Stochastic Blind Equalization Based on PDF Fitting Using Parzen Estimator

Marcelino Lázaro, *Member, IEEE*, Ignacio Santamaría, *Member, IEEE*, Deniz Erdogmus, *Member, IEEE*, Kenneth E. Hild, *Member, IEEE*, Carlos Pantaleón, *Member, IEEE*, and Jose C. Principe, *Fellow, IEEE*

***Abstract*—This paper presents a new blind equalization approach that aims to force the probability density function (pdf) at the equalizer output to match the known constellation pdf. Quadratic distance between pdf's is used as the cost function to be minimized. The proposed method relies on the Parzen window method to estimate the data pdf and is implemented by a stochastic gradient descent algorithm. The kernel size of the Parzen estimator allows a dual mode switch or a soft switch between blind and decision-directed equalization. The proposed method converges faster than the constant modulus algorithm (CMA) working at the symbol rate, with a similar computational burden, and reduces the residual error of the CMA in multilevel modulations at the same time. A comparison with the most common blind techniques is presented.**

***Index Terms*—Blind equalization, CMA, information theory, PDF, Parzen windowing.**

I. Introduction

CHANNEL equalization plays a key role in digital communication systems. Typically, the physical channel introduces a distortion to the transmitted signal that can make it difficult to recover the original data. In this case, an equalizer is necessary to reduce, or ideally to completely eliminate, the introduced intersymbol interference (ISI). Conventional equalization techniques rely on the transmission of a reference (training) sequence that is known at the equalizer. This sequence allows adaptation of the equalizer parameters to minimize some cost function that measures the distance between the actual equalizer output and the desired reference signal. For instance, when the equalizer is implemented by means of a linear filter, the filter coefficients can be easily adapted by using the well-known least mean squares (LMS) [1], which minimizes the expectation of the squared error. A detailed study of conventional adaptive equalization can be found in [2].

When a training sequence is not available at the receiver, the problem at hand is named blind equalization. Blind equalization has received a great amount of attention during the last years because of its importance in communication systems [3]–[5]. Without a reference sequence, the only knowledge about the transmitted sequence is limited to its probabilistic or statistical properties. Two broadly defined classes of blind algorithms can be used to exploit this knowledge [6]: methods based on second-order statistics (SOS) and methods based on higher order statistics (HOS). SOS methods [7], [8] exploit *cyclostationarity* of the channel output. These techniques require to sample the received signal at a rate faster than the symbol rate (thus introducing temporal diversity). Although adaptive versions have been proposed, SOS-based methods are typically block-oriented and computationally intensive algorithms. SOS methods provide a fast convergence. Unlike SOS methods, the blind algorithms based on higher order statistics can operate at the symbol rate. HOS algorithms typically minimize a cost function that is able to indirectly extract the higher order statistics of the signal or the current level of ISI at the equalizer output [3]. Usually, the cost function is minimized by means of stochastic gradient algorithms working at the symbol rate, which are simple to implement. The main drawback is that these algorithms usually need a high number of data symbols to achieve convergence. The different characteristics (complexity versus convergence speed) of SOS and HOS methods make these two approaches complementary [6]. For this reason, despite the slower convergence, HOS are currently used in a variety of communication systems, such as the ATM-LAN [9] and VDSL [10] standards. Examples of HOS-based algorithm include the Sato algorithm [11], which was the first blind technique for multilevel PAM signals, and the Godard algorithms [12]. The Constant Modulus Algorithm (CMA) [13], which is probably the most popular blind equalization technique, is a particular case of the Godard algorithms. For multimodulus modulations, such as M-QAM, in addition to convergence speed, CMA has another drawback: It shows a relatively high residual error.

There are several approaches in the literature that attempt to improve the convergence speed of conventional HOS blind techniques without substantially increasing complexity. Normalized-CMA (NCMA) [14] accelerates convergence by estimating the optimal step size at each iteration considering that a constant modulus constellation, like quadrature phase shift keying (QPSK), is employed (the output modulus is known). However, it fails for multimodulus constellations. Renyi's entropy has also been introduced as a cost function for blind equalization of constant modulus signals [15]. This approach is an application of information theoretic criteria [16] to equalization and it uses the Parzen window method (a nonpara-

Manuscript received October 7, 2003; revised February 24, 2004. This work was supported in part by Spanish Ministry of Science and Technology (MCYT) under Grant TIC2001-0751-C04-03 and by the National Science Foundation under Grant ECS-0 300 340. The associate editor coordinating the review of this paper and approving it for publication was Prof Zhi Ding.

M. Lázaro is with the Department of Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain (e-mail: marce@ieee.org).

I. Santamaría and C. Pantaleón are with the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Santander, Spain (e-mail: nacho@gtas.dicom.unican.es; carlos@gtas.dicom.unican.es).

D. Erdogmus, K. E. Hild, and J. C. Principe are with the Electrical and Computer Engineering Department, University of Florida, Gainesville, FL 32611 USA (e-mail: derdogmus@ieee.org; k.hild@ieee.org; princip@cnel.ufl.edu).

Digital Object Identifier 10.1109/TSP.2004.840767

metric method) to estimate the underlying probability density function (pdf). Although this method provides excellent results for some channels, it fails to equalize other ones (especially channels with zeros on the unit circle), and it is very sensitive to noise. Similar results have been obtained when using Fisher's information, instead of entropy, as the cost function.

An interesting alternative consists in trying to force the probability density at the output of the equalizer to match the known constellation pdf. Several approaches have been developed following this idea. Linear [17] and nonlinear (neural-networks-based) equalizers [18] use the Kullback–Leibler divergence between densities as a cost function. A new method based on the quadratic distance between pdf's has been proposed in [19]. A simple Gaussian model is used for the target pdf to be fitted (it was designed only for constant modulus constellations), and it defines computationally efficient stochastic gradient expressions, which are much simpler than the method in [17]. In [20], we presented a method based on fitting the pdf of the equalizer output at some relevant points, which are the points determined by the modulus of the constellation symbols. Since the cost function only considers these points, we called it sampled-pdf (SPDF) fitting.

Concerning the high residual error of CMA, typically, this is solved by the so-called dual mode switching techniques. CMA works until the eye is opened, and then, a hard switch to a technique providing a low residual error is performed. The classical choice is the well-known decision-directed equalization (DDE), but several alternatives have been recently proposed [21], [22]. Another option consists in a soft switch from blind to decision-directed mode. Some examples are the classical Benveniste-Goursat algorithm [23], which is used in digital TV systems [24], or the recently introduced Dual Mode CMA (DM-CMA) [25] and its Stop-and-Go extension (SAG-DM-CMA) [26]. In this case, a simple rule decides at each iteration whether the CMA or Radius Directed Equalization (RDE), which is a decision-directed-like method [27], must be employed.

In this paper, we present an adaptive algorithm that converges faster than the CMA working at the symbol rate, with a low computational burden and that reduces the residual error of the CMA in multilevel modulations at the same time. The proposed method aims to force the probability density of the equalizer output to match the known constellation pdf. Unlike the method in [20], which only considers a reduced set of sampling points where the pdf is fitted, now, the whole pdf is considered. This method is designed for multilevel modulations and works at the symbol rate, admitting a simple stochastic gradient-based implementation. In this way, its complexity is similar to CMA. The Parzen window method is used to estimate the underlying pdf. The proposed method improves the performance of CMA in two senses. First, it provides a faster convergence than CMA until the eye is opened. Second, just by changing the size of Parzen kernel, it presents a lower residual error when convergence is achieved, similar to decision-directed equalization. Moreover, the adaptive control of the kernel size allows a soft switch between blind and decision-directed equalization.

The paper is organized as follows. Section II formulates the blind equalization problem and outlines the most typical blind algorithms. In Section III, the proposed cost function and its corresponding stochastic gradient expression are presented. Section IV discusses some implementation details, and Section V presents the results obtained with the proposed method. Finally, Section VI discusses the main conclusions.

II. Blind Equalization Formulation and Conventional Approaches

In general, in a digital communication system, a sequence $\{s_k\}$ of i.i.d. complex symbols, belonging to the constellation of any digital modulation, is sent through a channel. Usually, the channel is described by means of its discrete-time complex coefficients h_k, assuming a finite impulse response (FIR) channel. Therefore, the channel output is obtained by

$$x_k = \sum_{n=0}^{L_h-1} h_n s_{k-n} + e_k \tag{1}$$

where L_h is the channel length, and e_k is the noise sequence that typically is modeled by a white Gaussian noise process. The blind equalizer will operate on the channel output to reduce the intersymbol interference introduced by the channel. In this paper, a linear equalizer will be implemented by means of an FIR filter. In this case, the equalizer output is given by

$$y_k = \sum_{n=0}^{L_w-1} w_n x_{k-n} = \mathbf{w}^T \mathbf{x}_k \tag{2}$$

where $\mathbf{w}$ is the vector of filter coefficients to be adapted by the blind equalization algorithm to minimize ISI.

Without a reference sequence, the blind algorithms must make use of some *a priori* knowledge of the statistics of the underlying modulation. In the following, we will outline some of the most common blind algorithms and the *a priori* assumptions they use.

The Godard algorithms [12] minimize the following cost function:

$$J_G(\mathbf{w}) = E[(|y_k|^p - R_p)^2] \tag{3}$$

where the ratio R_p contains the *a priori* knowledge about the current modulation. In this case

$$R_p = \frac{E[|s_k|^{2p}]}{E[|s_k|^p]}. \tag{4}$$

CMA is the Godard algorithm for $p = 2$.

The Sato algorithm [11] defines the following error function:

$$E_k^S = y_k - \alpha(\mathrm{sgn}(\mathrm{real}(y_k)) + j\mathrm{sgn}(\mathrm{imag}(y_k))) \tag{5}$$

where $j = \sqrt{-1}$, and

$$\alpha = \frac{E[|\mathrm{real}(s_k)|^2]}{E[|\mathrm{real}(s_k)|]} = \frac{E[|\mathrm{Imag}(s_k)|^2]}{E[|\mathrm{Imag}(s_k)|]}. \tag{6}$$

This error function is employed in the adaptation procedure instead of the error with respect to the desired solution under the MSE criterion.

The Benveniste-Goursat algorithm [23] is an extension of the Sato algorithm. It proposes the following error function:

$$E_k^{\mathrm{BG}} = k_1 E_k^D + k_2 \left|E_k^D\right| E_k^S \tag{7}$$

where $E_k^D = y_k - \mathrm{decis}(y_k)$ is the decision error. This algorithm implements a soft transition between blind equalization and decision-directed equalization. When E_k^D is large, the second term (the Sato term) dominates and it works in blind mode. When E_k^D is small, both terms have the same order of magnitude, and the noise due to E_k^S is removed (note that $|E_k^D| = 0$ for perfect equalization).

The Dual Mode CMA (DM-CMA) decides at each iteration if the adaption is performed by using the error in CMA $E_{\mathrm{CMA}} = |y_k|^2 - R_p$ or the error for Radius Directed Equalization (RDE) [27], $E_{\mathrm{RDE}} = |y_k|^2 - R_i^2$, where R_i is the closest radius of the underlying constellation. E_{RDE} is selected when y_k is close enough to $R_i(||y_k| - R_i| < d)$; otherwise, E_{CMA} is selected. The Stop-and-Go extension of this method (SAG-DM-CMA) [26] proposes to adapt the weights only when the error estimates are supposed to be reliable, which means that the sign of both errors is the same. This can accelerate the convergence for low noise environments.

III. Proposed Cost Function and Stochastic Algorithm

Information theory is an interesting alternative to extract as much information as possible from the available data. The data distribution contains more information than the simple statistics employed in the CMA or Sato algorithms. The proposed algorithm aims to force the probability density of the equalizer output to match the known constellation pdf. The method works in the space of the modulus of the symbols raised to the power p, i.e., it tries to fit the pdf of $S^p = \{|s_k|^p\}$. In practice, the best results are usually obtained for $p = 2$, but the method admits a general formulation.

A. Quadratic Distance (QD) Cost Function

The proposed approach is an extension of the method proposed in [19] to consider multilevel modulations. We use the quadratic distance between pdf's as the cost function, which is given by

$$J(\mathbf{w}) = \int_{-\infty}^{+\infty} \left(f_{Y^p}(z) - f_{S^p}(z)\right)^2 dz \tag{8}$$

where $Y^p = \{|y_k|^p\}$, $S^p = \{|s_k|^p\}$, and $f_Z(z)$ denotes the pdf of Z at z.

The Parzen window method [28] is used to estimate the current data pdf. Using this nonparametric estimator with a window of the L previous symbols, the estimate of the pdf $f_{Y^p}(z)$ at time k is

$$\hat{f}_{Y^p}(z) = \frac{1}{L}\sum_{i=0}^{L-1} K_{\sigma_o}(z - |y_{k-i}|^p) \tag{9}$$

where $K_\sigma(x)$ is the Parzen window kernel of size σ. This kernel has to be a suitable pdf function. We use Gaussian kernels with standard deviation σ

$$K_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \tag{10}$$

For consistency, the target pdf must consider the effect of the estimator. To guarantee the cost function is zero at perfect equalization, the same estimator must be used to estimate $f_{Y^p}(z)$ and to compute the target pdf $f_{S^p}(z)$. Therefore, the pdf of the original constellation must be the convolved with the kernel of the Parzen estimator being used to estimate $f_{Y^p}(z)$

$$\hat{f}_{S^p}(z) = \sum_{i=1}^{N_s} p_S(s_i) K_{\sigma_o}(z - |s_i|^p) \tag{11}$$

where N_s is the number of complex symbols in the constellation of the corresponding modulation, and $p_S(s_i)$ denotes the probability of symbol s_i. In the following, symbols are supposed to be equally likely, and therefore, $p_S(s_i) = 1/N_s$. With this target pdf, the cost function equals zero only at perfect equalization (for a large L). Consequently, the Benveniste–Goursat–Ruget Theorem [29] ensures that (8) is a suitable cost function for blind equalization.

Finally, substituting (9) and (11) in (8), rearranging terms, and taking into account that for Gaussian kernels

$$\int_{-\infty}^{+\infty} K_\sigma(y - C_1) K_\sigma(y - C_2)\, dy = K_{\sigma\sqrt{2}}(C_1 - C_2) \tag{12}$$

we obtain the following expression for the cost function

$$\begin{aligned} J(\mathbf{w}) = {} & \frac{1}{L^2}\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} K_\sigma(|y_{k-j}|^p - |y_{k-i}|^p) \\ & + \frac{1}{N_s^2}\sum_{i=1}^{N_s}\sum_{j=1}^{N_s} K_\sigma(|s_j|^p - |s_i|^p) \\ & - \frac{2}{LN_s}\sum_{i=1}^{N_s}\sum_{j=0}^{L-1} K_\sigma(|y_{k-j}|^p - |s_i|^p). \end{aligned} \tag{13}$$

For the sake of simplicity in the notation, we have denoted $\sigma_o\sqrt{2}$ as σ. Analyzing this expression, it can be seen that the first term on the right side is the estimator of the information potential associated with Renyi's entropy of order 2, i.e., the cost function in [15]. However, in this case $J(\mathbf{w})$ must be minimized, whereas in [15], the information potential must be maximized. The second term on the right side, which does not depend on $\mathbf{w}$, can be neglected in the optimization process. Finally, the third term pulls the output of the equalizer toward the desired pdf. We want to remark that for constant modulus modulations $J(\mathbf{w})$ becomes the cost function proposed in [19].

B. Stochastic Gradient Algorithm

We have considered a stochastic gradient approach using a window length $L = 1$, which, in the following, we will call stochastic gradient algorithm (SQD). We will focus on $p = 2$. Under these assumptions, the derivative of (13) with respect to the equalizer weights (neglecting the constant 2) is given by

$$\nabla_{\mathbf{w}} J(\mathbf{w}) = -\frac{1}{N_s}\sum_{i=1}^{Ns} K'_\sigma(|y_k|^2 - |s_i|^2) y_k \mathbf{x}_k^* \tag{14}$$

where $K'(x)$ is the derivative of the kernel $K(x)$. It is interesting to compare this expression with the classical updating expression of CMA algorithm. Under a constant modulus signal, we can write $R_p = |s_i|^2$, and the updating expression using a Gaussian kernel becomes

$$\nabla_{\mathbf{w}} J(\mathbf{w}) = \frac{(|y_k|^2 - R_p)}{\sqrt{2\pi}\sigma^3} y_k \mathbf{x}_k^* e^{-\frac{(|y_k|^2 - R_p)^2}{2\sigma^2}}. \quad (15)$$

Up to constant terms, this is the CMA updating expression multiplied by the exponential term. Therefore, in terms of computational burden the increment with respect to CMA is reduced to the evaluation of this exponential term.

Once the derivative of the cost function $J(\mathbf{w}_k)$ has been evaluated, the equalizer coefficients weights are adapted by

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_\sigma \nabla_{\mathbf{w}} J(\mathbf{w}). \quad (16)$$

In all cases, a normalized step size $\mu_\sigma = \mu\sigma^3$ has been introduced to compensate for the $1/\sigma^3$ term that appears in $K'_\sigma(x)$ for Gaussian kernels. In the following, we will only consider μ when referring to the step size. The proposed algorithm begins with a tap-centered equalizer and, for each received sample, estimates the gradient (14) and updates the equalizer by (16).

C. Alternative Cost Function

It is interesting to note that the stochastic expression (14) can also be obtained from a different cost function

$$J^c(\mathbf{w}) = \int_{-\infty}^{+\infty} f_{Y^p}(z) f_{S^p}(z)\, dz. \quad (17)$$

$J^c(\mathbf{w})$ is the correlation between both pdf's, and a maximum is obtained when the pdf of Y^p matches $f_{S^p}(z)$. Expanding the QD cost function (8), we get three terms

$$J(\mathbf{w}) = \int_{-\infty}^{+\infty} \left(f_{Y^p}^2(z) + f_{S^p}^2(z) - 2 f_{Y^p}(z) f_{S^p}(z)\right) dz. \quad (18)$$

The term responsible for measuring the distance is the crossed term, while the others act as normalization for different pdf's. Since these entities are used as a cost function for which we want to find an extreme, it is intuitive to use just the cross term, hoping that the minima of the two criteria occur at the same parameter value. Some evidence of this behavior is pointed out in [30], where the cost function (17) provides outstanding results in an application of information theory to clustering.

If the Parzen window method is used again to estimate the current pdf and it is included into the target pdf for the sake of consistency, substituting (9) and (11) in (17), we obtain the following expression for the cost function [31]:

$$J^c(\mathbf{w}) = \frac{1}{LN_s} \sum_{i=1}^{N_s} \sum_{j=0}^{L-1} K_\sigma(|y_{k-j}|^p - |s_i|^p). \quad (19)$$

Again, we denote $\sigma_o\sqrt{2}$ as σ for simplicity. The stochastic gradient of (19) with $L = 1$ is equal to (14) (up to the negative sign because now the goal is to maximize $J^c(\mathbf{w})$).

The previously proposed sampled-pdf method (SPDF) [20] employs a similar cost function, which fits the constellation pdf at a number N_p of representative points (the sampling points). The cost function is

$$J^s(\mathbf{w}) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(f_{Y^p}(r_i) - T_i\right)^2 \quad (20)$$

where T_i are the target values of the pdf at r_i ($T_i = f_{S^p}(r_i)$). The sampling points are $r_i = |s_i|^2$. Using the Parzen estimator, the stochastic gradient for $L = 1$ is

$$\Delta_{\mathbf{w}} J^s(\mathbf{w}) = -\frac{2}{N_p} \sum_{i=1}^{N_p} (K_\sigma(r_i - |y_k|^2) - K_\sigma(0)) \times K'_\sigma(r_i - |y_k|^2) y_k \mathbf{x}_k^*. \quad (21)$$

Basically, it is (14) multiplied by $(K_\sigma(r_i - |y_k|^2) - K_\sigma(0))$. Therefore, its computational requirement is slightly higher than the requirements for SQD.

IV. Implementation Details

In this section, some implementation details about the proposed algorithm are discussed. We focus on $L = 1$.

A. Kernel Size

The kernel size of the Parzen window estimator plays a key role in the proposed algorithm. It is responsible for both the convergence speed and the accuracy of the final solution. For the sake of speed, a large kernel size is necessary. A large kernel size allows the interaction of each symbol with all the constellation symbols (in the space of $|s_k|^p$), and this produces a fast convergence. However, the contrary is necessary when the goal is accuracy. In this case, a small kernel size allows only the interaction of each symbol with the closest symbol of the constellation. This reduced interaction produces, in practice, a decision-directed equalization that is able to produce an accurate final solution.

In communication systems, convergence speed is the main requirement for blind equalizers. Typically, the blind algorithm (for instance CMA) reduces ISI until the eye of the constellation is opened. At this point, a switch to decision-directed equalization is performed, which yields an accurate final solution. Under these premises, the natural choice is to employ a large initial kernel size to reinforce convergence speed. Then, when the eye of the constellation is opened, a switch to a small kernel size, or to decision-directed equalization, can be performed, similarly to CMA based systems. However, an adjustable size allows a more interesting approach. By adaptively controlling the kernel size, a soft switch from blind to decision-directed like equalization can be implemented. This option is detailed in Section IV-C.

We would want to point out another feature related with the kernel size. Based on our previous experience with the Parzen estimator [32], which suggest links with convolution smoothing, and on the results we have obtained in this application, we can suggest that a large kernel size has also the advantage of producing a cost surface with fewer local minima. Fig. 1 shows the normalized cost surface for a simple case: channel $h_k = \delta_k$ and an equalizer with a single coefficient, which controls the gain. Cost surfaces using $\sigma = 1$ and $\sigma = 15$ for a 16-QAM,

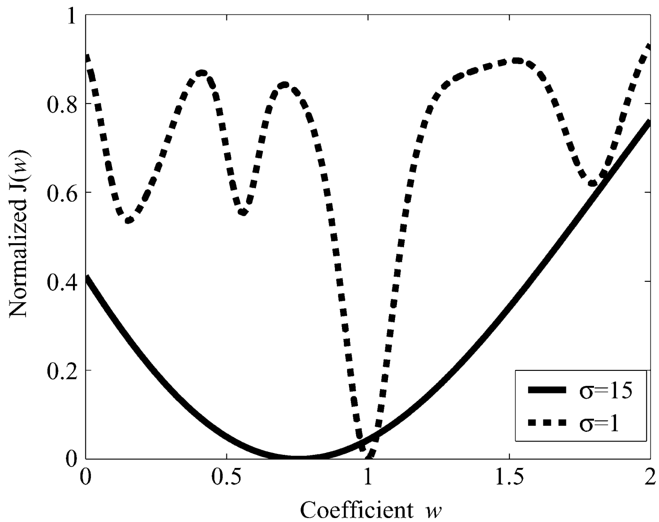


Fig. 1. Normalized SQD cost surface as function of a single equalizer coefficient for different kernel sizes in a 16-QAM modulation.

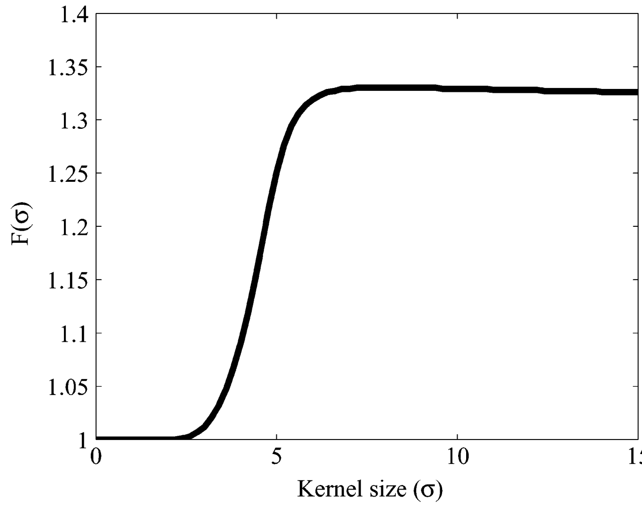


Fig. 2. Numerically obtained compensation factor $F(\sigma)$.

are plotted. The surface with $\sigma = 15$ has only one minimum in the plotted range, whereas with $\sigma = 1$, several local minima appear. It seems reasonable to assume a similar behavior in a more complex case, and this intuition has been corroborated by experimental results, as will be shown in Section V.

B. Gain Identification

Fig. 1 also reveals another interesting feature. For $\sigma = 1$, the minimum of the cost function is obtained at unity gain. However, for $\sigma = 15$, the minimum is not located at $w = 1$. In general, the proposed algorithm equalizes a channel up to gain identification. When a small kernel size is employed, the perfect equalization situation with exact gain corresponds to a minimum of the stochastic cost function. However, when a large kernel size is used, the analysis of the stochastic cost function reveals that a minimum of the cost function for perfect equalization corresponds to a slightly scaled-down constellation (inexact gain). This is due to the stochastic nature of the algorithm. Using a single sample to estimate the pdf by Parzen windowing produces this undesired scaling, while the batch algorithm, which uses a large number of samples to estimate the pdf's, provides equalization with gain identification. In general, $L > 10$ is sufficient for the bias to be negligible.

In order to ensure gain identification, the adapting expression (14) has to be analyzed, and some modification has to be included. Taking into account that the minimum of the stochastic cost function is a scaled version of the desired constellation, an obvious modification is to substitute the original symbols $|s_i|^2$ by some *precompensated* symbols

$$|s_i^c|^2 = F(\sigma)|s_i|^2 \tag{22}$$

where $F(\sigma)$ is the compensation factor that depends on the kernel size. In this case, the compensated adaptation term is

$$\nabla_{\mathbf{w}} J(\mathbf{w}) = \frac{1}{N_s} \sum_{i=1}^{Ns} K'_\sigma(|y_k|^2 - F(\sigma)|s_i|^2) y_k \mathbf{x}_k^*. \tag{23}$$

For the sake of gain identification, it is necessary to ensure that the zero-ISI solution $(y_k = s_k)$ is a minimum of the cost function. This is equivalent to ensuring that

$$E[\nabla_{\mathbf{w}} J(\mathbf{w})] = \frac{1}{N_s} \sum_{i=1}^{Ns} E[K'_\sigma(|s_k|^2 - F(\sigma)|s_i|^2) s_k \mathbf{x}_k^*] = 0. \tag{24}$$

Considering an infinite length equalizer, and taking into account relationship (1), we obtain

$$\sum_{i=1}^{N_s} E[K'_\sigma(|s_k|^2 - F(\sigma)|s_i|^2)|s_k|^2] = 0. \tag{25}$$

Although it is not possible to find an analytical expression for $F(\sigma)$, because of the nonlinearity of equations, it is simple to find a numerical solution for a specific modulation. Fig. 2 shows the numerically obtained compensation factor as a function of the kernel size, for a 16-QAM modulation. This demonstrates the need for a compensation factor. It must be noted that $F(\sigma)$ becomes 1 for small kernel sizes, which means that, in this case, the compensated symbols correspond to the original symbols.

C. Soft Switch From Blind to Decision-Directed Equalization

We have discussed the fact that the kernel size controls the convergence speed and the final accuracy of the solution and the fact that both requirements are in opposition to each other. A dual mode technique can be used to switch from blind to decision-directed modes. However, an interesting approach consists of adaptively controlling the kernel size to have a large value in the blind stage and to progressively decrease it during convergence to obtain a more accurate final equalization.

An error measure, which is based on the variance of the error with respect to the closest target, has been employed to control the kernel size. This measure is iteratively adapted, using a forgetting factor α, by means of

$$E_{k+1} = \alpha E_k + (1-\alpha) \min_{\{i=1,\ldots,N_s\}} ((|y_k|^2 - |s_i|^2)^2). \tag{26}$$

A linear relationship between the kernel size and this error measure has been assumed. In this case

$$\sigma_k = aE_k + b \tag{27}$$

where a and b are empirically determined constants. For instance, for a 16-QAM modulation, we have found, after testing in a large number of channels, that $a = 3.5$ and $b = -9.5$ provide very good results.

In order to obtain a suitable soft transition, the compensated symbols $|s_i^c|^2$ have to be adapted at each iteration with the current kernel size to guarantee a convergence with gain identification. A look-up table has been used to evaluate $F(\sigma)$.

Taking this into account, the soft transition algorithm can be summarized in the following steps:

1) Initialize μ, E_1, α, and $\mathbf{w}_0$ (tap-centered).
2) For k $= 1, 2, \ldots$ (sample-by-sample adaptation)
   a) Evaluate σ_k by (27)
   b) Update $\mu_\sigma = \mu\sigma_k^3$
   c) Obtain $F(\sigma_k)$ (by look-up table)
   d) Update $|s_i^c|^2 = F(\sigma_k)|s_i|^2$
   e) Evaluate $\nabla_{\mathbf{w}} J(\mathbf{w})$ by (14) using $|s_i^c|^2$
   f) Update $\mathbf{w}_{k+1}$ by (16)
   g) Estimate E_{k+1} by (26)

End

TABLE I
CONVERGENCE OF CMA, SPDF, AND THE SQD METHOD FOR 200 RANDOM CHANNELS

Method	Noiseless		SNR=10 dB	
	Iterations	Resid. ISI	Iterations	Resid. ISI
CMA	18680	-16.47 dB	42175	-9.66 dB
SPDF	17126	-16.45 dB	38005	-9.69 dB
SQD	16625	-16.53 dB	37110	-9.83 dB

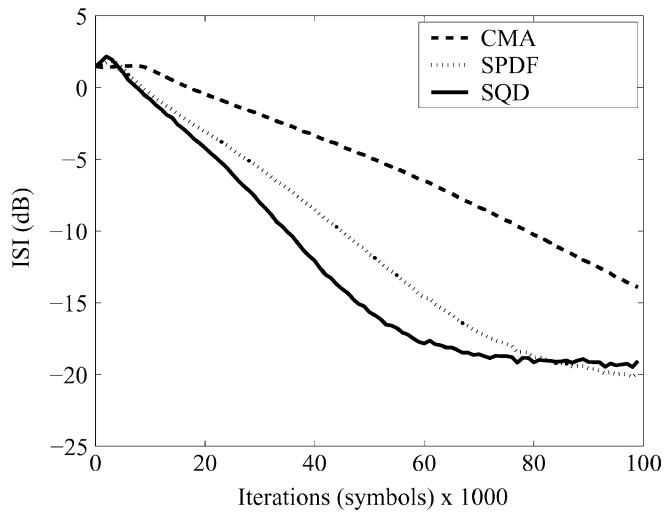


Fig. 3. Convergence curves for $H_1(z)$ with SNR $= 30$ dB.

V. RESULTS

In this section, some results obtained with the proposed methods are presented.

A. *Blind Mode*

We start by analyzing the performance of the proposed algorithm in blind mode. To maximize convergence speed a fixed, large kernel size is employed. The proposed method is compared with CMA, which is the most commonly used blind algorithm for QAM modulations, and SPDF. ISI will be used as the figure of merit, i.e.,

$$\mathrm{ISI} = 10\log_{10}\frac{\sum_n|\theta_n|^2 - \max_n|\theta_n|^2}{\max_n|\theta_n|^2} \tag{28}$$

where $\boldsymbol{\theta} = \mathbf{h} * \mathbf{w}$ is the combined channel-equalizer impulse response.

To assess the blind convergence properties of the SQD method, we equalized 200 random channels having seven coefficients by using a 21-tap linear equalizer with tap-centered initialization. The proposed algorithm converged for all channels. Table I compares the proposed method with CMA and SPDF in terms of the mean number of iterations needed to achieve convergence and in terms of the mean residual ISI at convergence. By achieving convergence, we mean to achieve the final residual ISI level. These results correspond to the optimal fixed step sizes for both methods for each channel. Noiseless and noisy environments have been tested. The noisy case corresponds to a signal-to-noise ratio (SNR) of 10 dB. The proposed method requires a lower mean number of iterations to achieve convergence in both cases. Moreover, it presents a similar residual ISI (even slightly better) than CMA and SPDF.

In general, we have observed that in channels where CMA provides a fast convergence, SQD converges at the same rate. However, for channels where the CMA is slow, SQD usually provides a faster convergence. For instance, a 16-QAM modulation (using $\pm\{1, 3\}$ levels for in-phase and quadrature components) and the following channel:

$$H_1(z) = (0.2258 + 0.5161z^{-1} + 0.6452z^{-2} + 0.5161z^{-3}),$$

have been considered in the following example. A filter with $L_w = 21$ taps is employed for the equalizer. The taps are initialized using the tap-centered strategy. Fig. 3 compares the average results obtained in 100 Monte Carlo trials for a signal-to-noise ratio (SNR) of 30 dB. The following parameters have been employed: $\sigma = 15$ and step sizes (μ) of 1e-5, 1e-2, and 1e-4 for CMA, SPDF, and SQD, respectively. These step sizes are the largest ones that guarantee stable convergence for each method (obtained by cross-validation). For this channel, the proposed method clearly converges faster than SPDF and much faster than CMA.

B. *Decision-Directed Mode*

In this section, we will show that the proposed method can also provide a low residual error for multilevel signals. Reducing the kernel size, we have observed for large SNRs that the proposed method behaves almost exactly like DDE in terms of both convergence speed and residual error. However, the proposed method has demonstrated a lower noise sensitivity. For

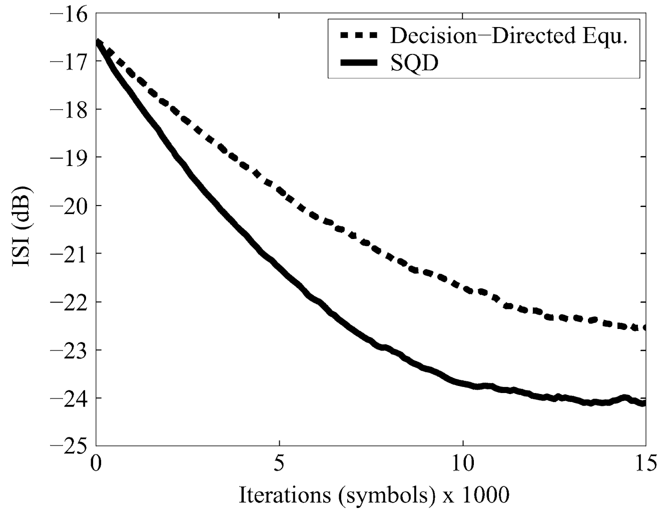


Fig. 4. DDE against the SQD for $H_2(z)$.

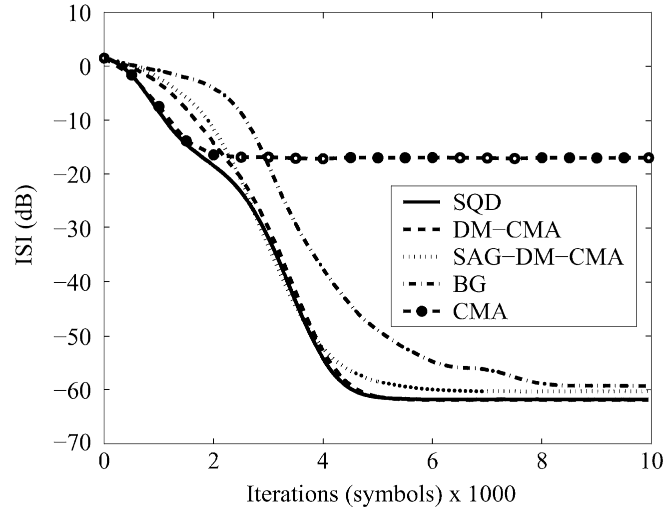


Fig. 5. Convergence curves for $H_2(z)$ without noise.

small SNRs, it converges faster than DDE. Fig. 4 illustrates this point. Noise has been added to the output of the channel

$$H_2(z) = 0.2258 + 0.5161z^{-1} + 0.6452z^{-2} - 0.5161z^{-3} \quad (29)$$

to obtain a SNR of 10 dB. After the initial blind convergence, DDE and SQD have been applied to reduce the residual ISI. The results of 100 experiments have been averaged, and results obtained with the optimum step size (the one resulting in the fastest stable convergence) are plotted. The proposed method again converges faster than DDE in this situation; furthermore, it reaches a lower residual ISI.

C. Soft Switch Between Blind and Decision-Directed Equalization

The proposed method has the ability to implement a soft transition between blind and decision-directed-like equalization using (26) and (27) to adaptively control the kernel size. The compensation factor $F(\sigma)$ plays an important role in this strategy because it allows exact gain identification during the whole process, independent of the current kernel size. $F(\sigma)$ can be easily precomputed for any modulation. In this case, it has been evaluated by means of a look-up table. Now, the performance of the proposed method is compared with the following methods: CMA, Benveniste-Goursat (labeled BG), the DM-CMA, which adaptively decides between CMA and RDE at each iteration, and its Stop-And-Go extension (SAG-DM-CMA).

To avoid the known sensitivity of the BG method to phase rotation, in this example, we used channel $H_2(z)$, which is a channel without phase rotation. The following parameters have been selected: $\alpha = 0.995$ and step sizes $5e-5, 2e-4, 2e-5, 6e-5$, and $3e-4$ for CMA, BG, DM-CMA, SAG-DM-CMA, and the proposed method, respectively (the optimal values in terms of convergence speed and stability obtained by cross-validation). E_1 has been initialized to start with a kernel size $\sigma_1 = 15$. In addition, $k_1 = 4$ and $k_2 = 1$ have been employed for BG (recommended values in [24] for digital TV channels). Fig. 5 compares the mean convergence curves in 100 Monte Carlo simulations for a noiseless case. The SQD is, along with CMA, the fastest method in the initial blind stage. After the initial convergence, CMA is not able to reduce the residual ISI, unlike the other methods. At this stage, the proposed method becomes similar to DM-CMA and SAG-DM-CMA, whereas the BG method is clearly the slowest one. On the other hand, all these methods provide an equivalent final accuracy.

For a noisy environment, the proposed method exhibits a clearly better behavior. Fig. 6 compares the convergence in 100 Monte Carlo simulations for a SNR of 30 dB and channel $H_2(z)$. The optimal step sizes in this case are $5e-5, 2e-4, 2e-5, 1.5e-5$, and $1e-4$ for CMA, BG, DM-CMA, SAG-DM-CMA, and the proposed method, respectively (optimal in terms of convergence speed and stability). In this case, with respect to the CMA, the proposed method exhibits a similar rate of convergence in the blind mode. Again, it obtains a lower residual error. With respect to the other methods, the SQD converges faster than all of them obtaining a lower residual error (of course, the other methods are capable of obtaining this residual ISI level but at the cost of even slower convergence).

This advantage becomes even more pronounced at lower SNRs. Fig. 7 shows the convergence curves for SNR = 15 dB. The optimal step sizes in this case are $2.5e-5, 1e-4, 2e-5, 1.5e-5$, and $1e-4$ for CMA, BG, DM-CMA, SAG-DM-CMA, and SQD, respectively. The advantage is more evident in this case, especially when compared with SAG-DM-CMA; this algorithm showed to be the more noise sensitive.

Finally, if a phase rotation is included, the performance of the BG method is clearly slowed down. This does not happen with the other methods in this comparison (including the proposed one), which are phase-rotation invariant. To avoid including this sensitivity in the comparison, the plotted results correspond to a channel without phase rotation. We want to remark that the improved performance of the proposed method has been observed in many different channels.

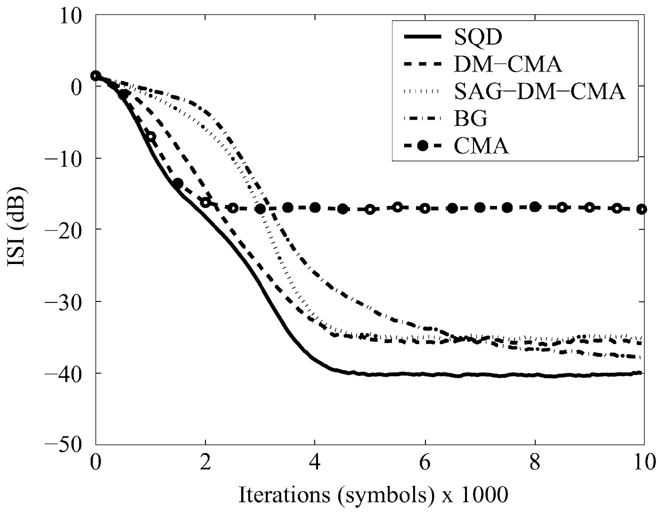


Fig. 6. Convergence curves for $H_2(z)$ with an SNR of 30 dB.

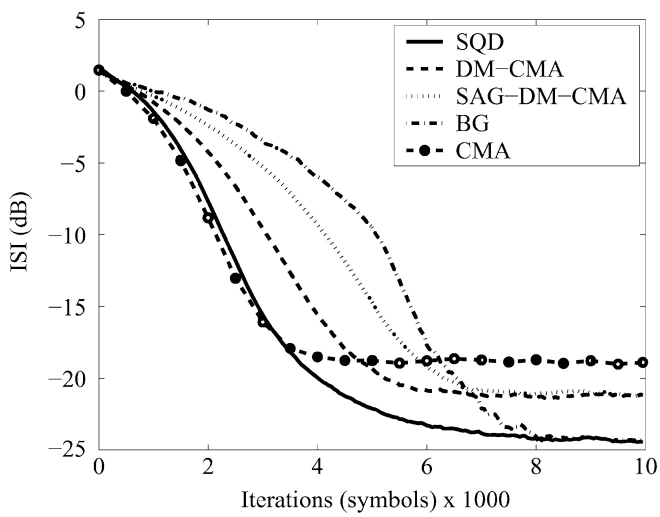


Fig. 7. Convergence curves for $H_2(z)$ with an SNR of 15 dB.

VI. Conclusion

A new method for blind equalization of multilevel modulations has been proposed. This method forces the pdf at the equalizer output to match that of the known constellation by employing the Parzen window method to estimate the data pdf. The kernel size of the Parzen method controls both the convergence speed and the accuracy of the final solution. To minimize the computational burden, which facilitates the online implementation, a stochastic gradient descent algorithm has been developed.

The proposed method has been compared with CMA, showing several advantages for multilevel signals. First, it converges faster than CMA by using a large kernel size until the eye of the constellation is opened. Second, after the initial convergence a small kernel size allows the reduction of the residual error, unlike CMA (this is the main reason to use dual mode switching techniques). In this case, it also outperforms decision-directed equalization in noisy environments. Consequently, the proposed method can be used as a dual mode technique, with the advantage of only requiring a change in the kernel size instead of a change in the algorithm. This feature makes the implementation of the dual mode switching much easier than in conventional CMA-based dual mode switching methods.

The adaptive kernel size produces a soft transition from blind to decision-directed equalization. The kernel size is progressively reduced as convergence is achieved, thus combining the convergence speed and the low residual error properties of the method. This approach was shown to behave better than the Benveniste-Goursat, DM-CMA, and SAG-DM-CMA methods. Moreover, the error measure (26) to control the kernel size endows the proposed method with the ability to recover from a drastic change in the channel, as compared to the above mentioned methods.

In the paper, we focused on $L = 1$ to work in a one sample basis. For $L > 1$, the proposed algorithm has a similar behavior but with a higher computational burden. However, the scaling-down of the constellation for large kernel sizes tends to disappear (for $L \geq 10$ is not further noticeable).

References

[1] B. Widrow and S. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
[2] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp. 1349–1387, 1985.
[3] S. Haykin, *Blind Deconvolution*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[4] Z. Ding and Y. Li, *Blind Equalization and Identification*. New York: Marcel Dekker, 2001.
[5] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds., *Signal Processing Advances in Wireless and Mobile Communications. Vol. I: Trends in Channel Estimation and Equalization*. Upper Saddle River, NJ: Prentice-Hall, 2001.
[6] S. Haykin, *Adaptive Filter Theory*, Fourth ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
[7] L. Tong and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.
[8] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind identification and equalization based on second-order statistics: A frequency domain approach," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 329–334, Jan. 1995.
[9] G. H. Im, D. D. Harman, G. Huang, A. V. Mandzik, M. H. Nguyen, and J. J. Werner, "51.84 Mb/s 16-CAP ATM LAN standard," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 4, pp. 620–632, May 1995.
[10] V. Oksman, Ed., *VDSL Draft Specification*, Jun. 1998, VDSL Coalition T1E1.4/98-045R1.
[11] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation," *IEEE Trans. Commun.*, vol. COM-23, pp. 679–682, Jun. 1975.
[12] D. Godard, "Self-recovering equalization and carrier tracking in two dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.
[13] J. R. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, pp. 349–372, Nov. 1983.
[14] D. L. Jones, "A normalized constant-modulus algorithm," in *Proc. 29th Asilomar Conf. Signals, Syst., Comput.*, vol. 1, Nov. 1995, pp. 694–697.
[15] I. Santamaría, C. Pantaleón, L. Vielva, and J. C. Principe, "A fast algorithm for adaptive blind equalization using order-α Renyi's entropy," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. III, Orlando, FL, USA, 2002, pp. 2657–2660.
[16] J. Principe, D. Xu, Q. Zhao, and J. W. Fisher III, "Learning from examples with information theoretic criteria," *J. VLSI Signal Process.*, vol. 26, pp. 61–77, 2000.
[17] J. Sala-Alvarez and G. Vazquez-Grau, "Stastistical reference criteria for adaptive signal processing in digital communications," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 14–31, Jan. 1997.
[18] T. Adali, X. Liao, and M. K. Sommez, "Conditional distribution learning with neural networks and its application to channel equalization," *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 1051–1064, Apr. 1997.

[19] I. Santamaría, C. Pantaleón, L. Vielva, and J. C. Principe, "Adaptive blind equalization through quadratic pdf matching," in *Proc. Eur. Signal Process. Conf.*, vol. II, Toulouse, France, Sept. 2002, pp. 289–292.

[20] M. Lázaro, I. Santamaría, C. Pantaleón, D. Erdogmus, K. E. H. II, and J. C. Principe, "Blind equalization by sampled pdf fitting," in *Proc. Fourth Int. Symp. Ind. Component Anal. Blind Equalization*, Nara, Japan, Sep. 2003, pp. 1041–1046.

[21] K. N. Oh and J. H. Park, "Property restoral approach to blind equalization of digital transmission channels," *IEEE Trans. Consum. Electron.*, vol. 43, no. 3, pp. 544–549, Aug. 1997.

[22] L. He and M. Amin, "A dual-mode technique for improved blind equalization for QAM signals," *IEEE Signal Process. Lett.*, vol. 10, no. 2, pp. 29–31, Feb. 2003.

[23] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.*, no. COM-32, pp. 871–883, Aug. 1984.

[24] "SMATV Reference Channel Model for Digital TV," DIGISMATV, DIGISMATV-NT-D1-031-HSA, Tech. Rep., 1995.

[25] V. Weerackody and S. A. Kassam, "Dual-mode type algorithm for blind equalization," *IEEE Trans. Commun.*, vol. 42, pp. 22–28, Jan. 1994.

[26] C.-H. Tseng and C.-B. Lin, "A stop-and-go dual-mode algorithm for blind equalization," in *Proc. Global Telecommun. Conf.*, vol. 2, Nov. 1996, pp. 1427–1431.

[27] M. Ready and R. Gooch, "Blind equalization based on radius directed adaptation," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Albuquerque, NM, 1990, pp. 1699–1702.

[28] E. Parzen, "On the estimation of a probability density function and the mode," *Ann. Math. Statist.*, vol. 33, pp. 1065–76, 1962.

[29] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications," *IEEE Trans. Autom. Control*, vol. AC-25, no. 3, pp. 385–399, Jun. 1980.

[30] E. Gokcay and J. C. Principe, "Information theoretic clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 2, pp. 158–171, 2002.

[31] M. Lázaro, I. Santamaría, C. Pantaleón, D. Erdogmus, and J. C. Principe, "Matched pdf-based blind equalization," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. IV, Hong Kong, Apr. 2003, pp. 297–300.

[32] D. Erdogmus and J. C. Principe, "Generalized information potential criterion for adaptive system training," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1035–1044, Sep. 2002.

Marcelino Lázaro (M'01) was born in Carriazo (Cantabria), Spain, in 1972. He received the Telecommunication Engineer degree and the Doctor (Ph.D.) degree from the Universidad de Cantabria, in 1996 and 2001 respectively.

From 1996 to 2002, he was with the Departamento de Ingenieria de Comunicaciones, Universidad de Cantabria. In 2003, he joined the Departamento de Teoria de la Senal y Comunicaciones, Universidad Carlos III, Madrid, Spain. His research interest includes digital signal processing and neural networks.

Ignacio Santamaría (M'96) was born in Vitoria, Spain, in 1967. He received the Telecommunication Engineer degree and the Doctor degree from the Universidad Politecnica de Madrid (UPM), Madrid, Spain, in 1991 and 1995, respectively.

In 1992, he joined the Departamento de Ingenieria de Comunicaciones, Universidad de Cantabria, Cantabria, Spain, where he is currently an Associate Professor. His research interests include digital signal processing, nonlinear systems, and neural networks.

Deniz Erdogmus (M'00) received the B.S. degree in electrical and electronics engineering and mathematics in 1997 and the M.S. degree in electrical and electronics engineering, with emphasis on systems and control, in 1999, all from the Middle East Technical University, Ankara, Turkey. He received the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, in 2002.

His current research interests include applications of information theory to adaptive systems, signal processing, communications, and control.

Dr. Erdogmus is a member of Tau Beta Pi and Eta Kappa Nu.

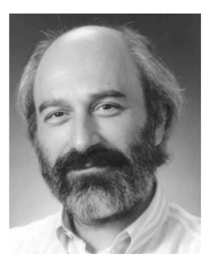
Kenneth E. Hild (M'03) received the B.S. degree in electrical engineering in 1992 and the M.S. degree in electrical engineering in 1996, both from the University of Oklahoma (OU), Norman. The emphasis during his graduate work at OU was in the areas of communications, signal processing, and controls. Currently, he is pursuing the Ph.D. degree at the University of Florida, Gainesville, where he is studying information theoretic learning and blind source separation.

From 1993 to 1995, he served in several different assistantships with OU, one of which involved conducting research for Seagate Technologies, Inc., Oklahoma City, OK. From 1995 to 1999, he worked full-time with Seagate, where he served as an advisory development engineer in the Advanced Concepts group.

Mr. Hild is a member of Tau Beta Pi and Eta Kappa Nu.

Carlos Pantaleón (M'95) was born in Badajoz, Spain, in 1966. He received the Telecommunication Engineer degree and the Ph.D. degree from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1990 and 1994, respectively.

In 1990, he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Cantabria, Spain, where he is currently an Associate Professor. His research interests include statistical signal processing, nonlinear systems, and neural networks.



Jose C. Principe (F'00) is a Distinguished Professor of Electrical and Computer Engineering and Biomedical Engineering at the University of Florida, Gainesville, where he teaches advanced signal processing, machine learning and artificial neural networks (ANN's) modeling. He is BellSouth Professor and the Founder and Director of the University of Florida Computational NeuroEngineering Laboratory (CNEL). His primary area of interest is processing of time-varying signals with adaptive neural models. The CNEL Lab has been studying signal and pattern recognition principles based on information theoretic criteria (entropy and mutual information).

Dr. Principe is a member of the ADCOM of the IEEE Signal Processing Society, Member of the Board of Governors of the International Neural Network Society, and Editor-in-Chief of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING. He is a member of the Advisory Board of the University of Florida Brain Institute and has more than 90 publications in refereed journals, ten book chapters, and 200 conference papers. He has directed 35 Ph.D. dissertations and 45 Master theses. He recently wrote an interactive electronic book entitled *Neural and Adaptive Systems: Fundamentals Through Simulation* (New York: Wiley).