existing adaptive MMSE and MBER equalizers. Section III details our proposed novel OMG-MBER nonlinear equalizer. The simulation results are presented in Section IV and our conclusions are given in Section V.

II. Preliminaries

We first introduce the channel model and then review the adaptive equalizers based on the MMSE and MBER criteria.

A. System Model

We assume, without loss of generality, that the system suffers from one CCI. The received signal sample at symbol index k is given by [42]

$$\begin{aligned} r(k) &= \bar{r}(k) + n(k) = \bar{r}_0(k) + \bar{r}_1(k) + n(k) \\ &= \sum_{i=0}^{n_0-1} a_{0,i} b_0(k-i) + \sum_{i=0}^{n_1-1} a_{1,i} b_1(k-i) + n(k) \end{aligned} \tag{1}$$

where $n(k)$ is the additive white Gaussian noise (AWGN) with zero mean and variance $E[n^2(k)] = \sigma_n^2$; $\bar{r}(k)$, $\bar{r}_0(k)$, and $\bar{r}_1(k)$ are the noise-free received signal, desired signal, and interfering signal, respectively; $a_{0,i}$, $0 \le i \le n_0 - 1$ and $a_{1,i}$, $0 \le i \le n_1 - 1$, are the coefficients for the channel and the cochannel, respectively; and $b_0(k)$ and $b_1(k)$ are the uncorrelated binary desired and interfering data, respectively, taking value from the set $\{\pm 1\}$.

We assume that the equalizer length is m, namely, the equalization is based on the received vector of the m most recently received signal samples given by

$$\mathbf{r}(k) = [r(k)\, r(k-1) \cdots r(k-m+1)]^T. \tag{2}$$

Then, (1) can be expressed in the matrix form as

$$\begin{aligned} \mathbf{r}(k) &= \bar{\mathbf{r}}(k) + \mathbf{n}(k) = \bar{\mathbf{r}}_0(k) + \bar{\mathbf{r}}_1(k) + \mathbf{n}(k) \\ &= \mathbf{H}_0 \mathbf{b}_0(k) + \mathbf{H}_1 \mathbf{b}_1(k) + \mathbf{n}(k) \end{aligned} \tag{3}$$

where $\mathbf{n}(k) = [n(k)n(k-1) \cdots n(k-m+1)]^T$ is the AWGN vector; $\mathbf{b}_0(k) = [b_0(k)b_0(k-1) \cdots b_0(k-n_0-m+2)]^T$ and $\mathbf{b}_1(k) = [b_1(k)b_1(k-1) \cdots b_1(k-n_1-m+2)]^T$ are the desired and interfering signal vectors, respectively; and $\mathbf{H}_0 \in \mathbb{R}^{m \times (m+n_0-1)}$ and $\mathbf{H}_1 \in \mathbb{R}^{m \times (m+n_1-1)}$ are the desired and interfering channel matrices, respectively, given by

$$\mathbf{H}_i = \begin{bmatrix} a_{i,0} & a_{i,1} & \cdots & a_{i,n_i-1} & 0 & \cdots & 0 \\ 0 & a_{i,0} & a_{i,1} & \cdots & a_{i,n_i-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & a_{i,0} & a_{i,1} & \cdots & a_{i,n_i-1} \end{bmatrix} \tag{4}$$

for $i = 0$ and 1. The equalization output can be expressed as

$$\widehat{b}_0(k-d) = \operatorname{sgn}(y(k)) \quad \text{with} \quad y(k) = f(\mathbf{r}(k); \mathbf{w}) \tag{5}$$

where $f(\cdot;\cdot)$ denotes the equalizer function, $\mathbf{w}$ is the equalizer parameter vector of an appropriate dimension containing all the equalizer's adjustable parameters, d is the decision delay, and $\operatorname{sgn}(y) = 1$ if $y \ge 0$ and $\operatorname{sgn}(y) = -1$ if $y < 0$.

As $b_0(k)$ and $b_1(k)$ are binary signals, there are $N_0 = 2^{m+n_0-1}$ and $N_1 = 2^{m+n_1-1}$ possible combinations or "states" for $\mathbf{b}_0(k)$ and $\mathbf{b}_1(k)$, respectively. The jth state of $\mathbf{b}_0(k)$ is denoted $\mathbf{b}_{0,j}$, where $1 \le j \le N_0$, whereas the lth state of $\mathbf{b}_1(k)$ is denoted $\mathbf{b}_{1,l}$, where $1 \le l \le N_1$. Accordingly, there are N_0 and N_1 states for $\bar{\mathbf{r}}_0(k)$ and $\bar{\mathbf{r}}_1(k)$, which are denoted $\{\bar{\mathbf{r}}_{0,j} = \mathbf{H}_0 \mathbf{b}_{0,j}\}_{j=1}^{N_0}$ and $\{\bar{\mathbf{r}}_{1,l} = \mathbf{H}_1 \mathbf{b}_{1,l}\}_{l=1}^{N_1}$, respectively. Consequently, there are $N_r = N_0 \times N_1$ states for $\bar{\mathbf{r}}(k)$, which we denote $\mathbf{R} = \{\bar{\mathbf{r}}_i\}_{i=1}^{N_r}$. $\mathbf{R}$ can be divided into two subsets, $\mathbf{R}_+$ and $\mathbf{R}_-$, corresponding to the dth element of $\mathbf{b}_{0,j}$, which is denoted $b_{0,j}^{(d)}$, being $+1$ and -1, respectively. Thus, $b_{0,j}^{(d)}$ serves as the "class" label for $\mathbf{r}(k) \in \mathbb{R}^m$, and the equalization (5) is equivalent to a classification process, which partitions the m-dimensional observation space by the decision boundary defined as

$$f(\mathbf{r}; \mathbf{w}) = 0. \tag{6}$$

The optimal Bayesian equalizer $y(k) = f_B(\mathbf{r}(k); \mathbf{w})$, given the channel and cochannel, is expressed by [42], [43]

$$\begin{aligned} f_B(\mathbf{r}(k); \mathbf{w}) &= \sum_{j=1}^{N_0} \sum_{l=1}^{N_1} \frac{b_{0,j}^{(d)}}{(2\pi\sigma_n^2)^{m/2} N_0 N_1} e^{-\frac{\left\|\mathbf{r}(k) - \bar{\mathbf{r}}_{0,j} - \bar{\mathbf{r}}_{1,l}\right\|^2}{2\sigma_n^2}} \\ &= \sum_{i=1}^{N_r} \frac{b_i^{(d)}}{(2\pi\sigma_n^2)^{m/2} N_r} e^{-\frac{\|\mathbf{r}(k) - \bar{\mathbf{r}}_i\|^2}{2\sigma_n^2}} \end{aligned} \tag{7}$$

where $b_i^{(d)}$ is the class label for $\bar{\mathbf{r}}_i$. It can be seen that the dimension of the optimal Bayesian equalizer's parameter vector $\mathbf{w}$ is very large as $\mathbf{w}$ consists of all the $\bar{\mathbf{r}}_i$. Thus, the computational complexity of this Bayesian solution is often too high because the size of $\mathbf{R}$ is typically huge.

B. Equalizer Based on the MMSE Criterion

Due to its mathematical tractability, the most widely adopted criterion for training an equalizer is the MMSE criterion, and the well-known LMS algorithm [11] offers an adaptive algorithm for training a linear equalizer based on the MMSE criterion. The LMS algorithm can be also extended to train a nonlinear equalizer according to

$$\mathbf{w}(k)=\mathbf{w}(k-1)+\eta\left(b_0(k-d)-y(k)\right)\frac{\partial f\left(\mathbf{r}(k);\mathbf{w}(k-1)\right)}{\partial \mathbf{w}} \tag{8}$$

where η is a small positive step size.

The adaptive nonlinear equalizer design based on the radial basis function (RBF) network was well studied [43], and the RBF equalizer output is given by

$$y(k)=f\left(\mathbf{r}(k);\,\mathbf{w}\right)=\sum_{j=1}^{n_c}\alpha_j e^{-\frac{\|\mathbf{r}(k)-\mathbf{c}_j\|^2}{\sigma_j^2}} \tag{9}$$

where n_c is the number of RBF nodes employed, and the RBF equalizer parameter vector $\mathbf{w}$ consists of all the RBF weights α_j, variances σ_j^2, and centers $\mathbf{c}_j$. The dimension of $\mathbf{w}$ is therefore $n_c(m+2)$ or $\mathbf{w}\in\mathbb{R}^{n_c(m+2)}$. The partial derivatives of the equalizer output with respect to the equalizer's parameters $\mathbf{w}$ in (8) are given for $1\le j\le n_c$ by

$$\begin{cases}\frac{\partial f(\mathbf{r}(k);\,\mathbf{w})}{\partial \alpha_j}=e^{-\frac{\|\mathbf{r}(k)-\mathbf{c}_j\|^2}{\sigma_j^2}}\\ \frac{\partial f(\mathbf{r}(k);\mathbf{w})}{\partial \sigma_j^2}=\alpha_j e^{-\frac{\|\mathbf{r}(k)-\mathbf{c}_j\|^2}{\sigma_j^2}}\frac{\|\mathbf{r}(k)-\mathbf{c}_j\|^2}{\sigma_j^4}\\ \frac{\partial f(\mathbf{r}(k);\,\mathbf{w})}{\partial \mathbf{c}_j}=2\alpha_j e^{-\frac{\|\mathbf{r}(k)-\mathbf{c}_j\|^2}{\sigma_j^2}}\frac{\mathbf{r}(k)-\mathbf{c}_j}{\sigma_j^2}.\end{cases} \tag{10}$$

For the linear equalizer $y(k)=f_L(\mathbf{r}(k);\,\mathbf{w})$, the equalizer parameter vector $\mathbf{w}$ has a dimension of m or $\mathbf{w}\in\mathbb{R}^m$. Thus, for the linear equalizer

$$y(k)=\mathbf{w}^T\mathbf{r}(k) \tag{11}$$

equation (8) returns to its original LMS form, i.e.,

$$\mathbf{w}(k)=\mathbf{w}(k-1)+\eta\left(b_0(k-d)-y(k)\right)\mathbf{r}(k). \tag{12}$$

C. *Equalizer Based on MBER Criterion*

However, the ultimate goal of equalization is to minimize the BER [12]–[16]. Many studies [20]–[23], [33]–[35] have demonstrated that the MBER equalizer can significantly improve the performance compared with the traditional equalizer design based on the MMSE criterion. Specifically, for the linear equalizer (11), the MBER design has been extensively studied [13]–[16], [18], [20], and adaptive MBER linear equalizer has been proposed based on the LBER algorithm [21], [23]. Moreover, the MBER design and its adaptive LBER algorithm have been extended to the generic nonlinear equalizer of (5) [33], [34].

The error probability of the equalizer (5) can be obtained as [33]

$$P_E(\mathbf{w})=\mathrm{Prob}\left\{\mathrm{sgn}\left(b_0(k-d)\right)y(k)<0\right\}. \tag{13}$$

By defining the signed decision variable as $y_s(k)=\mathrm{sgn}(b_0(k-d))y(k)$, we have

$$P_E(\mathbf{w})=\int_{-\infty}^{0}p_y(y_s)dy_s \tag{14}$$

where $p_y(y_s)$ is the pdf of $y_s(k)$. Linearizing the equalizer around $\bar{\mathbf{r}}(k)$ yields [33]

$$\begin{aligned}y(k)&=f\left(\bar{\mathbf{r}}(k)+\mathbf{n}(k);\,\mathbf{w}\right)\\ &\approx f\left(\bar{\mathbf{r}}(k);\,\mathbf{w}\right)+e(k)=\bar{y}(k)+e(k)\end{aligned} \tag{15}$$

where $e(k)$ is an equivalent zero-mean Gaussian noise with variance ρ^2. Noting that the noise-free received signal $\bar{y}(k)$ takes value from the finite set, namely, $\bar{y}(k)\in Y_f=\{\bar{y}_i=f(\bar{\mathbf{r}}_i;\,\mathbf{w})\}_{i=1}^{N_r}$, the pdf of $y_s(k)$ can be approximated as [33]

$$p_y(y_s)\approx\frac{1}{N_r\sqrt{2\pi}\rho}\sum_{i=1}^{N_r}e^{-\frac{\left(y_s-\mathrm{sgn}\left(b_i^{(d)}\right)\bar{y}_i\right)^2}{2\rho^2}}. \tag{16}$$

Substituting (16) into (14) then yields the following error probability of the equalizer:

$$P_E(\mathbf{w})\approx\frac{1}{N_r\sqrt{2\pi}}\sum_{i=1}^{N_r}\int_{\bar{g}_i(\mathbf{w})}^{\infty}e^{-\frac{x_i^2}{2}}dx_i=\frac{1}{N_r}\sum_{i=1}^{N_r}Q(\bar{g}_i(\mathbf{w})) \tag{17}$$

where

$$Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{y^2}{2}}dy \tag{18}$$

$$\bar{g}_i(\mathbf{w})=\frac{\mathrm{sgn}\left(b_i^{(d)}\right)f(\bar{\mathbf{r}}_i;\,\mathbf{w})}{\rho}. \tag{19}$$

Then, the gradient of $P_E(\mathbf{w})$ can be approximated as

$$\begin{aligned}\nabla P_E(\mathbf{w})&\approx-\frac{1}{N_r\sqrt{2\pi}}\sum_{i=1}^{N_r}e^{-\frac{\bar{g}_i^2}{2\rho^2}}\frac{\partial\bar{g}_i(\mathbf{w})}{\partial\mathbf{w}}\\ &=-\frac{1}{N_r\sqrt{2\pi}}\sum_{i=1}^{N_r}e^{-\frac{\bar{y}_i^2}{2\rho^2}}\mathrm{sgn}\left(b_i^{(d)}\right)\frac{\partial f(\bar{\mathbf{r}}_i;\mathbf{w})}{\partial\mathbf{w}}.\end{aligned} \tag{20}$$

Thus, if the channel and cochannel are known, the equalizer parameters can be updated by minimizing the approximate BER in (17), leading to the MBER updating rule, i.e.,

$$\mathbf{w}(l)=\mathbf{w}(l-1)-\eta\nabla P_E\left(\mathbf{w}(l-1)\right). \tag{21}$$

In practice, the channel and cochannel are unknown. Thus, set $\mathbf{R}$ and, hence, set Y_f are unknown. The key to developing an effective adaptive algorithm for implementing the MBER design is to estimate the pdf $p_y(y_s)$ of the signed decision variable $y_s(k)$ based on the training data $D_K=\{\mathbf{r}(k),\,b_0(k-d)\}_{k=1}^K$ [33]. Specifically, given D_K, the Parzen window or the kernel density estimate of the pdf $p_y(y_s)$ is given as [33]

$$\widehat{p}_y(y_s)=\frac{1}{K\sqrt{2\pi}\rho}\sum_{k=1}^{K}e^{-\frac{(y_s-\mathrm{sgn}(b_0(k-d))y(k))^2}{2\rho^2}}. \tag{22}$$

Substituting (22) into (14) yields the estimated error probability $\widehat{P}_E(\mathbf{w})$, the gradient of which is given by

$$\nabla \widehat{P}_E(\mathbf{w}) = -\frac{1}{K\sqrt{2\pi}\rho} \sum_{k=1}^{K} e^{-\frac{y^2(k)}{2\rho^2}} \operatorname{sgn}\left(b_0(k-d)\right) \times \frac{\partial f\left(\mathbf{r}(k); \mathbf{w}\right)}{\partial \mathbf{w}}. \quad (23)$$

This leads to the block adaptive MBER algorithm [33], i.e.,

$$\begin{aligned}\mathbf{w}(l) &= \mathbf{w}(l-1) - \eta \nabla \widehat{P}_E(\mathbf{w}(l-1)) \\ &= \mathbf{w}(l-1) + \frac{\eta}{K\sqrt{2\pi}\rho} \sum_{k=1}^{K} e^{-\frac{y^2(k)}{2\rho^2}} \operatorname{sgn}\left(b_0(k-d)\right) \\ &\quad \times \frac{\partial f\left(\mathbf{r}(k); \mathbf{w}(l-1)\right)}{\partial \mathbf{w}}. \end{aligned} \quad (24)$$

To realize a sample-by-sample adaptation, the nonlinear LBER algorithm [33] only uses a single data point to estimate the pdf, i.e.,

$$\widetilde{p}_y(y_s, k) = \frac{1}{\sqrt{2\pi}\rho} e^{-\frac{(y_s - \operatorname{sgn}(b_0(k-d))y(k))^2}{2\rho^2}}. \quad (25)$$

With this "instantaneous" pdf estimate, we have a "one-sample" error probability estimate $\widetilde{P}_E(\mathbf{w}, k)$ with the stochastic gradient given by

$$\nabla \widetilde{P}_E(\mathbf{w}, k) = -\frac{1}{\sqrt{2\pi}\rho} e^{-\frac{y^2(k)}{2\rho^2}} \operatorname{sgn}(b_0(k-d)) \frac{\partial f(\mathbf{r}(k); \mathbf{w})}{\partial \mathbf{w}}. \quad (26)$$

This leads to the stochastic gradient adaptive algorithm known as the LBER algorithm [33], i.e.,

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \frac{\eta}{\sqrt{2\pi}\rho} e^{-\frac{y^2(k)}{2\rho^2}} \operatorname{sgn}\left(b_0(k-d)\right) \times \frac{\partial f\left(\mathbf{r}(k); \mathbf{w}(k-1)\right)}{\partial \mathbf{w}} \quad (27)$$

in which the equalizer parameter vector is adjusted sample by sample, in a similar manner to the LMS algorithm for the nonlinear equalizer given in (8).

III. Online Mixture of Gaussian-Estimator-Aided MBER Equalizer

To realize a sample-by-sample adaptive MBER nonlinear equalizer, the LBER algorithm [33] uses a single sample to estimate the pdf online. Our OMG-MBER adopts a novel OMG pdf estimator. Specifically, we derive the kernel pdf estimator based on a mixture of Gaussians with a small number of mixtures, which is capable of adapting the pdf estimate sample by sample. With the aid of this online kernel density estimation algorithm, our adaptive MBER equalizer is capable of updating its equalizer's parameters sample by sample, similar to the LBER algorithm. Since our OMG-MBER algorithm relies on a much more accurate one-line pdf estimate, unlike the one-sample pdf of the LBER algorithm, we expect our OMG-MBER will outperform the LBER, at the expense of a negligibly small increase in complexity.

A. *Online Kernel-Based PDF Estimator*

Our goal is to derive an online pdf estimator that updates the estimate of $p_y(y_s)$ as each new data $y_s(k)$ is received. We consider the pdf estimator based on the mixture of M Gaussians [37] given by

$$\begin{aligned}\widehat{p}_y(y_s) &= \widehat{p}^{(M)}\left(y_s; \boldsymbol{\lambda}_M, \boldsymbol{\mu}_M, \boldsymbol{\rho}_M\right) \\ &= \sum_{j=1}^{M} \lambda_j G(y_s; \mu_j, \rho_j) \end{aligned} \quad (28)$$

$$\text{s.t.} \quad \lambda_j > 0, \ 1 \le j \le M, \ \text{and } \boldsymbol{\lambda}_M^T \mathbf{1}_M = 1 \quad (29)$$

where λ_j, μ_j, and ρ_j are the weight, mean, and standard deviation of the jth Gaussian kernel, respectively; $\boldsymbol{\lambda}_M = [\lambda_1\lambda_2 \cdots \lambda_M]^T$; $\boldsymbol{\mu}_M = [\mu_1\mu_2 \cdots \mu_M]^T$; $\boldsymbol{\rho}_M = [\rho_1\rho_2 \cdots \rho_M]^T$; $\mathbf{1}_M$ is the M-dimensional vector whose elements are all equal to one; and

$$G(y_s; \mu_j, \rho_j) = \frac{1}{\sqrt{2\pi}\rho_j} e^{-\frac{(y_s-\mu_j)^2}{2\rho_j^2}}. \quad (30)$$

At time $k=0$, we have the initial estimate for $p_y(y_s)$ given as

$$\widehat{p}^{(M)}\left(y_s; \boldsymbol{\lambda}_M(0), \boldsymbol{\mu}_M(0), \boldsymbol{\rho}_M(0)\right) = \sum_{j=1}^{M} \frac{1}{M} G\left(y_s; \mu_j(0), \rho_0\right) \quad (31)$$

where $\lambda_j(0) = 1/M$ and $\rho_j(0) = \rho_0$ for $1 \le j \le M$, with ρ_0 being a predetermined kernel width, whereas $\mu_j(0)$ for $1 \le j \le M$ are some randomly drawn positive values. The initialization of the OMG-MBER-based nonlinear equalizer will be discussed in more detail in Section III-C.

At time step k, the new data point $y_s(k)$ is received, and the triplets $\{\boldsymbol{\lambda}_M, \boldsymbol{\mu}_M, \boldsymbol{\rho}_M\}$ in the pdf estimate, i.e.,

$$\begin{aligned}&\widehat{p}^{(M)}\left(y_s; \boldsymbol{\lambda}_M(k-1), \boldsymbol{\mu}_M(k-1), \boldsymbol{\rho}_M(k-1)\right) \\ &\quad = \sum_{j=1}^{M} \lambda_j(k-1) G\left(y_s; \mu_j(k-1), \rho_j(k-1)\right) \end{aligned} \quad (32)$$

need to be updated accordingly, while keeping the same number of mixtures M. A natural way is to place a Gaussian kernel on $y_s(k)$ and to merge this new kernel with its nearest existing mixture component $G(y_s; \mu_{j^*}(k-1), \rho_{j^*}(k-1))$, where

$$j^* = \arg \min_{1 \le j \le M} |y_s(k) - \mu_j(k-1)|. \quad (33)$$

This can be realized in the following two steps.

1) Create a temporary $(M+1)$-order model by adding the newly created $(M+1)$th Gaussian kernel based on $y_s(k)$ to the estimate (32) according to

$$\begin{aligned}&\widehat{p}^{(M+1)}\left(y_s; \boldsymbol{\lambda}_{M+1}(k), \boldsymbol{\mu}_{M+1}(k), \boldsymbol{\rho}_{M+1}(k)\right) \\ &\quad = \frac{M}{M+1} \sum_{j=1}^{M} \lambda_j(k-1) G\left(y_s; \mu_j(k-1), \rho_j(k-1)\right) \\ &\quad + \frac{1}{M+1} G\left(y_s; y_s(k), \rho_0\right). \end{aligned} \quad (34)$$

Clearly, we set $\lambda_{M+1}(k)=1/(M+1)$, and for $1\leq j\leq M$

$$\lambda_j(k) = \frac{M\lambda_j(k-1)}{M+1} \tag{35}$$

to satisfy the constraints $\lambda_j(k) > 0$ for $1 \leq j \leq M+1$ and $\boldsymbol{\lambda}_{M+1}^T(k)\mathbf{1}_{M+1} = 1$, while using $\mu_{M+1}(k) = y_s(k)$ and $\rho_{M+1}(k) = \rho_0$, and for $1 \leq j \leq M$

$$\mu_j(k) = \mu_j(k-1) \tag{36}$$

$$\rho_j(k) = \rho_j(k-1). \tag{37}$$

2) Merge the j^*th, where j^* is determined in (33), and $(M+1)$th mixtures in the temporary estimate (34) into the single new j^*th mixture, so that

$$\begin{aligned}\lambda_{j^*}(k)G\left(y_s;\, \mu_{j^*}(k),\, \rho_{j^*}(k)\right) &\approx \frac{M\lambda_{j^*}(k-1)}{M+1}G\left(y_s;\, \mu_{j^*}(k-1),\, \rho_{j^*}(k-1)\right)\\ &\quad+\frac{1}{M+1}G\left(y_s;\, y_s(k),\, \rho_0\right).\end{aligned} \tag{38}$$

Thus, the new j^*th weight $\lambda_{j^*}(k)$ is given by

$$\lambda_{j^*}(k) = \frac{M\lambda_{j^*}(k-1)}{M+1} + \frac{1}{M+1} \tag{39}$$

whereas the new j^*th mean and standard deviation $\mu_{j^*}(k)$ and $\rho_{j^*}(k)$, respectively, are updated by matching the mean and standard deviation of the two mixtures with the new single Gaussian. As shown in Appendix A, we have the following adaptation rules:

$$\mu_{j^*}(k) = \frac{M\lambda_{j^*}(k-1)}{M\lambda_{j^*}(k-1)+1}\mu_{j^*}(k-1) + \frac{1}{M\lambda_{j^*}(k-1)+1}y_s(k) \tag{40}$$

$$\rho_{j^*}^2(k) = \frac{M\lambda_{j^*}(k-1)\left(\rho_{j^*}^2(k-1)+\mu_{j^*}^2(k-1)\right)}{M\lambda_{j^*}(k-1)+1} + \frac{\rho_0^2+y_s^2(k)}{M\lambda_{j^*}(k-1)+1} - \mu_{j^*}^2(k). \tag{41}$$

The pdf of the signed decision variable at sample time k can thus be approximated by

$$\begin{aligned}\widehat{p}_y(y_s,\, k) &= \widehat{p}^{(M)}\left(y_s;\, \boldsymbol{\lambda}_M(k),\, \boldsymbol{\mu}_M(k),\, \boldsymbol{\rho}_M(k)\right)\\ &= \sum_{j=1}^{M}\lambda_j(k)G\left(y_s;\, \mu_j(k),\, \rho_j(k)\right)\end{aligned} \tag{42}$$

in which the j^*th weight, mean, and standard deviation are given in (39)–(41), respectively, whereas the jth weight, mean, and standard deviation, where $1 \leq j \leq M$ and $j \neq j^*$, are given in (35)–(37), respectively. Note that only $\mu_{j^*}(k)$ and $\rho_{j^*}^2(k)$ contain the new information provided by $y_s(k)$. This online estimator is a much more accurate estimate of the true pdf $p_y(y_s)$ than the one-sample estimate of (25), at the cost of a modest increase in computational requirements, as the number of mixtures M is a very small number.

B. Adaptive MBER Equalizer

Based on the estimated pdf of the signed decision variable at sample k, as given in (42), the error probability of the equalizer can be expressed as

$$\begin{aligned}\widehat{P}_E(\mathbf{w},\, k) &= \int_{-\infty}^{0}\widehat{p}_y(y_s,\, k)\, dy_s\\ &= \int_{-\infty}^{0}\left(\sum_{j=1,\, j\neq j^*}^{M}\lambda_j(k)G\left(y_s;\, \mu_j(k), \rho_j(k)\right) + \lambda_{j^*}(k)G\left(y_s;\, \mu_{j^*}(k), \rho_{j^*}(k)\right)\right)dy_s\\ &= \widehat{R}_E(\mathbf{w},\, k) + \sum_{j=1,\, j\neq j^*}^{M}\lambda_j(k) \times \int_{-\infty}^{0} G\left(y_s;\, \mu_j(k), \rho_j(k)\right)\, dy_s\end{aligned} \tag{43}$$

where

$$\begin{aligned}\widehat{R}_E(\mathbf{w},\, k) &= \frac{\lambda_{j^*}(k)}{\sqrt{2\pi}\rho_{j^*}(k)}\int_{-\infty}^{0} e^{-\frac{\left(y_s-\mu_{j^*}(k)\right)^2}{2\rho_{j^*}^2(k)}}\, dy_s\\ &= \frac{\lambda_{j^*}(k)}{\sqrt{2\pi}}\int_{g_{j^*}(\mathbf{w},\, k)}^{\infty} e^{-\frac{y^2}{2}}\, dy\end{aligned} \tag{44}$$

with

$$g_{j^*}(\mathbf{w},\, k) = \frac{\mu_{j^*}(k)}{\rho_{j^*}(k)}. \tag{45}$$

In (43), only $\mu_{j^*}(k)$ and $\rho_{j^*}(k)$ depend on the current equalizer's parameter vector $\mathbf{w}$ since only $\mu_{j^*}(k)$ and $\rho_{j^*}(k)$ depend on $y_s(k) = b_0(k-d)y(k) = b_0(k-d)f(\mathbf{r}(k);\, \mathbf{w})$. Therefore, we have the "instantaneous" gradient of $\widehat{P}_E(\mathbf{w},\, k)$ given by

$$\begin{aligned}\triangledown\widehat{P}_E(\mathbf{w},\, k) &= \triangledown\, \widehat{R}_E(\mathbf{w},\, k)\\ &= -\,\frac{\lambda_{j^*}(k)}{\sqrt{2\pi}}e^{-\frac{\mu_{j^*}^2(k)}{2\rho_{j^*}^2(k)}}\frac{\partial g_{j^*}(\mathbf{w},\, k)}{\partial\mathbf{w}}\end{aligned} \tag{46}$$

where

$$\frac{\partial g_{j^*}(\mathbf{w},\, k)}{\partial\mathbf{w}} = \frac{\left(\rho_{j^*}^2(k)+\mu_{j^*}^2(k)\right)\mathrm{sgn}\left(b_0(k-d)\right)-\mu_{j^*}(k)y(k)}{\rho_{j^*}^3(k)\left(M\lambda_{j^*}(k-1)+1\right)} \times\frac{\partial f(\mathbf{r}(k);\mathbf{w})}{\partial\mathbf{w}} \tag{47}$$

the derivation of which is given in Appendix B.

At sample k, given the equalizer's signed decision variable $y_s(k) = b_0(k-d)f(\mathbf{r}(k); \mathbf{w}(k-1))$, the online parameter updating for the adaptive MBER equalizer based on "stochastic" gradient descent is therefore expressed by

$$\mathbf{w}(k) = \mathbf{w}(k-1) - \eta \bigtriangledown \widehat{P}_E(\mathbf{w}(k-1), k) = \mathbf{w}(k-1) + \frac{\eta\lambda_{j^*}(k)}{\sqrt{2\pi}} e^{-\frac{\mu_{j^*}^2(k)}{2\rho_{j^*}^2(k)}} \frac{\partial g_{j^*}(\mathbf{w}(k-1), k)}{\partial \mathbf{w}} \quad (48)$$

where $\eta > 0$ is a small step size.

C. OMG-MBER Algorithm Summary

It is well known that the mixture of Gaussians with a small M is capable of accurately estimating an arbitrary pdf [36], [37], [44], [45]. Therefore, the number of mixtures M in the online kernel density estimator (28) can be chosen as a very small number, and typically $M = 4$–6 is sufficient for most applications. Too large an M may cause slow convergence of the adaptive process, whereas a too small M may result in a poor steady-state performance. The initial means $\mu_j(0)$ for $1 \le j \le M$ of the pdf estimator can be simply set to some small positive and random values. Our extensive simulation experience suggests that the choice of the initial means $\mu_j(0)$ for $1 \le j \le M$ are not critical at all to the performance of the adaptive equalizer. This is because the propose OMG pdf estimator has an excellent adaptation capability and is capable of adapting to the underlying true data distribution from different choices of $\mu_j(0)$. By the same reason, the choice of the initial kernel width ρ_0 does not have a major impact on the achievable performance.

The nonlinear equalizer we employed is the RBF equalizer (9). The number of the RBF nodes n_c is problem dependent. Specifically, n_c is related to the size N_r of the state set $\mathbf{R}$. With n_c set to N_r, the RBF equalizer has the potential of attaining the full optimal performance of the Bayesian equalizer (7). However, this is rarely achievable as N_r is typically huge. In general, there exists a tradeoff between achievable performance and affordable cost in choosing n_c. A large n_c has the potential of achieving a better performance at the cost of higher complexity and slower adaptation process. In practice, we often choose a very small n_c, compared with the problem size N_r. We are now ready to summarize the proposed OMG-MBER RBF equalizer.

Initialization ($k \le 0$):

1) Given the equalizer order m and the RBF equalizer size n_c, initialize the RBF equalizer, namely, set the initial equalizer's parameter vector $\mathbf{w}(0)$.
2) Given the number of mixtures M and the initial kernel width ρ_0, initialize the mixture-of-Gaussian-based pdf estimator (31), namely, set the initial means $\mu_j(0)$, $1 \le j \le M$, of the M Gaussians.

Adaptation ($k \ge 1$):

1) Given $\mathbf{w}(k-1)$ and the new data $\{\mathbf{r}(k), b_0(k-d)\}$, calculate the RBF equalizer output $y(k) = f(\mathbf{r}(k); \mathbf{w}(k-1))$ using (9) and compute the signed decision variable $y_s(k) = b_0(k-d)y(k)$.
2) Update the mixture-of-Gaussian-based pdf estimator from (32) into (42). Specifically,
   - Find the index j^* according to (33).
   - Update the j^*th weight, mean, and standard deviation according to (39)–(41), respectively.
   - Update the jth weights, means, and standard deviations for $1 \le j \le M$ and $j \ne j^*$ according to (35)–(37), respectively.
3) Given the step size η, update the RBF equalizer's parameter vector from $\mathbf{w}(k-1)$ to $\mathbf{w}(k)$. Specifically
   - Calculate $(\partial f(\mathbf{r}(k); \mathbf{w}(k-1))/\partial \mathbf{w})$ according to (47).
   - Calculate $(\partial g_{j^*}(\mathbf{w}(k-1), k)/\partial \mathbf{w})$ according to (47).
   - Update the RBF equalizer's parameter vector according to (48).

Because the initial choice of the RBF equalizer's parameter vector $\mathbf{w}(0)$ has some influence on the convergence speed of the online adaptive procedure, we choose the initial centers $\mathbf{c}_j(0)$, $1 \le j \le n_c$, from the data $\{(\mathbf{r}(k), b_0(k-d)), k \le 0\}$ according to the following heuristic rules. Half of the initial centers are chosen from the data points $\mathbf{r}(k)$ that correspond to the desired signal $b_0(k-d) = +1$, whereas the other half of the initial centers are chosen from the data points $\mathbf{r}(k)$ that have the desired signal value $b_0(k-d) = -1$. Moreover, the chosen initial centers must have a minimum distance apart. Specifically, for $1 \le i, j \le n_c$ and $i \ne j$

$$\|\mathbf{c}_j(0) - \mathbf{c}_i(0)\| > \frac{\zeta_0}{n_c} \quad (49)$$

where the preset positive parameter ζ_0 controls the distribution of the initial RBF nodes. The initial weights $\alpha_j(0)$ are set either to $+\alpha_0$ or to $-\alpha_0$, depending on whether $\mathbf{c}_j(0)$ are related to $b_0(k-d) = +1$ or $b_0(k-d) = -1$, where $\alpha_0 > 0$ is a preset parameter. All the initial variances $\sigma_j^2(0)$, $1 \le j \le n_c$, are set to the same value of (σ_0^2/n_c^2), where the preset positive value σ_0^2 controls the influencing field of the initial RBF nodes. This RBF equalizer initialization procedure is similar to the one given in [33].

The values of ζ_0, α_0, and σ_0 and the step size η are carefully chosen based on trial-and-error to achieve a best possible performance.

D. Complexity Analysis

We now explicitly compare the computational complexity of the OMG-MBER RBF equalizer with that of the LBER RBF equalizer. The latter is well known to have a low complexity. Since both the adaptive RBF equalizers need to compute the RBF equalizer output $y(k)$ or the signed output $y_s(k)$ and the partial derivative vector of the RBF function $(\partial f(\mathbf{r}(k); \mathbf{w}(k-1))/\partial \mathbf{w})$, we simply denote the computational complexity of computing $y(k)$ or $y_s(k)$ and $(\partial f(\mathbf{r}(k); \mathbf{w}(k-1))/\partial \mathbf{w})$ by C_{RBF}.

TABLE I
COMPLEXITY COMPARISON FOR THE OMG-MBER-BASED AND LBER-BASED RBF EQUALIZERS

	Computing output and partial derivative vector	updating equalizer parameter vector
LBER	C_{RBF}	one $e^{\{\cdot\}}$, $3+n_c(m+2)$ multiplications, and $n_c(m+2)$ additions
OMG-MBER	C_{RBF}	one $e^{\{\cdot\}}$, $18+M+n_c(m+2)$ multiplications, one square root, and $9+n_c(m+2)$ additions

Noting the dimension of $\mathbf{w}$ is $n_c(m+2)$, from (27), it is a simple matter to work out that the parameter vector updating of the LBER algorithm additionally requires one $e^{\{\cdot\}}$ evaluation, $3+n_c(m+2)$ multiplications, and $n_c(m+2)$ additions.

For the OMG estimator, determining which mixture component, i.e., (33), to update requires M comparisons, and its complexity is negligible. Updating the mixture components, i.e., (35) and (39)–(41), requires additionally $M+9$ multiplications and 7 additions. From (47) and (48), it can be shown that the OMG-MBER algorithm updates the parameter vector by further adding one $e^{\{\cdot\}}$ evaluation, one square root computation, $9+n_c(m+2)$ multiplications, and $2+n_c(m+2)$ additions.

The computational complexity of these two adaptive RBF equalizers are listed in Table I, where it can be clearly seen that the complexity of the OMG-MBER-based RBF equalizer is only slightly more than that of the LBER-based RBF equalizer.

IV. SIMULATION STUDY

Two examples were used to compare the proposed OMG-MBER RBF equalizer with the LBER RBF equalizer given in [33] and the conventional linear MMSE equalizer based on the LMS algorithm. For the both examples, the number of mixtures was set to $M=6$ with the initial kernel parameters $\rho_0=1$ and $\mu_j(0)$, $1\le j\le M$, randomly chosen from [0.5, 1.5] for the proposed OMG pdf estimator, whereas the step size of the OMG-MBER RBF equalizer was set to $\eta=0.3$.

Example 1: In the first experiment, the transfer functions of the channel and cochannel were given as $A_0(z)=0.5+1.0z^{-1}$ and $A_1(z)=\lambda(1.0+0.5z^{-1})$, respectively, where λ was set to yield the signal-to-interference ratio (SIR) of 12 dB. The equalizer order and decision delay were set to $m=2$ and $d=1$, respectively; therefore, the size of the state set $\mathbf{R}$ was $N_r=64$. For both the OMG-MBER- and LBER-based RBF equalizers, we had $n_c=4$. For the OMG-MBER RBF equalizer, the initialization was carried out with $\zeta_0=8$, $\sigma_0=6$, and $\alpha_0=10$. For the LBER RBF equalizer, the same initialization as in [33] was used; moreover, the LBER algorithmic parameters were set to $\eta_k=0.5k^{-0.25}$ and $\rho^2=20\sigma_n^2$, the same as in [33].

Fig. 1 compares the decision boundaries of the linear MMSE equalizer, the LBER RBF equalizer, and the proposed OMG-MBER RBF equalizer using the optimal Bayesian equalizer (7) as the benchmark, under the SNR condition of 15 dB. It is shown in Fig. 1 that the linear MMSE equalizer could only realize a linear decision boundary, whereas the LBER and OMG-MBER RBF equalizers with only four RBF nodes were capable of approximating well the optimal nonlinear decision boundary of the 64-node Bayesian equalizer. The learning curves of the four-center LBER and OMG-MBER RBF equalizers and the two-tap linear MMSE equalizer based on the LMS algorithm are compared in Fig. 2 given SNR $=15$ dB, where the BER results were averaged over 100 runs. As expected, the linear MMSE equalizer had the worst BER performance. In Fig. 2, it is shown that the OMG-MBER RBF equalizer considerably outperformed the LBER RBF equalizer both in convergence speed and achievable BER performance. The achievable BER performance of the three adaptive equalizers is shown in Fig. 3 together with the BER of the optimal Bayesian equalizer. The results of Fig. 3 again confirm that the proposed OMG-MBER RBF equalizer had the superior performance over the LBER RBF equalizer.

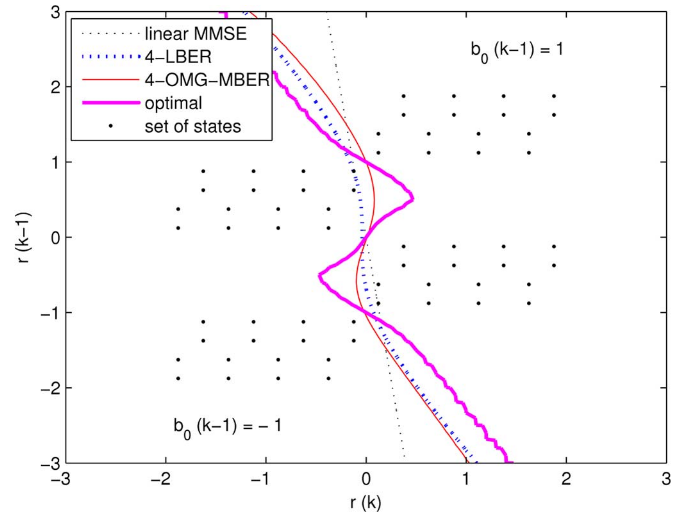


Fig. 1. Comparison of decision boundaries. (thin dotted) Linear MMSE. (thick dotted) LBER RBF. (thin solid) Proposed OMG-MBER RBF. (thick solid) Optimal Bayesian. *Example 1* given SNR $=15$ dB.

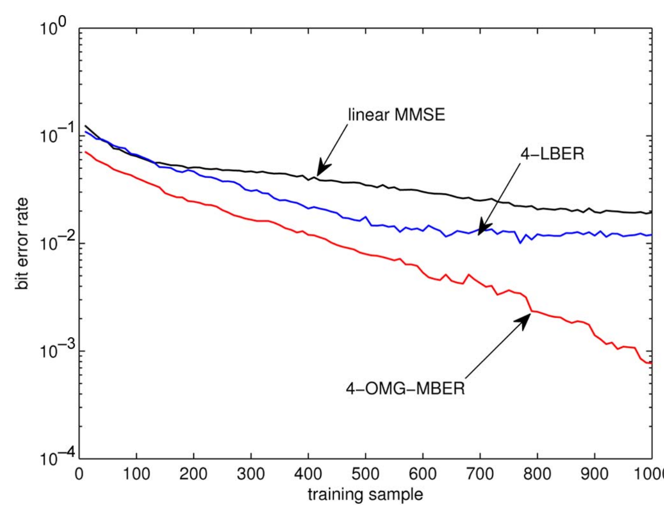


Fig. 2. Comparison of learning curves for the linear MMSE equalizer based on the LMS, the LBER RBF equalizer, and the OMG-MBER RBF equalizer, in terms of BER performance, and averaged over 100 runs for *Example 1* given SNR $=15$ dB.

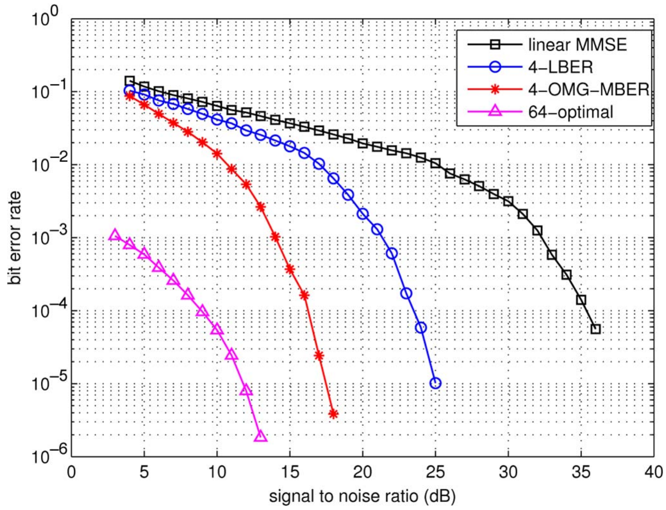


Fig. 3. BER performance comparison of the linear MMSE equalizer based on the LMS, the LBER RBF equalizer, the OMG-MBER RBF equalizer, and the optimal Bayesian equalizer for *Example 1*.

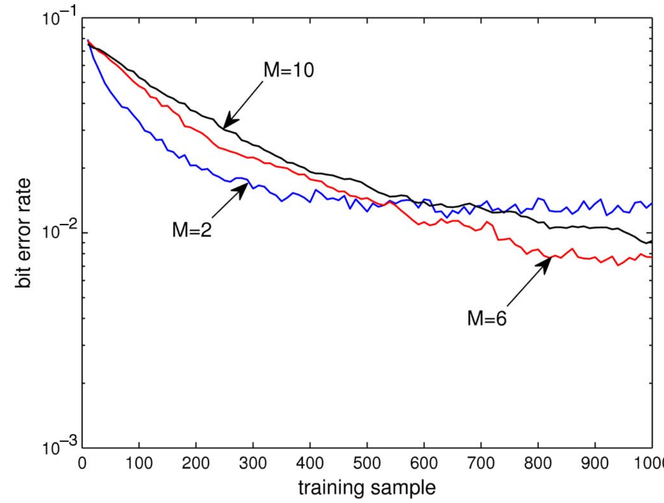


Fig. 5. Learning curves of the OMG-MBER RBF equalizer with different M, in terms of BER performance, and averaged over 100 runs for *Example 1* given $n_c = 4$ and SNR $= 12$ dB.

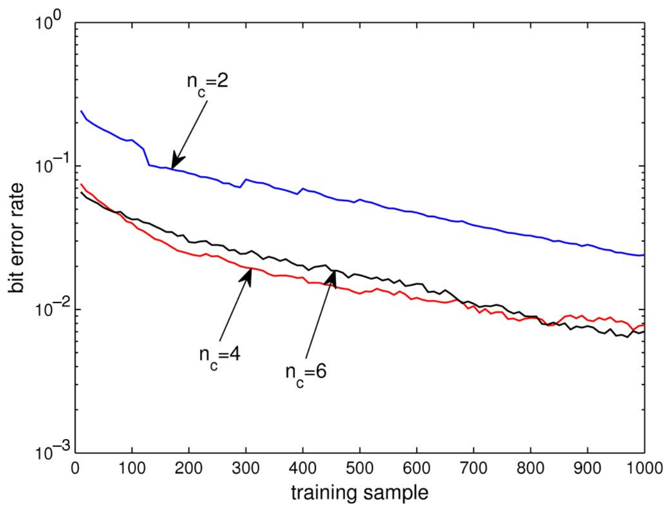


Fig. 4. Learning curves of the OMG-MBER RBF equalizer with different n_c, in terms of BER performance, and averaged over 100 runs for *Example 1* given $M = 6$ and SNR $= 12$ dB.

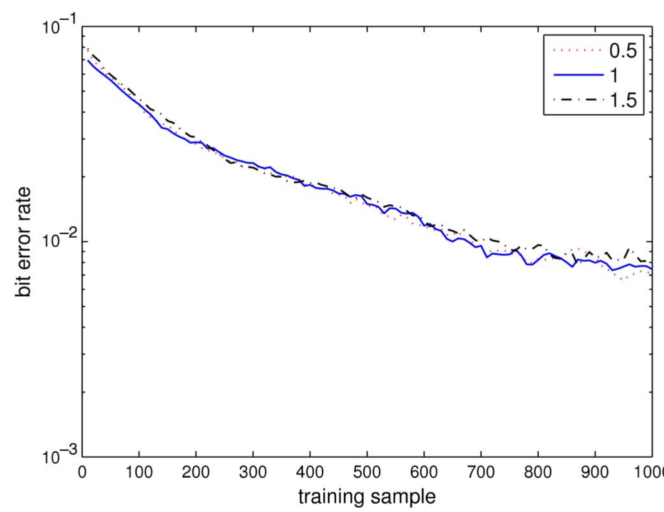


Fig. 6. Learning curves of the OMG-MBER RBF equalizer with different initial means for the OMG density estimator, in terms of BER performance, and averaged over 100 runs for *Example 1* given $n_c = 4$, $M = 6$, and SNR $= 12$ dB.

As explained previously, the achievable performance of the OMG-MBER RBF equalizer is influenced by the number of RBF nodes employed. For this example, $n_c = 4$ is appropriated. Too few RBF nodes may not achieve an adequate equalization performance, whereas a larger n_c has the potential of achieving a better performance at the cost of higher complexity and slower adaptation process. Fig. 4 shows the learning curves of the OMG-MBER RBF equalizer with different n_c, given $M = 6$ and SNR $= 12$ dB. The results of Fig. 4 clearly confirm our analysis. Similarly, there exists a tradeoff between achievable steady-state performance and convergence speed in choosing an appropriate number of mixture components M. Fig. 5 shows the learning curves of the OMG-MBER RBF equalizer with different M, given $n_c = 4$, and SNR $= 12$ dB, where it can be seen that $M = 6$ is an appropriate choice for this example. The choice of the initial means $\mu_j(0)$, $1 \le j \le M$, for the OMG density estimator on the other hand has little influence on the performance of the OMG-MBER RBF equalizer. To demonstrate this, given $n_c = 4$, $M = 6$, and SNR $= 12$ dB, we simply set $\mu_j(0) = a$, $1 \le j \le M$, with three different values of $a =$ 0.5, 1, and 1.5, and the results obtained are shown in Fig. 6.

Example 2: The transfer functions of the channel and cochannel were specified by $A_0(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$ and $A_1(z) = \lambda(0.6 + 0.8z^{-1})$, respectively, where the value of λ was chosen to yield the SIR $= 20$ dB. The equalizer order was set to $m = 4$ and the decision delay $d = 1$. The number of the signal states in $\mathbf{R}$ was $N_r = 2048$. We used $n_c = 16$ for both the LBER and OMG-MBER RBF equalizers. For the initialization of the OMG-MBER RBF equalizer, we used $\zeta_0 = 12$, $\sigma_0 = 16$, and $\alpha_0 = 10$. For the LBER RBF equalizer, $\eta_k = 0.3k^{-0.25}$ and $\rho^2 = 10\sigma_n^2$ and the same initialization as given in [33] were used.

The learning curves of the three adaptive equalizers are shown in Fig. 7 given the SNR $= 15$ dB, where the results were averaged over 100 runs. As expected, the four-tap linear MMSE equalizer trained by the LMS algorithm had the fastest

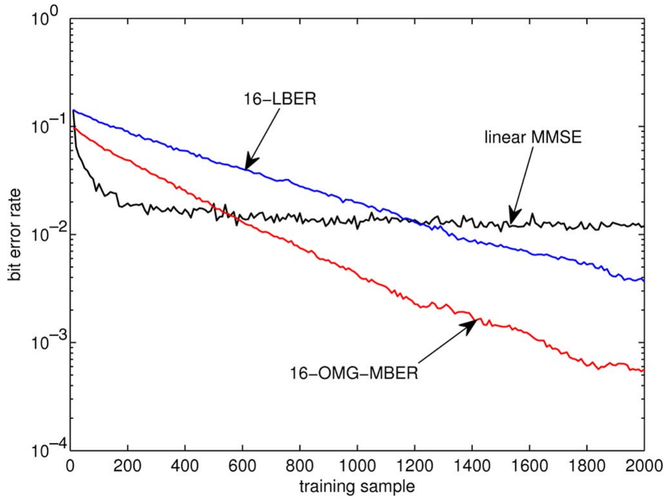


Fig. 7. Comparison of learning curves for the linear MMSE equalizer based on the LMS, the LBER RBF equalizer, and the OMG-MBER RBF equalizer, in terms of BER performance, and averaged over 100 runs for *Example 2* given SNR = 15 dB.

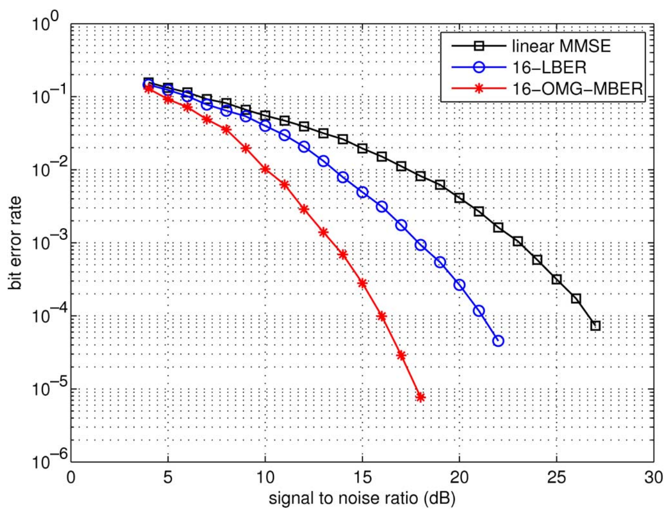


Fig. 8. BER performance comparison of the linear MMSE equalizer based on the LMS, the LBER RBF equalizer, and the OMG-MBER RBF equalizer for *Example 2*.

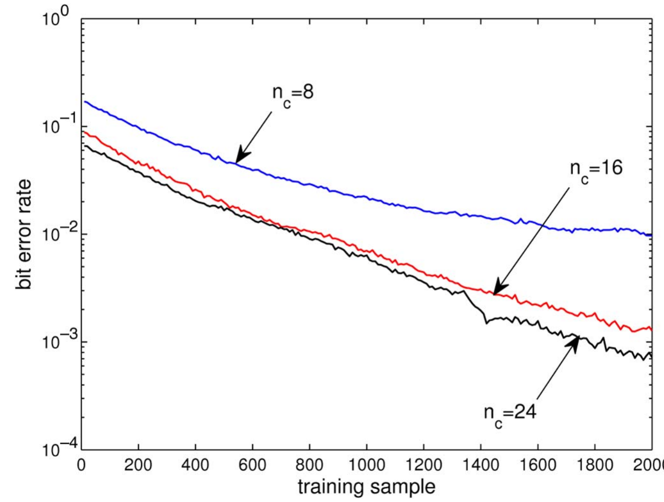


Fig. 9. Learning curves of the OMG-MBER RBF equalizer with different n_c, in terms of BER performance, and averaged over 100 runs for *Example 2* given $M = 6$ and SNR $= 12$ dB.

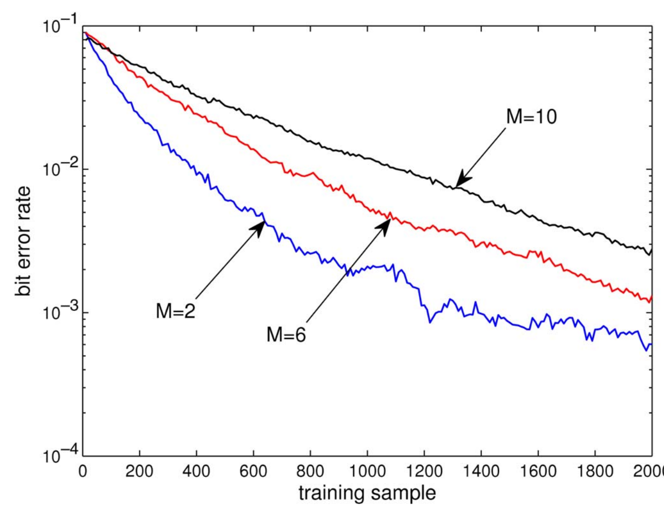


Fig. 10. Learning curves of the OMG-MBER RBF equalizer with different M, in terms of BER performance, and averaged over 100 runs for *Example 2* given $n_c = 16$ and SNR $= 12$ dB.

convergence speed, but its achievable BER performance was the worst. Fig. 8 shows the BER performance of the three adaptive equalizers tested. In both Figs. 7 and 8, it can be clearly seen that the proposed OMG-MBER RBF equalizer significantly outperformed the LBER RBF equalizer, in terms of both convergence speed and achievable BER. It is worth noting that, for this example, the optimal full Bayesian solution would require 2048 nodes. Thus, this optimal Bayesian equalizer becomes impractical when the channel and the interfering cochannel both have high orders.

The influence of the number of RBF nodes n_c to the achievable performance of the OMG-MBER RBF equalizer is investigated in Fig. 9, where it is shown that $n_c = 8$ is inadequate for this example and $n_c = 16$ is appropriate. Although a larger OMG-MBER RBF equalizer with $n_c = 24$ achieves a slightly better equalization performance than the one with $n_c = 16$, it considerably increases the computational complexity imposed. The influence of the number of mixture components M to the achievable performance of the OMG-MBER RBF equalizer is shown in Fig. 10, where it is seen that the choice of $M = 6$ is better than $M = 10$. The choice of the initial means for the OMG density estimator has little influence on the performance of the OMG-MBER RBF equalizer, as clearly shown from the results shown in Fig. 11.

V. Conclusion

The performance of an adaptive MBER nonlinear equalizer critically depends on the online pdf estimator used for estimating the signed decision variable's distribution. In this paper, a new online small-size Gaussian-mixture-based pdf estimator has been proposed to aid an adaptive MBER nonlinear equalizer. The resulting online MBER nonlinear equalizer, referred to as the OMG-MBER nonlinear equalizer, has been introduced under the generic setting of intersymbol interference

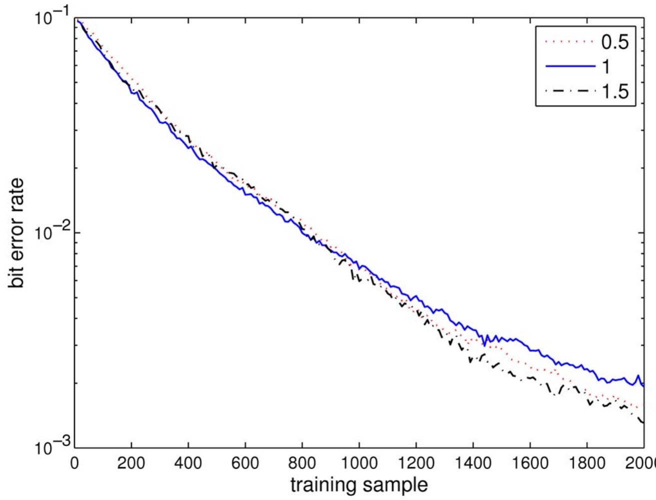


Fig. 11. Learning curves of the OMG-MBER RBF equalizer with different initial means for the OMG density estimator, in terms of BER performance, and averaged over 100 runs for *Example 2* given $n_c = 16$, $M = 6$, and SNR = 12 dB.

channel plus CCI. Compared with the existing LBER nonlinear equalizer, which relies on a one-sample pdf estimator, our online mixture-of-Gaussian-based pdf estimator is capable of providing a much more accurate estimate of the signed decision variable's pdf while maintaining very low computational requirements for online pdf estimation. Consequently, our proposed OMG-MBER nonlinear equalizer significantly outperforms the LBER nonlinear equalizer, in terms of both convergence speed and achievable BER performance.

APPENDIX A
MERGING TWO GAUSSIANS AS ONE

Consider merging a mixture of two Gaussians as follows:

$$\widehat{p}^{(2)}\left(y_s;\, \boldsymbol{\lambda}_2,\, \boldsymbol{\mu}_2,\, \boldsymbol{\rho}_2\right) = \sum_{j=1}^{2} \lambda_j G(y_s;\, \mu_j,\, \rho_j) \tag{50}$$

into one mixture by matching the resultant mean and variance. The mean μ of the two mixtures is given by

$$\begin{aligned}\mu &= \int y_s \widehat{p}^{(2)}(y_s;\, \boldsymbol{\lambda}_2,\, \boldsymbol{\mu}_2,\, \boldsymbol{\rho}_2)\, dy_s \\ &= \sum_{j=1}^{2} \lambda_j \int y_s G(y_s;\, \mu_j,\, \rho_j)\, dy_s = \sum_{j=1}^{2} \lambda_j \mu_j \end{aligned}\tag{51}$$

whereas the variance ρ^2 of the two mixtures is

$$\begin{aligned}\rho^2 &= \int (y_s - \mu)^2 \widehat{p}^{(2)}(y_s;\, \boldsymbol{\lambda}_2,\, \boldsymbol{\mu}_2,\, \boldsymbol{\rho}_2)\, dy_s \\ &= \int y_s^2 \widehat{p}^{(2)}(y_s;\, \boldsymbol{\lambda}_2,\, \boldsymbol{\mu}_2,\, \boldsymbol{\rho}_2)\, dy_s - \mu^2 \\ &= \sum_{j=1}^{2} \lambda_j \int y_s^2 G(y_s,\, \mu_j,\, \rho_j)\, dy_s - \mu^2 \\ &= \sum_{j=1}^{2} \lambda_j \left(\rho_j^2 + \mu_j^2\right) - \mu^2. \end{aligned}\tag{52}$$

APPENDIX B
DERIVATION OF $(\partial g_{j^*}(\mathbf{w}, k)/\partial \mathbf{w})$

Noting $y_s(k) = \mathrm{sgn}(b_0(k-d)) f(\mathbf{r}(k);\, \mathbf{w})$, (40) becomes

$$\begin{aligned}\mu_{j^*}(k) = &\frac{M\lambda_{j^*}(k-1)}{M\lambda_{j^*}(k-1)+1}\mu_{j^*}(k-1) \\ &+ \frac{\mathrm{sgn}\left(b_0(k-d)\right)}{M\lambda_{j^*}(k-1)+1} f\left(\mathbf{r}(k);\, \mathbf{w}\right). \end{aligned}\tag{53}$$

The partial derivative of $\mu_{j^*}(k)$ with respect to the equalizer's parameter vector $\mathbf{w}$, which is denoted $\mu'_{j^*}(k)$, is then given by

$$\mu'_{j^*}(k) = \frac{\partial \mu_{j^*}(k)}{\partial \mathbf{w}} = \frac{\mathrm{sgn}\left(b_0(k-d)\right)}{M\lambda_{j^*}(k-1)+1} \frac{\partial f(\mathbf{r}(k);\, \mathbf{w})}{\partial \mathbf{w}}. \tag{54}$$

From (41), we also have

$$\begin{aligned}\rho_{j^*}^2(k) = &\frac{M\lambda_{j^*}(k-1)\left(\rho_{j^*}^2(k-1)+\mu_{j^*}^2(k-1)\right)}{M\lambda_{j^*}(k-1)+1} \\ &+ \frac{\rho_0^2 + \left(\mathrm{sgn}\left(b_0(k-d)\right) f\left(\mathbf{r}(k);\, \mathbf{w}\right)\right)^2}{M\lambda_{j^*}(k-1)+1} - \mu_{j^*}^2(k) \\ = &\frac{M\lambda_{j^*}(k-1)\left(\rho_{j^*}^2(k-1)+\mu_{j^*}^2(k-1)\right)}{M\lambda_{j^*}(k-1)+1} \\ &+ \frac{\rho_0^2 + f^2\left(\mathbf{r}(k);\, \mathbf{w}\right)}{M\lambda_{j^*}(k-1)+1} - \mu_{j^*}^2(k). \end{aligned}\tag{55}$$

Thus, the partial derivation of $\rho_{j^*}^2(k)$ with respect to the equalizer's parameter vector $\mathbf{w}$ is

$$\begin{aligned}\frac{\partial \rho_{j^*}^2(k)}{\partial \mathbf{w}} = &\frac{2y(k)}{M\lambda_{j^*}(k-1)+1} \frac{\partial f\left(\mathbf{r}(k); \mathbf{w}\right)}{\partial \mathbf{w}} \\ &- 2\mu_{j^*}(k)\frac{\partial \mu_{j^*}(k)}{\partial \mathbf{w}}. \end{aligned}\tag{56}$$

Substituting (54) into (56) yields

$$\begin{aligned}\frac{\partial \rho_{j^*}^2(k)}{\partial \mathbf{w}} = &\frac{2y(k)}{M\lambda_{j^*}(k-1)+1} \frac{\partial f(\mathbf{r}(k); \mathbf{w})}{\partial \mathbf{w}} \\ &- \frac{2\mu_{j^*}(k)\mathrm{sgn}\left(b_0(k-d)\right)}{M\lambda_{j^*}(k-1)+1} \frac{\partial f\left(\mathbf{r}(k); \mathbf{w}\right)}{\partial \mathbf{w}} \\ = &\frac{2\left(y(k)-\mu_{j^*}(k)\mathrm{sgn}\left(b_0(k-d)\right)\right)}{M\lambda_{j^*}(k-1)+1} \frac{\partial f\left(\mathbf{r}(k); \mathbf{w}\right)}{\partial \mathbf{w}}. \end{aligned}\tag{57}$$

By denoting $\rho'_{j^*}(k) = (\partial \rho_{j^*}(k))/(\partial \mathbf{w})$, then

$$\frac{\partial \rho_{j^*}^2(k)}{\partial \mathbf{w}} = 2\rho_{j^*}(k)\rho'_{j^*}(k). \tag{58}$$

From (57) and (58), we have

$$\rho'_{j^*}(k) = \frac{y(k) - \mu_{j^*}(k)\mathrm{sgn}\left(b_0(k-d)\right)}{\rho_{j^*}(k)\left(M\lambda_{j^*}(k-1)+1\right)} \frac{\partial f\left(\mathbf{r}(k);\, \mathbf{w}\right)}{\partial \mathbf{w}}. \tag{59}$$

Finally, by making use of (59) and (54), we have

$$\begin{aligned}\frac{\partial g_{j^*}(\mathbf{w},\, k)}{\partial \mathbf{w}} &= \frac{\mu'_{j^*}(k)\rho_{j^*}(k) - \mu_{j^*}(k)\rho'_{j^*}(k)}{\rho_{j^*}^2(k)} \\ &= \frac{\left(\rho_{j^*}^2(k)+\mu_{j^*}^2(k)\right)\mathrm{sgn}\left(b_0(k-d)\right) - \mu_{j^*}(k)y(k)}{\rho_{j^*}^3(k)\left(M\lambda_{j^*}(k-1)+1\right)} \\ &\quad \times \frac{\partial f\left(\mathbf{r}(k);\, \mathbf{w}\right)}{\partial \mathbf{w}}. \end{aligned}\tag{60}$$

REFERENCES

[1] S. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, no. 9, pp. 1349–1387, Sep. 1985.

[2] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. IT-318, pp. 363–378, May 1972.

[3] D. Williamson, R. A. Kennedy, and G. W. Pulford, "Block decision feedback equalization," *IEEE Trans. Commun.*, vol. 40, no. 2, pp. 255–264, Feb. 1992.

[4] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian equalizer with decision feedback," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2918–2927, Sep. 1993.

[5] S. Siu, G. J. Gibson, and C. F. N. Cowan, "Decision feedback equalisation using neural network structures and performance comparison with standard architecture," *Proc. Inst. Elect. Eng. I—Commun., Speech Vision*, vol. 137, no. 4, pp. 221–225, Aug. 1990.

[6] S. Chen, G. J. Gibson, and C. F. N. Cowan, "Adaptive channel equalisation using a polynomial-perceptron structure," *Proc. Inst. Elect. Eng. I—Commun., Speech Vision*, vol. 137, no. 5, pp. 257–264, Oct. 1990.

[7] G. J. Gibson, S. Siu, and C. F. N. Cowan, "The application of nonlinear structures to the reconstruction of binary signals," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1877–1884, Aug. 1991.

[8] S. Chen, G. J. Gibson, C. F. N. Cowan, and P. M. Grant, "Reconstruction of binary signals using an adaptive radial-basis-function equalizer," *Signal Process.*, vol. 22, no. 1, pp. 77–93, Jan. 1991.

[9] I. Cha and S. Kassam, "Channel equalization using adaptive complex radial basis function networks," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 1, pp. 122–131, Jan. 1995.

[10] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood, NJ, USA: Prentice-Hall, 1991.

[11] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *1960 IRE WESCON Convention Record*. New York, NY, USA: Inst. Radio Eng., 1960, pt. 4, pp. 96–104.

[12] E. Shamash and K. Yao, "On the structure and performance of a linear decision feedback equalizer based on the minimum error probability criterion," in *Proc. IEEE Int. Conf. Commun.*, Jun. 17–19, 1974, pp. 25F1–25F5.

[13] S. Chen, E. S. Chng, B. Mulgrew, and G. J. Gibson, "Minimum-BER linear-combiner DFE," in *Proc. IEEE Int. Conf. Commun.*, Jun. 23–27, 1996, vol. 2, pp. 1173–1177.

[14] C.-C. Yeh, R. R. Lopes, and J. R. Barry, "Approximate minimum bit-error rate multiuser detection," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 8–12, 1998, vol. 6, pp. 3590–3595.

[15] S. Chen, B. Mulgrew, E. S. Chng, and G. J. Gibson, "Space translation properties and the minimum-BER linear-combiner DFE," *Proc. Inst. Elect. Eng.—Commun.*, vol. 145, no. 5, pp. 316–322, Oct. 1998.

[16] S. Chen and B. Mulgrew, "Minimum-SER linear-combiner decision feedback equaliser," *Proc. Inst. Elect. Eng.—Commun.*, vol. 146, no. 6, pp. 347–353, Dec. 1999.

[17] I. N. Psaromiligkos, S. N. Batalama, and D. A. Pados, "On adaptive minimum probability of error linear filter receivers for DS-CDMA channels," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1092–1102, Jul. 1999.

[18] S. Chen, S. Gunn, and C. J. Harris, "Decision feedback equaliser design using support vector machines," *Proc. Inst. Elect. Eng.—Vision, Image Signal Process.*, vol. 147, no. 3, pp. 213–219, Jun. 2000.

[19] X. F. Wang, W.-S. Lu, and A. Antoniou, "Constrained minimum-BER multiuser detection," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2903–2909, Oct. 2000.

[20] C.-C. Yeh and J. R. Barry, "Adaptive minimum bit-error rate equalization for binary signaling," *IEEE Trans. Commun.*, vol. 48, no. 7, pp. 1226–1235, Jul. 2000.

[21] B. Mulgrew and S. Chen, "Stochastic gradient minimum-BER decision feedback equalisers," in *Proc. IEEE Adapt. Syst. Signal Process., Commun., Control Symp.*, Oct. 1–4, 2000, pp. 93–98.

[22] K. A. Phillips, J. H. Reed, and W. H. Tranter, "Minimum BER adaptive filtering," in *Proc. IEEE Int. Conf. Commun.*, Jun. 18–22, 2000, vol. 3, pp. 1675–1680.

[23] B. Mulgrew and S. Chen, "Adaptive minimum-BER decision feedback equalisers for binary signalling," *Signal Process.*, vol. 81, no. 7, pp. 1479–1489, Jul. 2001.

[24] S. Chen, L. Hanzo, and N. N. Ahmad, "Adaptive minimum bit error rate beamforming assisted receiver for wireless communications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 6–10, 2003, vol. 4, pp. 640–643.

[25] M. Y. Alias, A. K. Samingan, S. Chen, and L. Hanzo, "Multiple antenna aided OFDM employing minimum bit error rate multiuser detection," *Electron. Lett.*, vol. 39, no. 24, pp. 1769–1770, Nov. 2003.

[26] S. J. Yi, C. C. Tsimenidis, O. R. Hinton, and B. S. Sharif, "Adaptive minimum bit error rate multiuser detection for asynchronous MC-CDMA systems frequency selective Rayleigh fading channels," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 7–10, 2003, vol. 2, pp. 1269–1273.

[27] R. C. de Lamare and R. Sampaio-Neto, "Adaptive MBER decision feedback multiuser receivers in frequency selective fading channels," *IEEE Commun. Lett.*, vol. 7, no. 2, pp. 73–75, Feb. 2003.

[28] D. Gesber, "Robust linear MIMO receivers: A minimum error-rate approach," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2863–2871, Nov. 2003.

[29] F.-J. Chen, S.-K. Xiong, and G. Wei, "A new constrained minimum-BER multiuser detection algorithm," in *Proc. IEEE Circuits Syst. Symp. Emerging Technol., Frontiers Mobile Wireless Commun.*, May 31–Jun. 2, 2004, vol. 3, pp. 365–368.

[30] S. Chen, N. N. Ahmad, and L. Hanzo, "Adaptive minimum bit error rate beamforming," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 341–348, Mar. 2005.

[31] S. Chen, A. Livingstone, and L. Hanzo, "Minimum bit-error rate design for space-time equalization-based multiuser detection," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 824–832, May 2006.

[32] J. Riani, S. van Beneden, J. W. M. Bergmans, and A. Immink, "Near-minimum bit-error rate equalizer adaptation for PRML systems," *IEEE Trans. Commun.*, vol. 55, no. 12, pp. 2316–2327, Dec. 2007.

[33] S. Chen, B. Mulgrew, and L. Hanzo, "Least bit error rate adaptive nonlinear equalisers for binary signalling," *Proc. Inst. Elect. Eng.—Commun.*, vol. 150, no. 1, pp. 29–36, Feb. 2003.

[34] S. Chen, "Adaptive minimum bit-error-rate filtering," *Proc. Inst. Elect. Eng.—Vision, Image Signal Process.*, vol. 151, no. 1, pp. 76–85, Feb. 2004.

[35] L. Kovacs, J. Levendovszky, A. Olah, and G. Treplan, "Approximate minimum bit error rate equalization for fading channels," *EURASIP J. Adv. Signal Process.*, vol. 2010, pp. 615623-1–615623-9, Feb. 2010.

[36] J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," Int. Comput. Sci. Inst., Berkeley, CA, USA, 1998.

[37] G. McLachlan and D. Peel, *Finite Mixture Models*. Hoboken, NJ, USA: Wiley, 2004.

[38] S. Chen, X. Hong, and C. J. Harris, "Sparse kernel density construction using orthogonal forward regression with leave-one-out test score and local regularization," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 4, pp. 1708–1717, Aug. 2004.

[39] X. Hong, S. Chen, and C. J. Harris, "A forward-constrained regression algorithm for sparse kernel density estimation," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 193–198, Jan. 2008.

[40] S. Chen, X. Hong, and C. J. Harris, "An orthogonal forward regression technique for sparse kernel density estimation," *Neurocomputing*, vol. 71, no. 4–6, pp. 931–943, Jan. 2008.

[41] S. Chen, X. Hong, and C. J. Harris, "Regression based D-optimality experimental design for sparse kernel density estimation," *Neurocomputing*, vol. 73, no. 4–6, pp. 727–739, Jan. 2010.

[42] S. Chen, S. McLaughlin, B. Mulgrew, and P. M. Grant, "Bayesian decision feedback equaliser for overcoming co-channel interference," *Proc. Inst. Elect. Eng.—Commun.*, vol. 143, no. 4, pp. 219–225, Aug. 1996.

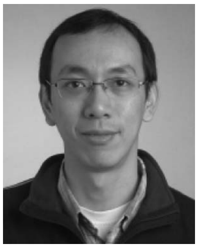
[43] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 4, no. 4, pp. 570–590, Jul. 1993.

[44] S. Chen, X. Hong, B. L. Luk, and C. J. Harris, "Construction of tunable radial basis function networks using orthogonal forward selection," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 2, pp. 457–466, Apr. 2009.

[45] S. Chen, X. Hong, and C. J. Harris, "Particle swarm optimization aided orthogonal forward regression for unified data modelling," *IEEE Trans. Evol. Comput.*, vol. 14, no. 4, pp. 477–499, Aug. 2010.

Hao Chen (S'11) received the B.Eng. degree in automatic control from National University of Defense Technology, Changsa, China, in 2006 and the M.Sc. degree (with distinction) in control systems from the University of Sheffield, Sheffield, U.K., in 2009. He is currently working toward the Ph.D. degree in cybernetics with the School of Systems Engineering, University of Reading, Reading, U.K., sponsored by the Engineering and Physical Sciences Research Council and Defence Science and Technology Laboratory from the British Government.

His research interests include online learning, nonlinear system identification, artificial neural networks, machine learning, and signal processing.



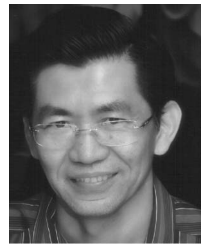
Yu Gong (M'07) received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Electronics and Science Technology of China, Chengdu, China, in 1992 and 1995, respectively, and the Ph.D. degree in communications from the National University of Singapore, in 2002.

After his Ph.D. studies, he took several research positions with the Institute for Infocomm Research, Singapore, and Queen's University, Belfast, U.K., respectively. From 2006 and 2012, he was an academic staff member with the School of Systems Engineering, University of Reading, Reading, U.K. Since July 2012, he has been with the School of Electronic, Electrical, and Systems Engineering, Loughborough University, Loughborough, U.K. His research interests include signal processing and communications, particularly wireless communications, cooperative networks, nonlinear and nonstationary system identification, and adaptive filters.

Xia Hong (SM'02) received the B.Sc. and M.Sc. degrees from the National University of Defense Technology, Changsa, China, in 1984 and 1987, respectively, and the Ph.D. degree from the University of Sheffield, Sheffield, U.K., in 1998, all in automatic control.

From 1987 to 1993, she worked as a Research Assistant with Beijing Institute of Systems Engineering, Beijing, China. From 1997 to 2001, she was a Research Fellow with the Department of Electronics and Computer Science, University of Southampton, Southampton, U.K. She is currently a Professor with the School of Systems Engineering, University of Reading. She is the author of over 100 research papers and co-author of a research book. Her research interests include nonlinear system identification, data modeling, estimation and intelligent control, neural networks, pattern recognition, and learning theory and their applications.

Dr. Hong received the Donald Julius Groen Prize from the Institution of Mechanical Engineers in 1999.



Sheng Chen (M'90–SM'97–F'08) received the B.Eng. degree in control engineering from East China Petroleum Institute, Dongying, China, in January 1982; the Ph.D. degree in control engineering from City University London, London, U.K., in September 1986; and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005.

From 1986 to 1999, he held research and academic appointments with the University of Sheffield, Sheffield, U.K.; the University of Edinburgh, Edinburgh, U.K; and the University of Portsmouth, Portsmouth, U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he is currently a Professor of intelligent systems and signal processing. He is the author of over 500 research papers. He is also a Distinguished Adjunct Professor with King Abdulaziz University, Jeddah, Saudi Arabia. He is an ISI highly cited researcher in engineering (March 2004). His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural networks and machine learning, intelligent control system design, and evolutionary computation methods and optimization.

Dr. Chen is a Fellow of the Institution of Engineering and Technology.