

A Blind Decision Feedback Equalizer Incorporating Fixed Lag Smoothing

Sylvie Perreau, Langford B. White, and Pierre Duhamel, *Fellow, IEEE*

Abstract—A new type of blind decision feedback equalizer (DFE) incorporating fixed lag smoothing is developed in this paper. The structure is motivated by the fact that if we make full use of the dependence of the observed data on a given transmitted symbol, delayed decisions may produce better estimates of that symbol. To this end, we use a hidden Markov model (HMM) suboptimal formulation that offers a good tradeoff between computational complexity and bit error rate (BER) performance. The proposed equalizer also provides estimates of the channel coefficients and operates adaptively (so that it can adapt to a fading channel for instance) by means of an online version of the expectation-maximization (EM) algorithm. The resulting equalizer structure takes the form of a linear feedback system including a quantizer, and hence, it is easily implemented. In fact, because of its feedback structure, the proposed equalizer shows some similarities with the well-known DFE. A full theoretical analysis of the initial version of the algorithm is not available, but a characterization of a simplified version is provided. We demonstrate that compared to the zero-forcing DFE (ZF-DFE), the algorithm yields many improvements. A large range of simulations on finite impulse response (FIR) channels and on typical fading GSM channel models illustrate the potential of the proposed equalizer.

I. INTRODUCTION

SEVERAL methods based on a hidden Markov model (HMM) formulation were recently proposed for non-linear blind equalization. An optimal off-line algorithm was presented in [1], performing alternatively forward and backward recursions on a whole block of data. After several iterations, the algorithm provides a reliable estimate of the channel parameters as well as a detection of the emitted sequence of symbols. Such methods can become intractable when dealing with long channels since the computational cost is exponentially increasing with the channel memory. Moreover, the off-line method has large memory requirements and cannot perform real-time processing, which can become an issue in the case of time-varying channels. An “online” algorithm was derived in [2], thus allowing a real-time computation. The derivation of this recursive algorithm was made feasible by applying the backward recursions on a fixed number of “future” observations, i.e., the algorithm is applied with a

delay. However, the issue of the computational complexity of the HMM formulation still remained. This was the motivation for the proposal of a suboptimal algorithm in [3]. The idea was to apply the forward recursion on a reduced number of symbols stored in the channel memory, where the decisions on the other symbols were no longer being questioned. This approach has strong connections with the DFE since, on some symbols, the decision that has been taken is no longer questioned and is used to remove the ISI. As a result, the algorithm in [3] has the same type of behavior as the DFE (although of reduced order) since a wrong detection may result in error propagation, especially when dealing with nonminimum-phase channels.

This paper proposes another suboptimal HMM formulation with a computational complexity linear in the channel memory. When compared with the DFE, it can be seen that both algorithms feedback symbol estimates before detection. The key difference is that, although the DFE never questions the estimates later on, the proposed algorithm produces *smoothed* estimates, that is, updates at each time the estimates of each symbol stored in the channel memory and feeds these updates back at the next iteration. Finally, the decision on a symbol is made the last time this symbol is seen by the channel memory, which justifies the term *fixed-lag smoothing*.

The paper is organized as follows. We first propose a general description of the algorithm briefly recalling the HMM definitions and recursions used for the symbol detection, as well as the steps involved in the expectation-maximization (EM) algorithm, allowing the estimation of the channel parameters. In Section IV, we derive a simplified version of the proposed algorithm, which allows a theoretical bit error rate (BER) analysis, and we finally propose, in Section V, a large range of simulations, in terms of BER and channel tap estimation accuracy.

II. GENERAL DESCRIPTION OF THE ALGORITHM

A. Signal Model and Problem Definition

We assume that symbols taken from a finite alphabet \mathcal{A} of size M , are transmitted through an FIR channel with transfer function $H(z) = \sum_{i=0}^{N-1} h_i z^{-i}$ where the h_i are complex-valued coefficients. We assume additive Gaussian white noise with zero mean and variance σ^2 . The received signal is thus modelled by

$$y_t = H^T X_t + n_t \quad (1)$$

where H and X_t are defined by

$$\begin{aligned} H^T &= [h_0 \quad h_1 \quad \cdots \quad h_{N-1}] \\ X_t &= [x_t \quad x_{t-1} \quad \cdots \quad x_{t-N+1}]^T. \end{aligned} \quad (2)$$

Manuscript received August 13, 1996; revised February 19, 1999. This work was partly supported by Alcatel Telecom, France, the Australian Government through the Co-operative Research Centres Program, and the Department of Defence of Australia. The associate editor coordinating the review of this paper and approving it for publication was Dr. Petar M. Djuric.

S. Perreau is with the University of South Australia, The Levels, Australia.

L. B. White is with the Department of Electrical and Electronic Engineering, the University of Adelaide, Adelaide, Australia.

P. Duhamel is with Department Signal, Telecom Paris, Paris, France.

Publisher Item Identifier S 1053-587X(00)02376-X.

Here, the operator $()^T$ denotes the transposition operation. The task of the algorithm is thus to estimate at each time t the transmitted symbols $x_t, x_{t-1}, \dots, x_{t-N+1}$, as well as providing an estimate of the channel impulse response, using the measurements $Y_t = \{y_1, \dots, y_t\}$.

B. Approximate MAP Symbol Detection

In [3], an optimal formulation for maximum *a posteriori* estimation of the transmitted symbols was presented. This formulation is based on an HMM formulation. Indeed, the vector X_t can be seen as the state vector of the Markov process described by the state equation

$$X_{t+1} = AX_t + x_{t+1} * [1 \ 0 \ \dots \ 0]^T \quad (3)$$

where A is a shift matrix with $A_{i,j} = 1 \Leftrightarrow i = j + 1$. This Markov process is only observable through the observation equation (1). Suppose that \hat{H}_t , which is the current estimate of the channel, is available at time t . As in [2], define the so-called forward variable, expressing the probability that the state X_t is equal to some realization $[q_{i_0} \dots q_{i_{(N-1)}}]^T$ according to the current channel estimate \hat{H}_t , and the set of measurements Y_t by

$$\alpha_t(i_0, i_1, \dots, i_{N-1}) = \Pr(X_t^T = [q_{i_0} \dots q_{i_{(N-1)}}] | \hat{H}_t, Y_t). \quad (4)$$

The exact computation of this probability involves the so-called forward recursion. See [1] for more details. This recursion requires the calculation of the probability in (4) for every possible realization of the stochastic process X_t . Such an evaluation obviously requires the computation of M^N probabilities at each step. In [3], a reduced computation algorithm based on state truncation was introduced. This formulation, however, appeared to perform rather poorly in comparison with the optimal HMM filter, in particular when operating in an adaptive mode. This property seems to be linked to the fact that a classical DFE is the coarsest approximation of the algorithm in [3], whereas the superior performances of the HMM may be due to the fact that past decisions are revisited several times over the duration of the channel memory and, hence, may be corrected. Thus, error propagation as often observed with DFE's is more unlikely to happen. This is even more important in an adaptive mode, where too many errors can lead to anomalous adaptive behavior.

Thus, it is desirable to seek a simplified algorithm that permits state revisiting but does not have the exponential complexity in N of the approaches of [1] and [2]. Such an algorithm was presented in [5]. This algorithm uses the *marginal* posterior probabilities of the symbols in the channel, rather than the joint probabilities described by (4), as in [2], or a partially quantized joint probability, as in [3]. Such an algorithm has linear computational complexity in the channel duration. We now describe the forward recursion for this algorithm.

Assume that the following quantities are available at time t :

- the approximate filtered probabilities $\alpha_{t-1|t-1}^{(n)}(i)$ denoting the probability that the $n + 1^{\text{th}}$ symbol in the channel memory at time $t - 1$ are equal to q_i , knowing the observations up to time $t - 1$, which is the estimate of the

channel parameters denoting (\hat{H}_{t-1}) , and the prediction of the other symbols stored in the channel memory at time $t - 1$ $[X_{t-1}^{(m)}, \forall m \neq n, X_{t-1}^{(m)}$ denoting the $m + 1$ th component of vector X_{t-1}]

$$\alpha_{t-1|t-1}^{(n)}(i) = P(X_{t-1}^{(n)} = q_i | \hat{H}_{t-1}, Y_{t-1}, X_{t-1}^{(m)} = \hat{X}_{t-1|t-2}^{(m)} \forall m \neq n \quad \forall n = 0, \dots, N - 1, q_i \in \mathcal{A}; \quad (5)$$

- the current estimate \hat{X}_{t-1} of the vector X_{t-1} as given by the previous recursion. A prediction $\hat{X}_{t|t-1}$ of vector X_t is easily obtained by taking advantage of the shift structure of the process X_t . Clearly, we have, for $n = 1 \dots N - 1$

$$\hat{X}_{t|t-1}^{(n)} = \hat{X}_{t-1|t-1}^{(n-1)}. \quad (6)$$

Then, by substituting $\hat{X}_{t|t-1}^{(n)}$ for $X_t^{(n)} \forall n = 1, \dots, N - 1$, we obtain the approximate filtered probability at time t of the only component of the state vector on which (6) does not provide information

$$\alpha_{t|t}^{(0)}(i) = P(X_t^{(0)} = q_i | \hat{H}_t, Y_t, X_t^{(n)} = \hat{X}_{t|t-1}^{(n)}). \quad (7)$$

Substituting from (1) yields

$$\alpha_{t|t}^{(0)}(i) = c_t^{(0)} N(y_t - \hat{H}_t^T [q_i, \hat{X}_{t|t-1}^{(1)}, \dots, \hat{X}_{t|t-1}^{(N-1)}]^T) \quad (8)$$

where $c_t^{(0)}$ is a normalizing constant, and $N(\cdot)$ is a zero mean Gaussian function with variance σ^2 . In the forthcoming, $L_t^{(n)}(i)$ denotes the quantity

$$L_t^{(n)}(i) = N(y_t - \hat{H}_t^T [\hat{X}_{t|t}^{(0)}, \hat{X}_{t|t-1}^{(1)}, \dots, q_i, \dots, \hat{X}_{t|t-1}^{(N-1)}]^T) \quad (9)$$

where $\hat{X}_{t|t-1}^{(n)}$ has been replaced by q_i .

The remaining updated probabilities involved in (5) are also approximated by applying the classical forward recursion of the HMM formulation on conditional instead of joint probabilities. The quantities $\alpha_{t|t}^{(n)}(i)$ recorded as smoothed probabilities are thus obtained as

$$\alpha_{t|t}^{(n)}(i) = c_t^{(n)} \alpha_{t-1|t-1}^{(n-1)}(i) L_t^{(n)}(i). \quad (10)$$

Hence, the conditional mean (CM) estimate at time t of symbol x_{t-n} denoted by $\hat{x}_{t-n|t}$ is

$$\hat{x}_{t|t}^{(n)} = \hat{x}_{t-n|t} = \sum_{i=1}^M q_i \alpha_{t|t}^{(n)}(i) \quad (11)$$

and the new state estimate of vector X_t is given by

$$\hat{X}_{t|t} = [\hat{x}_{t|t}, \dots, \hat{x}_{t-n|t}, \dots, \hat{x}_{t-N+1|t}]^T. \quad (12)$$

Note that (12) takes into account the probabilities computed on the previous step (which are used as predictions here) as well as the effect that a particular symbol may have on the observed data at time t , given other filtered and predicted estimates. Equations (10) and (11) illustrate one key difference between our algorithm and the DFE [6]. In the DFE, once the current symbol has been estimated, this estimate never changes, whereas our algorithm puts the decisions back into question, as long as the

considered symbol is seen by the channel. Thus, in principle, error propagation should occur less often.

Consider the case of large SNR. In this situation, the Gaussian functions should be sufficiently peaked in order to ensure the probability estimates in (9) to be very close to unity for one symbol and close to zero for the other ones. Under this approximation, CM estimates behave like maximum *a posteriori* probability (MAP) estimates and correspond to one of the symbol values. From (11), this symbol value is clearly the closest symbol to the current measurement with the *estimated* ISI removed

$$\begin{aligned}\hat{X}_{t|t}^{(0)} &= \hat{x}_{t|t} = \arg \max_{q_i \in \mathcal{A}} N(y_t - \hat{H}_t^T(q_i, \hat{X}_{t|t-1}^{(1)}, \dots, \hat{X}_{t|t-1}^{(N-1)})) \\ &= \arg \min_{q_i \in \mathcal{A}} |y_t - \hat{H}_t^T(q_i, \hat{X}_{t|t-1}^{(1)}, \dots, \hat{X}_{t|t-1}^{(N-1)})|. \quad (13)\end{aligned}$$

Now, we concentrate on the smoothed probabilities (10) at time t . Substituting for $\alpha_{t-1|t-1}^{(n)}(i)$, we have

$$\begin{aligned}\alpha_{t|t}^{(n)}(i) &= c_t^{(n)} c_{t-1}^{(n-1)} L_t^{(n)}(i) L_{t-1}^{(n-1)}(i) \alpha_{t-2|t-2}^{(n-2)} \\ &= \prod_{k=0}^n c_{t-k}^{(n-k)} L_{t-k}^{(n-k)}(i). \quad (14)\end{aligned}$$

Thus, maximizing over i , we obtain a refined delayed estimate of (13) as given below. This is called the “*smoothed estimate*”

$$\begin{aligned}\hat{X}_{t|t}^{(n)} &= \hat{x}_{t-n|t} = \arg \max_{q_i \in \mathcal{A}} \prod_{k=0}^n c_{t-k}^{(n-k)} L_{t-k}^{(n-k)}(i) \\ &= \arg \min_{q_i \in \mathcal{A}} \sum_{k=0}^n |y_{t-k} - H_{t-k}^T \mathcal{X}_{t-k}^{(n-k)}(i)|^2 \\ &= \arg \min_{q_i \in \mathcal{A}} C_{t-n|t}(q_i) \quad (15)\end{aligned}$$

where

$$\begin{aligned}\mathcal{X}_{t-k}^{(n-k)}(i) &= (\hat{x}_{t-k|t-k}, \hat{x}_{t-k-1|t-k-1}, \dots, \hat{x}_{t-k-n+1|t-k-1} \\ &\quad q_i, \hat{x}_{t-k-n-1|t-k-1}, \dots, \hat{x}_{t-k-N+1|t-k-1})^T. \quad (16)\end{aligned}$$

C. Approximate ML Estimation of the Channel Impulse Response

Once the symbol estimates are obtained, it is well known that since the noise model is Gaussian, the ML estimate of the channel can be obtained through a least squares solution. However, we have just seen that the estimate $\hat{x}_{t|t}$ can further be improved by estimates $\hat{x}_{t|t+k}$. Rather than directly utilising the classical recursive least squares (RLS) (with shift invariant property), we derive below the precise form of the estimation algorithm by use of the EM algorithm. We briefly recall here the various steps involved in this EM algorithm. See [2], [3], [5], and [7] for a complete description of this algorithm.

1) *Expectation Step*: This step corresponds to the computation of an estimate of the Kullback–Leibler (KL) function at time t , which is defined as the expectation of the logarithm of

the likelihood function for the complete data (see [7] for the terminology)

$$\begin{aligned}Q_t(\hat{H}_t, H) &= \sum_{i_0, \dots, i_{N-1}} P(X_t = [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}] | \hat{H}_t, Y_t) \\ &\quad \cdot \log(P(X_t = [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}] | y_t, H)) \\ &= \left(-\frac{1}{2\sigma^2}\right) \sum_{i_0, \dots, i_{N-1}} P(X_t = [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}] | \hat{H}_t, Y_t) |y_t \\ &\quad - [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}] H|^2. \quad (17)\end{aligned}$$

Now, focus on the computation of $P(X_t = [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}] | \hat{H}_t, Y_t)$.

In order to be compatible with the approximation derived in the previous section, we approximate this joint probability of every component of X_t by the product of the conditional marginal probabilities of each component, given the one-step predictions of the other components according to (5)

$$\begin{aligned}P(X_t = [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}] | \hat{H}_t, Y_t) &\approx \prod_{n=0}^{N-1} P(X_t^{(n)} = q_{i_n} | \hat{H}_t, Y_t, \hat{X}_{t|t-1}^{(m)}) \\ &= \prod_{n=0}^{N-1} \alpha_{t|t}^{(n)}(i_n). \quad (18)\end{aligned}$$

Appendix A shows the expansion of the calculation leading to the following expression of the expected *pseudo-likelihood* at time t

$$Q_t(\hat{H}_t, H) = \left(\frac{-1}{2\sigma^2}\right) |y_t - H^T \hat{X}_{t|t}|^2. \quad (19)$$

2) *Maximization Step*: This consists of maximizing $L_t(\hat{H}_0, \hat{H}_1, \dots, \hat{H}_t, H)$ as a function of H , since it has been shown [2], [7] that this maximization increases the likelihood of the observations. The function to be maximized $L(\hat{H}_0, \hat{H}_1, \dots, \hat{H}_t, H)$ is evaluated in [2] as

$$L_t(\hat{H}_0, \hat{H}_1, \dots, \hat{H}_t, H) = \left(\frac{-1}{2\sigma^2}\right) \sum_{k=0}^t Q_k(\hat{H}_k, H). \quad (20)$$

An online algorithm has the ability of tracking time-varying parameters. This property will be ensured by introducing a forgetting constant $0 \ll \lambda < 1$ in the estimation of $L_t(\hat{H}_0, \hat{H}_1, \dots, \hat{H}_t, H)$

$$L_t(\hat{H}_0, \hat{H}_1, \dots, \hat{H}_t, H) = \left(\frac{-1}{2\sigma^2}\right) \sum_{k=0}^t \lambda^{(t-k)} Q_k(\hat{H}_k, H). \quad (21)$$

The maximization step is realized recursively [2] by

$$\hat{H}_{t+1} = \hat{H}_t + R_t^{-1} S(y_t, \hat{H}_t) \quad (22)$$

where R_t is the Fisher information matrix of the complete data defined below, and the score S is defined by [5]

$$S(y_t, \hat{H}_t) = (y_t - \hat{H}_t^T \hat{X}_{t|t}) \hat{X}_{t|t}^* \quad (23)$$

where $()^*$ and $()^\dagger$, respectively, stand for the conjugation and transposed conjugation operations. This results in the following recursions:

$$\begin{aligned}\hat{H}_{t+1} &= \hat{H}_t + R_t^{-1}S(y_t, \hat{H}_t) \\ R_t &= \lambda R_{t-1} + \hat{X}_{t|t} \hat{X}_{t|t}^\dagger.\end{aligned}\quad (24)$$

Equation (24) correspond to the recursive estimation of the channel parameters by means of an EM sequential algorithm. Since the algorithm requires a recursive computation involving an exponential factor, it is not astonishing that the recursions show some similarity with the ones used in a classical RLS adaptive algorithm, where the so-called missing data X_t are replaced by their estimate $\hat{X}_{t|t}$. Note that an LMS (gradient) like version of the algorithm is also feasible, although we do not address its performance here.

D. The Proposed Algorithm

To summarize, the overall algorithm runs the following recursions:

- 1) Suppose a current estimate of X_{t-1} is available: $\hat{X}_{t-1|t-1}$
- 2) Predict vector X_t by (6): $\hat{X}_{t|t-1}$
- 3) Estimate vector X_t by the quantizations in (16) and (21): $\hat{X}_{t|t}$
- 4) Use (23) and (24) for estimating, recursively, the channel impulse response
- 5) $t = t + 1$, and return to step 1.

E. Initialization Issues

Since the proposed algorithm is an iterative one, initialization has to be addressed. As far as the estimates of the transmitted symbols are concerned, as no training sequence is available, we choose to initialize the marginal probabilities as uniform.

The initialization of the channel can be an issue. As the algorithm is based on the maximum likelihood approach, it is well known that it might converge to some local maxima of the likelihood function. This issue is not treated here but has been considered in [8], where this problem is overcome by coupling the proposed algorithm to a subspace-based method in the case of spatial diversity reception; in [9] in case of encoded data, where the knowledge of the convolutional code is shown to reduce significantly this initialization issue; or in [10], using a Gibbsian technique that significantly decreases the possibility of convergence to the basin of attraction of a nonglobal maximum of the likelihood. In any case, as the algorithm derived in this paper performs online, the issue of local maxima is significantly reduced compared with the offline context; in fact, in an online setting, as the algorithm does not increase the likelihood of the complete data at every step, it does not necessarily converge to the closest local maximum. Simulations showed very few cases (treated in [8]–[10]) where the algorithm did not converge to the expected solution. In the simulation part, unless stated, all taps

were initialized to the value 0, which means that we did not use any knowledge on the channel.

III. COMPUTATIONAL COMPLEXITY

Denoting by N the length of the channel and by M the size of the input alphabet, the computation of the scheme using MAP estimates of the symbols requires the following at each iteration:

- $N - 1$ multiplications to compute $H^T \hat{X}_{t|t-1}$;
- M multiplications to produce $\hat{X}_{t|t}^{(0)}$ via (13);
- M multiplications for each symbol estimate to be updated in the channel memory via (15), that is, a total of $M*(N - 1)$ multiplications;
- the update of the channel parameters requiring the same computational complexity as for the RLS algorithm: $O(N^2)$.

We have a total of $M*N + N - 1$ multiplications required at each iteration plus those required for the parameter estimation. Note that the computational complexity of a DFE would require $N - 1$ multiplications. The next sections compare this algorithm with the DFE in terms of error propagation and BER. This study is undertaken under the assumption that the channel is properly estimated.

IV. SIMPLIFIED “SHIFT-INVARIANT” ALGORITHM BER ANALYSIS

In this section, we propose an exact calculation of a lower bound for the probability to recover from an error, in the general case, and asymptotic error probabilities for every possible first-order channels. These exact calculations are performed using a simplification of the proposed algorithm that yields a shift-invariant structure. With this shift-invariant structure, we obtain a first-order Markov chain model for the error process, which allows for a theoretical analysis. This new algorithm does not provide as good results as the original one, but we show that the corresponding error probabilities are always smaller than those for the ZF-DFE. Note that the second analysis only holds for first-order channels because the mathematical derivation for a n th-order channel (channel with $n + 1$ taps) is intractable. However, it is useful because when an error starts to propagate for any kind of linear channels, the expressions of the probabilities are the same as the ones for a first-order channel. The analysis below is then interesting for understanding how an error will start propagating throughout the channel memory. In addition, BER simulations in Section V confirm some claims made in this analysis.

For an easier analysis, all variables are now supposed to be real valued, and the emitted symbols belong to the set $\{-1, +1\}$.

A. The Shift-Invariant Algorithm

This algorithm is simply obtained by using the approximation

$$\hat{X}_{t|t-1}^{(i)} \approx \hat{X}_{t|t-i}^{(i)} \quad (25)$$

which amounts to use the “filtered” estimates in the feedback paths rather than the “smoothed” estimates in (13), (15), and (16). Then, considering (13), we have

$$\begin{aligned} \hat{x}_{t|t} &= \arg \min_{q_i \in \{-1; +1\}} |y_t - \hat{H}_t^T(q_i, \hat{X}_{t|t-1}^{(1)}, \hat{X}_{t|t-1}^{(N-1)})|. \\ &= \arg \min_{q_i \in \{-1; +1\}} \left\{ y_t^2 + \left(\sum_{k=1}^{N-1} h_k \hat{X}_{t|t-1}^{(k)} \right)^2 \right. \\ &\quad \left. - 2h_0 q_i \left(y_t - \sum_{k=1}^{N-1} h_k \hat{X}_{t|t-1}^{(k)} \right) \right\}. \end{aligned} \quad (26)$$

Assuming h_0 is a positive quantity (the symmetric argument can be used in case it is negative), the above criterion is minimized when q_i has the same sign as the quantity $(y_t - \sum_{k=1}^{N-1} h_k \hat{X}_{t|t-1}^{(k)})$. Thus, applying the same procedure for $\hat{x}_{t-n|t}$, we get

$$\begin{aligned} \hat{x}_{t-n|t} &= \text{sign } z_1(t) \quad \text{where} \\ z_1(t) &= \sum_{k=0}^n h_{n-k} y_{t-k} - \sum_{k=0}^{2n} p_{n,k} \hat{x}_{t-k|t-k} \end{aligned} \quad (27)$$

$$\hat{x}_{t|t} = \text{sign } z_2(t) \quad \text{with } z_2(t) = y_t - \sum_{k=1}^n h_k \hat{x}_{t-k|t-k} \quad (28)$$

where the coefficients $p_{n,k}$ are given by

$$p_k = \sum_{l=0}^k h_{n-l} h_{n-k}$$

for $k = 0, \dots, n-1$, $h_n = 0$, and $h_{2n-k} = h_k$ for $k = n+1, \dots, 2n$.

In the frequency domain, denoting by $\tilde{H}(z)$ the reversed polynomial of $H(z)$, e.g., $\tilde{H}(z) = z^n H(z^{-1})$ and $H_0(z)$ defined as $H_0(z) = H(z) - h_0$, (27) and (28) are written as

$$\begin{aligned} Z_1(z) &= \tilde{H}(z) Y(z) - P(z) \hat{X}(z) \\ Z_2(z) &= Y(z) - H_0(z) \hat{X}(z). \end{aligned} \quad (29)$$

This allows us to represent the structure of the proposed algorithm as a two-input two-output linear system denoted by $T(z)$ with a feedback nonlinearity, as shown in Fig. 1. Evidently, $T(z)$ is a 2×2 matrix polynomial of degree $2n$ and can be computed for approximately four times the computation of the usual DFE.

Note that in this shift-invariant structure, error propagation may occur in the zero lag estimation, e.g., DFE, as usual. The shift-invariant smoother provides other subsequent estimates by simply weighting the posterior likelihoods associated with each delay. This has the effect of reducing overall error rate, as will

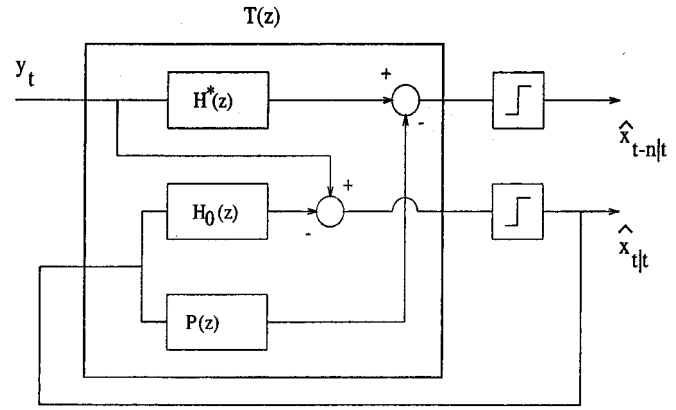


Fig. 1. Structure of lag n DFE.

be demonstrated with a simple example. It has to be pointed out that this shift-invariant structure is not as effective as the algorithm proposed in the first section in that it does not take advantage of the updates of the decisions before retrieving the corresponding ISI. However, we need the simplified structure in order to perform the error probability calculations.

B. Lower Bound on the Probability to Recover from an Error for a BPSK Modulation

1) *General Case Analysis:* This section provides a general expression for the probability to recover from an error without making any assumption on the channel. This general expression shows that the BER calculation in the general case is intractable, but it does permit a better understanding of the advantage of the proposed algorithm.

Assume that the estimate of symbol x_t at time t is wrong, i.e., $|e_{t|t}| = |\hat{x}_{t|t} - x_t| = 2$, whereas the previous symbols were correctly estimated. What is the probability for the next estimate of the same symbol at time $t+n$ to be correct? Denote this probability as $P(e_{t|t+n} = 0 | e_{t|t} \neq 0)$. The corresponding cost function to be minimized according to q_i , which is defined in (15), can be written as

$$\begin{aligned} C_{t|t+n}(q_i) &= (h_0(x_t - q_i) + n_t)^2 + (h_0 e_{t+1|t+1} \\ &\quad + h_1(x_t - q_i) + n_{t+1})^2 + \dots + (h_0 e_{t+n|t+n} \\ &\quad + \dots + h_n(x_t - q_i) + n_{t+n})^2. \end{aligned} \quad (30)$$

Let us denote

$$\Gamma_i = h_0 e_{t+i|t+i} + h_1 e_{t+i-1|t+i-1} + \dots + h_{i-1} e_{t+1|t+1}.$$

Then, the condition for a good detection on symbol x_t would be that the cost function $C_{t|t+n}(q_i)$ calculated with $q_i = x_t$ is less than the one obtained for $q_i = -x_t$, i.e.,

$$\sum_{i=0}^n (\Gamma_i + n_{t+i})^2 < \sum_{i=0}^n (\Gamma_i + n_{t+i} + h_i e_{t|t})^2. \quad (31)$$

Assume, as well, that $x_t = 1$ (the case $x_t = -1$ leads to the same probability because of the symmetry of the problem); then, $e_{t|t} = 2$, and the last equation becomes

$$\sum_{i=0}^n h_i^2 + \sum_{i=0}^n h_i (\Gamma_i + n_{t+i}) > 0 \quad (32)$$

or, equivalently

$$\sum_{i=0}^n h_i n_{t+i} > -\sum_{i=0}^n h_i^2 - \sum_{i=0}^n \Gamma_i h_i. \quad (33)$$

If we call $N_{t+1} = \sum_{i=1}^n h_i b_{t+i}$, which is a Gaussian random variable of mean 0, and variance $\sigma^2 \sum_{i=1}^n h_i^2$, we want to calculate the probability to recover x_t at time $t+n$, given that $e_{t|t} = 2$. First, denote by $P_1(\xi_i)$ the probability to recover x_t at time $t+n$ under the following assumptions.

- $e_{t|t} = 2$.
- The random process of errors $E_{t+1} = (e_{t+n|t+n}, \dots, e_{t+1|t+1})$ is equal to some ξ_i , which is one of the 3^n possible realizations of E_{t+1} (each component of the vector ξ_i can take the values 2, -2, or 0). Thus

$$P_1(\xi_i) = P \left(N_{t+1} > \underbrace{-\sum_{i=0}^n h_i^2 - h_0 n_t + \sum_{i=0}^n h_i \Gamma_i}_{\delta(\xi_i)} \middle| \underbrace{n_t}_{e_{t|t}=2} < -h_0 \right) \quad (34)$$

Then, we have to evaluate the probability for the error sequence E_{t+1} to be equal to ξ_i

$$\begin{aligned} & P((e_{t+n|t+n}, \dots, e_{t+1|t+1}) \\ &= \xi_i | e_{t|t} = 2) = \prod_{j=1}^n P(e_{t+n-j+1|t+n-j+1} \\ &= \xi_i^{(j)} | e_{t+n-j|t+n-j} = \xi_i^{(j+1)} \dots e_{t+1|t+1} = \xi_i^{(n)}, e_{t|t} = 2). \end{aligned} \quad (35)$$

Finally, we obtain, as a result, an evaluation of the probability $P(\hat{x}_{t|t+n} = x_t | e_{t|t} \neq 0)$ (that a symbol estimate is correct at time $t+n$, knowing it was wrong at time t) by

$$\begin{aligned} & P(\hat{x}_{t|t+n} = x_t | e_{t|t} \neq 0) \\ &= \sum_i P_1(\xi_i) P((e_{t+n|t+n}, \dots, e_{t+1|t+1}) = \xi_i | e_{t|t} = 2). \end{aligned} \quad (36)$$

Using this general result, an approximation valid only for high SNR is given below.

2) *Derivation of a Lower Bound for High SNR Levels:* Note that all the probabilities involved in (36) are Gaussian functions.

Hence, the following approximations are valid for high SNR levels.

- As the distribution functions of a very narrow Gaussian function can be accurately approximated by 1, $\frac{1}{2}$, or 0, we just take into account the sequences E_{t+1} leading to non-zero values of the exponential functions. We denote such sequences as $\xi_{\min}(i)$.
- On the other hand, we show in Appendix B that the high SNR level assumption allows the derivation of a lower bound for the probabilities $P((e_{t+n|t+n}, \dots, e_{t+1|t+1}) = \xi_i | e_{t|t} = 2)$. We find out that

$$P((e_{t+n|t+n}, \dots, e_{t+1|t+1}) = \xi_i | e_{t|t} = 2) \geq (\frac{1}{2})^{(n)}.$$

Finally, we have, from (36)

$$P(\hat{x}_{t|t+n} = x_t | e_{t|t} \neq 0) \geq (\frac{1}{2})^{(n)} \sum_i P_1(\xi_{\min}(i)). \quad (37)$$

We can comment on these results.

- This lower bound will hold more tightly when the channel is “pathological,” i.e., when the only possibility for a symbol to be correctly detected is that the sign of the filtering error is the same one as that of the symbol.
- The bound explains why the exact calculation in general is intractable without any assumption on the channel taps; the functions $\delta(\xi_{\min}(i))$, as well as the “plausible” error sequences $\xi_{\min}(i)$, depend explicitly on the channel parameters values.
- We can also find a particular case of channels for which we know that *this probability will be close to unity*, which means that nearly all the errors detected at time t will be corrected later on.
- The high SNR requirement is necessary to approximate the Gaussian functions by 1, $\frac{1}{2}$, or 0. For lower SNR’s, these approximations become less accurate. However, in this case, we can still provide a lower and upper bounds for the probability in (36), as explained below.

Suppose, for instance, the second channel tap absolute value is much greater than the first one. It is useful to deal with this situation for two reasons: First, as shown by simulations in [12], it is met, in particular, for mobile communications transmissions affected with fading and Doppler effects; second, this configuration would lead the worst error propagation schemes for a ZF-DFE when a wrong detection is made on a symbol. Therefore, it is necessary to check how the proposed algorithm performs in such a pathological situation. The opposite situation (second tap smaller than the first one) is not of interest because it would not lead to error propagation.

Suppose an error has occurred at time t . The remaining of this section provides order of magnitudes of this recovery probability in one iteration, which are checked against simulations. The condition (36) to recover from this error at time $t+1$ (that is, for $n=1$ according to the previous calculations) becomes

$$P(\hat{x}_{t|t+1} = x_t | e_{t|t} \neq 0) \geq (\frac{1}{2})(P_1(\xi_{\min}(0)) + P_1(\xi_{\min}(1))). \quad (38)$$

First, determine each possible realization of vector E_{t+1} , that is, assuming that $e_{t|t} = 2$, we have the permitted values of $e_{t+1|t+1}$. Recall that symbol x_{t+1} is estimated at time $t+1$ via

$$\hat{x}_{t+1|t+1} = \text{sign}(y_{t+1} - h_1 \hat{x}_{t|t}) = \text{sign}(h_0 x_{t+1} + 2h_1). \quad (39)$$

We have to take into account two different possibilities for the error process, corresponding to the two possible values for $e_{t+1|t+1}$: i) $e_{t+1|t+1} = 0$, corresponding to the error sequence ξ_0 and ii) $e_{t+1|t+1} = -2$ if x_{t+1} is equal to -1 , corresponding to the error sequence ξ_1 . Now, determine the values $\delta(\xi_i)$

$$\begin{aligned} \delta(\xi_0) &= -h_0^2 - h_1^2 - h_0 n_t + 2h_0^2 \\ &= h_0^2 - h_1^2 - h_0 n_t \end{aligned} \quad (40)$$

$$\begin{aligned} \delta(\xi_1) &= -h_0^2 - h_1^2 - h_0 n_t + 2h_0^2 - 2h_0 h_1 - 2h_1^2 \\ &= (h_0 - h_1)^2 - h_0 n_t - 2h_1^2. \end{aligned} \quad (41)$$

Now, to evaluate the corresponding probabilities

$$P_1(\xi_0) = P(h_1 n_{t+1} > \delta(\xi_0) | n_t < -h_0) \quad (42)$$

$$P_1(\xi_1) = P(h_1 n_{t+1} > \delta(\xi_1) | n_t < -h_0). \quad (43)$$

Then, for high SNR, and given the assumption that h_1 is much greater than h_0 , we can note that ξ_0 is almost surely negative, as is ξ_1 , so that both $P_1(\xi_0)$ and $P_1(\xi_1)$ will be close to unity. For poorer SNR levels, we can say that $P_1(\xi_0)$ satisfies

$$0 < P_1(\xi_1) \leq \frac{1}{2}.$$

As a consequence, for high SNR levels, (38) becomes

$$P(\hat{x}_{t|t+1} = x_t | e_{t|t} \neq 0) \approx 1.$$

For lower SNR, we have

$$\frac{1}{2} \leq P(\hat{x}_{t|t+1} = x_t | e_{t|t} \neq 0) \leq \frac{3}{4}.$$

These results are checked by simulations on Table I for a typical GSM real-valued channel corresponding to a hilly area $h = [0.5, 1, 0, 0, 0.63, 0, 0, 0, 0.25, 0, 0, 0, 0.16]$ fed by a white unit amplitude BPSK input constellation.

It is clearly seen that high SNR allows an almost sure correction of an error, whereas a 7-dB SNR still allows half the errors to be recovered.

3) *Error Probability Calculations for the Shift-Invariant Structure—First-Order Channel:* Now, consider a first-order channel $H(z) = 1 + hz^{-1}$. The decision equations for the DFE and lag one shift-invariant smoother are

$$\hat{x}_{t|t} = \text{sign}(y_t - h\hat{x}_{t-1|t-1}) \quad (44)$$

$$\hat{x}_{t-1|t} = \text{sign}(y_{t-1} - h\hat{x}_{t-2|t-2} + h(y_t - \hat{x}_{t|t})). \quad (45)$$

Substituting for y_t yields

$$\hat{x}_{t|t} = \text{sign}(x_t + n_t + h e_{t-1|t-1}) \quad (46)$$

$$\begin{aligned} \hat{x}_{t-1|t} = \text{sign}(h(e_{t|t} + e_{t-2|t-2}) \\ + (1 + h^2)x_{t-1} + h n_t + n_{t-1}) \end{aligned} \quad (47)$$

where $e_{t-1|t-1} = x_{t-1} - \hat{x}_{t-1|t-1}$ and similarly for other quantities.

Now, we will evaluate the asymptotic error performances. We define this ‘‘asymptotic error’’ by $\lim_{t \rightarrow \infty} e_{t|t}$. Let $S_t = (e_{t|t}, e_{t-1|t-1}, e_{t-2|t-2})$; then, we claim that S_t and $e_{t-1|t}$ form the states and observations, respectively, of a hidden Markov process. This is easily seen by observing from (46) that $e_{t|t}$ is a first-order scalar Markov chain. Clearly, if we denote by Π_t the vector

$$\Pi_t = (P(e_{t|t} = 0), P(e_{t|t} = 2), P(e_{t|t} = -2))^T \quad (48)$$

we can write the state transition equation

$$\Pi_t = A^T \Pi_{t-1}. \quad (49)$$

The transition probabilities, which are the coefficients of the state transition matrix A and the asymptotic error probabilities, are determined in Appendix C. These asymptotic error probabilities are the coefficients of the vector $\Pi_\infty = \lim_{t \rightarrow \infty} \Pi_t$ solution of

$$\Pi_\infty = A^T \Pi_\infty.$$

TABLE I
PERCENTAGE FOR RECOVERING AT TIME
 $t + 1$ FROM AN ERROR COMMITTED AT TIME t FOR DIFFERENT SNR

| SNR (dB) | 15 | 13 | 8 | 7 | 5 |
|---------------------|-------|-------|-------|-------|-------|
| error recovery in % | 97.7% | 92.8% | 74.4% | 69.6% | 51.6% |

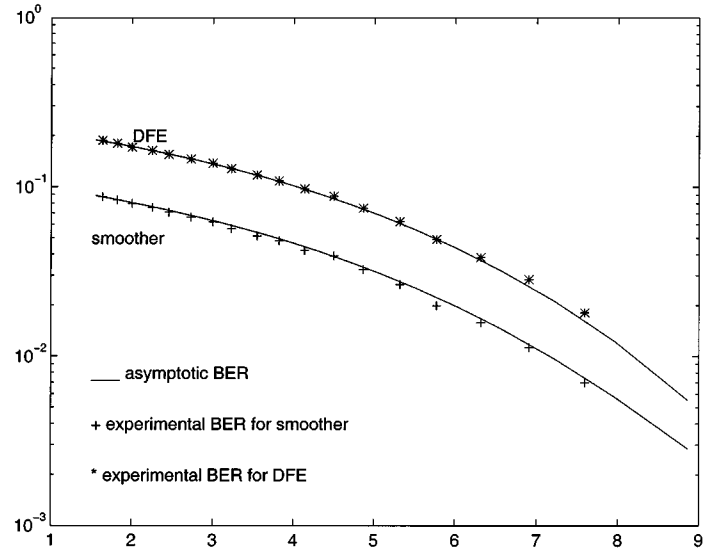


Fig. 2. Asymptotic and experimental BER on the channel $h = [1 \ 2]$.

Due to the symmetries of the error process, Π_∞ has the form

$$\Pi_\infty = \left(\pi_0, \frac{1 - \pi_0}{2}, \frac{1 - \pi_0}{2} \right)^T \quad (50)$$

where π_0 is defined in Appendix C.

Now, consider the shift-invariant smoother (47). Observe that the detected symbol depends only on the sum $f_t = e_{t|t} + e_{t-2|t-2}$ via

$$\hat{x}_{t-1|t} = \text{sign}(h f_t + (1 + h^2)x_{t-1} + m_t) \quad (51)$$

where m_t is the colored noise process $m_t = h n_t + n_{t-1}$. Note that a given state S_t defines f_t , although the mapping so generated is not one to one. The corresponding error probabilities for the shift-invariant smoother are

$$\begin{aligned} P(e_{t-1|t}) &= \sum_{i,j,k} P(e_{t-1|t} = 0, e_{t|t} = d_i, e_{t-1|t-1} \\ &= d_j, e_{t-2|t-2} = d_k) \\ &= \sum_{i,j,k} P(e_{t-1|t} = 0, e_{t|t} \\ &= d_i | e_{t-1|t-1} = d_j, e_{t-1|t-1} \\ &= d_j | e_{t-2|t-2} = d_k) * P(e_{t-2|t-2} = d_k). \end{aligned} \quad (52)$$

Then, according to (46), we have the equivalence between

$$e_{t|t} = d_i | e_{t-1|t-1} = d_j \Leftrightarrow N_{i,j} < n_t < M_{i,j} \quad (53)$$

$$e_{t-1|t-1} = d_j | e_{t-2|t-2} = d_k \Leftrightarrow N_{j,k} < n_{t-1} < M_{j,k} \quad (54)$$

which leads us to the error recovery probability

$$\begin{aligned}
 P(e_{t-1}|t=0) &= \sum_{i,j,k} P(n_{t-1} + hn_t > -1 - h^2 - h(d_i + d_k) \\
 &\quad N_{i,j} < n_t < M_{i,j}, N_{j,k} < n_{t-1}M_{j,k}) \\
 &\quad * P(e_{t-2} = 0) \\
 &= \iint_{\mathcal{D}_{i,j,k}} f_{n_t}(n_1) f_{n_{t-1}}(n_2) dn_1 dn_2. \quad (55)
 \end{aligned}$$

Here, f_{n_t} and $f_{n_{t-1}}$, respectively, denote the probability densities of n_t and n_{t-1} , and $\mathcal{D}_{i,j,k}$ is the domain defined by the joint realization of

- $N_{i,j} < n_t < M_{i,j}$;
- $N_{j,k} < n_{t-1} < M_{j,k}$;
- $n_{t-1} + hn_t > -1 - h^2 - h(d_i + d_k)$.

Fig. 2 illustrates the asymptotic error probabilities for the DFE and the shift-invariant smoother as a function of the SNR for $h = 2$. Note that the performances of the smoother are all the better as the channel is nonminimum phase, which is a pathological case for the DFE.

Theorem 1: The asymptotic error probability for the smoother is less than that for the DFE for every possible first-order channel at high SNR.

Proof: We give here some elements of the proof, which requires some approximations on the Gaussian asymptotic probabilities. The proof relies on the two following observations:

- 1) The most likely event occurs when $e_{t-2}|t-2$ is equal to zero for high SNR. Then, suppose an error has occurred at time $t-1$. Then, whatever may happen at time t , the error at time $t-1$ cannot be corrected by the DFE, although here, it is possible.
- 2) Then, we may wonder if the possibility to change a decision, whether it is correct or not, poses a problem. In fact, suppose detection on symbols x_{t-2} and x_{t-1} were correct and that x_t was estimated incorrectly. Can the wrong decision on x_t result in an incorrect update of x_{t-1} at time t ?

Appendix D compares both situations. We show that the first situation, which is the favorable one, has a much greater probability ($o(\sigma^2)$) than the second one ($o(\sigma^4)$)

4) *Noiseless Error Recovery:* First, consider the set of length-1 channels (parameterized by a single real number h) for which the lag 1 smoother exhibits no pathological behavior. By this, we mean that there are no input sequences for which the equalizer cannot recover from an error in the noise-free case [6]. For the DFE, such sequences exist if $|h| > \frac{1}{2}$ since ISI may then occur. Now, consider the smoother. It is easily seen [from (51)] that under the condition that $f_t = 0$ or $f_t = 2, -2$ (with $f_t = e_{t|t} + e_{t-2|t-2}$), the smoother output is always correct (in the absence of noise). It is only under the condition that $f_t = 4, -4$ that the smoother output may be in error. This condition occurs if and only if the DFE decisions at both times t and $t-2$ are neither correct nor equal. Hence, the ranges of the channel parameters ensuring error recovery are $|h| < 2 - \sqrt{3}$ and $|h| > 2 + \sqrt{3}$.

5) *Average Error Performance:* We follow here an approach similar to that of [11]. The results can apply to the initial algo-

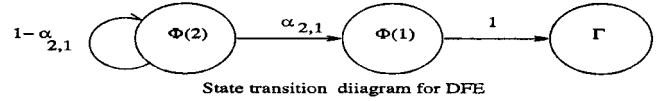
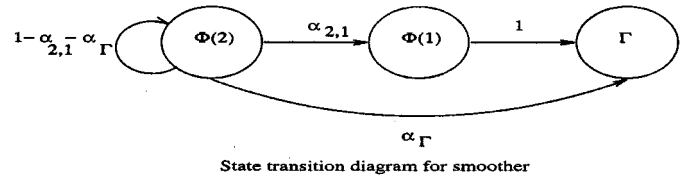


Fig. 3. State transition diagrams.

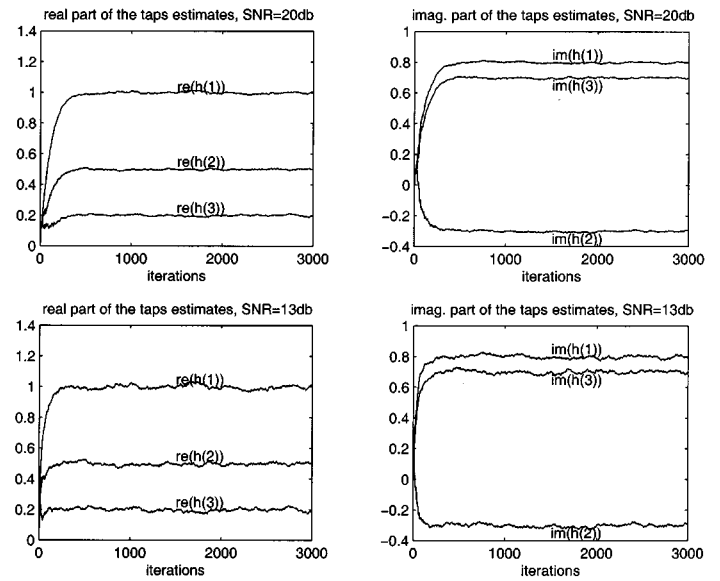


Fig. 4. Adaptive behavior for complex channel taps.

rithm of Section II. Consider a noiseless channel, and assume that an error has occurred at time t on symbol x_t . Define the state error as $E_t = (e_{t|t}, e_{t-1|t})$. This state error can take three values.

$$\Gamma = (0, 0) \quad (56)$$

$$\Phi(1) = (0, E) \quad (57)$$

$$\Phi(2) = (E, X) \quad (58)$$

where

- E error has occurred ($E = \pm 2$);
- 0 correct decision;
- X either a correct or an incorrect decision.

We focus here on the mean duration (the mean number of iterations) for the algorithm to reach state Γ (no error) starting with state $\Phi(2)$ (an error just occurred). We can show that this duration is reduced for the smoother, compared with the DFE. This is easily understood from Fig. 3; for the smoother, the transition from state $\Phi(2)$ to state Γ is allowed; therefore, this absorbing

state may be reached in only one iteration, whereas the DFE has to first transit through state $\Phi(1)$ before going to Γ .

V. SIMULATIONS

All simulations in this section are concerned with the initial algorithm of Section II, unless otherwise stated.

A. Adaptive Behavior of the Proposed Algorithm

Second-Order Channel Example: The algorithm was run on a 4PSK modulation going through a complex-valued channel with coefficients $H = [(1 + 0.8i), (0.5 - 0.3i), \text{ and } (0.2 + 0.7i)]$. The algorithm was initialized to the value $h_0 = [0, 0, 0]$.

Fig. 4 (20 dB SNR) and (13 dB SNR) show the channel taps adaptation. We can note that the convergence is faster as the SNR decreases, which can be explained by the fact that the Fisher matrix is better conditioned in the case of noisy estimates of the input sequence.

Nonminimum-Phase Channel and Comparison with the ZF-DFE (see Fig. 5): The channel considered in the section $H = [0.5 \ 0.7 \ 0.5]^T$ has a spectral null. Fig. 5 shows the quadratic error on the coefficients of the impulse response for the ZF-DFE and the smoother for a 13-dB SNR. This shows that the ZF-DFE (without any feed-forward filter) does not converge to the true values of the channel taps, whereas our algorithm does. The fact that the ZF-DFE converges to another solution is due to the error propagation phenomenon that prevents the adaptation to be properly performed. For the smoother, as an error may be corrected thanks to more reliable future observations, the error propagation is limited and does not adversely affect the adaptation process. It has to be pointed out that a DFE with a smoothing lag may have improved performances over a ZF-DFE if the smoothing lag is chosen large enough. However, this would result in an increased computational complexity.

B. BER Results

These results were obtained with real-valued BPSK signals and channels.

Checking the Theoretical Results: Fig. 2 compares on the first-order channel $H = [1; 2]$ simulations results with the theoretical results derived in Section IV-B1 concerning the asymptotic error probabilities for the ZF-DFE and the shift-invariant smoother. This simulation confirms the theoretical results on the comparison between the proposed method and the ZF-DFE, especially the fact that the performance of the smoother is better when the channel is nonminimum phase.

a) *Comparison with Other Equalizers:* Simulations have been performed on the channel $H = [0.5 \ 0.7 \ 0.5]^T$ to allow comparison with other equalizer structures. In Fig. 6, we compare the smoother to different kind of DFE's and to the linear MMSE. The DFE's involved here are

- the nonadaptive ZF-DFE (where the channel taps are known and fixed);
- the nonblind MMSE-DFE, which gives the best performances;

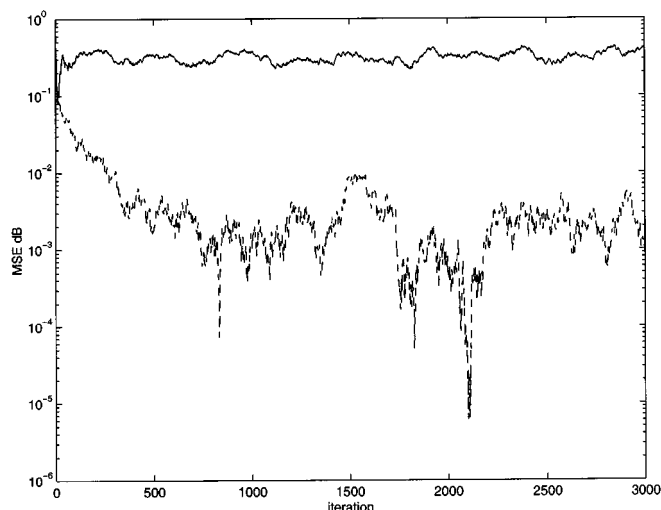


Fig. 5. Quadratic error on the channel taps estimates of the ZF-DFE (dashed curve) and smoother (plain curve) for a 15-dB SNR.

- the blind predictive constant-modulus DFE derived in [13].

The last two algorithms involve a 32-tap feedforward filter and a two-tap feedback filter. We can see that the performance of the smoother is very close to that obtained with the blind PCM-DFE. It has to be pointed out that although the MMSE-DFE performs better, the number of taps involved are such that the computational cost in terms of the number of multiplications is higher. It has been seen by simulations in [13] that the time of convergence of these DFE's involving a feedforward filter on this particular channel is on order of 5000 symbols, although we can check in Fig. 5 that it is only 500 symbols for the smoother. It has to be pointed out that this particular channel is a insidious one as the zeros of the impulse response are on the unit circle, which explains why the feedforward filters of the DFE's have so many taps and, therefore, converge slowly. This is, however, a good example to show that the proposed algorithm is not affected by the location of the zeros while providing BER results that are very close to the ones of the MMSE-DFE after convergence.

b) *Comparison with Existing Suboptimal HMM Algorithms:* Fig. 7 compares the performance of the proposed algorithm with the algorithm described in [3], leading to a computational complexity of M^{N-k} , where k is able to take the values in the set $\{1, \dots, N - 1\}$. Note that the value $k = 0$ corresponds to the optimal case, whereas $k = N - 1$ corresponds to the DFE. We performed this suboptimal algorithm for all possible values for k and compared the corresponding BER with the proposed algorithm, which has a computational complexity of $M * L$. We chose the following fifth-order channel: $h_2 = [0.28, -0.62, 0.2354, 0.4, 0.47, 0.3]$. This figure emphasizes the interest of revisiting past decisions; for instance, our algorithm, with a complexity of $2 * 4$ multiplications per sample, has improved performances compared with the other suboptimal algorithm with a complexity of 2^4 multiplications. This is due to the fact that we “revisit” the two last decisions rather than taking them for granted. It is interesting to point out that these results, which are, of course, specific to the channel

considered here, may be well explained by the fact that the energy of the channel taps is spread over the whole set of taps, which makes revisiting the decisions advantageous.

C. Performances of the Algorithm on Fading and Time Varying Channels: A GSM Application

In order to validate the proposed equalizer, we propose to perform simulations on various standard channel models chosen to mimic different downlink scenarios of GSM. We give a brief description of these scenarios and the characteristics of the channels to be estimated. After checking the estimates of the channel impulse response provided by the algorithm, we focus on BER performances by comparing the proposed blind equalizer with the one recommended by the GSM standard, which operates with a training sequence and produced an identification of the channel impulse response from each packet of data.

A Simulation on a GSM Channel Model: The source delivers a sequence of random bits that are modulated according to a Gaussian minimum shift keying (GMSK). We will assume valid the approximation that the generated symbols belong to the following alphabet: $\mathcal{A} = \{1; \exp(i\pi/2); -1; \exp(-i\pi/2)\}$. The oversampling rate at the output of the modulator is 36 samples per symbol, where the symbol rate is 270 kbits/s. This oversampling is used for simulation purposes only, and only one sample per symbol is considered for the equalization purposes. The system also incorporates the usual equipment filters at the transmitter and receiver. The data is then passed through the propagation channel. It has to be pointed out that we did not consider the burst structure (continuous transmission) as in the GSM standard. In addition, we did not use any frequency hopping in the simulations. The models implemented are the ones defined in the GSM recommendations 05.05 [14]. They correspond to six tap models. One important feature of these propagation models is that they are time varying, where the variations are mainly due to the local diffractions (leading to rapid and small variations of the signal power) as well as the Doppler and shadowing effect (resulting in slower but deeper variations). We show here the performances on the typical urban (Tu) channel, but similar results corresponding to other channels can be found in [12]. In the simulation, the SNR is set to 10 dB at the output of the channel. It is worth mentioning that the exact number of channel taps is *a priori* unknown for these kinds of applications. For instance, for the typical urban channel, only the three first taps have a significant value, whereas for the hilly terrain case, five taps have a non-neglectable contribution. In our simulations, the length of the equalizer is equal to 6, which means that the order of the true channel will be overestimated. In that case, simulations show that the extra taps converge to zero. The equalizer (of length 6) has been initialized to the value

$$\Theta_0 = [0.1 \ 0 \ 0 \ 0 \ 0 \ 0]^T.$$

Fig. 8 shows the two first tap estimates (real and imaginary parts) of the channel provided by the smoother highlighting the improved adaptation performances of the smoother.

Fig. 9 compares the BER performances of the smoother with that of the standard equalizer of the GSM device, operating with

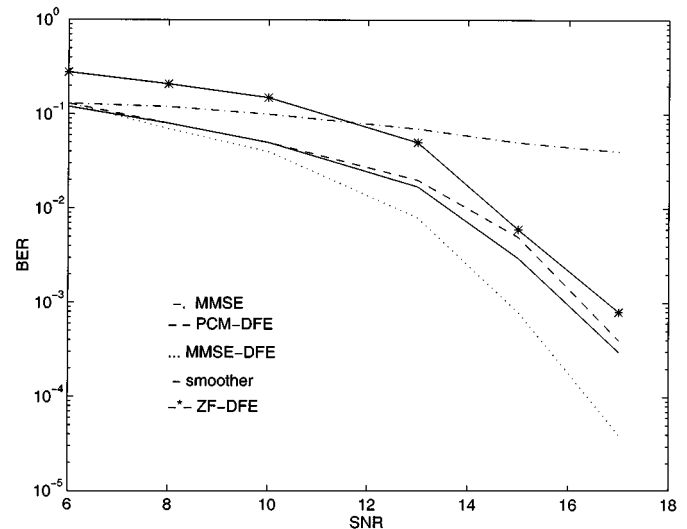


Fig. 6. BER results.

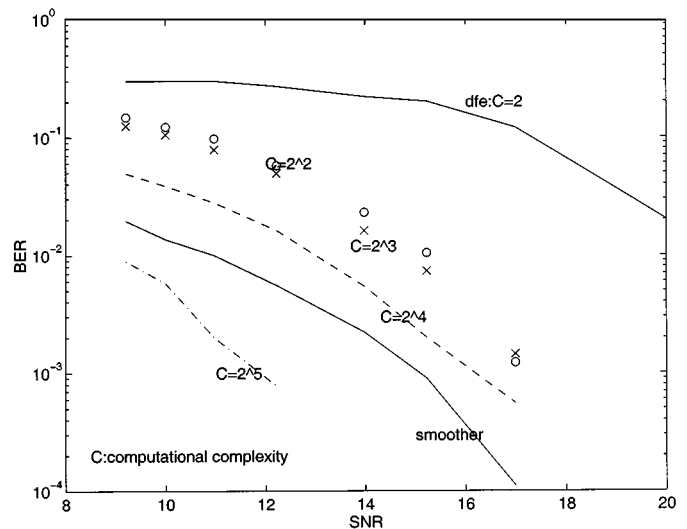


Fig. 7. BER results for the smoother and an other suboptimal algorithm based on HMM with different values of computational complexity.

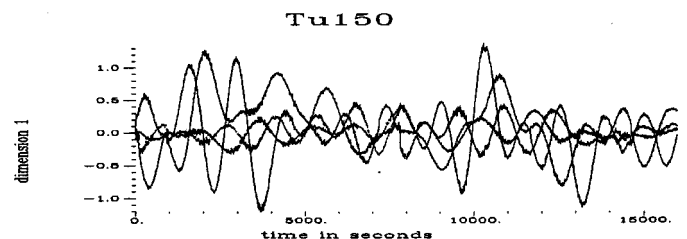


Fig. 8. Tu channel taps estimates for a 150km/h speed.

a training sequence of 26 bits on each packet, allowing an estimation of the channel impulse response. We can note that the BER's provided by the smoother are comparable with the ones of the nonblind equalizer. The fact that the performances seem better for poor SNR are due to the fact that in the case of fading, the 26 bits of the training sequence are not enough to provide a reliable estimate of the channel on some bursts. On the other hand, when the smoother is an RLS-type algorithm, it can take

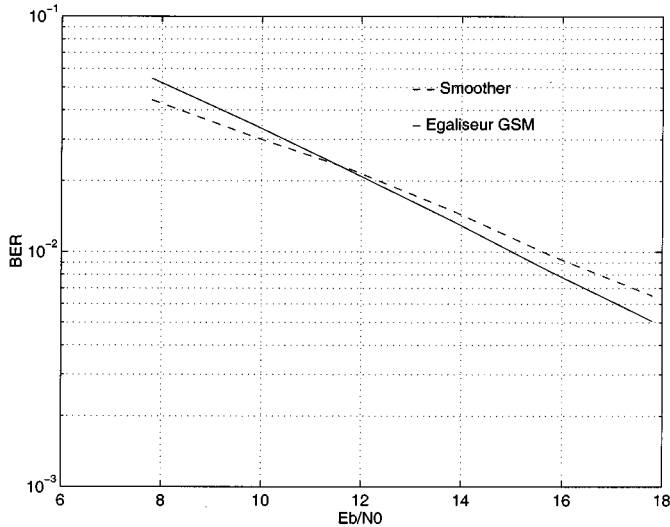


Fig. 9. BER results for Tu50.

into account more reliable information from the previous samples that are not affected by fading.

VI. CONCLUSION

A new decision feedback equalizer (DFE) incorporating fixed lag smoothing has been derived in this paper. This algorithm has been derived using a suboptimal hidden Markov model (HMM) formulation. The advantage of such a formulation is that it allows the complexity of the HMM formulation to be significantly reduced; this new DFE has complexity linear with the channel duration instead of exponential, as in the optimal case. Despite the approximations, we observe considerable benefits in term of bit error rate (BER), compared with the ZF-DFE. Moreover, simulations show that the smoother provides better results in term of BER than other suboptimal algorithms based on HMM, even if the latter algorithms have a higher computational complexity cost. Compared with simulations to DFE's incorporating a feedforward filter, the proposed algorithm seems to offer a better tradeoff between BER, computational complexity, and time of convergence. Finally, it looks well suited to the problem of blind equalization of fading and time-varying channels.

APPENDIX A

CALCULATION OF THE APPROXIMATE LOG-LIKELIHOOD

From (17) and taking into account (18), we show how (19) is obtained. Altogether, (17) and (18) read

$$Q(\hat{H}_t, H) = -\frac{1}{2\sigma^2} \sum_{i_0, i_1, \dots, i_{N-1}} \prod_{n=0}^{N-1} \alpha_{t|t}^{(n)}(i_n) (y_t - H^T [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}])^2. \quad (59)$$

We first show that

$$\sum_{i_0, i_1, \dots, i_{N-1}} \prod_{n=0}^{N-1} \alpha_{t|t}^{(n)}(i_n) [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}] = \hat{X}_t \quad (60)$$

and

$$\begin{aligned} & \sum_{i_0, i_1, \dots, i_{N-1}} \prod_{n=0}^{N-1} \alpha_{t|t}^{(n)}(i_n) [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}]^T \\ & \quad \cdot [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}] \\ & = X_t \hat{X}_t^T = \hat{X}_t \hat{X}_t^T. \end{aligned} \quad (61)$$

Proof of (60)

We have to calculate, for every $m = 1, \dots, M$

$$S_m = \sum_{i_0, i_1, \dots, i_{N-1}} \prod_{n=0}^{N-1} \alpha_{t|t}^{(n)}(i_n) q_{i_m}. \quad (62)$$

Then, by directly expanding (71), we have

$$S_1 = \underbrace{\sum_{i_m=1}^M \alpha_{t|t}^{(n)}(i_m) q_{i_m}}_{\text{Conditional Mean estimate}} \prod_{n \neq m} \sum_{i_n=1}^M \alpha_{t|t}^{(n)}(i_n).$$

Taking into account the fact that $\sum_{i_n=1}^M \alpha_{t|t}^{(n)}(i_n) = 1$ and using the following conjecture

$$\sum_{i_m=1}^M \alpha_{t|t}^{(m)}(i_m) q_{i_m} = \hat{X}_t^{(m)} \quad (63)$$

we obtain

$$S_m = \hat{X}_t^{(m)}. \quad (64)$$

Performing the same calculation for every m , we find (60).

Proof of (61):

First, compute the coefficients of matrix $[X_t \hat{X}_t^T]$. Let $[X_t \hat{X}_t^T]_{m, m'}$ denote the element on the m th row and the m' th column.

1) $m \neq m'$

$$\begin{aligned} & \sum_{i_0, i_1, \dots, i_{N-1}} \prod_{n=0}^{N-1} \alpha_{t|t}^{(n)}(i_n) q_{i_m}, q_{i_{m'}} \\ & = \sum_{i_m} q_{i_m} \alpha_{t|t}^{(m)}(i_m) \sum_{i_{m'}} q_{i_{m'}} \alpha_{t|t}^{(m')}(i_{m'}) \\ & \quad \cdot \prod_{n \neq m, m'} \sum_{i_n=1}^M \alpha_{t|t}^{(n)}(i_n) \end{aligned}$$

whereas using (72) gives

$$[X_t \hat{X}_t^T]_{m, m'} = \hat{X}_t^{(m)} \hat{X}_t^{(m')} = [\hat{X}_t \hat{X}_t^T]_{m, m'}. \quad (65)$$

2) $m = m'$.

In that case, we need to calculate

$$\sum_{i_0, i_1, \dots, i_{N-1}} \prod_{n=0}^{N-1} \alpha_{t|t}^{(n)}(i_n) q_{i_m}^2.$$

Expanding this expression in the same way as for the previous case, we obtain

$$[X_t \hat{X}_t^T]_{m,m} = \sum_{i_{m'}}^M q_{i_{m'}}^2 \alpha_{t|t}^{(m')}.$$

Now, assuming the following conjecture: At high SNR, the Gaussian functions involved in the computation of each $\alpha_{t|t}^{(n)}(i_n)$ are sufficiently peaked so that one of these variable is equal to unity, and the other is equal to zero. Then

$$\sum_{i_{m'}}^M q_{i_{m'}}^2 \alpha_{t|t}^{(m')} = (\hat{X}_t^{(m)})^2 = [\hat{X}_t \hat{X}_t^T]_{m,m} \quad (66)$$

i.e.,

$$X_t \hat{X}_t^T = \hat{X}_t \hat{X}_t^T. \quad (67)$$

Now, by directly expanding (17), and using (60) and (67)

$$\sum_{i_0, i_1, \dots, i_{N-1}} \prod_{n=0}^{N-1} \alpha_{t|t}^{(n)}(i_n) (y_t - H^T [q_{i_0}, q_{i_1}, \dots, q_{i_{N-1}}])^2$$

$$= y_t - 2H^T \hat{X}_t + H^T X_t \hat{X}_t^T H \quad (68)$$

$$= y_t - 2H^T \hat{X}_t + H^T \hat{X}_t \hat{X}_t^T H \quad (69)$$

$$= (y_t - H^T \hat{X}_t)^2. \quad (70)$$

APPENDIX B

LOWER BOUND OF ERROR PROPAGATION PROBABILITY

Under the assumption of high SNR, we show that

$$P((e_{t+n}, \dots, e_{t+1}) = \xi_i | e_t = 2) \geq (\frac{1}{2})^{(n)}.$$

Recall that Section IV-B1 has provided the equality

$$P((e_{t+n}, \dots, e_{t+1}) = \xi_i | e_t = 2)$$

$$= \prod_{j=1}^n P(e_{t+n-j+1} = \xi_i^{(j)} | e_{t+n-j} = \xi_i^{(j+1)} \dots e_{t+1} = \xi_i^{(n)}, e_t = 2). \quad (71)$$

Now, concentrate on the term $P(e_{t+n-j+1} = \xi_i^{(j)} | e_{t+n-j} = \xi_i^{(j+1)} \dots e_{t+1} = \xi_i^{(n)}, e_t = 2)$. As explained in Section IV-A, the decision at time $t+n-j+1$ on symbol $x_{t+n-j+1}$ is taken according to

$$\hat{x}_{t+n-j+1} = \text{sign} \left\{ h_0 x_{t+n-j+1} + \sum_{k=1}^{n-j} h_k e_{t+n-j+1-k} + b_{t+n-j+1-k} \right\}. \quad (72)$$

Assuming high SNR, we neglect the effect of the noise. We have two situations to examine:

- No error has occurred at time $t+n-j+1$: $\xi_i^{(j)} = 0$
If $h_0 x_{t+n-j+1}$ and $\sum_{k=1}^{n-j} h_k e_{t+n-j+1-k}$ have the same sign, which occurs with the probability $\frac{1}{2}$ (probability for $x_{t+n-j+1}$ to be equal to 1 or -1), in that case, $e_{t+n-j+1} = 0$.
If $h_0 x_{t+n-j+1}$ and $\sum_{k=1}^{n-j} h_k e_{t+n-j+1-k}$ do not have the same sign, then the value taken by $e_{t+n-j+1}$ depends on the absolute values of each of the two above terms.
- This implies that

$$P(e_{t+n-j+1} = 0 | e_{t+n-j} = \xi_i^{(j+1)} \dots e_{t+1} = \xi_i^{(n)}, e_t = 2) > \frac{1}{2}. \quad (73)$$

- An error has occurred at time $t+n-j+1$ $|\xi_i^{(j)}| = 2$

Note that $x_{t+n-j+1}$ has the same sign as $\xi_i^{(j)}$ and that this occurs with probability $\frac{1}{2}$. Another necessary condition for an error to occur is $|h_0| < |\sum_{k=1}^{L+j} h_k e_{t+n-j+1-k}|$. If this is not the case, the error sequence ξ_i has a null probability and is not involved in the calculation of probability derived in Section IV-B1. Then, for every plausible sequence

$$P(|e_{t+n-j+1}| = 2 | e_{t+n-j} = \xi_i^{(j+1)} \dots e_{t+1} = \xi_i^{(n)}, e_t = 2) = \frac{1}{2}. \quad (74)$$

Then, we obtain

$$P((e_{t+n}, \dots, e_{t+1}) = \xi_i | e_t = 2)$$

$$= \prod_{j=1}^n P(e_{t+n-j+1} = \xi_i^{(j)} | e_{t+n-j} = \xi_i^{(j+1)} \dots e_{t+1} = \xi_i^{(n)}, e_t = 2)$$

$$\geq (\frac{1}{2})^{(n)}. \quad (75)$$

APPENDIX C

CALCULATION OF DFE ERRORS PROBABILITIES

First, suppose $e_{t-1|t-1} = 0$, i.e., the DFE estimate was correct at time $t-1$. Then, $\hat{x}_t = \text{sgn}(x_t + n_t)$ so that

$$P(\hat{x}_t = x_t | e_{t-1|t-1} = 0)$$

$$= \frac{1}{2} (P(\hat{x}_t = x_t | e_{t-1|t-1} = 0, x_t = -1) + P(\hat{x}_t = x_t | e_{t-1|t-1} = 0, x_t = 1))$$

$$= \frac{1}{2} (P(n_t < 1) + P(n_t > -1)) = \chi(\sigma^{(-1)}) \triangleq \epsilon_0. \quad (76)$$

where χ denotes the pdf of a $N(0, 1)$ random variable. The transitions to the error states have probabilities

$$P(e_{t|t} = 2 | e_{t-1|t-1} = 0)$$

$$= \frac{1}{2} P(x_t | e_{t-1|t-1} = 0, x_t = 1)$$

$$= \frac{1}{2} P(n_t < -1) \triangleq \frac{1}{2}(1 - \epsilon_0) \quad (77)$$

and similarly for $e_{t|t} = -2$.

Similarly, if an error was made at time $t-1$ so that $e_{t-1|t-1} \pm 2$, then $\hat{x}_t = \text{sgn}(x_t + n_t \pm 2h)$, and the probability of recovering on the next step is

$$\begin{aligned}
 & P(\hat{x}_t = x_t | e_{t-1|t-1} = 2) \\
 &= \frac{1}{2} (P(\hat{x}_t = x_t | e_{t-1|t-1} = 2, x_t = -1) \\
 &\quad + P(\hat{x}_t = x_t | e_{t-1|t-1} = 2, x_t = 1)) \\
 &= \frac{1}{2} (P(n_t < 1 - 2h) \\
 &\quad + P(n_t > -1 - 2h)) \\
 &= \frac{1}{2} \left(\chi \left(\frac{1 - 2h}{\sigma} \right) + 1 - \chi \left(\frac{-1 - 2h}{\sigma} \right) \right) \\
 &\triangleq 1 - (\epsilon_1 + \epsilon_2)
 \end{aligned} \tag{78}$$

and similarly for $P(\hat{x}_t = x_t | e_{t-1|t-1} = 2)$. Similarly

$$\begin{aligned}
 & P(e_{t|t} = 2 | e_{t-1|t-1} = 2) \\
 &= \frac{1}{2} P(\hat{x}_{t|t} = -1 | e_{t-1|t-1} = 2, x_t = 1) \\
 &= \frac{1}{2} P(n_t < -1 - 2h) = \frac{1}{2} \chi \left(\frac{-1 - 2h}{\sigma} \right) \triangleq \epsilon_1
 \end{aligned} \tag{79}$$

$$\tag{80}$$

and

$$\begin{aligned}
 & P(e_{t|t} = -2 | e_{t-1|t-1} = 2) \\
 &= \frac{1}{2} P(\hat{x}_{t|t} = 1 | e_{t-1|t-1} = 2, x_t = -1) \\
 &= \frac{1}{2} P(n_t > 1 - 2h) = \frac{1}{2} \left(1 - \chi \left(\frac{1 - 2h}{\sigma} \right) \right) \triangleq \epsilon_2.
 \end{aligned} \tag{81}$$

$$\tag{82}$$

By symmetry, (79) is also the probability of remaining in the -2 error state, and (81) is also the probability of transition from -2 to $+2$. Enumerating the states according to the ordering $0, 2, -2$, we therefore have the state transition matrix

$$A = \begin{bmatrix} \epsilon_0 & \frac{1 - \epsilon_0}{2} & \frac{1 - \epsilon_0}{2} \\ 1 - \epsilon_1 - \epsilon_2 & \epsilon_1 & \epsilon_2 \\ 1 - \epsilon_1 - \epsilon_2 & \epsilon_2 & \epsilon_1 \end{bmatrix}. \tag{83}$$

From (49) and defining $\Pi_\infty = \lim_{t \rightarrow \infty} \Pi_t$, we have

$$\Pi_\infty = A^T \Pi_\infty \tag{84}$$

which means that Π_∞ is an eigenvector of A^T associated to the eigenvalue 1. Hence, we obtain the following asymptotic error probability

$$1 - \pi_0 = \frac{1 - \epsilon_0}{2 - \epsilon_0 - (\epsilon_1 + \epsilon_2)}. \tag{85}$$

APPENDIX D

SOME ELEMENTS FOR THE PROOF OF THEOREM 1

A. Calculation of the Probability to Recover from an Error Committed at Time $t - 1$

First, suppose that $x_{t-1} = 1, e_{t-2|t-2} = 0, e_{t-1|t-1} = 2$. Under these assumptions, we aim at calculating the joint probability for $e_{t-1|t}$ to be equal to zero.

$$\begin{aligned}
 & P(e_{t-1|t} = 0, e_{t-2|t-2} = 0, e_{t-1|t-1} = 2) \\
 &= \frac{1}{2} \sum_{i=1, -1} P(e_{t-1|t} = 0, e_{t-2|t-2} = 0, e_{t-1|t-1} = 2 \\
 &\quad x_t = i).
 \end{aligned} \tag{86}$$

Note that if $x_t = 1$, then $e_t = 0$ arises with a probability close to one. Therefore, we may write

$$\begin{aligned}
 & P(e_{t-1|t} = 0, e_{t-2|t-2} = 0, e_{t-1|t-1} = 2) \\
 &> \frac{1}{2} P(e_{t-1|t} = 0, e_{t-2|t-2} = 0, e_{t-1|t-1} = 2, x_t = 1) \\
 &> \frac{1}{2} P(e_{t-1|t} = 0, e_t = 0 | e_{t-1|t-1} = 2, x_t = 1 \\
 &\quad e_{t-1|t-1} = 2 | e_{t-2|t-2} = 0) * P(e_{t-2|t-2} = 0) \\
 &> \frac{1}{2} P(-1 > n_{t-1} > -1 - h^2 - hn_t) P(e_{t-2} = 0)
 \end{aligned} \tag{87}$$

and since $P(-1 > n_{t-1} > -1 - h^2 - hn_t) = o(\sigma^2)$, we have

$$P(e_{t-1|t} = 0, e_{t-2|t-2} = 0, e_{t-1|t-1} = 2) > o(\sigma^2).$$

B. Calculation of the Error Probability on a Previously Well-Estimated Symbol

First, assume that $x_{t-1} = 1$ and that $h > 0$, (the other cases would provide the same results by symmetries). We have to deal with the following events: $e_{t-2|t-2} = 0, e_{t-1|t-1} = 0, e_{t|t} = -2$. We aim at calculating the joint probability to have $e_{t-1|t} = 2$ and the above assumptions

$$\begin{aligned}
 & P(e_{t-1|t} = 2, e_{t-2|t-2} = 0, e_{t-1|t-1} = 0, e_{t|t} = -2) \\
 &= \frac{1}{2} P(n_t > 1, -1 < n_{t-1} < -1 - h^2 + h(2 - n_t)) \\
 &\quad * P(e_{t-2|t-2} = 0).
 \end{aligned} \tag{88}$$

Note that $|2 - n_t| < 1$, so that $-1 - h^2 + h(2 - n_t)$ will be negative. For high SNR, we have

$$P(e_{t-1|t} = 2, e_{t-2|t-2} = 0, e_{t-1|t-1} = 0, e_{t|t} = -2) = o(\sigma^4). \tag{89}$$

Then, taking into account (89), we have

$$\begin{aligned}
 & P(e_{t-1|t} = 2, e_{t-2|t-2} = 0, e_{t-1|t-1} = 0, e_{t|t} = -2) \\
 &\ll P(e_{t-1|t} = 0, e_{t-2|t-2} = 0, e_{t-1|t-1} = 2).
 \end{aligned}$$

ACKNOWLEDGMENT

The authors wish to thank B. D. O. Anderson of the Australian National University, Canberra, for fruitful discussions.

REFERENCES

- [1] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown dispersive channels," *IEEE Trans. Commun.*, vol. 42, Feb. 1994.
- [2] V. Krishnamurthy and J. B. Moore, "On-line estimation of hidden markov model parameters based on the Kullback-Leibler information measure," *IEEE Trans. Signal Processing*, vol. 41, Aug. 1993.
- [3] L. B. White and V. Krishnamurthy, "Adaptive blind equalization of FIR channels using hidden markov models," in *Proc. IEEE ICC*, Geneva, Switzerland, 1993.
- [4] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Proc. IEEE*, vol. 77, Feb. 1989, pp. 257–285.
- [5] L. B. White, S. Perreau, and P. Duhamel, "Reduced computation blind equalization for FIR channel input markov models," in *Proc. ICC*, Seattle, WA, June 1995.
- [6] R. Kennedy and B. D. O. Anderson, "Recovery times of decision feedback equalizers on noiseless channels," *IEEE Trans. Commun.*, vol. COMM-35, pp. 1012–1025, Oct. 1987.
- [7] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Soc.*, vol. 6, 1977.
- [8] S. Perreau, L. White, and P. Duhamel, "A reduced computation multi-channel adaptive equalizer based on HMMs," in *Proc. SSAP*, Corfu, Greece, 1996.
- [9] —, "An equalizer including a soft channel decoder of convolutional codes," in *Proc. SPAWC*, Paris, France, 1997.
- [10] S. Perreau and P. Duhamel, "On-line blind equalization of FIR channels using a Gibbsian technique," in *Proc. ICASSP*, Munich, Germany, 1997.
- [11] N. C. Beaulieu, "Bounds on recovery times of decision feedback equalizers," *IEEE Trans. Commun.*, vol. 42, pp. 2786–2794, Oct. 1994.
- [12] S. Perreau, "Application des Methodes de Maximum de Vraisemblance a L'egalisation Autodidacte," Ph.D. dissertation, Telecom Paris, Paris, France, Feb. 1997.
- [13] L. Tong and D. Liu, "Blind predictive decision-feedback equalization via the constant modulus algorithm," in *Proc. ICASSP*, 1997, pp. 3901–3904.
- [14] COST 207, "Digital land mobile radio communications," *Comm. Euro. Communities*, 1989.
- [15] D. L. Duttweiler, J. E. Mazo, and D. G. Messerschmitt, "An upper bound on the probability in decision feedback equalization," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 440–497, July 1974.



Sylvie Perreau was born in 1969 in France. She received the engineering and DEA degrees from the National School for Applied Electronics (ENSEA), Cergy, France, in 1993 and the Ph.D. degree from Telecom Paris (ENST), Paris, France, in 1997.

Since her graduation, she was a Research Assistant at the University of Connecticut, Storrs, in 1997 and has been a Lecturer at the University of South Australia, Adelaide, Australia, since 1998. Her main research interests are in channel estimation, data detection, smart antennas, and interference mitigation.



Langford B. White received the B.Sc. degree in mathematics and the B.E. and PhD degrees in electrical engineering from the University of Queensland, Brisbane, Australia, in 1984, 1985, and 1989, respectively.

From 1986 to 1999, he worked for the Defence Science and Technology Organisation, Adelaide, Australia, in the areas of communications EW and signal analysis. Since March 1999, he has been Professor of Communications Networks Analysis.

Dr. White is a member of the UMF of America.

Pierre Duhamel (M'98) was born in France in 1953. He received the B.Eng. degree in electrical engineering from the National Institute for Applied Sciences (INSA), Rennes, France, in 1975 and the Dr.Eng. degree in 1978 and the Doctorat es Sciences degree in 1986, both from Orsay University, Orsay, France.

From 1975 to 1980, he was with Thompson CSF, Paris, France, where his research interests were in circuit theory and signal processing, including digital filtering and analog fault diagnosis. In 1980, he joined the National Research Center in Telecommunications (CNET), where his research activities were first concerned with the design of recursive CDD filters. Later, he worked on fast Fourier transforms and convolution algorithms and applied similar techniques to adaptive filtering, spectral analysis, and wavelets transforms with applications in channel equalization and source coding. Since 1993, he has been Professor at Telecom Paris with research activities in the same area.