Adaptive Bayesian Equalizer with Decision Feedback

Sheng Chen, *Member, IEEE,* Bernard Mulgrew, *Member, IEEE,* and Stephen McLaughlin

***Abstract*—A Bayesian solution is derived for digital communication channel equalization with decision feedback. This is an extension to the maximum *a posteriori* probability symbol-decision equalizer to include decision feedback. A novel scheme of utilizing decision feedback is proposed which not only improves equalization performance but also reduces computational complexity dramatically. It is shown that the Bayesian equalizer has an equivalent structure to the radial basis function network, the latter being a one-hidden-layer artificial neural network widely used in pattern classification and many other areas of signal processing. Two adaptive approaches are developed to realize the Bayesian solution. The maximum likelihood Viterbi algorithm and the conventional decision feedback equalizer are used as two benchmarks to assess the performance of the Bayesian decision feedback equalizer.**

I. Introduction

MANY digital communication channels are subject to intersymbol interference (ISI) and can be characterized by a finite impulse response (FIR) filter with an additive noise source [1]–[3]. At the receiver, the ISI must be compensated in order to reconstruct the transmitted symbols, and this is referred to as channel equalization. The digital communication scenario is illustrated in the baseband discrete time model depicted in Fig. 1.

The channel, which is a convolution of the transmitter filter, the transmission medium, and the receiver filter, is modeled as a FIR filter with a transfer function

$$A(z) = \sum_{i=0}^{n_a-1} a_i z^{-i} \tag{1}$$

where n_a is the length of the channel impulse response. The symbol sequence $s(k)$ and the channel taps a_i can be complex valued. In this study, channels and symbols are restricted to be real valued. This corresponds to the use of multilevel pulse amplitude modulation (M-ary PAM) [2] with a symbol constellation defined by

$$s_i = 2i - M - 1, \qquad 1 \le i \le M. \tag{2}$$

Concentration on the simpler real case allows us to highlight the basic principles and concepts. In particular, the case of binary symbols ($M = 2$) provides a very useful geometric visualization of equalization process. $s(k)$ is

Manuscript received October 3, 1991; revised November 3, 1992. The associate editor coordinating the review of this paper and approving it for publication was Prof. S. Y. Kung. This work was supported by the U.K. Science and Engineering Research Council. The work of S. McLaughlin was supported by the U.K. Royal Society.

The authors are with the Department of Electrical Engineering, University of Edinburgh, Edinburgh EH9 3JL, Scotland.

IEEE Log Number 9210119.

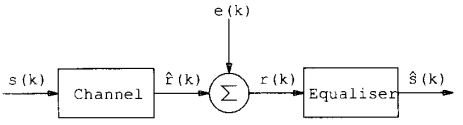


Fig. 1. Schematic of a digital communication system.

further assumed to be an equiprobable and independent sequence with the statistics

$$\left.\begin{aligned} E[s(k)] &= 0 \\ E[s(k_1)s(k_2)] &= \sigma_s^2 \delta(k_1 - k_2) \end{aligned}\right\} \tag{3}$$

where $E[\cdot]$ denotes the expectation operator, σ_s^2 is the symbol variance, and $\delta(k)$ is the Dirac delta function. The channel output is corrupted by an additive Gaussian white noise $e(k)$ specified by the statistics

$$\left.\begin{aligned} E[e(k)] &= 0 \\ E[e(k_1)e(k_2)] &= \sigma_e^2 \delta(k_1 - k_2) \end{aligned}\right\} \tag{4}$$

where σ_e^2 is the noise variance. $s(k)$ and $e(k)$ are assumed to be uncorrelated. Let the channel transfer function be normalized. Then the signal-to-noise ratio (SNR) is defined by

$$\text{SNR} = \sigma_s^2/\sigma_e^2. \tag{5}$$

The task of the equalizer is to reconstruct the transmitted symbols as accurately as possible based on the noisy channel observations $r(k)$. Various equalizers can be classified into two categories, namely, the symbol-decision equalizer and the sequence-estimation equalizer.

Two well-known symbol-decision structures are the transversal equalizer (TE) and the decision feedback equalizer (DFE). The structure of a generic TE is shown in Fig. 2. Essentially, the equalization process defined in Fig. 2 is one of using the information present in the observed channel output vector

$$\boldsymbol{r}(k) = [r(k) \cdots r(k - m + 1)]^T \tag{6}$$

to produce an estimate $\hat{s}(k - d)$ of $s(k - d)$, where the integers m and d are known as the equalizer feedforward order and the decision delay, respectively. Traditionally this is viewed as an inverse filtering in which the equalizer forms an approximation to the inverse of the distorting channel [1]. This consideration leads to the linear TE

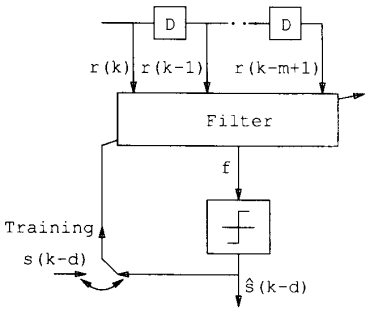


Fig. 2. Schematic of a generic transversal equalizer.

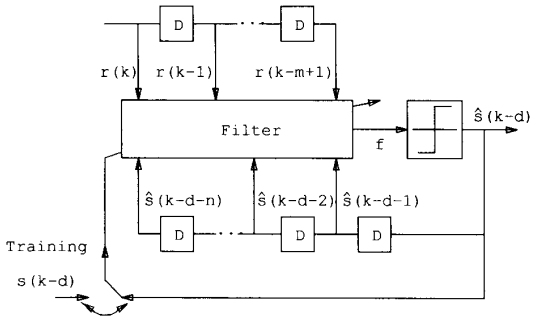


Fig. 3. Schematic of a generic decision feedback equalizer.

(LTE), which consists of a linear filter

$$f_l(\boldsymbol{r}(k)) = \boldsymbol{w}^T\boldsymbol{r}(k) = \sum_{i=1}^{m} w_i r(k - i + 1) \tag{7}$$

where $\boldsymbol{w} = [w_1 \cdots w_m]^T$ are the feedforward filter's coefficients. A decision slicer can then quantize the filter output $f_l(\boldsymbol{r}(k))$ into one of the symbol points defined in (2). A powerful technique to improve the equalization performance is to employ decision feedback. The structure of a generic DFE is illustrated in Fig. 3, where the integer n is known as the equalizer feedback order. This structure is obtained by expanding the inputs of the TE to include past detected symbols

$$\hat{\boldsymbol{s}}_f(k - d) = [\hat{s}(k - d - 1) \cdots \hat{s}(k - d - n)]^T. \tag{8}$$

The conventional DFE [1] is based on a linear filtering of this expanded equalizer input vector and is defined by

$$f_l(\boldsymbol{r}(k), \hat{\boldsymbol{s}}_f(k - d)) = \boldsymbol{w}^T\boldsymbol{r}(k) + \boldsymbol{b}^T\hat{\boldsymbol{s}}_f(k - d) \tag{9}$$

where $\boldsymbol{b} = [b_1 \cdots b_n]^T$ are the feedback filter's coefficients.

A different class of equalizers is based on the principle of maximum likelihood detection of the entire transmitted symbol sequence. The optimal solution for this class of equalizers is the maximum likelihood Viterbi algorithm (MLVA) (see [4], [5], and [2, Ch. 6]), which determines the estimated symbol sequence $\{\hat{s}(1), \hat{s}(2), \cdots, \hat{s}(k), \cdots\}$ by minimizing the cost

$$J = \sum_{k=1}^{\infty} \left(r(k) - \sum_{i=0}^{n_a-1} a_i \hat{s}(k - i) \right)^2. \tag{10}$$

The MLVA provides the lowest error rate attainable for any equalizer when the channel is known but it is computationally very expensive. In practice, the adaptive MLVA employs a channel estimator to estimate a channel model and, consequently, it will deviate from theoretical performance due to errors in the channel estimate. For stationary channels, a fairly accurate channel estimate can be obtained, and it is possible to employ a long decision delay. Under such conditions, loss of optimality is negligible. A new study [6] has shown that performance degradation of the adaptive MLVA becomes very serious for highly nonstationary channels, and this degradation is inherent in the structure of the MLVA.

The LTE does not achieve the full potential of the equalization process given in Fig. 2. The optimal solution for this structure can be derived using Bayes decision theory [7], [8], and this is known as the maximum *a posteriori* probability symbol-decision equalizer [9]. This Bayesian TE can be implemented using a variety of adaptive nonlinear structures based on artificial neural networks [10]–[13]. In particular, the radial basis function (RBF) network [14]–[17] can be shown to have an equivalent structure to the Bayesian solution and, therefore, it is an ideal means to realize the latter [12], [13]. The Bayesian TE offers significant performance gain over the LTE at the expense of a considerable increase in computational complexity. In general, the Bayesian TE cannot achieve the performance bound set by the MLVA since it is only a symbol-decision equalizer. However the Bayesian equalizer is very robust in a highly nonstationary environment [6]. Using an adaptive clustering algorithm, it is capable of compensating nonlinear channel distortion [13].

This present study derives the Bayesian solution for the symbol-decision structure with decision feedback defined in Fig. 3. A work published recently [18] describes a similar DFE scheme to our Bayesian DFE. We highlight how decision feedback can be utilized to reduce computational complexity dramatically. By adopting the Bayesian view, why decision feedback improves equalization performance has a clear geometric explanation. Issues of DFE design are discussed. Simulation is employed to compare the performance of the Bayesian DFE with those of the conventional DFE and the MLVA under the assumption of perfect knowledge of the channel. This provides theoretical performance bounds for the corresponding adaptive equalizers. The connections between the Bayesian equalizer and the RBF network are highlighted, and two adaptive schemes are presented for implementing the Bayesian equalizer. The first method identifies a channel model and uses this channel estimate to compute the channel states required in the Bayesian solution. The second method estimates these channel states directly based on clustering technique.

II. Bayesian Equalizer without Decision Feedback

The optimal Bayesian solution for the structure of Fig. 2 is summarized. This will provide insights into the issues such as why decision feedback improves performance and how it can be utilized to reduce computational complexity. The equalization process depicted in Fig. 2 can be viewed as a classification problem, which seeks to classify an observation vector $\boldsymbol{r}(k)$ into one of the symbol points s_i, $1 \le i \le M$. The transmitted symbols that influence the equalizer decision at k are

$$\boldsymbol{s}(k) = [s(k) \cdots s(k - m - n_a + 2)]^T. \quad (11)$$

This channel input sequence has $n_s = M^{m+n_a-1}$ combinations. In the absence of noise, the noise-free channel output vector

$$\hat{\boldsymbol{r}}(k) = [\hat{r}(k) \cdots \hat{r}(k - m + 1)]^T \quad (12)$$

has n_s corresponding outcomes. The set of these states, denoted as $R_{m,d}$, can be partitioned into M subsets according to the value of $s(k - d)$

$$R_{m,d} = \bigcup_{1 \le i \le M} R_{m,d}^{(i)} \quad (13)$$

where

$$R_{m,d}^{(i)} = \{\hat{\boldsymbol{r}}(k) | s(k - d) = s_i\}, \qquad 1 \le i \le M. \quad (14)$$

The number of states in each $R_{m,d}^{(i)}$ is $n_s^{(i)} = n_s/M$. Because of the additive noise, the observation vector $\boldsymbol{r}(k)$ is a stochastic vector. Each channel state is a conditional mean vector of $\boldsymbol{r}(k)$ given $\boldsymbol{s}(k)$.

Bayes decision theory [7], [8] provides the optimal solution to the general decision problem and is applicable here. Compute M Bayesian decision variables

$$\eta_i(k) = \sum_{j=1}^{n_s^{(i)}} p_j^{(i)} p_e(\boldsymbol{r}(k) - \boldsymbol{r}_j^{(i)}), \qquad 1 \le i \le M \quad (15)$$

where $\boldsymbol{r}_j^{(i)} \in R_{m,d}^{(i)}$, $p_j^{(i)}$ are the *a priori* probabilities of $\boldsymbol{r}_j^{(i)}$ and $p_e(\cdot)$ is the probability density function (pdf) of $\boldsymbol{e}(k) = [e(k) \cdots e(k - m + 1)]^T$. Each $\eta_i(k)$ is the conditional pdf of $\boldsymbol{r}(k)$ given $s(k - d) = s_i$. Since all the channel states can be assumed to be equiprobable, all the $p_j^{(i)}$ are equal, and the noise distribution is assumed to be Gaussian, (15) can be expressed explicitly as

$$\eta_i(k) = \sum_{j=1}^{n_s^{(i)}} \alpha \exp(-\|\boldsymbol{r}(k) - \boldsymbol{r}_j^{(i)}\|^2/2\sigma_e^2), \quad 1 \le i \le M \quad (16)$$

where α is $p_j^{(i)}(2\pi\sigma_e^2)^{-m/2}$ multiplied by an arbitrary positive constant. The minimum-error-probability decision is defined by

$$\hat{s}(k - d) = s_i^* \quad \text{if} \quad \eta_i^*(k) = \max\{\eta_i(k), 1 \le i \le M\}. \quad (17)$$

The Bayesian decision procedure effectively partitions the m-dimensional observation space into M decision regions. When $\boldsymbol{r}(k)$ appears in the ith region, the decision $\hat{s}(k - d) = s_i$ is made.

In particular, the Bayesian solution for $M = 2$ reveals some geometric insights of the equalization process. In the binary case, (17) can be rearranged as

$$\hat{s}(k - d) = \text{sgn}(f_B(\boldsymbol{r}(k))) = \text{sgn}(\eta_2(k) - \eta_1(k)) \quad (18)$$

where sgn $(\cdot)$ is the signum function and $f_B(\cdot)$ can be referred to as the Bayesian decision function. The set

$$\{\boldsymbol{r} | f_B(\boldsymbol{r}) = 0\} \quad (19)$$

defines the optimal decision boundary, which is a hypersurface in the observation space. Because the LTE can only realize a hyperplane decision boundary, a performance gap always exists between the Bayesian TE and the LTE. The previous simulation results [13] suggested that the Bayesian decision boundary is relatively insensitive to the noise variance. This is because a small change in σ_e^2 hardly causes any change in the decision boundary (19). This is important because in practice only an estimate of the noise variance is available.

It is interesting to compare the Bayesian TE with the RBF network. The RBF network [14]–[17] is a one-hidden-layer neural network. Each hidden node in the network has a radially symmetric response around a node parameter vector called a center, and an output node is a linear combiner. The overall response of the RBF network with m inputs and a single output node is defined as

$$f_r(\boldsymbol{r}(k)) = \sum_{i=1}^{n_h} w_i \phi(\|\boldsymbol{r}(k) - \boldsymbol{c}_i\|^2/\rho_i) \quad (20)$$

where n_h is the number of hidden nodes, $\boldsymbol{c}_i$ are the m-dimensional RBF centers, ρ_i are the positive scalars known as the widths, w_i are the weights, and $\phi(\cdot)$ is the node's nonlinearity. The Bayesian TE in the binary case has an identical structure to the above RBF network. Given the channel, it is known how to assign all the network parameters. The number of hidden nodes is equal to the number of channel states and RBF centers are placed at these states. The node's nonlinearity is chosen as $\phi(x) = \exp(-x)$ and all the widths are set to $\rho = 2\sigma_e^2$. The weights can be fixed to either α if the corresponding centers $\boldsymbol{c}_i$ belong to $R_{m,d}^{(2)}$ or $-\alpha$ if $\boldsymbol{c}_i$ belong to $R_{m,d}^{(1)}$. Each hidden node implements a conditional pdf for a given state and the overall network response realizes precisely the Bayesian decision function $f_B(\cdot)$. The Bayesian TE for the general M-ary PAM case can be realized by an extended RBF network depicted in Fig. 4. The network has two layers. The first layer consists of M subnets, each realizing a Bayesian variable in (16). The second layer is a MAXNET [19], which selects the maximum Bayesian decision variable and thus implements the Bayesian decision (17).

An example is given to illustrate the characteristics of the Bayesian TE. Consider the channel

$$A(z) = 0.4084 + 0.8164z^{-1} + 0.4084z^{-2} \quad (21)$$

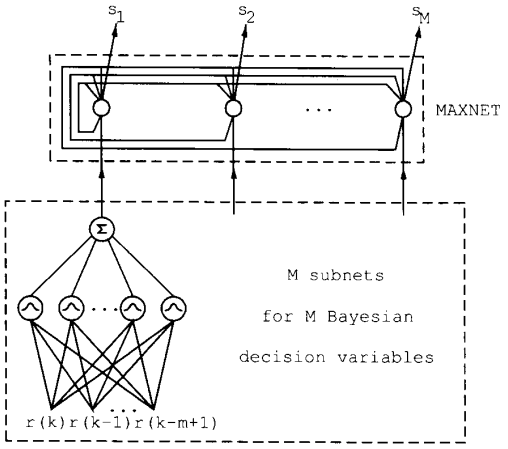


Fig. 4. Network implementation of Bayesian equalizer for M-ary PAM.

with 2-ary PAM symbols. For the purpose of a graphical display, assume $m = 2$ and let $d = 1$. All the combinations of $s(k)$ and the channel states are listed in Table I. The states of $R_{2,1}^{(2)}$ and $R_{2,1}^{(1)}$ are plotted in Fig. 5 using the "square" □ and "cross" × respectively. The Bayesian decision boundaries for two given SNR's are also plotted in Fig. 5. It is evident from Fig. 5 that the decision boundary of the Bayesian TE does not depend critically on the noise variance. This example represents a kind of worst scenario. The performance for this channel is mainly limited by the fact that a state of $R_{2,1}^{(2)}$ and a state of $R_{2,1}^{(1)}$ are coincident. In the next section it is shown how decision feedback overcomes such a difficulty.

The superior performance of the Bayesian TE over the LTE is at the cost of computational complexity. The dominant factor is that n_s can be quite large, and a comparison of (16) with (7) confirms that the Bayesian TE requires significantly more computation than the LTE. This problem can be alleviated using the parallel RBF network implementation. It is highly desirable if only a small subset of the channel states is required to compute the Bayesian decision at each sample k, and this can actually be achieved with decision feedback.

III. Bayesian Equalizer with Decision Feedback

The equalization process defined in Fig. 3 can obviously be considered as a classification problem, and Bayes decision theory can be applied to derive its optimal solution. First note that it is sufficient to employ a feedback order

$$n = n_a + m - 2 - d. \tag{22}$$

The feedback vector $\hat{s}_f(k-d)$ has $n_f = M^n$ states. Denote these feedback states as $s_{f,j}$, $1 \le j \le n_f$. A subset of the channel states $R_{m,d}^{(i)}$ defined in (14) can further be partitioned into n_f subsets according to the feedback state

$$R_{m,d}^{(i)} = \bigcup_{1 \le j \le n_f} R_{m,d,j}^{(i)} \tag{23}$$

TABLE I
Symbol and Channel States for Channel $0.4084 + 0.8164z^{-1} + 0.4084z^{-2}$ with Binary Symbols. $d = 1$, $m = 2$, and $n = 2$, Boldfaced Numbers: Given Feedback $s_{f,j} = [1 \quad 1]^T$

$s(k)$	$s(k-1)$	$s(k-2)$	$s(k-3)$	$\hat{r}(k)$	$\hat{r}(k-1)$
1	1	**1**	**1**	**1.6332**	**1.6332**
1	1	1	−1	1.6332	0.8164
1	1	−1	1	0.8164	0.0004
1	1	−1	−1	0.8164	−0.8164
−1	1	**1**	**1**	**0.8164**	**1.6332**
−1	1	1	−1	0.8164	0.8164
−1	1	−1	1	−0.0004	0.0004
−1	1	−1	−1	−0.0004	−0.8164
1	−1	**1**	**1**	**0.0004**	**0.8164**
1	−1	1	−1	0.0004	−0.0004
1	−1	−1	1	−0.8164	−0.8164
1	−1	−1	−1	−0.8164	−1.6332
−1	−1	**1**	**1**	**−0.8164**	**0.8164**
−1	−1	1	−1	−0.8164	−0.0004
−1	−1	−1	1	−1.6332	−0.8164
−1	−1	−1	−1	−1.6332	−1.6332

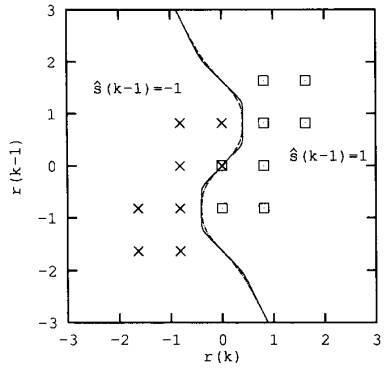


Fig. 5. Bayesian decision boundaries for channel $0.4084 + 0.8164z^{-1} + 0.4084z^{-2}$ with binary symbols. Dashed line: SNR = 10 dB, solid: SNR = 15 dB.

with

$$R_{m,d,j}^{(i)} = \{\hat{r}(k) | s(k-d) = s_i \cap \hat{s}_f(k-d) = s_{f,j}\},$$
$$1 \le j \le n_f. \tag{24}$$

The number of states in each $R_{m,d,j}^{(i)}$ is $n_{s,j}^{(i)} = n_s^{(i)}/n_f$. Under the assumption of correct feedback vector $\hat{s}_f(k-d)$, the M Bayesian decision variables given $\hat{s}_f(k-d) = s_{f,j}$ are

$$\eta_i(k | \hat{s}_f(k-d) = s_{f,j}) = \sum_{l=1}^{n_{s,j}^{(i)}} \alpha \exp\left(-\|\boldsymbol{r}(k) - \boldsymbol{r}_l^{(i)}\|^2 / 2\sigma_e^2\right),$$
$$1 \le i \le M \tag{25}$$

where $\boldsymbol{r}_l^{(i)} \in R_{m,d,j}^{(i)}$. The conditional Bayesian decision is obtained by substituting $\eta_i(k | \hat{s}_f(k-d) = s_{f,j})$ for $\eta_i(k)$ in (17).

It is interesting to see that the feedback vector is used to reduce the number of channel states needed in decision

making. Without feedback, all the n_s channel states are required to compute the M decision variables. As a result of feedback, only a fractional number of these states, $n_s/n_f = M^{d+1}$, are needed to compute the decision variables. The reduction factor in computational complexity owing to decision feedback is actually larger than n_f because a DFE requires a smaller feedforward order m than that which is required without decision feedback. The improvement offered by decision feedback has a simple geometric explanation. A group of the M conditional subsets $R^{(i)}_{m,d,j}$, $1 \le i \le M$, is related to any other group of the M conditional subsets $R^{(i)}_{m,d,l}$, $1 \le i \le M$, by some linear transformation. Specifically, a point in $R^{(i)}_{m,d,j}$ is related to a corresponding point in $R^{(i)}_{m,d,l}$ by a coordinate translation. This transformation is determined by $s_{f,j}$ and $s_{f,l}$ via the channel model (1), and it does not alter geometric distance. It is thus sufficient to consider just a group of the M conditional subsets when examining the symbol error rate of the Bayesian DFE. It is apparent that the minimum distance among the M conditional subsets $R^{(i)}_{m,d,j}$, $1 \le i \le M$, is larger than that among the M full subsets $R^{(i)}_{m,d}$, $1 \le i \le M$. This is why decision feedback improves performance.

These geometric properties of the DFE are illustrated using the channel and the equalizer structure specified in Table I. The subset states given feedback $s_{f,j} = [1 \quad 1]^T$ are emphasized in boldface in Table I. These subset states and the corresponding conditional Bayesian decision boundaries for two given SNR's are depicted in Fig. 6. The cases for the other three feedback states can be obtained by three coordinate translations of Fig. 6. Because of feedback, considerably fewer states are used in computing $f_B(\cdot)$. The degree of nonlinearity of the optimal decision boundary is milder and therefore easier to realize. Comparing Fig. 6 with Fig. 5, it is clear that decision feedback greatly increases the minimum distance between the two classes of channel states. In particular, there is no coincidence of channel states corresponding to the different classes. For this example, when the noise level is extremely high and is therefore the dominant factor over ISI, the conditional Bayesian decision boundary becomes almost linear. This suggests that, for this example, the conventional DFE will have a similar performance to the Bayesian DFE under extremely poor SNR conditions (<10 dB for binary symbols). For more realistic SNR levels, the Bayesian DFE always performs better than the conventional DFE because the latter can only form a linear boundary.

The structure of a DFE is specified by equalizer delay d, feedforward order m, and feedback order n. Given the channel characteristics, there are practical rules in designing a Bayesian DFE. The basic structure parameter of the Bayesian DFE is d, which specifies the number of states required for computing decision variables and thus determines computational complexity. Given d, it can be proven that $m = d + 1$ is sufficient. That is, a Bayesian DFE of $m = d + 1$ has the same performance as those of $m > d + 1$ (see the Appendix). Substituting this result

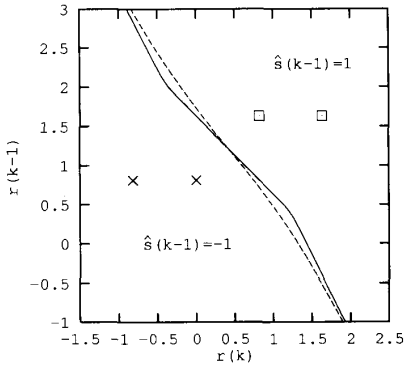


Fig. 6. Conditional Bayesian decision boundaries for channel $0.4084 + 0.8164z^{-1} + 0.4084z^{-2}$ with binary symbols and given feedback $[1 \quad 1]^T$. Dashed lines SNR = 5 dB, solid: SNR = 15 dB.

into (22) gives rise to the corresponding feedback order $n = n_a - 1$.

An implication of $m = d + 1$ is that, when equalizer delay is $d = 0$, $m = 1$ is sufficient for the Bayesian DFE. In this case, only the current observation $r(k)$ is used as the equalizer feedforward input. The channel states become scalars and each conditional subset given $\hat{s}_f(k) = s_{f,j}$, $R^{(i)}_{1,0,j}$, contains only a scalar $r^{(i)}_j$. The Bayesian DFE then reduces to a simple form, consisting of M decision variables

$$\eta_i(k|\hat{s}_f(k) = s_{f,j}) = (r(k) - r^{(i)}_j)^2, \qquad 1 \le i \le M \tag{26}$$

and a minimum detector

$$\hat{s}(k) = s^*_i \quad \text{if } \eta^*_i(k|\hat{s}_f(k) = s_{f,j}) = \min\{\eta_i(k|\hat{s}_f(k) = s_{f,j}), \quad 1 \le i \le M\}. \tag{27}$$

In particular, for the binary case, the conditional Bayesian decision function is

$$f_B(r(k)|\hat{s}_f(k) = s_{f,j}) = (r(k) - r^{(1)}_j)^2 - (r(k) - r^{(2)}_j)^2. \tag{28}$$

In the binary case, it can be verified that

$$|r^{(2)}_j - r^{(1)}_j| = 2|a_0|, \qquad 1 \le j \le n_f. \tag{29}$$

The channel is assumed to be normalized so that $|a_0| < 1$. The theoretical error probability of the simple Bayesian DFE (28) under the assumption of correct decisions being fed back can be shown to be

$$Q(|a_0|/\sigma_e) = \int_{|a_0|/\sigma_e}^{\infty} (2\pi)^{-1/2} \exp(-r^2/2)\, dr. \tag{30}$$

For the ideal channel having no ISI and a unit channel gain, the error probability in detecting binary symbols is known to be $Q(1/\sigma_e)$. Compared with this ideal case, the Bayesian DFE (28) assuming that correct decisions are fed back has a degradation of $10 \log_{10}(a_0^{-2})$ dB in SNR due to ISI.

A pragmatic rule in selecting equalizer delay, which is often adopted in the conventional DFE, is to set $d = n_a$

-1. For the Bayesian DFE, this rule of setting d has a rational explanation. In the extreme case of $d = 0$, the decision delay covers the first channel tap a_0 and the equalizer performance is shown to depend on energy of a_0. It can be imagined that in general the equalizer performance depends on energy of the channel taps a_0 to a_d covered by the decision delay. Increasing d improves performance but there is no need to use $d > n_a - 1$. The maximum delay $d = n_a - 1$ is sufficient to achieve all the performance potential. However, reducing d in turn reduces equalizer complexity, and the minimum complexity is achieved when $d = 0$. In practice, a compromise between complexity and performance can be obtained as follows. If most of the channel energy can be counted in the channel taps a_0 to a_u, the equalizer delay is set to $d = u$, where $u \leq n_a - 1$.

Computer simulation is used to compare the performance of the Bayesian DFE with those of the MLVA and the conventional DFE. The MLVA is implemented as a Viterbi algorithm with a fixed decision delay. The coefficients of the conventional DFE are set to their optimal values defined by the Wiener solution [20]. In the simulation, perfect knowledge of the channel is assumed, and the symbol error rates obtained are the theoretical bounds for the corresponding adaptive equalizers.

In the first example, binary symbols are used and the channel is defined by

$$A(z) = -0.2052 - 0.5131z^{-1} + 0.7183z^{-2} + 0.3695z^{-3} + 0.2052z^{-4}. \quad (31)$$

The Bayesian and conventional DFE's both had a structure of $d = 4$, $m = 5$ and $n = 4$, respectively, and the decision delay of the MLVA was set to 4 and 15, respectively. Fig. 7 compares the error rates of the three equalizers. The number of samples used for computing a symbol error rate was 10^6 to 10^7 depending on the SNR. Under severe noise conditions (SNR < 7 dB), the Bayesian DFE and the conventional DFE had a similar performance but the former generally performed better than the latter. For example, at an error probability level of 10^{-4}, the Bayesian DFE offered almost a 1.5 dB improvement in SNR over the conventional DFE. From Fig. 7, it can be seen that the Bayesian DFE achieved the same performance as the MLVA with the same delay of 4. The latter achieved a slightly better performance with a larger decision delay. The error rates of the Bayesian and conventional DFE's given in Fig. 7 were obtained with detected symbols being fed back. Clearly this is the more realistic scenario. The equalizer decisions cannot be guaranteed to be 100% correct, thus when an error is made, error propagation will result. The effects of error propagation can be investigated in simulation by comparing the error rates obtained using correct symbols and detected symbols as feedback, respectively. This is demonstrated in Fig. 8, where it is seen that error propagation resulted in a very small performance loss for this example.

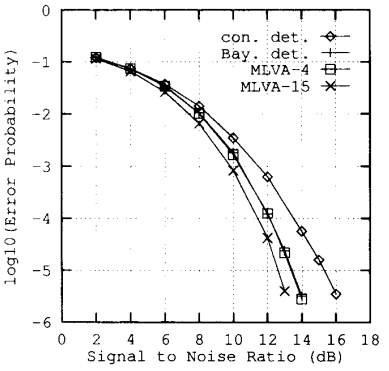


Fig. 7. Performance comparison for channel $-0.2052 - 0.5131z^{-1} + 0.7183z^{-2} + 0.3695z^{-3} + 0.2052z^{-4}$ and 2-ary PAM symbols. con./Bay. det.: conventional/Bayesian DFE with $d = 4$, $m = 5$, $n = 4$, and detected symbols being fed back. MLVA-d: MLVA with decision delay d.

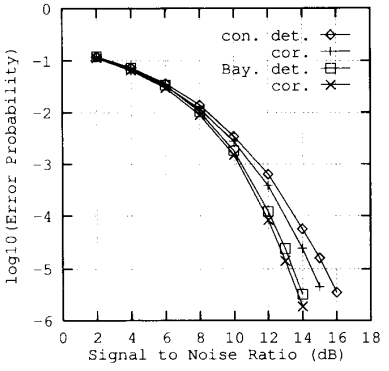


Fig. 8. Effects of error propagation for channel $-0.2052 - 0.5131z^{-1} + 0.7183z^{-2} + 0.3695z^{-3} + 0.2052z^{-4}$ and 2-ary PAM symbols. con./Bay.: conventional/Bayesian DFE with $d = 4$, $m = 5$, and $n = 4$. det./cor.: detected/correct symbols being fed back.

The second example employs 4-ary PAM symbols and the channel

$$A(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}. \quad (32)$$

The Bayesian and conventional DFE's were defined by $d = 2$, $m = 3$, and $n = 2$, and the decision delay of the MLVA was chosen as 2 and 10, respectively. Fig. 9 depicts the error rates of the three equalizers while Fig. 10 illustrates the effects of error propagation on the Bayesian and conventional DFE's. For this example, the Bayesian DFE had almost 3 dB improvement in SNR over the conventional DFE at an error probability level of 10^{-4}. The MLVA offered superior performance when a long decision delay was employed. From Fig. 10, it can be again seen that the performance loss due to error propagation was not very serious and that the Bayesian DFE was less affected by error propagation compared with the conventional DFE.

IV. Adaptive Implementation

The realization of the Bayesian equalization performance depends on knowing the channel states. Two adap-

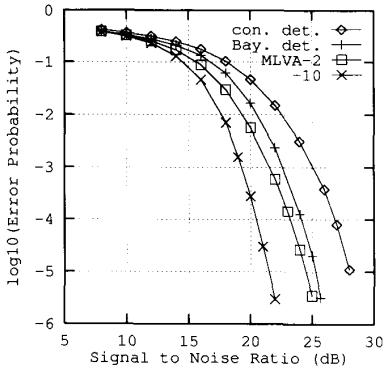


Fig. 9. Performance comparison for channel $0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$ and 4-ary PAM symbols. con./Bay. det.: conventional/Bayesian DFE with $d = 2$, $m = 3$, $n = 2$, and detected symbols being fed back. MLVA-d: MLVA with decision delay d.

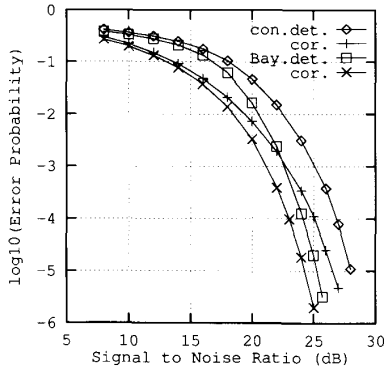


Fig. 10. Effects of error propagation for channel $0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$ and 4-ary PAM symbols. con./Bay.: conventional/Bayesian DFE with $d = 2$, $m = 3$, and $n = 2$. det./cor.: detected/correct symbols being fed back.

tive schemes are available to obtain these vector states [12], [13]. The first method estimates the channel model explicitly based on, for example, the least mean square (LMS) algorithm and uses the channel estimate to calculate the channel states. The second approach estimates the channel states directly based on a clustering algorithm.

The convergence properties and computational requirements of the LMS channel estimator are well known [3], [20]. Define the channel estimate as

$$\hat{\boldsymbol{a}}(k) = [\hat{a}_0(k) \cdots \hat{a}_{n_a-1}(k)]^T \tag{33}$$

and introduce the estimator input vector

$$\boldsymbol{s}_a(k) = [s(k) \cdots s(k - n_a + 1)]^T. \tag{34}$$

The LMS channel estimator is then given as

$$\left.\begin{aligned} \epsilon(k) &= r(k) - \hat{\boldsymbol{a}}^T(k-1)\boldsymbol{s}_a(k) \\ \hat{\boldsymbol{a}}(k) &= \hat{\boldsymbol{a}}(k-1) + \mu\epsilon(k)\boldsymbol{s}_a(k) \end{aligned}\right\} \tag{35}$$

where μ is an adaptive gain. During data transmission, a decision-directed and delayed version of (35) can be employed to track time-varying channels [1], [3]. Given the channel model $\hat{a}$, it is straightforward to compute the channel states required in the Bayesian DFE.

The number of combinations of $\boldsymbol{s}_a(k)$ is $n_{s,a} = M^{n_a}$. Denote these input states as $\boldsymbol{s}_{a,i}$ and the corresponding one-dimensional channel states as r_i, where $1 \le i \le n_{s,a}$. A scalar state r_i is just the conditional mean of the noisy observation $r(k)$ given $\boldsymbol{s}_a(k) = \boldsymbol{s}_{a,i}$, and a clustering procedure can be used to update the scalar channel states. At sample k, if $\boldsymbol{s}_a(k) = \boldsymbol{s}_{a,i}$, r_i is adjusted according to

$$r_i(k) = r_i(k-1) + \mu^*(r(k) - r_i(k-1)) \tag{36}$$

where the adaptive gain $0 \le \mu \le 1.0$. For time-varying channels, it is necessary to continuously update r_i during data transmission. This can be achieved using the following decision-directed version of (36). If $\hat{\boldsymbol{s}}_a(k-d) = \boldsymbol{s}_{a,i}$,

$$r_i(k) = r_i(k-1) + \mu^*(r(k-d) - r_i(k-1)) \tag{37}$$

where $\hat{\boldsymbol{s}}_a(k-d) = [\hat{s}(k-d) \cdots \hat{s}(k-d-n_a+1)]^T$. Once the scalar states r_i, $1 \le i \le n_{s,a}$, are obtained, it is straightforward to expand them into the set of the vector states $R_{m,d}$.

The computational load for the LMS channel estimator or the clustering scheme is very low. Additional processing is required to convert the channel estimate into the vector channel states or to expand the scalar states into the vector states. This is not costly especially when it is implemented in parallel. The computational load for computing the set of the conditional Bayesian decision variables (25) is given in Table II. From Table II, it is seen that the computational complexity is an order of M^{d+1}, which represents a significant reduction in processing complexity compared with M^{n_s} which is required for the Bayesian TE (16). In the special case of $d = 0$, the computational requirements reduce to the minimum. Assume that the Bayesian DFE with $d = 0$ employs the adaptive clustering scheme (36), and the conventional DFE with the same structure of $d = 0$, $m = 1$, and $n = n_a - 1$ uses the LMS algorithm for adjusting its coefficients. Table III compares the computational requirements of these two adaptive equalizers.

Because the number of the channel taps is smaller than that of the scalar channel states, the adaptive scheme based on a channel estimator requires a shorter training period than the clustering approach. Thus the former is better suited for time-variant channels. The clustering scheme does not assume the linear model (1) and is immune from nonlinear distortion. When significant nonlinear distortion is present in the channel, the estimated channel states based on a linear model will deviate from the true states, causing performance loss. Real-time identification of a nonlinear channel model is a difficult task. This problem also exists in the adaptive MLVA where an adaptive channel estimator is required. The clustering approach does not suffer from this difficulty. It always converges to the set of the true channel states regardless of whether the channel is linear or nonlinear.

TABLE II
COMPUTATIONAL LOAD OF CONDITIONAL BAYESIAN DECISION VARIABLES ($d > 0$)

Additions	Multiplications	exp $(-r)$ Evaluations
$2 \times (d+1) \times M^{d+1} - M$	$(d+2) \times M^{d+1}$	M^{d+1}

TABLE III
COMPARISON OF COMPUTATIONAL COMPLEXITY ($d = 0$)

Equalizer	Multiplications	Additions
Conventional DFE with LMS	$2 \times n_a + 1$	$2 \times n_a$
Bayesian DFE with clustering	$M + 1$	$M + 2$

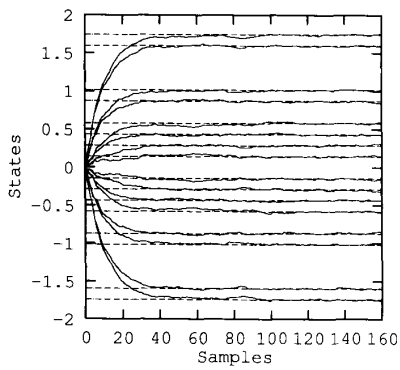


Fig. 11. Trajectories of estimated states based on channel-estimator scheme over ensembles of 50 different runs. Channel $0.7255 + 0.5804z^{-1} + 0.3627z^{-2} + 0.0724z^{-3}$ with binary symbols. SNR = 15 dB.

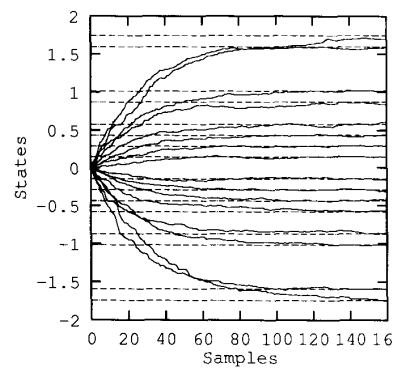


Fig. 12. Trajectories of estimated states based on clustering scheme over ensembles of 50 different runs. Channel $0.7255 + 0.5804z^{-1} + 0.3627z^{-2} + 0.0724z^{-3}$ with binary symbols. SNR = 15 dB.

The behaviors of the two adaptive schemes are illustrated using the channel

$$A(z) = 0.7255 + 0.5804z^{-1} + 0.3627z^{-2} + 0.0724z^{-3} \tag{38}$$

with binary symbols and SNR = 15 dB. The trajectories of the estimated scalar channel states obtained using the LMS channel estimator (35) with an adaptive gain 0.1 are plotted in Fig. 11. Fig. 12 shows the trajectories of the estimated scalar channel states obtained using the clustering algorithm (36) with an adaptive gain 0.6. In both cases, the dashed lines indicate the true values of the channel states and the trajectories were averaged over ensembles of 50 different runs.

V. CONCLUSIONS

A novel Bayesian DFE has been derived for digital communications channel equalization. It has been shown how decision feedback is utilized to improve equalizer performance as well as to reduce computational complexity. The selection of the three structure parameters, namely the equalizer delay, the feedforward order and the feedback order, has been discussed. The relationship between the Bayesian equalization solution and the RBF network has been highlighted. Two adaptive schemes have been presented for implementing the Bayesian DFE. The scheme based on a channel estimator requires a shorter training period while the clustering scheme offers greater immunity to nonlinear channel distortion.

In terms of computational complexity, the Bayesian DFE is generally more complex than the conventional DFE but is simpler than the MLVA. Theoretical performance of the Bayesian DFE assuming a known channel has been compared with those of the MLVA and the conventional DFE. Better performance of the Bayesian DFE over the conventional DFE has been demonstrated. The simulation results have suggested that the performance of the Bayesian DFE is comparable to that of the MLVA with an equivalent short decision delay. However, the MLVA provides the best theoretical performance when it employs a sufficiently large decision delay.

For highly nonstationary channels such as multipath mobile radio fading channels, a new study has demonstrated that the adaptive Bayesian DFE is actually superior over the adaptive MLVA. This has recently been submitted for publication [6]. This study also extends the current Bayesian DFE to the general case of complex-valued channels with quadrature amplitude modulation scheme.

APPENDIX
THE PROOFS OF MINIMAL FEEDFORWARD ORDER

Assume that $m > d + 1$. Let $\hat{s}_f(k - d) = s_{f,j}$ and consider $\boldsymbol{r}_l^{(i)} \in R_{m,d,j}^{(i)}$ where

$$\boldsymbol{r}_l^{(i)} = [r_{l,0}^{(i)} \cdots r_{l,m-1}^{(i)}]^T, \qquad 1 \le i \le M. \tag{A1}$$

The squared distance between $\boldsymbol{r}(k)$ and $\boldsymbol{r}_l^{(i)}$ is

$$\begin{aligned}\omega_{m,l}^{(i)}(k) &= \|\boldsymbol{r}(k) - \boldsymbol{r}_l^{(i)}\|^2 = \sum_{v=0}^{d} (r(k-v) - r_{l,v}^{(i)})^2 \\ &\quad + \sum_{v=d+1}^{m-1} (r(k-v) - r_{l,v}^{(i)})^2 \\ &= \omega_{d+1,l}^{(i)}(k) + \sum_{v=d+1}^{m-1} (r(k-v) - r_{l,v}^{(i)})^2, \\ &\quad 1 \le i \le M. \end{aligned} \tag{A2}$$

The feedback is assumed to be correct, that is,

$$\hat{\boldsymbol{s}}_f(k-d) = [s(k-d-1) \cdots s(k-d-n)]^T$$

where $n = n_a + m - 2 - d$. For any $\boldsymbol{r}_l^{(i)} \in R_{m,d,j}^{(i)}$ and $1 \le i \le M$,

$$r_{l,v}^{(i)} = \sum_{u=0}^{n_a-1} a_u s(k-v-u), \quad d+1 \le v \le m-1. \tag{A3}$$

For any $\boldsymbol{r}_l^{(i)} \in R_{m,d,j}^{(i)}$ and $1 \le i \le M$, introduce

$$\tilde{\omega}(k) = \sum_{v=d+1}^{m-1} (r(k-v) - r_{l,v}^{(i)})^2. \tag{A4}$$

Thus for any $\boldsymbol{r}_l^{(i)} \in R_{m,d,j}^{(i)}$ and $1 \le i \le M$,

$$\omega_{m,l}^{(i)}(k) = \omega_{d+1,l}^{(i)}(k) + \tilde{\omega}(k). \tag{A5}$$

The conditional Bayesian decision variables given $\hat{\boldsymbol{s}}_f(k-d) = \boldsymbol{s}_{f,j}$ are

$$\eta_i(k|\hat{\boldsymbol{s}}_f(k-d) = \boldsymbol{s}_{f,j}) = \sum_{l=1}^{n_{s,j}^{(i)}} \alpha \exp(-\omega_{m,l}^{(i)}(k)/\rho), \quad 1 \le i \le M \tag{A6}$$

where α is an arbitrary positive scalar, $\rho = 2\sigma_e^2$ and $n_{s,j}^{(i)}$ is the number of states in $R_{m,d,j}^{(i)}$. Substituting (A5) into (A6) yields

$$\begin{aligned}\eta_i(k|\hat{\boldsymbol{s}}_f(k-d) = \boldsymbol{s}_{f,j}) &= \sum_{l=1}^{n_{s,j}^{(i)}} \alpha \exp(-\tilde{\omega}(k)/\rho) \exp(-\omega_{d+1,l}^{(i)}(k)/\rho) \\ &= \sum_{l=1}^{\tilde{n}_{s,j}^{(i)}} \tilde{\alpha} \exp(-\omega_{d+1,l}^{(i)}(k)/\rho), \quad 1 \le i \le M\end{aligned} \tag{A7}$$

where $\tilde{n}_{s,j}^{(i)}$ is the number of states in $R_{d+1,d,j}^{(i)}$ and $\tilde{\alpha}$ is a positive scalar. This proves that the Bayesian DFE of a feedforward order $m = d + 1$ has the same conditional decision variables as those of feedforward order $m > d + 1$.

In the above proofs, the number of states in $R_{d+1,d,j}^{(i)}$ has first implicitly been multiplied by a factor of M^{m-d-1} so as to match the number of states in $R_{m,d,j}^{(i)}$, and then reduced to the original $\tilde{n}_{s,j}^{(i)}$. This is obviously allowed since $\tilde{\alpha}$ is an arbitrary positive scalar.

REFERENCES

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

[2] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1983.

[3] B. Mulgrew and C. F. N. Cowan, *Adaptive Filters and Equalizers*. Boston: Kluwer, 1988.

[4] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 3, pp. 363–378, 1972.

[5] F. R. Magee and J. G. Proakis, "Adaptive maximum-likelihood sequence estimation for digital signaling in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 1, pp. 120–124, 1973.

[6] S. Chen, S. McLaughlin, B. Mulgrew, and P. M. Grant, "Adaptive Bayesian decision feedback equalizer for dispersive mobile radio channels," *IEEE Trans. Commun.*, to be published, 1994.

[7] D. F. Specht, "Generation of polynomial discriminant functions for pattern recognition," *IEEE Trans. Electron. Comput.*, vol. EC-16, no. 3, pp. 308–319, 1967.

[8] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[9] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, vol. 58, no. 5, pp. 779–785, 1970.

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

[11] S. Chen, G. J. Gibson, and C. F. N. Cowan, "Adaptive channel equalization using a polynomial-perceptron structure," *Proc. Inst. Elec. Eng.*, pt. I, vol. 137, no. 5, pp. 257–264, 1990.

[12] S. Chen and B. Mulgrew, "Overcoming cochannel interference using an adaptive radial basis function equalizer," *EURASIP Signal Processing J.*, vol. 28, no. 1, pp. 91–107, 1992.

[13] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. Neural Networks*, July 1993.

[14] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Syst.*, vol. 2, pp. 321–355, 1988.

[15] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, no. 9, pp. 1481–1497, 1990.

[16] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks*, vol. 2, no. 2, pp. 302–309, 1991.

[17] J. Moody and C. J. Darken, "Fast-learning in networks of locally tuned processing units," *Neural Computation*, vol. 1, no. 2, pp. 281–294, 1989.

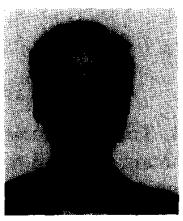
[18] D. Williamson, R. A. Kennedy, and G. W. Pulford, "Block decision feedback equalization," *IEEE Trans. Commun.*, vol. 40, no. 2, pp. 255–264, 1992.

[19] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Mag.*, vol. 4, 1987.

[20] S. Haykin, *Adaptive Filter Theory*, second ed. Englewood Cliffs, N.J: Prentice-Hall, 1991.

Sheng Chen (M'90) was born in Fujian Province, China, in 1957. He received the B.Sc. degree in control engineering from the East China Petroleum Institute in 1982 and the Ph.D. degree in control engineering from the City University at London in 1986.

From 1986 to 1989 he was a postdoctoral Research Associate in the Department of Control Engineering at the University of Sheffield. He joined the Signal Processing Group in the Department of Electrical Engineering at the University of Edinburgh in 1989 as a Research Fellow. His research interests are in modeling and identification of nonlinear systems, artificial neural networks, and adaptive signal processing for communications.

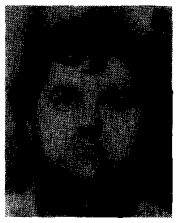


Bernard Mulgrew (M'88) received the Ph.D. degree in 1987 from the University of Edinburgh.

In 1979, he joined the Radar Systems Department at Ferranti Defence Systems, Edinburgh. From 1983 to 1986 he was a Research Associate in the Department of Electrical Engineering at the University of Edinburgh, studying the performance and design of adaptive filter algorithms. He is currently a lecturer in that department, teaching courses on signals and systems and signal processing. His research interests are in adaptive sig-

nal processing and estimation theory and in their application to radar and communications systems. He is a coauthor of two books on signal processing.

Dr. Mulgrew is an associate member of the IEE and a member of the Audio Engineering Society.



Stephen McLaughlin was born in Clydebank, Scotland, in 1960. He received the B.Sc. degree in electronics and electrical engineering from the University of Glasgow in 1981 and the Ph.D. degree from the University of Edinburgh in 1989.

From 1981 to 1984 he was a Development Engineer with Barr & Stroud Ltd., Glasgow, involved in the design and simulation of integrated thermal imaging and fire control systems. From 1984 to 1986 he worked on the design and development of high frequency data communication systems with MEL Ltd. In 1986 he joined the Department of Electrical Engineering at the University of Edinburgh as a Research Associate. There he studied the performance of linear adaptive algorithms in high noise and nonstationary environments. In 1988 he joined the teaching staff at Edinburgh, and in 1991 he was awarded a Royal Society University Research Fellowship to study nonlinear signal processing techniques. His research interests lie in the fields of adaptive signal processing and nonlinear dynamical systems theory and their applications to communication systems.

Dr. McLaughlin is an associate member of the Institute of Electrical Engineering and a member of the Institute of Electronics and Electrical Engineers.