

Asymptotic Bayesian Decision Feedback Equalizer Using a Set of Hyperplanes

Sheng Chen, *Senior Member, IEEE*, Bernard Mulgrew, *Member, IEEE*, and Lajos Hanzo

Abstract—We present a signal space partitioning technique for realizing the optimal Bayesian decision feedback equalizer (DFE). It is known that when the signal-to-noise ratio (SNR) tends to infinity, the decision boundary of the Bayesian DFE is asymptotically piecewise linear and consists of several hyperplanes. The proposed technique determines these hyperplanes explicitly and uses them to partition the observation signal space. The resulting equalizer is made up of a set of parallel linear discriminant functions and a Boolean mapper. Unlike the existing signal space partitioning technique of Kim and Moon, which involves complex combinatorial search and optimization in design, our design procedure is simple and straightforward, and guarantees to achieve the asymptotic Bayesian DFE.

Index Terms—Asymptotic decision boundary, Bayesian decision feedback equalizer, signal space partition.

I. INTRODUCTION

EQUALIZATION technique plays an ever-increasing role in combating distortion and interference in communication links [1], [2] and high-density data storage systems [3], [4]. The equalization topic is well researched, and a variety of solutions are available. The maximum *a posteriori* probability (MAP) sequence detector [5]–[7], although providing the lowest bit error rate (BER) attainable, finds little application in practice due to its computational complexity. The more popular sequence detector is the maximum likelihood sequence estimation (MLSE) based on the Viterbi algorithm [8], which is still computationally expensive and requires a sufficiently long decision delay to guarantee the optimal or near optimal performance. Many practical techniques employ a symbol-decision finite-memory structure with a fixed delay and are based on adaptive linear filter algorithms [1] with very low complexity. In particular, the conventional or linear-combiner DFE offers acceptable performance in many practical situations.

The BER gap between the MLSE and the conventional DFE for practical SNR conditions is large, and recent research in neural network equalizers [9]–[15] has attempted to fill in this gap and to strike a good balance between performance and complexity. A discussion on why it is worth considering the application of neural networks to equalization is given in [16]. It

is well known that the optimal solution for the symbol-decision finite-memory equalizer structure with a fixed delay is the Bayesian DFE [11], [17], and this equalization solution has an equivalent form to a radial basis function network. The computational complexity of the Bayesian DFE is, however, considerably more than that of the simple linear-combiner DFE. Various neural network equalizers can be regarded as realizations or approximations of this Bayesian solution with various degrees of complexity.

Geometrically, the conventional DFE partitions the observation space with a hyperplane. The optimal Bayesian decision boundary, however, is a hypersurface in the observation space [11]. It can be shown that asymptotically, as the SNR tends to infinity, the Bayesian hypersurface becomes piecewise linear and is made up of a set of hyperplanes [18]. In practice, at large rather than infinite SNR, the performance difference between Bayesian decision boundary and a piecewise linear approximation is negligible. This motivates research using multiple hyperplanes to partition the signal space. Kim and Moon [19], [20] recently developed a signal space partitioning technique for equalization. This technique basically determines a set of hyperplanes that separate clusters of noiseless channel states or signals. A combinatorial search and optimization process is carried out to find these hyperplanes, which requires extensive computational efforts. The convex regions associated with individual channel states are constructed by appropriately intersecting hyperplanes. The overall decision region is then formed from these convex regions.

Kim and Moon's design method results in an equalizer structure consisting of parallel linear discriminant functions and a many-to-one Boolean mapper. The decision complexity and performance of the detector are controlled during design by a specified minimum separating distance. Although it is possible to achieve the asymptotic Bayesian solution by an appropriate choice of the minimum separating distance, this is by no means guaranteed as the combinatorial search and optimization process does not necessarily produce the set of hyperplanes that form the asymptotic Bayesian decision boundary.

In this paper, we propose a much simpler alternative design. Specifically, the set of hyperplanes that form the asymptotic Bayesian decision boundary can straightforwardly be determined, thus avoiding the design complexity of Kim and Moon's method. Furthermore, our design guarantees to achieve, asymptotically, the optimal Bayesian solution. The decision complexity of the detector is solely determined by the channel geometric structure. Section II reviews the Bayesian DFE, which also serves to introduce the necessary notations

Manuscript received July 7, 1999; revised July 20, 2000. The associate editor coordinating the review of this paper and approving it for publication was Dr. Vikram Krishnamurthy.

S. Chen and L. Hanzo are with Department of Electronics and Computer Science, University of Southampton, U.K. (e-mail: sqc@ecs.soton.ac.uk; www-mobile.ecs.soton.ac.uk).

B. Mulgrew is with Department of Electronics and Electrical Engineering, University of Edinburgh, Edinburgh, U.K.

Publisher Item Identifier S 1053-587X(00)10164-3.

and definitions. The proposed new signal space partitioning technique is presented in Section III. Simulation results are provided in Section IV, and concluding remarks are given in Section V.

II. BAYESIAN DECISION FEEDBACK EQUALIZER

We will assume that the channel is modeled as a finite impulse response filter with an additive noise source. Specifically, the received signal at sample k is

$$r(k) = \bar{r}(k) + c(k) = \sum_{i=0}^{n_a-1} a_i s(k-i) + c(k) \quad (1)$$

where

- $\bar{r}(k)$ noiseless channel observation;
- n_a channel length;
- a_i channel tap weights;
- $\{c(k)\}$ Gaussian white noise sequence having zero mean and variance σ_c^2 ;
- $\{s(k)\}$ independently identically distributed symbol sequence and is uncorrelated with $c(k)$.

The SNR of the system is defined as

$$\text{SNR} = \left(\sum_{i=0}^{n_a-1} a_i^2 \right) \sigma_s^2 / \sigma_c^2 \quad (2)$$

where σ_s^2 is the symbol variance. In this study, for the purpose of easy geometrical visualization, we will assume that the channel is real-valued and that the transmitted symbol $s(k)$ is binary, taking value from the set $\{\pm 1\}$. There is no difficulty, however, to apply the current results to complex-valued channels and multilevel signaling schemes.

The structure of a generic DFE is depicted in Fig. 1. The equalization process defined in Fig. 1 uses the information present in the channel observation vector

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

and the past detected symbol vector

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

to produce an estimate $\hat{s}(k-d)$ of $s(k-d)$, where the integers d , m , and n are the decision delay and the feedforward and feedback orders, respectively. Without the loss of generality, we will choose $d = n_a - 1$, $m = n_a$ and $n = n_a - 1$ as this choice of the DFE structure parameters is sufficient to guarantee the linear separability (see Proposition 1 in this section). The observation vector (3) can be arranged as

$$\mathbf{r}(k) = F\mathbf{s}(k) + \mathbf{e}(k) \quad (5)$$

where $\mathbf{e}(k) = [c(k), \dots, c(k-m+1)]^T$, $\mathbf{s}(k) = [s_f^T(k) s_b^T(k)]^T$ with

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

and the $m \times (d+1+n)$ channel impulse response matrix F has the form

$$F = [F_1 \ F_2] \quad (7)$$

with the $m \times (d+1)$ matrix F_1 and $m \times n$ matrix F_2 defined by

$$F_1 = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n_a-1} \\ 0 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1 \\ 0 & \cdots & 0 & a_0 \end{bmatrix} \quad (8)$$

and

$$F_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{n_a-1} & 0 & \ddots & \vdots \\ a_{n_a-2} & a_{n_a-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \cdots & a_{n_a-2} & a_{n_a-1} \end{bmatrix}. \quad (9)$$

Past decisions on the symbols $s(k-d-i)$, $1 \leq i \leq n$ are used to cancel intersymbol interference terms from observation samples. In this process, past decisions $\hat{s}(k-d-i)$, $1 \leq i \leq n$ are assumed to be correct. Under this assumption, $\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k)$, and

$$\mathbf{r}(k) = F_1 \mathbf{s}_f(k) + F_2 \hat{\mathbf{s}}_b(k) + \mathbf{e}(k). \quad (10)$$

Thus, the decision feedback can be viewed as a translation of the original space $\mathbf{r}(k)$ into a new space $\mathbf{r}'(k)$:

$$\mathbf{r}'(k) \triangleq \mathbf{r}(k) - F_2 \hat{\mathbf{s}}_b(k). \quad (11)$$

Let $z^{-1}\{\cdot\}$ be the unit delay operator. The elements of $\mathbf{r}'(k)$ can be computed recursively according to

$$\begin{aligned} r'(k-i) &= z^{-1}\{r'(k-i+1)\} - a_{n_a-i} \hat{s}(k-d-1) \\ &\quad i = m-1, \dots, 2, 1 \\ r'(k) &= r(k). \end{aligned} \quad (12)$$

The DFE structure of Fig. 1 is therefore equivalent to the equalization structure of Fig. 2.

Let the $N_f = 2^{d+1}$ sequences or states of $\mathbf{s}_f(k)$ be $\mathbf{s}_{f,j}$, $1 \leq j \leq N_f$. Define the set of the noiseless channel states in the translated space by

$$R \triangleq \{\mathbf{r}_j = F_1 \mathbf{s}_{f,j}, 1 \leq j \leq N_f\}. \quad (13)$$

This set can be partitioned into two subsets conditioned on $s(k-d) = \pm 1$:

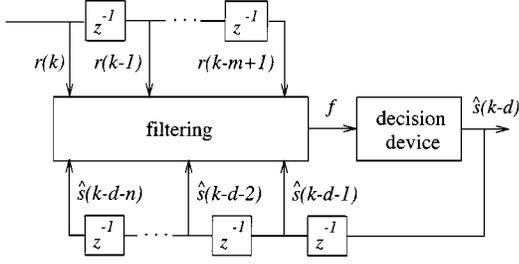
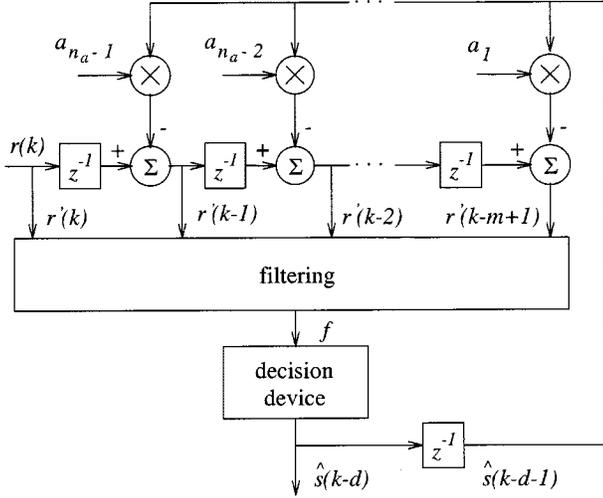
$$R^{(\pm)} \triangleq \{\mathbf{r}_j \in R: s(k-d) = \pm 1\}. \quad (14)$$

We have the following linear separability property.

Proposition 1: $R^{(+)}$ and $R^{(-)}$ are linearly separable.

The proof is straightforward. Choose the weights of a hyperplane $H(\mathbf{r}) = \mathbf{w}^T \mathbf{r} = 0$ to be

$$\mathbf{w}^T = \left[0, \dots, 0, \frac{1}{a_0} \right]. \quad (15)$$


 Fig. 1. Generic decision feedback equalizer. f is the decision variable.

 Fig. 2. Translated decision feedback equalizer. f is the decision variable.

For any $\mathbf{r}^{(+)} \in R^{(+)}$ and $\mathbf{r}^{(-)} \in R^{(-)}$, we have $\mathbf{w}^T \mathbf{r}^{(+)} = 1 > 0$ and $\mathbf{w}^T \mathbf{r}^{(-)} = -1 < 0$.

Proposition 1 states a well-known fact that it is always possible to construct a single hyperplane to correctly separate opposite-class channel states. What is interesting is that the weight vector (15) is the limit of the minimum mean square error (MMSE) solution for the conventional DFE with $\text{SNR} \rightarrow \infty$ [21]. This implies that the MMSE solution does not achieve the full performance potential of the linear-combiner DFE. How to construct a single optimal hyperplane to realize the minimum BER linear-combiner DFE has been addressed in [21] and [22].

The true optimal solution for the equalization structure of Fig. 2, however, cannot be realized using a single hyperplane. Assuming equiprobable channel states, the optimal solution for the equalization structure of Fig. 2 is given by the following Bayesian decision function [11], [15]:

$$f_B(\mathbf{r}'(k)) = \sum_{\mathbf{r}_j^{(+)} \in R^{(+)}} \exp\left(-\frac{\|\mathbf{r}'(k) - \mathbf{r}_j^{(+)}\|^2}{2\sigma_e^2}\right) - \sum_{\mathbf{r}_j^{(-)} \in R^{(-)}} \exp\left(-\frac{\|\mathbf{r}'(k) - \mathbf{r}_j^{(-)}\|^2}{2\sigma_e^2}\right) \quad (16)$$

with the minimum-error-probability decision defined by

$$\hat{s}(k-d) = \begin{cases} +1, & f_B(\mathbf{r}'(k)) \geq 0 \\ -1, & f_B(\mathbf{r}'(k)) < 0. \end{cases} \quad (17)$$

The decision boundary of this Bayesian DFE

$$\mathcal{D}_B \triangleq \{\mathbf{r}: f_B(\mathbf{r}) = 0\} \quad (18)$$

is generally a hypersurface. Before describing the asymptotic Bayesian decision boundary for $\text{SNR} \rightarrow \infty$ (or $\sigma_e^2 \rightarrow 0$), we introduce the following definition.

Definition 1: A pair of opposite-class states ($\mathbf{r}^{(+)}$, $\mathbf{r}^{(-)}$) is said to be *dominant* if $\forall \mathbf{r}_j \in R$, $\mathbf{r}_j \neq \mathbf{r}^{(+)}$, $\mathbf{r}_j \neq \mathbf{r}^{(-)}$:

$$\|\mathbf{r}_j - \mathbf{r}_0\|^2 > \|\mathbf{r}^{(+)} - \mathbf{r}_0\|^2 \quad (19)$$

where

$$\mathbf{r}_0 = \frac{\mathbf{r}^{(+)} + \mathbf{r}^{(-)}}{2}. \quad (20)$$

Proposition 2: Asymptotically, the decision boundary \mathcal{D}_B is piecewise linear and made up of a set of hyperplanes. Each of these hyperplanes is defined by a pair of dominant channel states, and the hyperplane is orthogonal to the line connecting the pair of dominant states and passes through the midpoint of the line.

Proof: Consider $f_B(\mathbf{r}) = 0$. As $\sigma_e^2 \rightarrow 0$, Itis [18] has shown that a necessary condition for a point $\mathbf{r}_B \in \mathcal{D}_B$ is

$$\mathbf{r}_B = \frac{\mathbf{r}^{(+)} + \mathbf{r}^{(-)}}{2} + \left[\frac{\mathbf{r}^{(+)} - \mathbf{r}^{(-)}}{2} \right]^\perp \quad (21)$$

where \mathbf{x}^\perp denotes an arbitrary vector in the subspace orthogonal to \mathbf{x} , $\mathbf{r}^{(+)}$, and $\mathbf{r}^{(-)}$ are a pair of dominant states; the sufficient conditions for $\mathbf{r}_B \in \mathcal{D}_B$ are

$$\|\mathbf{r}_B - \mathbf{r}^{(+)}\|^2 < \|\mathbf{r}_B - \mathbf{r}_i\|^2, \quad \forall \mathbf{r}_i \in R^{(+)}, \quad \mathbf{r}_i \neq \mathbf{r}^{(+)} \quad (22)$$

$$\|\mathbf{r}_B - \mathbf{r}^{(-)}\|^2 < \|\mathbf{r}_B - \mathbf{r}_j\|^2, \quad \forall \mathbf{r}_j \in R^{(-)}, \quad \mathbf{r}_j \neq \mathbf{r}^{(-)} \quad (23)$$

$$\|\mathbf{r}_B - \mathbf{r}^{(+)}\|^2 = \|\mathbf{r}_B - \mathbf{r}^{(-)}\|^2. \quad (24)$$

Proposition 2 follows as a direct consequence.

Remark 1: The section of the decision boundary \mathcal{D}_B within the ‘‘influencing domain’’ of a dominant pair ($\mathbf{r}^{(+)}$, $\mathbf{r}^{(-)}$), as defined by (22) and (23), is a section of the hyperplane that passes through the midpoint of the line connecting the pair, as indicated in (21) and (24). Although the boundary described in Proposition 2 is only exact in the limit case of infinite SNR, our empirical experience, and others such as [18], have shown that for the SNRs on the order of 10–20 dB, the true Bayesian boundary \mathcal{D}_B is closely approximated by the asymptotic multiple hyperplane form, and the two forms are often indistinguishable.

Remark 2: It should be emphasized that the pairs ($\mathbf{r}^{(+)}$, $\mathbf{r}^{(-)}$) needed to define the boundary hyperplanes are only a subset of all the possible signal states R . For the generic channel $\mathbf{a} = [a_0, a_1, \dots, a_{n_a-1}]^T$, it is impossible to determine a theoretic bound for the number of the dominant pairs L . Empirically, we have found that usually, $L \ll N_f$ for $n_a > 2$. In addition, L always appears to be odd. For the two-tap channel with the DFE structure defined by $m = 2$, $d = 1$, and $n = 1$, it can easily be shown that the asymptotic Bayesian boundary consists of either a single hyperplane or three hyperplanes, and there is no other possibility. Two typical scenarios for $n_a = 2$ are depicted in Fig. 3.

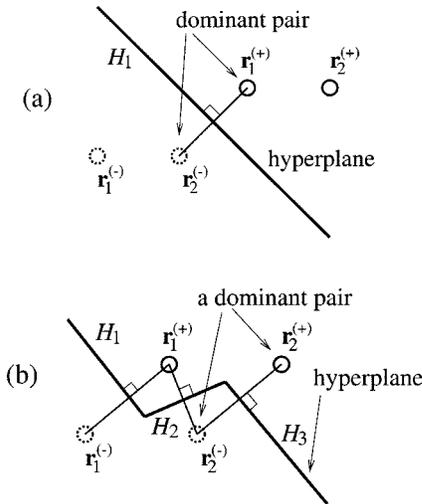


Fig. 3. Two typical cases of the asymptotic Bayesian decision boundary for channel $\mathbf{a} = [a_0, a_1]^T$.

III. ASYMPTOTIC BAYESIAN DFE USING HYPERPLANES

In a previous work [23], the Bayesian solution is approximated by only using the set of L dominant state pairs, which determine the asymptotic Bayesian decision boundary, in the computation of the Bayesian decision variable (16). In this study, we consider using the multiple-hyperplane detector structure of Kim and Moon [19], [20] (see Fig. 4) to realize the asymptotic Bayesian DFE. Our design procedure for constructing the detector, which is very different from that of Kim and Moon, is as follows:

- Step 1) Select all L pairs of dominant channel states from the state set R . For each pair, compute a hyperplane that separates these two opposite-class states.
- Step 2) A Boolean logic function is obtained to make a final decision based on the location of the observation vector relative to each hyperplane. This is achieved by first defining a convex region associated with each state in a given class and then forming a union of these regions.

From the proof of Proposition 2, it is easily seen that pairs of dominant states that define the asymptotic boundary can be selected using the following algorithm:

```

L = 0;
FOR  $\mathbf{r}_i^{(+)} \in R^{(+)}$ 
  FOR  $\mathbf{r}_j^{(-)} \in R^{(-)}$ 
     $\mathbf{x} = \frac{\mathbf{r}_i^{(+)} + \mathbf{r}_j^{(-)}}{2}$ ;  $\eta = \|\mathbf{r}_i^{(+)} - \mathbf{x}\|^2$ ;
    IF ( $\|\mathbf{r}_l^{(+)} - \mathbf{x}\|^2 > \eta, \forall \mathbf{r}_l^{(+)} \in R^{(+)}, l \neq i$ ) AND
       ( $\|\mathbf{r}_l^{(-)} - \mathbf{x}\|^2 > \eta, \forall \mathbf{r}_l^{(-)} \in R^{(-)}, l \neq j$ )
      L+ = 1;
       $R_{\text{Asym}} \leftarrow (\mathbf{r}_L^{(+)}, \mathbf{r}_L^{(-)}) \triangleq (\mathbf{r}_i^{(+)}, \mathbf{r}_j^{(-)});$ 
    END IF
  NEXT  $\mathbf{r}_j^{(-)}$ 
NEXT  $\mathbf{r}_i^{(+)}$ 

```

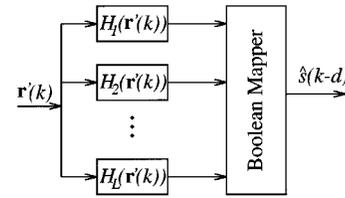


Fig. 4. Asymptotic Bayesian DFE using a set of hyperplanes.

Each pair $(\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)}) \in R_{\text{Asym}}$ determines a hyperplane

$$H_l(\mathbf{r}) = \mathbf{w}_l^T \mathbf{r} + b_l = 0 \quad (25)$$

that is a part of the asymptotic Bayesian decision boundary. The weight vector \mathbf{w}_l and bias b_l of the hyperplane can be computed straightforwardly as

$$\mathbf{w}_l = \frac{2(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)})}{\|\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\|^2} \quad (26)$$

and

$$b_l = -\frac{(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)})^T (\mathbf{r}_l^{(+)} + \mathbf{r}_l^{(-)})}{\|\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\|^2}. \quad (27)$$

Notice that we have applied the theory of support vector machines [22], [24], [25] in determining the hyperplane H_l with $(\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)})$ as its two support vectors. That is, the hyperplane defined by (26) and (27) is a *canonical* hyperplane [24] having the property $H_l(\mathbf{r}_l^{(+)}) = 1$ and $H_l(\mathbf{r}_l^{(-)}) = -1$. Let us introduce the definition of sufficient separability.

Definition 2: A state $\mathbf{r}_i \in R$ is said to be *sufficiently separable* by the hyperplane H_l if H_l can separate \mathbf{r}_i correctly with a “canonical distance” $|\mathbf{w}_l^T \mathbf{r}_i + b_l| \geq 1$.

Notice that $\mathbf{r}_i \in R^{(+)}$ is sufficiently separable by H_l if and only if $\mathbf{w}_l^T \mathbf{r}_i + b_l \geq 1$. Similarly, $\mathbf{r}_i \in R^{(-)}$ is sufficiently separable by H_l if and only if $\mathbf{w}_l^T \mathbf{r}_i + b_l \leq -1$. All the states in R are tested to see they can be separated sufficiently by H_l , $1 \leq l \leq L$. This generates the following matrix:

	$\mathbf{r}_1^{(-)}$	$\mathbf{r}_2^{(-)}$	\cdots	$\mathbf{r}_{N_s}^{(-)}$	$\mathbf{r}_1^{(+)}$	\cdots	$\mathbf{r}_{N_s}^{(+)}$
	\mathbf{r}_1	\mathbf{r}_2	\cdots	\mathbf{r}_{N_s}	\mathbf{r}_{N_s+1}	\cdots	\mathbf{r}_{N_f}
H_1	$h_{1,1}$	$h_{1,2}$	\cdots	h_{1,N_s}	h_{1,N_s+1}	\cdots	h_{1,N_f}
H_2	$h_{2,1}$	$h_{2,2}$	\cdots	h_{2,N_s}	h_{2,N_s+1}	\cdots	h_{2,N_f}
\vdots	\vdots	\vdots	\cdots	\vdots	\vdots	\cdots	\vdots
H_L	$h_{L,1}$	$h_{L,2}$	\cdots	h_{L,N_s}	h_{L,N_s+1}	\cdots	h_{L,N_f}

where $h_{l,i} \in \{0, 1\}$, $N_s = N_f/2$, and states are numbered such that $\mathbf{r}_i \in R^{(-)}$ for $1 \leq i \leq N_s$ and $\mathbf{r}_i \in R^{(+)}$ for $1 + N_s \leq i \leq N_f$. The rule in generating this matrix is as follows: If \mathbf{r}_i can *sufficiently* be separated by H_l , $h_{l,i} = 1$; otherwise, $h_{l,i} = 0$. Notice that in every row, there are at least two nonzero elements (associated with a dominant pair), and in every column, there is at least one nonzero element.

The half-space $\mathcal{H}_l^{(+)}$ defined by a hyperplane H_l is

$$\mathcal{H}_l^{(+)} \triangleq \{\mathbf{r}: H_l(\mathbf{r}) \geq 0\}. \quad (28)$$

To construct a convex region $\mathcal{R}_i^{(+)}$ corresponding to a state $\mathbf{r}_i^{(+)} \in R^{(+)}$, select those hyperplanes that can sufficiently separate $\mathbf{r}_i^{(+)}$ and denote

$$G_i^{(+)} \triangleq \{l: h_{l, i+N_s} = 1\}. \quad (29)$$

Then, $\mathcal{R}_i^{(+)}$ is obtained by the intersection of all the $\mathcal{H}_j^{(+)}$ with $j \in G_i^{(+)}$

$$\mathcal{R}_i^{(+)} = \bigcap_{j \in G_i^{(+)}} \mathcal{H}_j^{(+)}. \quad (30)$$

In fact, it may not be necessary to use every hyperplanes defined in $G_i^{(+)}$ to construct $\mathcal{R}_i^{(+)}$. A subset of these hyperplanes will be enough in the construction of $\mathcal{R}_i^{(+)}$, provided that every state in $R^{(-)}$ can sufficiently be separated by at least one hyperplane in the subset. The overall decision region $\mathcal{R}^{(+)}$ associated with the decision $\hat{s}(k-d) = 1$ is simply formed as the union of all the $\mathcal{R}_i^{(+)}$

$$\mathcal{R}^{(+)} = \bigcup_{i=1}^{N_s} \mathcal{R}_i^{(+)}. \quad (31)$$

For an illustration, consider Fig. 3(b). It is easily seen that $H_1(\mathbf{r}_1^{(-)}) = -1$, $H_1(\mathbf{r}_2^{(-)}) > 0$, $H_1(\mathbf{r}_1^{(+)}) = 1$, and $H_1(\mathbf{r}_2^{(+)}) > 1$. Thus, $h_{11} = h_{13} = h_{14} = 1$, and $h_{12} = 0$. The complete matrix of the separating hyperplanes and channel states is given in Table I. From Table I, it can be seen that the state $\mathbf{r}_1^{(+)}$ requires two hyperplanes H_1 and H_2 to be separated from the states in $R^{(-)}$. The region $\mathcal{R}_1^{(+)}$ for $\mathbf{r}_1^{(+)}$ is the intersection of the two half spaces $\mathcal{H}_1^{(+)}$ and $\mathcal{H}_2^{(+)}$ defined by H_1 and H_2 . The hyperplane H_3 can separate $\mathbf{r}_2^{(+)}$ from the states in $R^{(-)}$. The region $\mathcal{R}_2^{(+)}$ for $\mathbf{r}_2^{(+)}$ is thus the half-space $\mathcal{H}_3^{(+)}$ defined by H_3 . The overall decision region $\mathcal{R}^{(+)}$ is the union of $\mathcal{R}_1^{(+)}$ and $\mathcal{R}_2^{(+)}$.

The Boolean logic function for the detector depicted in Fig. 4 is now completely defined. Let a threshold detector output $\beta_j(\mathbf{r}'(k))$ for a linear discriminant function $H_j(\mathbf{r}'(k))$ have Boolean logic value 1 or 0 depending on $\mathbf{r}'(k) \in \mathcal{H}_j^{(+)}$ or not. A Boolean logic value $\theta_i^{(+)}(\mathbf{r}'(k))$ indicating whether $\mathbf{r}'(k) \in \mathcal{R}_i^{(+)}$ or not is obtained via a logic AND operation of $\{\beta_j(\mathbf{r}'(k)): j \in G_i^{(+)}\}$. A Boolean logic value indicating whether $\mathbf{r}'(k) \in \mathcal{R}^{(+)}$ [that is, $\hat{s}(k-d) = 1$] or not is obtained via a logic OR operation of $\{\theta_i^{(+)}(\mathbf{r}'(k))\}$ for all i . This detector achieves, asymptotically, the optimal Bayesian performance since it realizes exactly the asymptotic Bayesian decision boundary.

In our design, the determination of appropriate multiple hyperplanes for the observation space partitioning is straightforward. The required number of hyperplanes L is specified by the asymptotic Bayesian decision boundary and, therefore, is defined by the channel impulse response. The algorithm automatically selects the set of L hyperplanes, which specify the asymptotic Bayesian decision boundary, with very low computational efforts in this design stage. This should be compared with Kim and Moon's signal space partitioning technique [19], [20]. For purposes of comparison, their design

TABLE I
MATRIX OF SEPARATING HYPERPLANES AND CHANNEL STATES FOR THE EXAMPLE GIVEN IN FIG. 3(b)

hyperplane	$\mathbf{r}_1^{(-)}$	$\mathbf{r}_2^{(-)}$	$\mathbf{r}_1^{(+)}$	$\mathbf{r}_2^{(+)}$
H_1	1	0	1	1
H_2	0	1	1	0
H_3	1	1	0	1

TABLE II
COMPARISON OF DECISION COMPLEXITY FOR THE CONVENTIONAL DFE, FULL BAYESIAN, AND OUR MULTIPLE-HYPERPLANE DETECTORS. L (USUALLY $\ll 2^{n_a}$) IS THE NUMBER OF HYPERPLANES, AND n_a IS THE LENGTH OF CHANNEL IMPULSE RESPONSE. THE DFE STRUCTURE IS CHOSEN TO BE $m = n_a$, $d = n_a - 1$, AND $n = n_a - 1$

	Full Bayesian detector	Multiple-hyperplane detector	Conventional DFE
Multiplications	$(n_a + 1) \times 2^{n_a}$	$n_a \times L$	n_a
Additions	$n_a \times 2^{n_a+1} - 1$	$n_a \times L$	$n_a - 1$
Others	$2^{n_a} \exp(\cdot)$ function evaluations	logic ANDs $\leq 2^{n_a-1}$ a logic OR	

procedure is summarized in the Appendix, where it can be seen that a combinatorial search process with nonlinear gradient optimization and integer programming is performed to find a set of separating hyperplanes, given a specified minimum distance. Their design procedure therefore requires extensive computation. Furthermore, their procedure is not guaranteed to converge to the infinite-SNR asymptote of the optimal Bayesian detector. The performance of their detector will generally lie between the conventional DFE and the asymptotic Bayesian DFE, depending on the specified minimum distance in the design.

Table II compares decision complexity for the conventional DFE, the full Bayesian DFE, and our multiple-hyperplane detector. The conventional DFE has the lowest complexity as it is made up of a single hyperplane. Our detector generally has much simpler decision complexity than the full Bayesian detector since usually, $L \ll N_f$. Obviously, the decision complexity of Kim and Moon's multiple-hyperplane detector can be similar to the conventional DFE or similar to or higher than our detector, depending on the actual number of hyperplanes selected in the design.

IV. SIMULATION RESULTS

Two channels were used to demonstrate the proposed new signal space partitioning technique. The BER performance of the multiple-hyperplane detector was compared with those of the full Bayesian DFE and the conventional MMSE DFE. Except otherwise explicitly stated, all the BER results were obtained with detected symbols being fed back.

Channel A: The impulse response was specified by

$$\mathbf{a} = [a_0, a_1, a_2]^T = [0.4, 0.7, 0.4]^T. \quad (32)$$

The structure parameters of the DFE were accordingly set to $m = 3$, $d = 2$, and $n = 2$. The channel state set R had eight states. The algorithm selected five hyperplanes. The matrix of the separating hyperplanes and channel states is listed in Table III. The required Boolean logic function was easily obtained using Table III. The state $\mathbf{r}_1^{(+)}$ requires the two hyperplanes H_1 and H_2 to be separated from all the opposite-class

states $R^{(-)}$, and therefore, the convex region $\mathcal{R}_1^{(+)}$ for $\mathbf{r}_1^{(+)}$ is the intersection of the two half spaces $\mathcal{H}_1^{(+)}$ and $\mathcal{H}_2^{(+)}$. The states $\mathbf{r}_2^{(+)}$ and $\mathbf{r}_3^{(+)}$ are separated from $R^{(-)}$ by the two hyperplanes H_3 and H_4 . Thus, $\mathcal{R}_2^{(+)} = \mathcal{R}_3^{(+)}$ is the intersection of the half spaces $\mathcal{H}_3^{(+)}$ and $\mathcal{H}_4^{(+)}$. The state $\mathbf{r}_4^{(+)}$ is separated by the single hyperplane H_5 from all the opposite-class states, and the convex region $\mathcal{R}_4^{(+)}$ for $\mathbf{r}_4^{(+)}$ is the half-space $\mathcal{H}_5^{(+)}$ defined by H_5 . The overall decision region $\mathcal{R}^{(+)}$ is the union of $\mathcal{R}_1^{(+)}$, $\mathcal{R}_2^{(+)}$, and $\mathcal{R}_4^{(+)}$.

The resulting detector requires 15 multiplications and 15 additions to detect a symbol, compared with 32 multiplications, 47 additions, and eight $\exp(\cdot)$ evaluations required by the full Bayesian detector and only three multiplications and two additions required by the conventional DFE. The BERs of this multiple-hyperplane detector are compared with those of the full Bayesian DFE in Fig. 5 under different SNR conditions. It can be seen from Fig. 5 that there is hardly any BER performance difference between the two equalizers. The BER curve of the conventional MMSE DFE is also depicted in Fig. 5.

Channel B: Channel B had an impulse response given by

$$\mathbf{a} = [a_0, a_1, a_2, a_3]^T = [0.35, 0.8, 1.0, 0.8]^T. \quad (33)$$

The DFE structure was defined by $m = 4$, $d = 3$, and $n = 3$. The state set R had 16 states. The asymptotic Bayesian decision boundary in this case consisted of seven hyperplanes, and they were automatically selected by the algorithm. The matrix of the separating hyperplanes and channel states is given in Table IV. The states $\mathbf{r}_1^{(+)}$, $\mathbf{r}_3^{(+)}$, $\mathbf{r}_4^{(+)}$, and $\mathbf{r}_7^{(+)}$ require only the hyperplane H_1 to separate them from all the opposite-class states. Thus, $\mathcal{R}_1^{(+)}$, $\mathcal{R}_3^{(+)}$, $\mathcal{R}_4^{(+)}$, and $\mathcal{R}_7^{(+)}$ are defined by the half-space $\mathcal{H}_1^{(+)}$ specified by H_1 . As the states $\mathbf{r}_2^{(+)}$ and $\mathbf{r}_8^{(+)}$ are separated from $R^{(-)}$ by the two hyperplanes H_2 and H_3 , $\mathcal{R}_2^{(+)} = \mathcal{R}_8^{(+)}$ is the intersection of the two half-spaces $\mathcal{H}_2^{(+)}$ and $\mathcal{H}_3^{(+)}$. The state $\mathbf{r}_5^{(+)}$ requires the two hyperplanes H_4 and H_5 to separate it from all the opposite-class states, and the convex region $\mathcal{R}_5^{(+)}$ is thus the intersection of $\mathcal{H}_4^{(+)}$ and $\mathcal{H}_5^{(+)}$. Similarly, the convex region $\mathcal{R}_6^{(+)}$ for $\mathbf{r}_6^{(+)}$ is the intersection of $\mathcal{H}_6^{(+)}$ and $\mathcal{H}_7^{(+)}$. The overall decision region $\mathcal{R}^{(+)}$ is the union of the four convex regions $\mathcal{R}_1^{(+)}$, $\mathcal{R}_2^{(+)}$, $\mathcal{R}_5^{(+)}$, and $\mathcal{R}_6^{(+)}$.

This multiple-hyperplane-based detector requires 28 multiplications and 28 additions to make a decision, compared with 80 multiplications, 127 additions, and 16 $\exp(\cdot)$ evaluations required by the full Bayesian detector and four multiplications and three additions required by the conventional DFE. Fig. 6 compares the BER's of this multiple-hyperplane detector with those of the full Bayesian DFE and the conventional MMSE DFE. Again, there exists hardly any BER performance difference between our multiple-hyperplane detector and the full Bayesian detector.

With decision feedback, there is no performance advantage in implementing the full Bayesian detector over the piecewise linear approximation. At high to infinite SNR, the two decision boundaries are almost identical. At medium to low SNR, any theoretical performance advantage offered by the full Bayesian detector is offset by the effects of error propagation. This was

TABLE III
MATRIX OF SEPARATING HYPERPLANES AND CHANNEL STATES FOR CHANNEL A. $R^{(\pm)} = \{\mathbf{r}_1^{(\pm)}, \mathbf{r}_2^{(\pm)}, \mathbf{r}_3^{(\pm)}, \mathbf{r}_4^{(\pm)}\}$

hyperplane	$R^{(-)}$	$R^{(+)}$
H_1	1 0 0 0	1 1 1 1
H_2	0 1 1 1	1 0 0 0
H_3	1 1 1 0	0 1 1 1
H_4	0 0 0 1	1 1 1 0
H_5	1 1 1 1	0 0 0 1

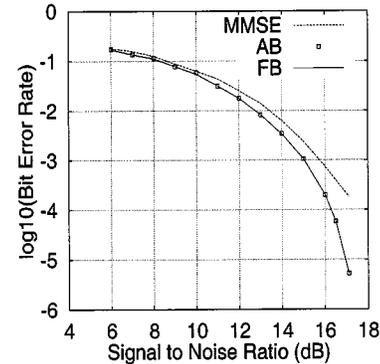


Fig. 5. Performance comparison of the conventional MMSE DFE (MMSE: dashed line), the multiple-hyperplane detector (AB: points), and the full Bayesian DFE (FB: solid line) with detected symbols being fed back for channel A.

TABLE IV
MATRIX OF SEPARATING HYPERPLANES AND CHANNEL STATES FOR CHANNEL B. $R^{(\pm)} = \{\mathbf{r}_1^{(\pm)}, \mathbf{r}_2^{(\pm)}, \mathbf{r}_3^{(\pm)}, \mathbf{r}_4^{(\pm)}, \mathbf{r}_5^{(\pm)}, \mathbf{r}_6^{(\pm)}, \mathbf{r}_7^{(\pm)}, \mathbf{r}_8^{(\pm)}\}$

hyperplane	$R^{(-)}$	$R^{(+)}$
H_1	1 1 1 1 1 1 1 1	1 0 1 1 0 0 1 0
H_2	1 1 1 0 1 1 0 0	0 1 1 1 1 1 1 1
H_3	1 1 0 1 1 1 1 1	1 1 1 1 0 0 1 1
H_4	1 1 1 1 1 1 0 0	0 0 1 1 1 1 1 1
H_5	1 1 0 0 1 1 1 1	1 1 1 1 1 0 1 1
H_6	1 1 1 1 1 1 1 0	0 0 1 1 0 1 1 1
H_7	0 1 0 0 1 1 0 1	1 1 1 1 1 1 1 1

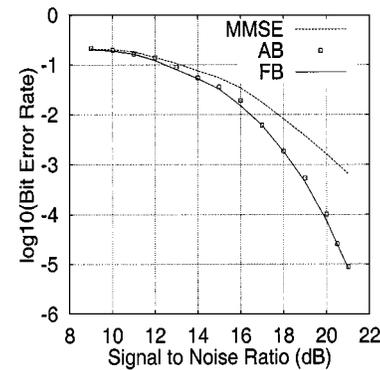


Fig. 6. Performance comparison of the conventional MMSE DFE (MMSE: dashed line), the multiple-hyperplane detector (AB: points), and the full Bayesian DFE (FB: solid line) with detected symbols being fed back for channel B.

confirmed in the two simulation examples. The BER gap between the two detectors can only be observed in low SNRs when the original transmitted symbols are artificially used in feedback (which is the original assumption of correct decision feedback).

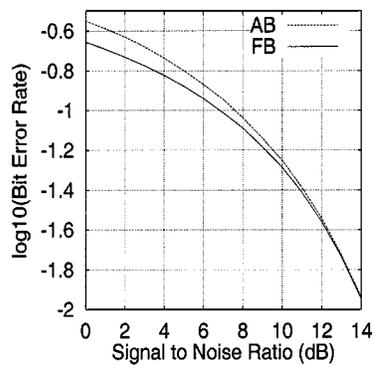


Fig. 7. Performance comparison of the multiple-hyperplane detector (AB: dashed line) and the full Bayesian DFE (FB: solid line) with correct symbols being fed back for channel B.

Fig. 7 shows the performance differences between the multiple-hyperplane detector and the full Bayesian DFE with correct symbols being fed back in low SNR conditions for channel B.

V. CONCLUSIONS

A novel equalizer has been derived based on the observation signal space partitioning. The equalizer consists of a set of linear discriminant functions and a Boolean logic function. The design procedure involves automatically finding the set of dominant opposite-class state pairs and constructing a separating hyperplane for each pair using support vector machines. The resulting decision boundary is exactly the asymptotic Bayesian decision boundary. Unlike the existing signal space partitioning technique due to Kim and Moon, which requires extensive computation in design, our design process involves little computational effort. Moreover, the resulting equalizer is guaranteed to achieve asymptotically the optimal Bayesian performance and has much lower decision complexity compared with the full Bayesian decision feedback equalizer.

APPENDIX

KIM AND MOON'S DESIGN PROCEDURE

- 1) Form all the possible pairs of subsets $(R_j^{(-)}, R_i^{(+)})$, where $R_j^{(-)} \subset R^{(-)}$, $R_i^{(+)} \subset R^{(+)}$, and the total number of states in $(R_j^{(-)}, R_i^{(+)})$ is in the range of 2 to $m + 1$.
- 2) For each subset pair $(R_j^{(-)}, R_i^{(+)})$, a separating hyperplane $H(\mathbf{r}) = \mathbf{w}^T \mathbf{r} + b$ is obtained to maximize the minimum distance from any state in $(R_j^{(-)}, R_i^{(+)})$ to the hyperplane. This is achieved by solving for the following nonlinear optimization problem¹

$$\max_{b, \mathbf{w}} \left(\min_{\mathbf{r} \in (R_j^{(-)}, R_i^{(+)})} (-1)^{I(\mathbf{r})} \left(\frac{\mathbf{w}^T \mathbf{r}}{\|\mathbf{w}\|} + b \right) \right) \quad (34)$$

subject to $\|\mathbf{w}\| = 1$, where

$$I(\mathbf{r}) = \begin{cases} 1, & \text{if } \mathbf{r} \in R_j^{(-)} \\ 0, & \text{if } \mathbf{r} \in R_i^{(+)} \end{cases} \quad (35)$$

¹A better way of obtaining such a hyperplane is to apply the method of support vector machines, which only requires solution of a simpler quadratic optimization problem; see [22].

A gradient algorithm is used, which requires many iterations to converge.

Only those hyperplanes that yield the minimum distance that is greater than or equal to a prescribed value β_{\min} are retained for the next step. The specified distance β_{\min} determines the performance of the detector.

- 3) From the chosen hyperplanes, a “minimum” number of hyperplanes are obtained by which every pair of opposite-class signal states can be separated with the prescribed distance β_{\min} . This is achieved by solving for an integer programming problem.
- 4) At this stage, a matrix of the separating hyperplane and channel states has been obtained. The design of the Boolean logic function for the detector is straightforward, as described in Section III.

REFERENCES

- [1] S. U. H. Qureshi, “Adaptive equalization,” *Proc. IEEE*, vol. 73, pp. 1349–1387, Sept. 1985.
- [2] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [3] J. Moon, “The role of SP in data-storage systems,” *IEEE Signal Processing Mag.*, vol. 15, pp. 54–72, Apr. 1998.
- [4] J. G. Proakis, “Equalization techniques for high-density magnetic recording,” *IEEE Signal Processing Mag.*, vol. 15, pp. 73–82, Apr. 1998.
- [5] R. W. Chang and J. C. Hancock, “On receiver structures for channels having memory,” *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 463–468, 1966.
- [6] K. Abend, T. J. Harley Jr., B. D. Fritchman, and C. Gumacos, “On optimum receivers for channels having memory,” *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 818–819, 1968.
- [7] K. Abend and B. D. Fritchman, “Statistical detection for communication channels with intersymbol interference,” *Proc. IEEE*, vol. 58, pp. 779–785, May 1970.
- [8] G. D. Forney, “Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference,” *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, Mar. 1972.
- [9] S. Siu, G. J. Gibson, and C. F. N. Cowan, “Decision feedback equalization using neural network structures and performance comparison with the standard architecture,” *Proc. Inst. Elect. Eng. I*, vol. 137, no. 4, pp. 221–225, 1990.
- [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, pp. 1877–1884, Aug. 1991.
- [11] S. Chen, B. Mulgrew, and S. McLaughlin, “Adaptive Bayesian equaliser with decision feedback,” *IEEE Trans. Signal Processing*, vol. 41, pp. 2918–2927, Sept. 1993.
- [12] S. Chen, S. McLaughlin, and B. Mulgrew, “Complex-valued radial basis function network—Part II: Application to digital communications channel equalization,” *EURASIP Signal Process. J.*, vol. 36, pp. 175–188, 1994.
- [13] S. Chen, S. McLaughlin, B. Mulgrew, and P. M. Grant, “Adaptive Bayesian decision feedback equaliser for dispersive mobile radio channels,” *IEEE Trans. Commun.*, vol. 43, pp. 1937–1946, May 1995.
- [14] I. Cha and S. A. Kassam, “Channel equalization using adaptive complex radial basis function networks,” *IEEE J. Select. Areas Commun.*, vol. 13, pp. 122–131, Jan. 1995.
- [15] S. Chen, S. McLaughlin, B. Mulgrew, and P. M. Grant, “Bayesian decision feedback equaliser for overcoming co-channel interference,” *Proc. Inst. Elect. Eng., Commun.*, vol. 143, no. 4, pp. 219–225, 1996.
- [16] B. Mulgrew, “Applying radial basis functions,” *IEEE Signal Processing Mag.*, vol. 13, pp. 50–65, Mar. 1996.
- [17] D. Williamson, R. A. Kennedy, and G. W. Pulford, “Block decision feedback equalization,” *IEEE Trans. Commun.*, vol. 40, pp. 255–264, Feb. 1992.
- [18] R. A. Iltis, “A randomized bias technique for the importance sampling simulation of Bayesian equalizers,” *IEEE Trans. Commun.*, vol. 43, pp. 1107–1115, Feb.-Apr. 1995.

- [19] Y. Kim and J. Moon, "Delay-constrained asymptotically optimal detection using signal-space partitioning," in *Proc. ICC'98*, Atlanta, GA, 1998.
- [20] —, "Multi-dimensional signal space partitioning using a minimal set of hyperplanes for detecting ISI-corrupted symbols," *IEEE Trans. Commun.*, 2000, to be published.
- [21] S. Chen, B. Mulgrew, E. S. Chng, and G. Gibson, "Space translation properties and the minimum-BER linear-combiner DFE," *Proc. Inst. Elect. Eng. Commun.*, vol. 145, no. 5, pp. 316–322, 1998.
- [22] S. Chen, S. Gunn, and C. J. Harris, "Decision feedback equalizer design using support vector machines," *Proc. Inst. Elect. Eng. Vision, Image, Signal Process.*, vol. 147, no. 3, pp. 213–219, 2000.
- [23] E. S. Chng, B. Mulgrew, S. Chen, and G. Gibson, "Optimum lag and subset selection for radial basis function equaliser," in *Proc. 5th IEEE Workshop Neural Networks Signal Process.*, Cambridge, MA, 1995, pp. 593–602.
- [24] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [25] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.



Sheng Chen (SM'97) received the Ph.D. degree in control engineering from the City University, London, U.K., in 1986.

From 1986 to 1999, he held various research and academic appointments at the Universities of Sheffield, Edinburgh, and Portsmouth, U.K. Since 1999, he has been with the Department of Electronics and Computer Science, the University of Southampton, U.K., where he currently holds a post of Readership in Communications Signal Processing. His research interests include modeling

and identification of nonlinear systems, adaptive nonlinear signal processing, artificial neural network research, finite-precision digital controller design, evolutionary computation methods, and optimization. He has published over 130 research papers.

Bernard Mulgrew (M'88) received the B.Sc. degree in electrical and electronic engineering in 1979 from Queen's University, Belfast, U.K. He received the Ph.D. degree in 1987 from the University of Edinburgh, Edinburgh, U.K.

After graduation from Queen's University, he worked for four years as a Development Engineer with the Radar Systems Department, GEC-Marconi Avionics, Edinburgh, U.K. From 1983 to 1986, he was a Research Associate with the Department of Electrical Engineering, University of Edinburgh. He became a Lecturer in 1986. Promotion to Senior Lectureship and Readership followed in 1994 and 1996, respectively. He was elected to the Personal Chair in Signals and Systems in 1999. His research interests are in adaptive signal processing and estimation theory and in their application to radar, audio, and communications systems. He is a co-author of three books on signal processing and has published over 150 research papers.

Dr. Mulgrew is a member of EURASIP, the IEE, and the Audio Engineering Society.



Lajos Hanzo has held various research and academic posts in Hungary, Germany, and the U.K.. Since 1986, he has been with the Department of Electronics and Computer Science, University of Southampton, U.K., where he currently holds the chair in Telecommunications. He co-authored five books on mobile radio communications, published over 200 research papers, organized and chaired conference sessions, presented overview lectures, and was awarded a number of distinctions. Currently, he is managing a research team that is working on a

range of research projects in the field of wireless multimedia communications under the auspices of the Engineering and Physical Sciences Research Council (EPSRC) U.K., the European IST Programme, and the Mobile Virtual Centre of Excellence (VCE).