

Adaptive Bayesian Decision Feedback Equalizer for Dispersive Mobile Radio Channels

Sheng Chen, *Member, IEEE*, Stephen McLaughlin, *Member, IEEE*, Bernard Mulgrew, *Member, IEEE*, and Peter M. Grant, *Member, IEEE*

Abstract—The paper investigates adaptive equalization of time-dispersive mobile radio fading channels and develops a robust high performance Bayesian decision feedback equalizer (DFE). The characteristics and implementation aspects of this Bayesian DFE are analyzed, and its performance is compared with those of the conventional symbol or fractional spaced DFE and the maximum likelihood sequence estimator (MLSE). In terms of computational complexity, the adaptive Bayesian DFE is slightly more complex than the conventional DFE but is much simpler than the adaptive MLSE. In terms of error rate in symbol detection, the adaptive Bayesian DFE outperforms the conventional DFE dramatically. Moreover, for severely fading multipath channels, the adaptive MLSE exhibits significant degradation from the theoretical optimal performance and becomes inferior to the adaptive Bayesian DFE.

I. INTRODUCTION

ADAPTIVE EQUALIZATION is an important technique for combatting distortion and interference in communication links. Equalizer design for future public land-mobile telecommunication systems, which include mobile radio service, must overcome problems that are much more complex than those encountered in fixed-link communications. In a fast frequency-selective fading environment, it is critical that an equalizer is able to adapt itself rapidly to changing channel conditions. From the viewpoint of signal detection, there are basically two categories of equalizers, namely sequence-estimation and symbol-by-symbol-decision equalizers.

The optimal sequence-estimation equalizer is the MLSE [1]. In practice the MLSE is implemented in the form of a Viterbi detector (VD) with a sufficiently large fixed decision delay and equipped with an adaptive channel estimator. For time-varying channels, tracking errors in the channel estimate can be considerable, and these errors will accumulate in the likelihood functions, causing serious degradation in performance. Moreover, decision-directed adaptation during actual data transmission is essential for rapidly time-varying channels. Because of the equalizer decision delay, the channel estimator can only provide a past channel estimate with this time delay while the current channel may have changed significantly. This estimate error due to decision delay can

further degrade performance, and it precludes the use of a large decision delay.

Symbol-by-symbol-decision equalizers are more commonly seen and are typically based on adaptive linear filter design. These include the conventional symbol and fractional spaced DFE's [2]. The linear filter approach has a very simple computational requirement but does not achieve the optimal solution for the symbol-by-symbol-decision equalizer structure. The realization of the optimal symbol-by-symbol-decision equalizer requires a nonlinear processing capability.

The optimal solution for the symbol-by-symbol-decision equalizer structure without decision feedback can be derived by adopting a Bayesian approach. This is known as the maximum a posteriori (MAP) symbol-by-symbol-decision equalizer [3]. This Bayesian equalization solution can be implemented by a variety of adaptive nonlinear structures based on artificial neural networks [4]–[7]. The same Bayesian approach can be extended to the decision feedback equalizer structure, and this is referred to as the Bayesian DFE [8], [9]. This Bayesian DFE is identical to a special case of the general block DFE proposed in [10]. How an equalizer is implemented is vital to adaptive applications. Simple and efficient adaptive algorithms have been developed [8], [9] which make it possible to apply the Bayesian DFE to mobile radio fading channels.

For stationary channels, the performance of the adaptive Bayesian DFE is better than that of the conventional adaptive DFE but is inferior to that of the adaptive MLSE [8], [9]. The current study shows that the adaptive Bayesian DFE has significant advantages over the adaptive MLSE for rapidly time-varying channels. Extensive simulation results demonstrate that the adaptive Bayesian DFE actually outperforms the adaptive MLSE considerably in terms of error rate for severely fading multipath channels. This observation is not entirely surprising. The Bayesian DFE makes decisions on a symbol-by-symbol basis and therefore does not accumulate tracking errors in channel estimation. Furthermore, the Bayesian DFE has a very short decision delay, typically 1 or 2 symbol periods for mobile radio channels, so it suffers less from the time delay in channel estimate.

In the remainder of this paper, a derivation of the Bayesian DFE is given, and the nonstationary performance of the adaptive Bayesian DFE is studied using time-dispersive mobile radio fading channels. Implementation issues of the Bayesian DFE, which include adaptive algorithm, equalizer design and computational complexity, are investigated and compared with those of the conventional DFE and the MLSE. A computer

Paper approved by M. J. Joindot, the Editor for Radio Communications, for the IEEE Communications Society. Manuscript received May 4, 1992; revised July 5, 1993. This work was supported by the UK Science and Engineering Research Council under Award GR/G/53095. The work of S. McLaughlin was supported by the UK Royal Society.

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

IEEE Log Number 9406232.

simulator for dispersive mobile radio fading channels is used to realize a realistic nonstationary channel environment and to study performance of various adaptive equalizers. Simulation results are provided and their implications discussed.

II. BAYESIAN DECISION FEEDBACK EQUALIZER

Consider the channel impulse response (CIR) $a(t)$ which includes the effects of the transmitter filter, the transmission medium and the receiver matched filter. Sampling $a(t)$ at the symbol rate gives rise to a symbol-spaced (SS) channel model $a(k)$, where k is the shorthand for kT_{sb} and T_{sb} is the symbol period. $a(k)$ is modelled by a finite impulse response (FIR) filter with transfer function

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

where n_a is the channel order, and the coefficients a_i are complex-valued and are generally time-varying. The SS sampled channel output is then defined by

$$r(k) = \hat{r}(k) + e(k) = \sum_{i=0}^{n_a-1} a_i s(k-i) + e(k) \quad (2)$$

where $\hat{r}(k)$ denotes the noiseless channel output, $e(k)$ is a complex-valued additive noise, and $\{s(k)\}$ is the transmitted symbol sequence.

In the current study, a 4-QAM signalling scheme is considered, that is, the constellation of $s(k)$ is given by

$$s(k) = \text{Re}[s(k)] + j \text{Im}[s(k)] = \begin{cases} s^{(1)} = 1 + j, \\ s^{(2)} = -1 + j, \\ s^{(3)} = 1 - j, \\ s^{(4)} = -1 - j, \end{cases} \quad (3)$$

where $j = \sqrt{-1}$. The discussion in this study however can readily be applied to other signalling schemes. The real and imaginary parts of $s(k)$ can generally be assumed to be equiprobable and independent sequences, and they are mutually independent. The real and imaginary parts of $e(k)$ are both white Gaussian noise with variance σ_e^2 , and they are mutually independent. $e(k)$ and $s(k)$ are also assumed to be uncorrelated. The task of the equalizer is to reconstruct the transmitted symbols as accurately as possible based on noisy observations $r(k)$.

The structure of a generic 5symbol-by-symbol DFE is depicted in Fig. 1, where the integers m , n , and d are the equalizer feedforward order, feedback order and decision delay respectively. Fig. 1 only shows the SS case where the feedforward section consists of SS sampled channel outputs. Channel outputs can also be sampled at a rate faster than the symbol rate and the feedforward section can include fractional spaced (FS) samples. The resulting equalizer is called a FS equalizer. The conventional DFE is a well-known example of the DFE depicted in Fig. 1. The filter operation within the conventional DFE is linear. The optimal solution for the structure of Fig. 1, however, requires nonlinear processing, and this optimal solution can be derived using Bayes decision theory [8]–[10].

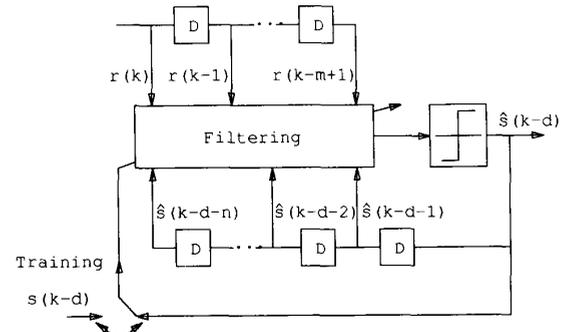


Fig. 1. Schematic of symbol-by-symbol decision feedback equalizer.

A. Bayesian Equalization Solution

For the channel (1) and a given m , the transmitted symbols that influence the equalizer decision at k are

$$\mathbf{s}(k) = [s(k) \cdots s(k-m-n_a+2)]^T. \quad (4)$$

The symbol vector $\mathbf{s}(k)$ has $N_s = 4^{n_a+m-1}$ combinations, and this gives rise to N_s states of the noise-free channel output vector

$$\hat{\mathbf{r}}(k) = [\hat{r}(k) \cdots \hat{r}(k-m+1)]^T. \quad (5)$$

The set of these N_s channel output states will be denoted as $R_{m,d}$. The oldest feedback symbol is $\hat{s}(k-d-n)$ and the oldest symbol in (4) is $s(k-m-n_a+2)$. Therefore it is sufficient to consider a feedback order

$$n = n_a + m - d - 2. \quad (6)$$

The feedback vector

$$\hat{\mathbf{s}}_f(k-d) = [\hat{s}(k-d-1) \cdots \hat{s}(k-d-n)]^T \quad (7)$$

has $N_f = 4^n$ combinations and let these N_f feedback states be labelled $s_{f,i}$, $1 \leq i \leq N_f$. $R_{m,d}$ can be divided into N_f subsets conditioned on $\hat{\mathbf{s}}_f(k-d) = s_{f,i}$

$$R_{m,d} = \bigcup_{1 \leq i \leq N_f} R_{m,d,i} \quad (8)$$

where

$$R_{m,d,i} = \{\hat{\mathbf{r}}(k) | \hat{\mathbf{s}}_f(k-d) = s_{f,i}\}, \quad 1 \leq i \leq N_f. \quad (9)$$

Each $R_{m,d,i}$ can further be divided into 4 subsets according to the value of $s(k-d)$

$$R_{m,d,i} = \bigcup_{1 \leq l \leq 4} R_{m,d,i}^{(l)} \quad (10)$$

where

$$R_{m,d,i}^{(l)} = \{\hat{\mathbf{r}}(k) | s(k-d) = s^{(l)} \cap \hat{\mathbf{s}}_f(k-d) = s_{f,i}\}, \quad 1 \leq l \leq 4. \quad (11)$$

Each subset $R_{m,d,i}$ contains $N_{s,i} = N_s/N_f = 4^{d+1}$ states, and the number of states in $R_{m,d,i}^{(l)}$ will be denoted as $N_{s,i}^{(l)} = 4^d$, $1 \leq l \leq 4$.

The optimal solution for the structure of Fig. 1 minimizes the average error probability in symbol detection and it must operate according to Bayes decision rule [11]. Under the assumption that the correct signal vector $\hat{\mathbf{s}}_f(k-d)$ is fed back, it can be shown that this optimal Bayesian DFE rule takes the form [9]

$$\hat{\mathbf{s}}(k-d) = \text{csgn}(f_B(\mathbf{r}(k)|\hat{\mathbf{s}}_f(k-d) = \mathbf{s}_{f,i})) \quad (12)$$

where

$$\text{csgn}(f) = \text{sgn}(\text{Re}[f]) + j \text{sgn}(\text{Im}[f]) \quad (13)$$

$\text{sgn}(\cdot)$ is the signum function, $\mathbf{r}(k)$ is the channel observation vector

$$\mathbf{r}(k) = [r(k) \cdots r(k-m+1)]^T \quad (14)$$

and the conditional Bayesian decision function given $\hat{\mathbf{s}}_f(k-d) = \mathbf{s}_{f,i}$ is

$$f_B(\mathbf{r}(k)|\hat{\mathbf{s}}_f(k-d) = \mathbf{s}_{f,i}) = \sum_{q=1}^4 h^{(q)} \sum_{l=1}^{N_{s,i}^{(q)}} \exp(-\|\mathbf{r}(k) - \mathbf{r}_l^{(q)}\|^2/\rho). \quad (15)$$

The 4 coefficients in (15) are $h^{(1)} = 1 + j$, $h^{(2)} = -1 + j$, $h^{(3)} = 1 - j$ and $h^{(4)} = -1 - j$; the 4 inner sums are over $\mathbf{r}_l^{(q)} \in R_{m,d,i}^{(q)}$, $1 \leq q \leq 4$, respectively; the real positive scalar ρ is equal to $2\sigma_e^2$. The derivation of (15) can be found in [9].

An immediate observation is that the feedback vector is used to reduce computational complexity. Without decision feedback, all of the N_s channel output states would be required in the computation of the Bayesian decision function at each sample k . As a result of decision feedback, only a small subset of $N_{s,i}$ states are needed in the computation. A geometric explanation of how decision feedback improves equalization performance is given in [9].

The parameter ρ in (15) should be set to $2\sigma_e^2$ thus perfect knowledge of the noise power at the receiver is theoretically required. However, ρ is not influential and need not be accurately set to $2\sigma_e^2$. This insensitivity to the value of ρ is demonstrated using the channel

$$A(z) = (0.4313 + j0.4311)(1 - (0.5 + j)z^{-1}) \cdot (1 - (0.35 + j0.7)z^{-1}). \quad (16)$$

The structure of the Bayesian DFE is defined by $d = 2$, $m = 3$ and $n = 2$. The value of ρ is set to $4\sigma_e^2$, σ_e^2 and $2\sigma_e^2$, respectively. The three symbol error rate (SER) curves obtained are depicted in Fig. 2, where it is seen that they are indistinguishable. In view of this insensitivity, ρ is set to the true noise power $2\sigma_e^2$ in all the simulations that follow.

During data transmission, $\hat{\mathbf{s}}_f(k-d)$ consists of detected symbols. When an error is made, error propagation will result. The effects of error propagation on the Bayesian DFE and the conventional DFE are demonstrated using the channel (16). Both equalizers have the same structure of $d = 2$, $m = 3$ and $n = 2$. The coefficients of the conventional DFE are set to the Wiener solution. The SER's obtained using correct symbols and detected symbols as feedback respectively are plotted in Fig. 3.

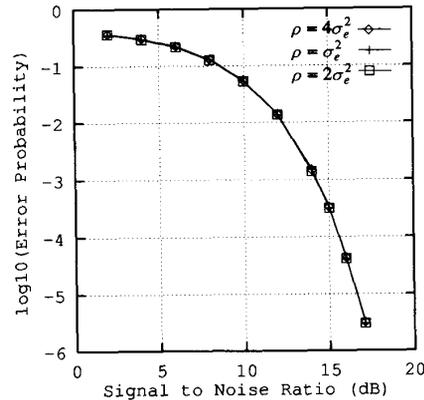


Fig. 2. Effects of parameter ρ on the performance of Bayesian DFE. Three-tap stationary channel and detected symbols being fed back.

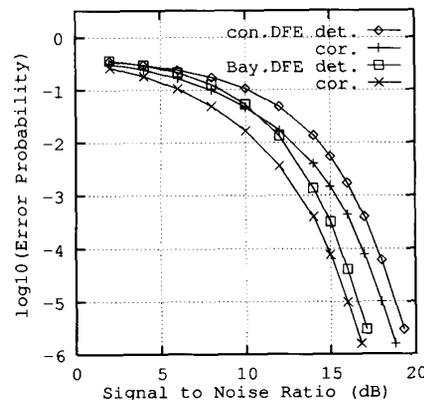


Fig. 3. Effects of error propagation, con./Bay. DFE: conventional/Bayesian DFE, det./cor.: detected/correct symbols being fed back.

B. Adaptive Algorithm

Two adaptive schemes have been developed for updating the channel output states required in the Bayesian DFE. The clustering approach [6]–[9] identifies these states directly using a clustering algorithm. This scheme is computationally very simple and is immune from nonlinear channel distortion. However, it requires a large amount of training data and is only suitable for stationary or slowly time-varying channels. The second approach estimates the channel model (1) by a conventional adaptive algorithm and uses the resulting channel estimate to calculate the subset states $R_{m,d,i}^{(q)}$, $1 \leq q \leq 4$. This approach, although computationally more complex than the clustering scheme, needs a much smaller training set and is suitable for rapidly time-varying channels. In this present application to severely fading channels, this second adaptive scheme is used.

In this paper the estimation of the SS CIR (1) is achieved using the least mean square (LMS) algorithm. The channel estimate, defined as

$$\hat{\mathbf{a}}(k) = [\hat{a}_0(k) \cdots \hat{a}_{n-1}(k)]^T \quad (17)$$

can alternatively be updated using the recursive least square (RLS) algorithm. Because the successive symbols $s(k)$ are assumed to be independent, the correlation matrix of the estimator input vector is diagonal and has an eigenvalue spreading ratio of 1. This is ideal for the LMS algorithm. The LMS algorithm not only has a much simpler computational complexity compared to the RLS algorithm but also has as good a steady state tracking performance as the RLS algorithm in this situation.

For rapidly time-varying channels, it is essential to continuously adapt the channel estimate during data transmission. This is achieved using decision-directed learning. The situation is slightly complicated by the equalizer decision delay and, at sample k , a feasible decision-directed LMS algorithm is

$$\left. \begin{aligned} \epsilon(k-d) &= r(k-d) - \hat{\mathbf{a}}^T(k-d-1)\hat{\mathbf{s}}_a(k-d), \\ \hat{\mathbf{a}}(k-d) &= \hat{\mathbf{a}}(k-d-1) + \mu\epsilon(k-d)\bar{\hat{\mathbf{s}}}_a(k-d), \end{aligned} \right\} \quad (18)$$

where μ is an adaptation gain, the bar above $\hat{\mathbf{s}}_a$ denotes complex conjugate and

$$\bar{\hat{\mathbf{s}}}_a(k-d) = [\hat{s}(k-d) \cdots \hat{s}(k-d-n_a+1)]^T. \quad (19)$$

At $k+1$, the equalizer must use this delayed estimate $\hat{\mathbf{a}}(k-d)$ as though it were the most recent estimate $\hat{\mathbf{a}}(k)$ to make a decision. The current channel model $\mathbf{a}(k+1)$ may have changed considerably. This tracking error owing to inherent decision delay will degrade the equalizer performance. Fortunately the Bayesian DFE can have a very short decision delay, typically 1 or 2 for mobile fading channels, to minimize the effect of decision delay.

It is interesting to consider this implication of decision-directed learning to the adaptive MLSE. Theoretically a very large decision delay ($> 5n_a$) is required to realize the optimal MLSE performance. Such a large decision delay would introduce unacceptable channel tracking errors during decision-directed adaptation. A compromise must be found to accommodate these two conflicting requirements, and this results in a modest decision delay. An alternative is to employ two decision delays, and this allows the VD to output two decision sequences. A low-delay preliminary decision sequence is used for channel estimation only while the final decision sequence adopts a sufficiently large decision delay [12].

C. Equalizer Design

The structure of the Bayesian DFE is specified by d , m , and n . The basic structure parameter is d , which specifies the number of states required for computing the conditional Bayesian decision function and thus determines the level of complexity. Given d , $m = d+1$ is sufficient for the Bayesian DFE. That is, a Bayesian DFE with $m = d+1$ has the same performance as those with $m > d+1$. The proofs of this conclusion can be found in [9]. Substituting this result into (6) gives rise to $n = n_a - 1$.

In the case of $d = 0$, $m = 1$ is sufficient and the Bayesian DFE reduces to a very simple form (see [9]). Assume that the channel is normalized, that is

$$\sum_{i=0}^{n_a-1} |a_i|^2 = 1. \quad (20)$$

The theoretical error probability of the Bayesian DFE with $d = 0$ and assuming correct $\hat{\mathbf{s}}_f(k)$ can be shown to be

$$P_e(|a_0|/\sigma_e) = 2Q(|a_0|/\sigma_e) - Q^2(|a_0|/\sigma_e) \quad (21)$$

where

$$Q(|a_0|/\sigma_e) = \int_{|a_0|/\sigma_e}^{\infty} (2\pi)^{-1/2} \exp(-\chi^2/2) d\chi. \quad (22)$$

For the ideal channel with no intersymbol interference and a unit channel gain, the error probability in symbol detection is known to be $P_e(1/\sigma_e)$.

A pragmatic rule of selecting d is to set $d = n_a - 1$, which has a heuristic explanation. In the case of $d = 0$, the decision delay covers the first channel tap a_0 , and the equalizer performance depends on the energy of a_0 as shown in (21). It can be imagined that in general performance should depend on the energy of the channel taps a_0 to a_d . Increasing d improves performance and $d = n_a - 1$ is sufficient to achieve the full performance potential. Since computational complexity of the adaptive Bayesian DFE for 4-QAM symbols is an order of 4^{d+1} as will be shown later, it is important to choose a d which is as small as possible without sacrificing too much performance. If most of the channel energy is contained in the taps a_0 to a_w , where $w < n_a - 1$, the equalizer delay can be set to $d = w$. For multipath mobile ratio fading channels, channel energy is usually concentrated near line-of-sight, and it is often sufficient to choose $d = 1$ or 2.

D. Computational Complexity

Computational load of the adaptive Bayesian DFE consists of three parts, namely the channel estimator based on the LMS algorithm, calculation of $R_{m,d,i}$ based on the channel estimate and computation of $f_B(\cdot)$. Computational requirements for these three subtasks are listed in Table I, where complex arithmetic has been converted into equivalent real arithmetic. For $d > 0$, an estimated upper bound is given for the computation of subset states. This upper bound is obtained by assuming that symbol combinations in the feedforward section are random. In reality these symbol combinations exhibit regular patterns, and this redundancy can be exploited leading to a substantial saving in computation.

As a comparison, computational complexity of the conventional DFE is also listed in Table I. For the conventional DFE, inputs to the adaptive algorithm contain channel outputs which are colored. Theoretically whether the RLS or the LMS is used can make a difference. Therefore both the cases of using the RLS and LMS algorithms are given. The complexity with the RLS is based on the full ordinary version of the algorithm (e.g. [13]). Because the adaptive algorithm is expected to continuously operate during both the training and transmission periods in a highly nonstationary environment, its numerical stability is vital. Many versions of the fast RLS algorithm may not be suitable for this purpose.

Consider a channel with $n_a = 4$. Assume that the conventional DFE is $T_{sb}/2$ FS. To cover the total channel dispersion, the feedforward section has $2n_a - 1 = 7$ coefficients and the feedback section has $n_a - 1 = 3$ SS coefficients. Such a $T_{sb}/2$

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITY. † ESTIMATED UPPER BOUND FOR COMPUTING SUBSET STATES. ‡ FULL ORDINARY VERSION OF RLS

| Bayesian DFE ($d > 0$) with LMS | | |
|---------------------------------------|---|-----------------|
| Channel Estimator | $8 \times n_a + 2$ | multiplications |
| | $8 \times n_a$ | additions |
| Subset States | $\sum_{i=0}^d \{4 \times (d+1-i) \times 4^{d+1} + 4 \times (n_a - d - 1 + i)\} \dagger$ | multiplications |
| Decision | $\sum_{i=0}^d \{4 \times (d+1-i) \times 4^{d+1} + 4 \times (n_a - d - 1 + i) - 2\} \dagger$ | additions |
| Function | 4^{d+1} | $\exp(\cdot)$ s |
| | $(2 \times d + 3) \times 4^{d+1}$ | multiplications |
| | $4 \times (d+1) \times 4^{d+1} + 2$ | additions |
| Bayesian DFE ($d = 0$) with LMS | | |
| Channel Estimator | $8 \times n_a + 2$ | multiplications |
| | $8 \times n_a$ | additions |
| Subset States | $4 \times n_a + 12$ | multiplications |
| Decision | $4 \times n_a + 10$ | additions |
| Function | 8 | multiplications |
| | 12 | additions |
| Conventional DFE (m, n) with LMS | | |
| | $8 \times (m+n) + 2$ | multiplications |
| | $8 \times (m+n)$ | additions |
| Conventional DFE (m, n) with RLS‡ | | |
| | $11 \times (m+n)^2 + 23 \times (m+n)$ | multiplications |
| | $10 \times (m+n)^2 + 13 \times (m+n) - 1$ | additions |

TABLE II
COMPUTATIONAL COMPLEXITY FOR $n_a = 4$

| Bayesian DFE ($d = 1$) with LMS | | |
|-----------------------------------|------|-----------------|
| Channel Estimator | 34 | multiplications |
| | 32 | additions |
| Subset States | 68 | multiplications |
| Decision | 112 | additions |
| Function | 16 | $\exp(\cdot)$ s |
| | 80 | multiplications |
| | 130 | additions |
| Total | 16 | $\exp(\cdot)$ s |
| | 182 | multiplications |
| | 274 | additions |
| Upper bound for subset states | 212 | multiplications |
| | 208 | additions |
| $T_{sb}/2$ FS DFE (7, 3) with LMS | | |
| | 82 | multiplications |
| | 80 | additions |
| $T_{sb}/2$ FS DFE (7, 3) with RLS | | |
| | 1330 | multiplications |
| | 1129 | additions |

FS DFE can be represented as (7, 3). For the Bayesian DFE, $d = 1$ is chosen. The computational requirements for these two adaptive equalizers are listed in Table II. For the subtask of computing subset states required in the Bayesian DFE, both the estimated upper bound and the real complexity taking into account redundancy are listed in Table II.

From Table I, it can be seen that the complexity of the adaptive Bayesian DFE for 4-QAM symbols is of the order of 4^{d+1} . Since a small d is used in practice, typically 1 or 2, the Bayesian DFE is only slightly more complex than the conventional DFE. The VD demands sophisticated processing capability while the implementation of the Bayesian DFE is straightforward. Even without counting this subtle difference, the basic complexity of the adaptive MLSE is more than that of the adaptive Bayesian DFE. The discussion can be extended to

the general M -QAM case. For high order M , the complexity of the conventional DFE remains more or less the same as the 4-QAM case. The complexity of the Bayesian DFE will increase quickly as M increases since its complexity is of order M^{d+1} . However the complexity of the MLSE will increase at an even more rapid pace.

III. PERFORMANCE FOR MOBILE RADIO CHANNELS

Equalization performance of the adaptive Bayesian DFE is investigated using severely fading multipath channels. The purpose of this investigation is to assess the nonstationary error rate performance of the adaptive Bayesian DFE and to compare the Bayesian DFE with the two other adaptive equalization schemes, namely the conventional DFE and the MLSE.

A. Computer Simulator for Mobile Radio Channels

A software simulator has been developed to simulate multipath mobile radio fading channels. The symbol source generates 4-QAM symbols at a symbol rate 300 kHz. The combined transfer function of the transmitter and receiver filters is of raised-cosine type with a rolloff factor 0.5. The transmitter and receiver filters are identical. These two filters are implemented as FIR filters, and the filter tap weights are samples of the truncated root-raised-cosine pulse with a sufficient pulse length of n_p symbol periods. The time-dispersive multipath fading channel is implemented as a tapped-delay-line model

$$\hat{y}(t) = \sum_{i=0}^{n_c-1} c_i(t)u(t - i\delta) \quad (23)$$

where n_c is the number of paths, δ is the path delay parameter, the time-varying tap weights $c_i(t)$ are zero mean complex-valued Gaussian random processes and they are mutually

uncorrelated, and $u(t)$ is the transmitter filter output. The receiver filter output is given by the convolution

$$r(t) = a_{Rx}(t) * (\hat{y}(t) + e(t)) \quad (24)$$

where $a_{Rx}(t)$ is the impulse response of the receiver filter and $e(t)$ is a complex-valued Gaussian white process. To realize a near analogue waveform, the system blocks of the transmitter filter, the multipath fading medium and the receiver filter are operated at a system sample rate of 4.8 MHz, 16 times faster than the symbol rate. Considerations of computational efficiency prevent us from using a higher system sample rate.

To realize Rayleigh fading characteristics, real and imaginary components of each tap weight $c_i(t)$ are generated by passing Gaussian white sequences through digital second-order low pass Butterworth filters. The bandwidth of the Butterworth filter is of the order of Doppler frequency caused by the motion of the mobile. In this simulation study, a Doppler frequency of 100 Hz is used. The output power of the Butterworth filter when driven by a white noise of unit variance is calculated using the method given in chapter 5 of [14], and the square root of this filter power is used to scale each tap weight component. Each tap weight component is further scaled by its designed root mean power (RMP).

The transmitted data are organized in blocks; a block consists of 20 training symbols and 60 information symbols. The SS sampled receiver filter outputs can be expressed as $r(k) = \hat{r}(k) + \tilde{e}(k)$, where $\tilde{e}(k)$ are SS samples of $a_{Rx}(t) * e(t)$. The signal to noise ratio (SNR) of the simulated system is defined as

$$\text{SNR} = \left(\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E_s[|\hat{r}(i)|^2] \right) / E_e[|\tilde{e}(k)|^2] \quad (25)$$

where $E_p[\cdot]$ denotes the expectation operator with respect to the random process p .

C. Decision-directed Adaptation

For the Bayesian DFE and the MLSE, the channel estimator updates $\hat{\mathbf{a}}(k)$ during a training period. The channel estimate may then be fixed throughout a transmission period. However, due to the time-varying nature of the medium, the underlying channel $\mathbf{a}(k)$ can change significantly in a transmission period and some performance improvement can be obtained by decision-directed updating of the equalizer coefficients during transmission. The significance of this decision-directed learning can be demonstrated by examining the two criteria, namely the mean square error (MSE) criterion

$$\text{MSE} = |r(k) - \hat{\mathbf{a}}^T(k-d-1)\mathbf{s}_a(k)|^2 \quad (26)$$

and the mean tap weight error (MTE) criterion

$$\text{MTE} = \|\mathbf{a}(k) - \hat{\mathbf{a}}(k-d-1)\|^2 / \|\mathbf{a}(k)\|^2 \quad (27)$$

where $\mathbf{s}_a(k) = [s(k) \cdots s(k-n_a+1)]^T$ is the channel input symbol vector. For fast frequency-selective fading channels, an equalizer that adopts decision-directed adaptation continuously during transmission performs considerably better than one which does not; this is illustrated using the following example.

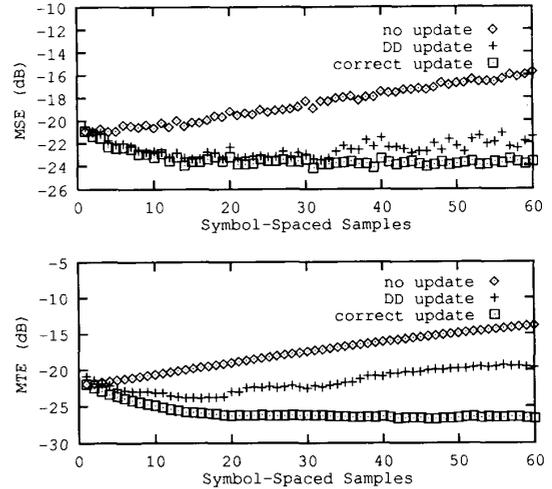


Fig. 4. Variations of mean square error and mean tap error throughout transmission. Channel has 3 symbol-spaced paths.

The transmission medium consists of 3 SS ($\delta = T_{sb}$) fading paths, and the RMP's of both the real paths and the imaginary paths are

$$[0.2463 \quad 0.6154 \quad 0.2463]. \quad (28)$$

It is straightforward to verify that the equivalent SS CIR model is defined as

$$A(z) = z^{-\tau}(a_0(k) + a_1(k)z^{-1} + a_2(k)z^{-2}) \quad (29)$$

where $\tau = n_p$ is the transmission delay. The channel model (29) is perfect with $a_i(k) = c_i(k)$, $0 \leq i \leq 2$. The Bayesian DFE is employed with $d = 1$, and the LMS channel estimator is used with $\mu = 0.05$. The MSE and MTE plots are averaged over ensembles of 500 data blocks. If the transmitted symbols were known, they could be used to update the channel estimate during transmission, and this provides the lower bounds for the MSE and MTE shown in Fig. 4 under the title "correct update" for SNR = 25 dB. If the channel estimate is fixed throughout transmission, the MSE and MTE will increase, departing away from their lower bounds as can be seen from Fig. 4. Fig. 4 also shows performance improvement achievable by decision-directed (DD) updating.

C. Symbol Error Rate Performance

A simulation study has been carried out to investigate SER's of the adaptive Bayesian DFE, the adaptive MLSE and the conventional DFE. All the SER plots are averaged over 2000 to 10^5 data blocks depending on the SNR. The LMS algorithm is used for each of the three equalizers studied. During transmission the adaptive algorithm operates continuously in decision-directed mode. A variety of different values for adaptive gain have been tested and it is found that $\mu = 0.05$ provides the best overall performance. The results shown have been obtained using this value of adaptive gain. The SER's for the two DFE's shown in this study have been obtained with detected symbols being fed back.

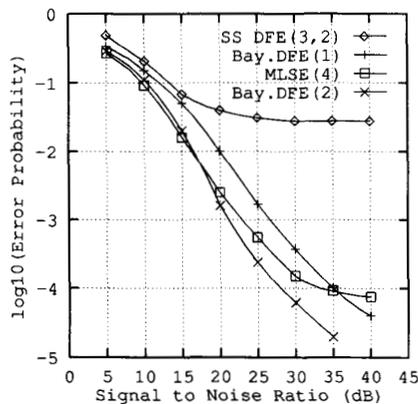


Fig. 5. Performance comparison for adaptive equalizer. Channel has 3 symbol-spaced paths.

For the Bayesian DFE and the MLSE, it is sufficient to use the LMS algorithm. For the conventional DFE, theoretically the RLS algorithm should have convergence advantages over the LMS. However, when we apply the conventional DFE based on the RLS to the channels simulated in this study, its performance is much worse than that based on the LMS. This is because, when the SER is not sufficiently small, error propagation during decision-directed learning often causes the RLS algorithm to diverge. The performance of the conventional DFE based on the RLS is not included in the present paper.

Two versions of the adaptive MLSE have been compared. In the first version, the VD uses a single decision delay, and the detected symbol sequence is simultaneously used for channel estimation. In the simulation, this delay is increased until it reaches the point that a further increase will not improve and may even worsen performance. For the channels simulated, it is found that a decision delay of 4 to 6 is appropriate for this version. In the second version, the VD adopts two decision delays. The low-delay preliminary decision sequence is used for channel estimation, and the final decision sequence has a sufficiently long delay. For both versions, the VD is reset to the correct initial conditions at the beginning of every data block. Without this procedure, it was observed that the performance quickly degrades as a result of error accumulation.

Example 1: This is the channel used in Fig. 4. The channel order is $n_a = 3$. Therefore the SS DFE has a structure of (3, 2) while a $T_{sb}/2$ FS DFE is defined by (5, 2). The SER's of the conventional SS DFE (3, 2), the version-one adaptive MLSE with decision delay 4, and the adaptive Bayesian DFE with $d = 1$ and $d = 2$ are depicted in Fig. 5. The SER of the $T_{sb}/2$ FS DFE (5, 2), not shown, is slightly better than that of the SS DFE (3, 2). The last channel tap $a_2(k)$ contains significant energy and, as expected, the adaptive Bayesian DFE with $d = 2$ is significantly better than the adaptive Bayesian DFE with $d = 1$.

If the transmission paths $c_i(t)$ are all known, $a_i(k)$ can be calculated at each k , and this true SS channel model can then be used to provide the theoretical SER of an equalizer. Fig. 6 shows how an adaptive MLSE deviates from its theoretical bound. The second version of the adaptive MLSE has a

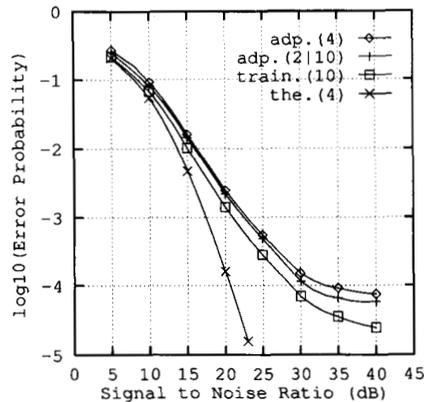


Fig. 6. Deviation of adaptive MLSE from its theoretical performance. Channel has 3 symbol-spaced paths, in the curve titled "train.(10)", channel estimator uses transmitted symbols all the time with no delay.

preliminary delay 2 and a final delay 10, and its SER is depicted under the title "adp.(2|10)" in Fig. 6. The results clearly demonstrate that the deviation of an adaptive MLSE from its theoretical bound is very serious. From Fig. 6, it can be seen that the second version of the adaptive MLSE is only marginally better than the first version. The main cause of the performance degradation is the accumulation of tracking errors in likelihood functions. To support this claim, a hypothetical situation is simulated where the channel estimator assumes perfect knowledge of the transmitted symbols. This removes any error caused by estimation delay and by decision-directed adaptation. Estimation errors in this case are associated solely with the fact that the channel estimator is incapable of tracking the fast time-varying channel precisely. The VD uses this channel estimate to decode the transmitted symbol sequence with a sufficiently large decision delay of 10 symbol periods. The resulting SER is plotted under the title "train.(10)" in Fig. 6, where it is seen that this hypothetical MLSE is only slightly better than the true adaptive MLSE.

Fig. 7 compares the performance of the adaptive Bayesian DFE with its theoretical bound. The graphs shown in Figs. 6 and 7 confirm that the theoretical MLSE is superior to the theoretical Bayesian DFE. However, the deviation of the adaptive Bayesian DFE from its theoretical performance is much less serious than that of the adaptive MLSE from its theoretical performance and, consequently, the adaptive Bayesian DFE outperforms the adaptive MLSE.

The theoretical conventional DFE assuming perfect knowledge of the channel is also calculated. This is achieved by solving the Wiener equation given $a(k)$ and noise variance at each k to derive the equalizer coefficients. Fig. 8 shows how the adaptive conventional DFE deviates from its theoretical bound. The curve entitled "adp.(3,2)" in Fig. 8 is the same curve of the conventional DFE shown in Fig. 5, where the equalizer coefficients are updated using the LMS algorithm. Because the input vector to the LMS algorithm is colored, the tracking performance of the LMS is considerably poorer compared with the case of channel estimation where the input vector to the LMS algorithm is white. This is an important

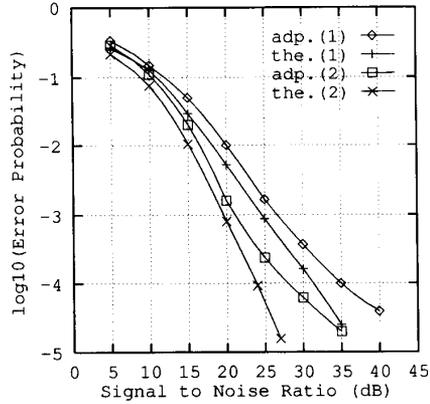


Fig. 7. Deviation of adaptive Bayesian DFE from its theoretical performance. Channel has 3 symbol-spaced paths.

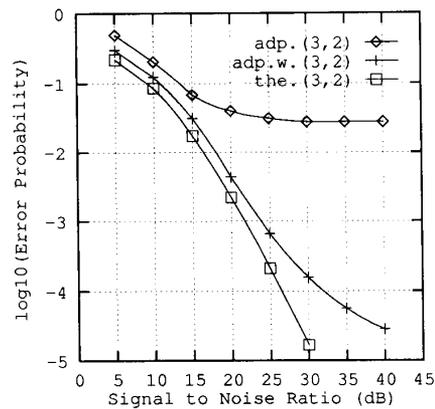


Fig. 8. Deviation of adaptive conventional symbol-spaced DFE from its theoretical performance. Channel has 3 symbol-spaced paths.

factor that causes the serious performance degradation. An alternative adaptive scheme is to estimate the channel model using the LMS algorithm and to solve for the coefficients of the conventional DFE using the channel estimate. This however requires the solution of the Wiener equation at every k and will increase computational complexity dramatically. The SER obtained using this adaptive scheme is depicted in Fig. 8 under the title "adp.w.(3,2)". It should be emphasized that real-time implementation of this adaptive scheme is very difficult.

Example 2: The transmission medium consists of $6 T_{sb}/2$ FS ($\delta = T_{sb}/2$) fading paths, and the RMP's of the real paths and the imaginary paths are

$$[0.4704 \quad 0.2582 \quad 0.1417 \quad 0.0778 \quad 0.0427 \quad 0.0234]. \quad (30)$$

It can be shown that the equivalent SS sampled CIR is given by

$$A(z) = z^{-\tau} \sum_{i=0}^{n_a-1} a_i(k) z^{-i} \quad (31)$$

where $\tau = n_p - 1$ and $n_a = 5$. There are other nonzero taps $a_i(k)$ but they are all very small at every k and can be neglected. For most practical purposes, the model structure (31) is very accurate. The SER plots of the $T_{sb}/2$ FS DFE (9,

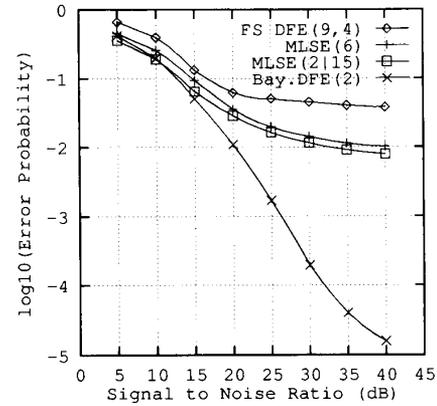


Fig. 9. Performance comparison for adaptive equalizers. Channel has 6 half-symbol-spaced paths, equalizers are designed based on an accurate channel structure.

4), the first version of the adaptive MLSE with decision delay 6, the second version of the adaptive MLSE with preliminary delay 2 and final delay 15, and the adaptive Bayesian DFE with $d = 2$ are given in Fig. 9. For the conventional DFE, the standard scheme of direct updating of the equalizer coefficients is adopted. It can be seen that the performance gap between the adaptive Bayesian DFE and the adaptive MLSE is very large. From Fig. 9, the second version of the adaptive MLSE is observed again to be marginally better than the first version. A more detailed analysis shows that most of the channel energy is concentrated in $a_1(k)$ and $a_2(k)$. Thus a delay of $d = 2$ is sufficient for the adaptive Bayesian DFE.

Since the average amplitude of $a_0(k)$ in (31) is much smaller compared with that of $a_1(k)$ or $a_2(k)$, it is interesting to consider performance of adaptive equalizers based on the simplified channel structure

$$\tilde{A}(z) = z^{-n_p} (\tilde{a}_0(k) + \tilde{a}_1(k)z^{-1} + \tilde{a}_2(k)z^{-2} + \tilde{a}_3(k)z^{-3}). \quad (32)$$

This simplified sampled CIR is constructed by neglecting $a_0(k)$ and re-numbering the rest of the taps in (31). Note that assuming $a_0(k)$ in (31) to be zero is equivalent to adding a transmission delay of one symbol duration. The SER graphs of three adaptive equalizers designed using this less accurate channel structure are shown in Fig. 10. The SER of the second version of the adaptive MLSE with preliminary delay 1 and final delay 14, not shown, is similar to that of the first version with decision delay 6. The advantage of employing this simplified channel structure is a substantial saving in computation. For example, $d = 1$ becomes sufficient for the adaptive Bayesian DFE. Furthermore, the deterioration in SER should not be notable since the original $a_0(k)$ neglected in (32) is not very significant. Comparing Fig. 10 with Fig. 9 confirms this view.

IV. CONCLUSION

A novel adaptive Bayesian DFE has been presented, and its performance in a nonstationary environment has been investigated using a mobile radio fading channel simulator.

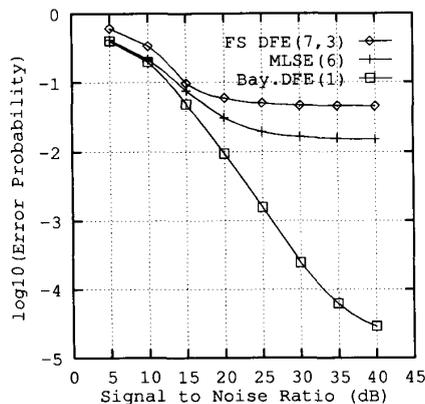


Fig. 10. Performance comparison for adaptive equalizers. Channel has 6 half-symbol-spaced paths, equalizers are designed based on a simplified channel structure.

Compared with the conventional DFE and the adaptive MLSE, the adaptive Bayesian DFE has significant performance and implementation advantages. For modulation schemes such as 4-QAM, the computational complexity of the adaptive Bayesian DFE is only slightly more than that of the conventional DFE. For fast frequency-selective fading channels, it has been shown that the performance of the conventional DFE with the standard scheme of direct adapting equalizer coefficients is poor. Although the theoretical MLSE provides the best attainable equalization performance, the adaptive Bayesian DFE actually outperforms the adaptive MLSE dramatically in a highly nonstationary environment. It has been suggested that the adaptive MLSE accumulates tracking errors which causes serious performance degradation. The adaptive Bayesian DFE in contrast appears to be very robust.

REFERENCES

- [1] 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, 1972.
- [2] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp. 1349-1387, 1985.
- [3] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, vol. 58, pp. 779-785, 1970.
- [4] G. J. Gibson, S. Siu, and C. F. N. Cowan, "The application of nonlinear structures to the reconstruction of binary signals," *IEEE Trans. Signal Proc.*, vol. 39, pp. 1877-1884, 1991.
- [5] S. Chen, G. J. Gibson, and C. F. N. Cowan, "Adaptive channel equalisation using a polynomial-perceptron structure," *IEE Proc.*, Pt. I, vol. 137, no. 5, pp. 257-264, 1990.
- [6] S. Chen and B. Mulgrew, "Overcoming co-channel interference using an adaptive radial basis function equaliser," in *EURASIP Signal Proc. J.*, vol. 28, no. 1, pp. 91-107, 1992.
- [7] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalisation using radial basis function networks," *IEEE Trans. Neural Net.*, vol. 4, pp. 570-579.
- [8] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian decision feedback equaliser based on a radial basis function network," in *Proc. ICC*, Chicago, IL, 1992, pp. 343.3.1-343.3.5.
- [9] ———, "Complex-valued radial basis function network, Part II: application to digital communications channel equalisation," *EURASIP Signal Proc. J.*, vol. 36, pp. 175-188, 1994.
- [10] D. Williamson, R. A. Kennedy, and G. W. Pulford, "Block decision feedback equalization," *IEEE Trans. Commun.*, vol. 40, pp. 255-264, 1992.

- [11] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [12] P. R. Chevillat and E. Eleftheriou, "Decoding of trellis-encoded signals in the presence of intersymbol interference and noise," *IEEE Trans. Commun.*, vol. 37, pp. 669-676, 1989.
- [13] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991, 2nd ed.
- [14] K. J. Aström, *Introduction to Stochastic Control Theory*. New York: Academic, 1970.



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. In November 1993 he took a senior lecturer post in the Department of Electrical and Electronic Engineering at the University of Portsmouth. His research interests are in modelling and identification of nonlinear systems, artificial neural networks and adaptive signal processing for communications.



Stephen McLaughlin (M'90) 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, where 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.



Bernard Mulgrew (M'90) received the Ph.D. degree in 1987.

He is currently a Senior Lecturer in the Department of Electrical Engineering at the University of Edinburgh, where he gives courses on signals & systems and signal processing. His research interests are in adaptive signal processing and estimation theory and in their application to radar and communication systems. After graduation in 1979, he worked for 4 years in the Radar Systems Department at GEC-Marconi Avionics, 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 a coauthor of two books on signal processing. He is an associate member of the IEE and a member of the Audio Engineering Society.



Peter M. Grant (M'77-SM'83) was born in St. Andrews, Scotland in 1944. He received the B.Sc. degree in electronic engineering from the Heriot-Watt University, Edinburgh, Scotland in 1966, and the Ph.D. degree from the University of Edinburgh in 1975.

He worked for the Plessey Company Ltd., before he was appointed to a Research Fellowship at the University of Edinburgh to study the applications of surface acoustic wave and charge-coupled devices in communication systems. He was subsequently

appointed to a Lectureship and promoted through to a Personal Chair in Electronic Signal Processing with responsibility for teaching signal processing and communication systems. He leads the signal processing research group with personal involvement in adaptive filtering, pattern recognition and spread spectrum communications.

Dr. Grant was the recipient of a James Caird Travelling Scholarship during the academic year 1977-1978, and as a Visiting Professor, he researched at the Ginzton Laboratory, Stanford University. From 1985 to 1986 he was appointed as a Visiting Staff Member at the MIT Lincoln Laboratory. He is a Fellow of the Institution of Electrical Engineers (London). He serves there as one of the Honorary Editors of the IEE Proceedings "Vision Image and Signal Processing," as the Senior Editor and Chairman of the IEE Proceedings Editorial Panel and as an Advisory Panel Member for the IEE Electronics and Communications Engineering Journal.