An Upper Bound on the Performance of a Novel Feedforward Perceptron Equalizer

Graham W. Pulford, *Member, IEEE,* Rodney A. Kennedy, *Member, IEEE,* and Brian D. O. Anderson, *Fellow, IEEE*

Abstract—**The emulation of a nonadaptive, binary decision feedback equalizer, operating on a noiseless, finite impulse response channel by a feedforward multilayer processor is considered. This feedforward perceptron equalizer comprises a triangular array of hard-limiting processing elements. The functional similarity between the two systems is exploited to obtain a tight upper bound on the probability of error as a function of the number of layers, using the theory of finite state Markov processes.**

Index Terms—**Equalization, feedforward perceptron equalizer, multilayer perceptron neural networks, finite-state Markov processes, error probability upper bound.**

I. Introduction

THIS paper is motivated by a current trend—the application of multilayer perceptron (MLP) neural networks in the equalization of channel for digital communication. As has been confirmed by experimental studies [1], [2], MLP neural networks are definite candidates for the development of new equalization strategies. Seeking to justify from a theoretical standpoint the use of nonlinear feedforward processors (like the conventional MLP neural network) in equalization, we consider a tuned, noiseless decision feedback equalizer (DFE) operating on a linear finite impulse response (FIR) channel. We derive a related feedforward multilayer processor or *feedforward perceptron equalizer* (FPE) that emulates this well-studied device. The analysis leading to the upper bound on the error probability relies heavily on the theory of finite-state Markov process [3].

The FPE structure we derive may be cast, in the noiseless case, as a multilayer perceptron feedforward neural network. In general, however, it has a nonstandard structure akin to a (systolic array) parallel processing architecture. The processing elements are hard-limiting perceptrons which compute the signum function sgn(·) of the weighted sum of their inputs. The matrix of interconnection weights is Toeplitz and upper triangular, and can be related directly to the vector of feedback tap gains in the DFE. Unlike MLP neural networks whose weights may be chosen freely, the parameters of the FPE are constrained so that only a few may vary independently. If we have knowledge of the communications channel, then we know what the weights must be if the FPE is to emulate a DFE.

In Section II, we present the conventional decision feedback equalizer and detail an *unwrapping* procedure which, when truncated, results in a multilayer processor with hard-limiting nodes and feedback. (We ignore the front-end linear filter which is usually incorporated in the DFE to cancel precursor intersymbol interference). By disconnecting the feedback of decisions made in the distant past, we obtain our feedforward perceptron equalizer. We give some low-order illustrations and show how the structure generalizes to an arbitrary number of layers.

In Section III-A, we introduce a finite-state Markov process (FSMP) description of the DFE [4], and specialize to the tuned (nonadaptive), noise-free case for simplicity. We extend the model to embrace both the DFE and the feedforward neural network by enlarging the state space. In Section III-B, we define a worst case channel class (i.e., channels guaranteeing the worst bit error rate performance for any channel of the same order). We subsequently apply FSMP theory (developed in Section III-C) to upper bound the noiseless error probability of the FPE (Section III-D). This bound is a function of the number of layers in the FPE. We also give a simple numerical example demonstrating that, in the presence of noise, an FPE with enough layers does indeed emulate a decision feedback equalizer.

Manuscript received July 30, 1991; revised April 27, 1992. This work was supported in part by the Australian Telecommunications and Electronics Research Board, the ANU Centre for Information Science Research, and the Australian Research Council. This paper was presented at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'91), Toronto, Ont., Canada, May 1991.

G. W. Pulford is with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, Victoria 3052 Australia.

R. A. Kennedy is with the Department of Engineering, Faculty of Engineering and Information Technology, The Australian National University, Canberra ACT 0200 Australia.

B. D. O. Anderson is with the Department of Systems Engineering, Research School of Physical Sciences and Engineering, The Australia National University, Canberra, ACT 0200 Australia.

IEEE Log Number 9213029.

II. Synthesis of the Feedforward Processor

A. Unwrapping the Decision Feedback Equalizer

Conventional MLP neural networks are highly interconnected systems whose analysis is, in general, very difficult. Our motivation in this work is to determine a specific MLP structure tailored to perform as a nonlinear equalizer. As will be seen, the new structure is amenable to analysis, and it is possible to obtain a bound on its performance. To this end, we consider a binary decision feedback equalizer, operating on a finite impulse response

0018-9448/93$03.00

channel, corrupted by additive zero-mean white Gaussian noise n_k. At the output of the channel, the measured signal is

$$y_k = \sum_{i=0}^{L} h_i u_{k-i} + n_k \tag{1}$$

where $\{h_i\}$ are the impulse response coefficients and $\{u_k\}$ is a sequence of equiprobable, independent binary data (not directly measurable). The DFE (Fig. 1) generates an estimate of the input signal (based on its past decisions) given by

$$\hat{u}_k = \text{sgn}\left(y_k - \sum_{j=1}^{L} d_j(k)\hat{u}_{k-j} \right) \tag{2}$$

where $\text{sgn}(x) = 1$ if $x \geq 0$. The feedback tap gains $d_j(k)$ are adapted to cancel the intersymbol interference introduced by the channel. We will only be concerned with the analysis of the nonadaptive system in which the $d_j(k) = d_j, \forall j$, are constant. We lose no generality in assuming that the cursor $h_0 = 1$.

We develop an alternative recursive representation of (2) by "unwrapping" the DFE, and in so doing, introduce a delay in the computation of the decisions. This proceeds as follows. At the first step, we write

$$\hat{u}_{k-1} = \text{sgn}\left(y_{k-1} - \sum_{j=1}^{L} d_j \hat{u}_{k-j-1} \right)$$

and substitute for $\hat{u}_{k-1}$ in (2), thus obtaining

$$\begin{aligned}\hat{u}_k &= \text{sgn}\left(y_k - d_1 \,\text{sgn}\left(y_{k-1} - \sum_{j=1}^{L} d_j \hat{u}_{k-j-1} \right) \right. \\ &\quad \left. - \sum_{j=2}^{L} d_j \hat{u}_{k-j} \right) \\ &\triangleq f_1^L(y_k, y_{k-1}; \hat{u}_{k-2}, \cdots, \hat{u}_{k-L-1}). \end{aligned} \tag{3}$$

At the next step, we eliminate $\hat{u}_{k-2}$ [which appears twice in (3)] to get

$$\begin{aligned}\hat{u}_k &= \text{sgn}\Bigg(y_k - d_1 \,\text{sgn}\Bigg(y_{k-1} - d_1 \\ &\quad \cdot \text{sgn}\left(y_{k-2} - \sum_{j=1}^{L} d_j \hat{u}_{k-j-2} \right) - \sum_{j=2}^{L} d_j \hat{u}_{k-j-1} \Bigg) \\ &\quad - d_2 \,\text{sgn}\left(y_{k-2} - \sum_{j=1}^{L} d_j \hat{u}_{k-j-2} \right) - \sum_{j=3}^{L} d_j \hat{u}_{k-j} \Bigg) \\ &\triangleq f_2^L(y_k, y_{k-1}, y_{k-2}; \hat{u}_{k-3}, \cdots, \hat{u}_{k-L-2}). \end{aligned} \tag{4}$$

After d such steps, we obtain a highly nested composition of sgn($\cdot$) functions whose functional form can be written as

$$\hat{u}_k = f_d^L(y_k, y_{k-1}, \cdots, y_{k-d}; \hat{u}_{k-d-1}, \cdots, \hat{u}_{k-L-d}). \tag{5}$$

There are, in fact, $d + 1$ degrees of nesting in this expression, and we can interpret these as the layers in a recursive multilayer processor whose external inputs are the $\{y_k, \cdots, y_{k-d}\}$, whose feedback inputs are the $\{\hat{u}_{k-d-1}, \cdots, \hat{u}_{k-L-d}\}$, and whose output is $\hat{u}_k$. The processing elements are nodes which compute the sign of the weighted sum of their inputs. Fig. 2 depicts a typical node.

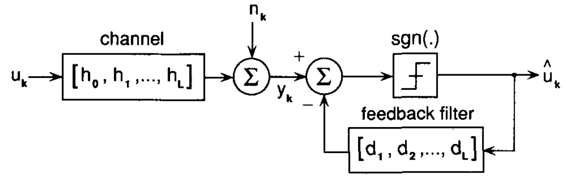

Fig. 1. Decision feedback equalizer.

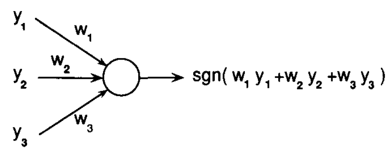

Fig. 2. Processing element (node).

Intuitively, it is reasonable to expect that the dependence of $\hat{u}_k$ on the past decisions $\hat{u}_{k-d-j}$ decreases as we make d larger (at least for $d > L$). This is equivalent to the effect of earlier decisions ceasing to influence later decisions in the DFE, given a large enough time delay, we give substance to this claim in the following sections.

Suppose, then, that we can ignore decisions in the distant past, we consider disconnecting the feedback implied in (5) by setting the arguments involving past decisions to zero, obtaining

$$\tilde{u}_k = f_d^L(y_k, y_{k-1}, \cdots, y_{k-d}; 0, \cdots, 0). \tag{6}$$

If our intuition is correct, then with high probability and given a large number of layers (d large), we would have $\hat{u}_k = \tilde{u}_k$. It is instructive to visualize (6) as a multilayer *feedforward* processor with signum nodes. We illustrate in Fig. 3 the corresponding system in the $d = 2$ case by setting $\hat{u}_{k-j} = 0, j \leq 3$ in (4).

Before proceeding with the general case embodied by (6), we make a short digression to examine the noiseless feedforward version of (4) with $d = 2$, which may be expressed as

$$\begin{aligned}\tilde{u}_k &= f_2^L(y_k, y_{k-1}, y_{k-2}; 0, \cdots, 0)|_{n_k = n_{k-1} = \cdots = 0} \\ &= \text{sgn}\Bigg(h_0 u_k + \sum_{i=1}^{L} h_i u_{k-i} - d_2 \\ &\quad \cdot \text{sgn}\left(h_0 u_{k-2} + \sum_{i=1}^{L} h_i u_{k-i-2} \right) - d_1 \\ &\quad \cdot \text{sgn}\Bigg(h_0 u_{k-1} + \sum_{i=1}^{L} h_i u_{k-i-1} - d_1 \\ &\quad \cdot \text{sgn}\left(h_0 u_{k-2} + \sum_{i=1}^{L} h_i u_{k-i-2} \right) \Bigg) \Bigg). \end{aligned}$$

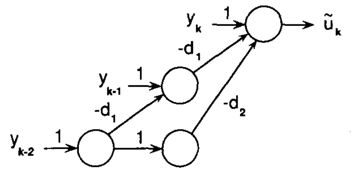

Fig. 3. Feedforward perceptron equalizer with three layers ($d = 2$).

This is depicted in Fig. 4 for a channel length $L = 2$. Of course, in practice, the input sequence $\{u_k\}$ and the channel parameters are unknown, but the purpose of Fig. 4 is only to show how Fig. 3 can be captured in a standard MLP framework.

Returning to the general $(d + 1)$-layer structure described before in (6), it is fairly easy to generalize the low-order cases to arrive at the diagram in Fig. 5, which will be referred to in the sequel as a $(d + 1)$-layer FPE. We have deliberately drawn Fig. 5 to accentuate its Toeplitz structure (i.e., the weight of the branch connecting node i in layer k to node $i + j$ in layer $k + 1$ is $-d_j$, independently of i). We alert the reader to the following important differences between Fig. 5 and a standard MLP neural network: 1) each diagonal node has one external input, this being a noisy channel output; 2) all horizontal connections have fixed weight 1; and 3) there are at most d distant weights.

Finally, the FPE is clearly self-similar. The d-layer FPE is embedded in the $(d + 1)$-layer FPE. This embedding property has interesting consequences in terms of monotonicity of performance (see Section III-F).

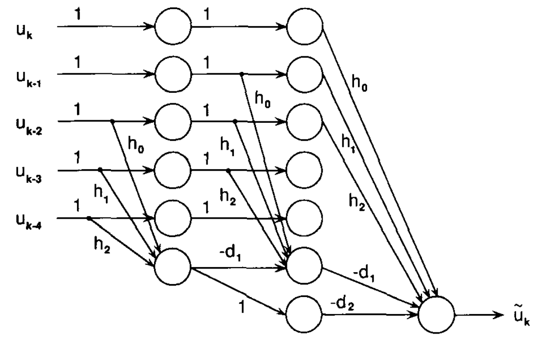

Fig. 4. Noiseless $d = 2$ MLP realization.

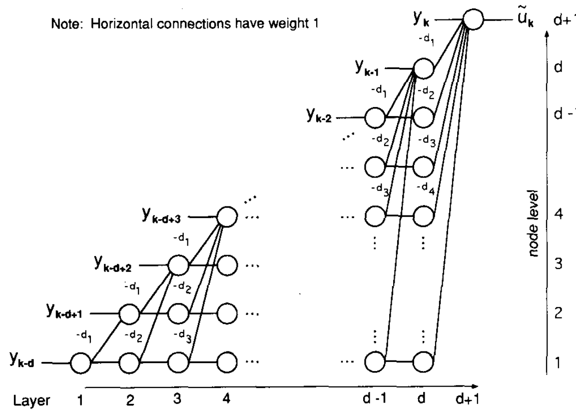

Fig. 5. General $(d + 1)$-layer feedforward perceptron equalizer.

III. Analysis of Noiseless Error Probability

A. Finite-State Markov Process Description

We can model the DFE, and consequently the FPE, using the theory of discrete time, finite-state Markov processes [3] (this follows from the independence assumption on the input sequence). Referring to (1) and (2), the input is u_k and we choose

$$X_k = [u_{k-1},\cdots,u_{k-L},\hat{u}_{k-1},\cdots,\hat{u}_{k-L}]'$$

as the state vector. There are 4^L states in total if all elements are binary. However, in this case, we can reduce the number of states to 3^L since the DFE is assumed to be *tuned* in the sense that $d_j = h_j, j \in \{1,\cdots, L\}$ by defining the state

$$E_k \triangleq [e_{k-1},\cdots,e_{k-L}]'$$

where each component $e_k = u_k - \hat{u}_k$ can take the values ± 2 or 0. Henceforth, we only deal with the tuned case. We denote by $\mathscr{E}$ the set of all 3^L E_k-states. By way of simplification, as is standard in the analysis of error propagation of DFE's [5], [6], we consider only the noise-free case ($n_k = 0\ \forall K$) so that the unique absorbing state of the FSMP is $E_k = 0$. To see this, note that the error state satisfies a simple shift register property:

$$E_{k+1} = \begin{bmatrix} 0'_{L-1} & 0 \\ I_{L-1} & 0_{L-1} \end{bmatrix} E_k + \begin{bmatrix} u_k - \operatorname{sgn}(u_k + H'E_k) \\ 0_{L-1} \end{bmatrix} \tag{7}$$

where

$$H \triangleq [h_1,\cdots,h_L]',$$

I_n is the identity matrix of order n, and 0_n is a column vector of n zeros. If $E_k = 0$, then for all inputs, we have

$$\operatorname{sgn}(u_k + H'E_k) = u_k \tag{8}$$

and the DFE remains in the zero-error state.

Returning to the recursive representation of the DFE described in the last section, we see that the output, in the absence of noise, can be viewed as depending on the sequence of inputs $u_{k-d},\cdots,u_k$ and the initial state E_{k-d}. Referring to (5), we introduce the notation

$$f_d^L(y_k,\cdots,y_{k-d};\hat{u}_{k-d-1},\cdots,\hat{u}_{k-d-L})|_{n_k=\cdots=n_{k-d}=0} \\ \triangleq g_d^L(u_k,\cdots,u_{k-d-L};\hat{u}_{k-d-1},\cdots,\hat{u}_{k-d-L}). \tag{9}$$

We also define an extended state $\overline{E}_k$ having the same form as E_k, but in which the decisions which appear in the initial state may take the additional value zero. Now, each element $\bar{e}_k$ of the extended error state $\overline{E}_k$ may take one of five possible values: 0, ± 1, ± 2, so there are a total of 5^L $\bar{E}_k$-states comprising the set $\bar{\mathscr{E}}$. Of course, a DFE with an initial state in $\bar{\mathscr{E}}$ reverts to a DFE with state in $\mathscr{E}$ after L time units because the binary decisions being fed back will displace the initial conditions. The concept of a DFE with an extended state-space will allow us to model the effect of omitting the feedback part, i.e., old decisions, in the recursive representation for the DFE (5), thus obtaining a *feedforward perceptron equalizer* structure generated by

$$\tilde{u}_k = g_d^L(u_k, \ldots, u_{k-d}, u_{k-d-1}, \ldots, u_{k-d-L}; \tilde{u}_{K-d-1} := 0, \ldots, \tilde{u}_{k-d-L} := 0) \quad (10)$$

in the noiseless case and by (6) in the noisy case. In the above, $\tilde{u}_k$ is binary, but by the notation $\tilde{u}_{k-d-j} := 0$, we imply that any feedback paths in the recursive representation (9) have been deleted. Note that the same $\tilde{u}_k$ as (10) would be generated by a DFE initialized in an abnormal initial state $\bar{E}_{k-d}$ with fictitious past decisions $\hat{u}_{k-d-1} = 0, \ldots, \hat{u}_{k-d-L} = 0$ and fed by the sequence of inputs $u_{k-d}, \ldots, u_k$. Thus, the FPE in (10) is effectively a "sliding window" version of the DFE which resets its initial conditions at each time instant. We shall have more to say about this in Section III-D.

B. Worst Case Channels

With a view to developing a tight performance bound, we now determine a class of channel on which the FPE has the worst possible performance. From (8), any channel $[h_0, h_1, \ldots, h_L]$ satisfying

$$\min_{\bar{E}_k \neq 0} |H'\bar{E}_k| > 1 \quad (11)$$

has the property that the error probability of the DFE $\Pr(\hat{u}_k \neq u_k) = \frac{1}{2}$ for any nonzero extended error state $\bar{E}_k$. (The equiprobability of the input sequence implies that u_k has a probability of $\frac{1}{2}$ of having the same sign as $H'\bar{E}_k$.) Since we are dealing with extended error states, this framework also includes the FPE. Channels satisfying (11) will be termed *worst case channels*. For example, any channel (with $h_0 = 1$) whose parameters belong to the set

$$\{h_1 > 1\} \bigcap_{j=2}^{L} \left\{h_j > 1 + 2\sum_{k=1}^{j-1} h_k\right\}$$

will fall into the worst case category. This is because the h_j have been spaced far enough apart that no linear combination with coefficients in the set $\{0, \pm 1, \pm 2\}$ has magnitude less than one.

As an illustration, consider the $L = 2$ case. We may take $h_1 = 1.2$ and $h_2 = 3.5$, so that

$$\min_{\bar{E}_k \neq 0} |[1.2, 3.5]\bar{E}_k| = 3.5 - 2 \times 1.2 = 1.1 > 1,$$

demonstrating that [1, 1.2, 3.5] is a worst case channel.

C. FSMP Aggregation

In what follows, we assume a worst case channel. Consequently, the order of the FSMP model can further be reduced by aggregating states. This is valid since all states in any given aggregated state transit with equal probability to states in the destination aggregated state [7]. In this case, we retain the Markov property if we choose to aggregate the $\bar{E}_k$-states according to the following rule [8].

Definition 1 (Aggregation Rule): $\bar{E}_k$ is in aggregated state $\epsilon_k = q$ at time k is there exists an input sequence $\{u_j\}_{j=k}^{k+q-1}$ such that the absorbing state $E_{k+q} = 0$ is reached in q steps (while it cannot be reached in fewer).

Appearing to the shift register property for the DFE (7) applied to $\bar{E}_k$ (by allowing $\hat{u}_k = 0$), it is clear that at most L successive correct decisions are needed to force an arbitrary nonzero $\bar{E}_k$-state to the zero $\bar{E}_k$-state. (We handle the FPE case in the next section.) Hence, there are $L + 1$ aggregated states ϵ_k in the new FSMP. From a given state $\epsilon_k = q (q \neq 0)$, there is a probability of $\frac{1}{2}$ (for equiprobable inputs) of transiting either to state $\epsilon_{k+1} = q - 1$ (with a correct decision) or to state $\epsilon_{k+1} = L$ (with an incorrect decision) at the next time instant (see Fig. 6). Subsequently, the transition probability matrix can be partitioned as

$$P \triangleq (p_{ij}) = \left[\begin{array}{c|c} Q & 0_L \\ \hline r' & 1 \end{array}\right] \in \mathbb{R}^{(L+1)\times(L+1)} \quad (12)$$

where

$$p_{ij} = \Pr(\epsilon_{k+1} = L + 1 - i | \epsilon_k = L + 1 - j)$$

and

$$Q = \left[\begin{array}{c|c} \frac{1}{2} \cdots \frac{1}{2} & \frac{1}{2} \\ \hline \frac{1}{2} I_{L-1} & 0_{L-1} \end{array}\right] \in \mathbb{R}^{L \times L} \quad (13)$$

$$r' = [0, \ldots, 0, \tfrac{1}{2}] \in \mathbb{R}^{1 \times L}.$$

Since $\epsilon_k = 0$ is the unique absorbing class (containing only $E_k = 0$), the eigenvalues of Q are less than one (in magnitude) [3], [7]. Let π_k be the $(L + 1)$-vector whose ith component $\pi_{k,i}$ is the probability of the aggregated system state at time k being $L + 1 - i$:

$$\pi_{k,i} = \Pr(\epsilon_k = L + 1 - i), \qquad i = 1, \ldots, L + 1.$$

This state distribution vector evolves according to

$$\pi_{k+1} = P\pi_k. \quad (14)$$

Now, suppose the initial error state $\bar{E}_0 \in \bar{\mathscr{E}}$ induces the state distribution π_l at time l. The probability of the system failing to reach the absorbing state ϵ_k at time $k > l$, while operating on a worst case channel, is given by

$$p_k(\pi_l) \triangleq \Pr(\epsilon_k \neq 0 | \pi_l) = \sum_{i=1}^{L} \Pr(\epsilon_k = i | \pi_l). \quad (15)$$

In other words, $p_k(\pi_l)$ is the sum of the first L components of the vector π_k. If we partition π_k conformably

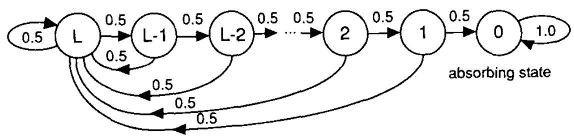

Fig. 6. Aggregated FSMP for a worst case channel with order L.

with (12) as

$$\pi_k = \begin{bmatrix} \overline{\pi}_k \\ \rho \end{bmatrix},$$

and make repeated use of (14), we have

$$p_k(\pi_l) = [\underbrace{1 \cdots 1\ 0}_{L+1}]\pi_k = [\underbrace{1 \cdots 1\ 0}_{L+1}]P^{k-l}\pi_l = [\underbrace{1 \cdots 1}_{L}]Q^{k-l}\overline{\pi}_l, \tag{16}$$

in which $\overline{\pi}_l$ is an initial distribution across nonzero aggregated error states ϵ_l.

We can obtain an upper bound on the noiseless error recovery probability for a DFE by setting $l = 0$ in (16) and applying the power method to the matrix Q. This yields a bound that behaves asymptotically (for large k) as $O(\lambda_1^k)$, where $\lambda_1 \in (0, 1)$ is the dominant eigenvalue of Q. The computation for the bound is relatively straightforward and can also be applied to the FPE by virtue of the extended state space FSMP model. In the FPE case, however, the bound is on the error probability.

D. Noiseless Error Probability Bound for the FPE

We now concretize the link between the DFE and the FPE with a view to applying the FSMP analysis described before, treating the noiseless case for simplicity. At each time instant k, the FPE (10) is equivalent (in the sense of producing the same output from a given sequence of inputs) to a (turned, noiseless) DFE that has been initialized in the nonzero error state

$$\overline{E}_{k-d} = [u_{k-d-1}, \cdots, u_{k-d-L}]' \in \overline{\mathscr{E}} \tag{17}$$

at time $k - d$. Recall that we have forced $\tilde{u}_{k-d-j} = 0$ for $j + 1, \cdots, L$.

We can reformulate (10) in a way which reflects more lucidly the internal structure of the FPE. We denote by $\gamma_k^i (1 \le i \le d + 1)$ the internal binary decision generated at the ith layer of the network, used in computing the eventual output $\tilde{u}_k$ at time k. (Hence, γ_k^i is the output of the ith diagonal node in the FPE of Fig. 5 when $d_i = h_i$ and there is no noise.) These preliminary decisions are obtained iteratively as follows:

$$\gamma_k^{i+1} = \begin{cases} \operatorname{sgn}\left(u_{k-d+i} + H'\overline{E}_{k-d+i}\right) & \text{if } 0 \le i \le d \\ 0 & \text{otherwise} \end{cases}$$
$$\overline{E}_{k-d+i} = \left[u_{k-d+i-1} - \gamma_k^i, \cdots, u_{k-d+i-L} - \gamma_k^{i+1-L}\right]' \in \overline{\mathscr{E}} \tag{18}$$

and $\tilde{u}_k = \gamma_k^{d+1}$. Note that we have assigned $\gamma_{k_i} = 0$ for $i < 1$ to match the initial conditions and produce the same $\tilde{u}_k$ as in (10). Thus, $\tilde{u}_k$ is the output of an FSMP with initial state $\overline{E}_{k-d}$ driven by the input sequence $\{u_{k-d}, \cdots, u_k\}$, and passing though the sequence of states $\overline{E}_{k-d+i}$ $(i = 0, \cdots, d)$. In what follows, we take $[1, H']$ to be a worst case channel and aggregate the states according to Definition 1.

We can calculate the probability that the FPE decision is correct at time k, supposing an "initial" state distribution π_{k-d}. We use Bayes' rule to condition on the aggregated state ϵ_k corresponding to the extended state $\overline{E}_k$ [defined by (18)] at the $(d + 1)$st layer (output) of the FPE, viz.

$$\begin{aligned} \Pr(\tilde{u}_k = u_k | \pi_{k-d}) &= \Pr(\tilde{u}_k = u_k | \pi_{k-d}, \epsilon_k \ne 0) \Pr(\epsilon_k \ne 0 | \pi_{k-d}) \\ &\quad + \Pr(\tilde{u}_k = u_k | \pi_{k-d}, \epsilon_k = 0) \Pr(\epsilon_k = 0 | \pi_{k-d}). \end{aligned}$$

Now, the ϵ_k-states have the transition diagram, Fig. 6, corresponding to the choice of a worst case channel. Hence, in the absorbing state $\epsilon_k = 0$, the probability of the correct decision is 1, whereas for all states $(\epsilon_k \ne 0)$ outside the absorbing state, this probability is $\frac{1}{2}$. We now have [see (15)]

$$\begin{aligned} \Pr(\tilde{u}_k = u_k | \pi_{k-d}) &= \tfrac{1}{2} p_k(\pi_{k-d}) + 1 - p_k(\pi_{k-d}) \\ &= 1 - \tfrac{1}{2} p_k(\pi_{k-d}). \end{aligned} \tag{19}$$

However, unlike the DFE, the "initial" distribution π_{k-d} (for each k) is not arbitrary. In fact, all "initial" error states $\overline{E}_{k-d}$ for the FPE (10) will belong to aggregated state $\epsilon_{k-d} = L$. To see this, recall the shift register property for the extended states (18). Clearly, we would need at least a sequence of L correct decision $\gamma_k^{i+1} = u_{k-d+i}$ $(i = 0, \cdots, L - 1)$ to attain the absorbing state $\overline{E}_{k-d+L-1} = 0$. Therefore, the particular "initial" distribution we seek is

$$\pi_{k-d}^* = [\underbrace{1, 0, \cdots, 0}_{L+1}]', \qquad \forall k$$

corresponding to $\epsilon_{k-d} = L$.

We now drop the conditioning on the left-hand side of (19) since only one π_{k-d} is possible in the FPE case. Applying (16) with $\pi_{k-d} = \pi_{k-d}^*$, we now have:

Theorem 1 (FPE): The noiseless decisions error probability for the $(d + 1)$-layer FPE generated by (10), with worst case channel parameters, is given by

$$\Pr(\tilde{u}_k \ne u_k) = \tfrac{1}{2} p_k(\pi_{k-d}^*) = \tfrac{1}{2}[\underbrace{1, \cdots, 1}_{L}]Q^d[\underbrace{1, 0, \cdots, 0}_{L}]'$$

where Q is given by (13) for channel order L.

As an immediate consequence of Theorem 1, we have the following.

Corollary 1 (Upper Bound): The noiseless error probability of a (d + 1)-layer feedforward perceptron equalizer operating on an arbitrary FIR channel with order L is upper bounded by

$$\Pr(\tilde{u}_k \neq u_k) \leq \tfrac{1}{2} p_k(\pi^*_{k-d})$$

where $p_k(\pi^*_{k-d})$ is given in Theorem 1.

The above bound is tight in the sense that certain channels that realize the bound exist, but on practical channels (e.g., with decaying impulse responses), fewer layers would be required to obtain the same noiseless error probability. In order to obtain a tight bound on the noiseless error probability for channels which are not worst case, we could still apply FSMP analysis. In general, however, it would not be possible to aggregate the FSMP model to obtain the simple L-state description used above, so the full 5^L-state nonaggregated FSMP would need to be used.

E. Asymptotic Form of the Bound

Applying the noiseless error bound result for the DFE found in [6], it is easy to deduce the following asymptotic form of the bound in Corollary 1 stated below without proof.

Corollary 2: The noiseless error probability of the feedforward perceptron equalizer, when operating on a worst case channel of order L, asymptotically satisfies

$$\Pr(\tilde{u}_k \neq u_k) \approx \tfrac{1}{2} \alpha^*_a \lambda_1^d$$

where

$$\alpha_1^* = \left(1 + \sum_{j=1}^{L-1} \left(\sum_{i=1}^{j} \mu^i\right)^2\right)^{-1/2},$$

$\mu \triangleq (2\lambda_1)^{-1}$, and λ_1 is the (unique) dominant eigenvalue of Q (13).

For long channel impulse responses, λ_1 approaches unity. We can obtain a further asymptotic expression by expanding the term α_1^* in Corollary 2. Since $\lambda_1 < 1$, we have $\mu > \frac{1}{2}$, and therefore $\sum_{i=1}^{j} \mu^i > 1 - (\frac{1}{2})^j$. After some manipulations, we obtain

$$\Pr(\tilde{u}_k \neq u_k) \lesssim \tfrac{1}{2}\left[L - \tfrac{5}{3} + (\tfrac{1}{2})^{L-2} - \tfrac{1}{3}(\tfrac{1}{4})^{L-1}\right]^{-1/2}.$$

The above formula overbounds the noiseless error probability of the FPE derived from a DFE tuned to the channel as a function of the channel length alone (for any number of layers). It would be possible to incorporate noise into the error bound in much the same way as the treatment in [9].

F. Nonadaptive Performance of the Presence of Noise

It will be apparent from the next example that the FPE structure can attain performance equaling that of a DFE when operating on an FIR channel in the presence of noise. In Fig. 7, the bit error rate of a DFE on a [1, 2, 1] binary partial response channel has been simulated (over 10^5 points) at a single-to-noise ratio of 10 dB (where the SNR is defined as $10 \log_{10} \sigma^2$ (dB) where σ^2 is the variance of the white Gaussian noise). On the same graph, we show the simulated error probability versus the number of layers (minus one) of an FPE whose weights have been tuned to the channel coefficients. For this example, the two systems perform with effectively the same probability of error when the number of layers is 50 or greater. We mention that an MLP equalizer with only three layers is capable of better performance than a DFE [2]. However, the FPE has much lower connectivity than the standard MLP neural network; in addition, many of the weights (d_j in Fig. 5 for $j > L$) can be set to zero with no appreciable effect on performance.

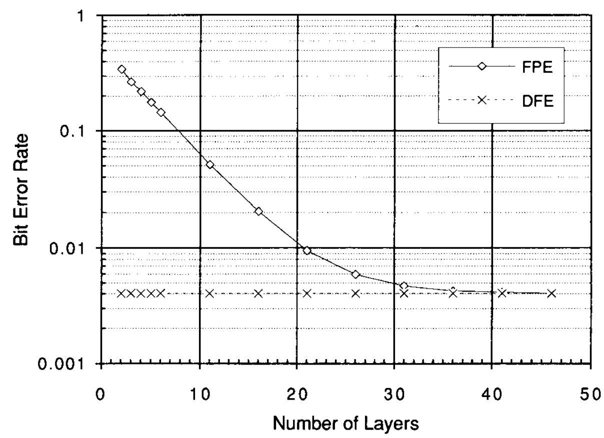

Fig. 7. Simulated BER of FPE versus number of layers ($d + 1$).

IV. Conclusions and Further Work

We have addressed the problem of deriving a feedforward perceptron equalizer which approximates a nonadaptive DFE on a given channel with desired accuracy. The computation of a tight bound on the noiseless error probability of such a system, as a function of the number of layers, was effected by using a finite-state Markov process for a worst case channel. The noiseless assumption was used to simplify the presentation. A numerical example showed that the asymptotic performance of an FPE operating on a noisy second-order channel is the same as that of the DFE from which it derives.

We stress here that the upper bound on the FPE's performance is in no way representative of the achievable performance of more general perceptron-based equalizers such as [1], [2]. Indeed, it is possible to devise MLP structures which outperform the DFE (although they may require a larger decision delay than the latter). As previously mentioned, however, the analysis of general MLP networks and equalizers is still an open problem.

References

[1] S. Chen, G. J. Gibson, and C. F. N. Cowan, "Adaptive channel equalization using a polynomial-perceptron structure," *Proc. IEE, Part I*, vol. 137, pp. 257–264, Oct. 1990.

[2] S. Siu, G. J. Gibson, and C. F. N. Cowan, "Decision feedback equalization neural network structures and performance comparison with the standard architecture," *Proc. IEE, Part I*, vol. 137, pp. 221–225, Aug. 1990.

[3] L. Isaacson, and R. W. Madsen, *Markov Chains, Theory and Applications*. New York: Wiley, 1985.

[4] D. L. Duttweiler, J. E. Mazo, and D. G. Messerschmitt, "An upper bound on the error probability in decision feedback equalization," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 490–497, July 1974.

[5] A. Cantoni and P. Butler, "Behavior of decision feedback inverses," *IEEE Trans. Commun.*, vol. COM-24, pp. 1064–1075, Sept. 1976.

[6] R. A. Kennedy and B. D. O. Anderson, "Recovery times of decision feedback equalizers on noiseless channels," *IEEE Trans. Commun.*, vol. COM-35, pp. 1012–1021, Oct. 1987.

[7] F. R. Gantmacher, *The Theory of Matrices, Vol.* 2. New York: Chelsea, 1974.

[8] A. M. de Oliveira Duarte, and J. J. O'Reilly, "Simplified techniques for bounding error statistics for DFB receivers," *Proc. IEE, Part F, Commun., Radar and Signal Processing*, vol. 132, no. 7, pp. 567–575, 1985.

[9] R. A. Kennedy, B. D. O. Anderson, and R. R. Bitmead, "Tight bounds on the error probabilities of decision feedback equalizers," *IEEE Trans. Commun.*, vol. COM-35, pp. 1022–1028, Oct. 1987.