Proceedings of the Second Australian Conference on Neural Networks

ACNN'91

Sydney, February 4–6, 1991

Edited by

Marwan Jabri

Systems Engineering and
Design Automation Laboratory

Sydney University Electrical
Engineering, NSW 2006, Australia

Conference Chairman

Dr Marwan Jabri
Director
Systems Engineering & Design Automation Laboratory (SEDAL)
School of Electrical Engineering
University of Sydney

Conference Co-Chairman

Prof Max Bennett
President
Australian Neurosciences Society
Neurobiology Research Centre
University of Sydney

Technical Program Chairs

A/Prof Ah Chung Tsoi, University of Queensland
Prof Bill Levick, Australian National University

Conference Steering Committee

Prof Yanni Attikiouzel, University of Western Australia
Prof Bob Bogner, University of Adelaide
Prof Richard Brent, Australian National University
Mr Jacob Cybulski, Amdahl Australian Intelligent Tools Program
Dr Tom Osborn, University of Technology, Sydney
Prof Steve Redman, Australian National University
A/Prof Sam Reisenfeld, OTC
Prof Graham Rigby, University of New South Wales
Dr Steve Schwartz, University of Queensland
Dr Charles Watson, D S T O
Dr Janet Wiles, University of Queensland
Dr Hong Yan, University of Sydney

A NEURAL NET STRUCTURE FOR DECISION FEEDBACK EQUALISERS[†]

Graham W. Pulford[‡], Rodney A. Kennedy[‡§] and Brian D. O. Anderson [‡]

[†] Work supported by the Australian Research Council, and ANU Centre for Information Science Research.
[‡] Dept. of Systems Eng., RSPhysS, Australian National University, G.P.O. Box 4, Canberra, A.C.T. 2601, AUSTRALIA
§ Interdisciplinary Engineering Program, Faculty of Science, Australian National University, G.P.O. Box 4, Canberra, A.C.T. 2601, AUSTRALIA

ABSTRACT: The emulation of a tuned decision feedback equaliser, operating on a noiseless finite impulse response channel, by a feedforward multilayer neural network is considered. The similarity between the two structures is exploited to obtain tight bounds on the probability of error for the neural net, as a function of the number of layers, using the theory of finite state Markov processes. A class of channels for which exact representation by a neural net of finite complexity is possible is established.

1 Introduction

Much current research on neural networks focuses on their functional aspects—the training and experimental verification of a system with desired input-output properties. Such simulation-based approaches seek to justify the use of neural nets in an empirical fashion, avoiding analysis. This has the disadvantage of giving little insight into why the system works, or even if the problem is well posed in the sense of functional representation.

In the field of high speed digital data transmission over channels introducing significant intersymbol interference, neural networks have been applied to the equalisation problem, but have met with limited success [8]. The inherent nonlinearity of these devices complicates the analysis and gives rise to heuristic approaches. There is no clearly defined method for choosing the number of layers or the number of nodes per layer for a specific problem. Even though the question of existence of desirable parameter settings has been addressed [3], the target values of connecting weights for a particular network are usually not known in advance.

Seeking to justify the use of neural networks in equalisation from a more theoretical standpoint, we consider a tuned noiseless decision feedback equaliser (DFE) operating on a linear finite impulse response (FIR) channel. We derive a related feedforward neural net that emulates this well studied system. We provide analysis that determines the conditions under which the two systems are effectively equivalent.

The paper is organised as follows: In section 2, we present the DFE and develop a representation that has the structure of a recursive multilayer neural net. In section 3, we introduce an extended-state finite state Markov process (FSMP) that is capable of modelling this neural net with the recursion turned off—generating a feedforward neural net. Section 4 describes a class of channels that guarantee the worst bit error rate performance for the DFE. In section 5, we examine the error recovery properties of a noiseless DFE under the extended state model. The results are applied to derive a tight bound (in terms of the number of layers) for the feedforward neural net, given in section 6. Section 7 examines the conditions on the channel under which the representation is exact (in the sense of the two systems producing the same output) and contains some numerical examples of neural nets that approximate a DFE to a given degree of accuracy. Conclusions are drawn in section 8.

2 The Nonadaptive Decision Feedback Equaliser

We consider a binary decision feedback equaliser operating on a finite impulse response channel corrupted by additive zero-mean white Gaussian noise n_k. At the output of the channel the measured signal is

$$y_k = \sum_{i=0}^{L} h_i u_{k-i} + n_k, \tag{2.1}$$

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

$$\widehat{u}_k = \mathrm{sgn}\Big(y_k - \sum_{j=1}^{L} d_j \widehat{u}_{k-j}\Big). \tag{2.2}$$

[1] $\mathrm{sgn}(x) = 1$ if $x \geq 0$ and -1 if $x < 0$

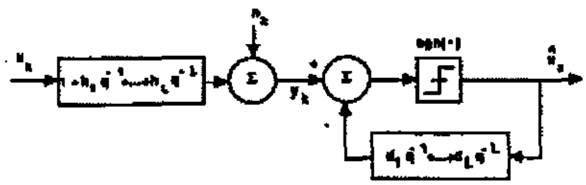

Figure 1. Decision Feedback Equaliser

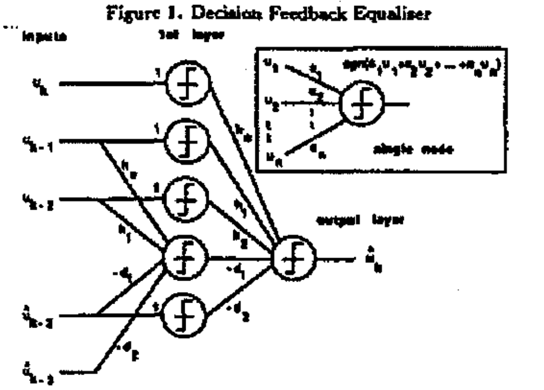

Figure 2. 2-layer ($d = 1$) recursive neural net for DFE

The feedback tap gains d_j are adjusted to cancel the intersymbol interference introduced by the channel. We assume that the cursor $h_0 = 1$.

We develop an alternative recursive representation of (2.2) by eliminating in turn $\widehat{u}_{k-1}$ down to $\widehat{u}_{k-d}$ ($d \geq 1$) and reexpressing $\widehat{u}_k$ as a function of the outputs $(y_k, \cdots, y_{k-d})$ and the past decisions $(\widehat{u}_{k-d-1}, \cdots, \widehat{u}_{k-d-L})$, symbolically

$$\widehat{u}_k = f_d^L(y_k, \cdots, y_{k-d}; \widehat{u}_{k-d-1}, \cdots, \widehat{u}_{k-d-L}). \tag{2.3}$$

For instance the first stage of this process yields

$$\begin{aligned} \widehat{u}_k &= f_1^L(y_k, y_{k-1}; \widehat{u}_{k-2}, \cdots, \widehat{u}_{k-L-1}) \qquad (2.4)\\ &= \mathrm{sgn}\Big(y_k - \sum_{j=2}^{L} d_j \widehat{u}_{k-j} - d_1 \mathrm{sgn}\big(y_{k-1} - \sum_{j=1}^{L} d_j \widehat{u}_{k-j-1}\big)\Big). \end{aligned}$$

After d such steps d nested sgn(·) functions are introduced as a new argument of (2.2). In this representation the DFE takes on the structure of a $(d+1)$-layer recursive neural network with sgn(·) functions replacing the usual *sigmoid* functions, as illustrated for $d = 1$ in Figure 2. Intuitively it is reasonable to expect that the dependence of $\widehat{u}_k$ on the past decisions $\widehat{u}_{k-d-j}$ decreases the larger we make d (at least for $d > L$). This is equivalent to the effect of earlier decisions ceasing to influence later decisions given a large enough time delay. We give substance to this claim in the following sections.

3 Finite State Markov Process Description

We can model the stochastic dynamics of the DFE using the theory of finite state Markov processes (FSMPs) [10] (this follows from the independence assumption on the input sequence). Referring to (2.1) and (2.2), the input is u_k and we choose

$$X_k = (u_{k-1}, \cdots, u_{k-L}, \widehat{u}_{k-1}, \cdots, \widehat{u}_{k-L})'$$

as the state vector. There are 4^L states in total if all elements are binary. We can in this case reduce the 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 $e_{k-i} = u_{k-i} - \widehat{u}_{k-i}$ can take the values ± 2 or 0. We denote by $\mathcal{E}$ the set of all 3^L E_k states. Consequently, the tuned DFE output can be reexpressed as

$$\widehat{u}_k = \text{sgn}(u_k + \sum_{i=1}^{L} h_i e_{k-i} + n_k).$$

In order to simplify the analysis, as is standard in the analysis of error propagation of DFEs [2,4], we consider only the noise-free case ($n_k = 0 \ \forall k$) so that the unique absorbing state of the FSMP is $E_k = \underline{0}$. To see this, note that the error vector or state satisfies the simple recursion

$$\begin{aligned} E_{k+1} &= \begin{bmatrix} 0 & 0 \\ I_{L-1} & 0 \end{bmatrix} E_k \\ &+ \begin{bmatrix} u_k - \text{sgn}(u_k + H'E_k) \\ 0_{L-1} \end{bmatrix}, \end{aligned} \tag{3.1}$$

where $H \triangleq (h_1, \cdots, h_L)'$, I_n is the identity matrix of order n and 0_n is a vector of n zeros. If $E_k = \underline{0}$ then for all inputs we have $\text{sgn}(u_k + H'E_k) = u_k$ and the DFE remains in the zero-error state regardless of the input. However, from an E_k state having a nonzero entry, there is a nonzero probability of reaching the absorbing state in M steps for any $M \geq L$ (since a sequence of L correct decisions will cause transition to the absorbing state). Also, the probability of ultimately reaching the absorbing state is 1 [2]. When noise is present, only a noise-induced decision error can cause a transition from 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, depends on the sequence of inputs $u_{k-d}, \ldots, u_k$ and the initial state X_{k-d} (or equivalently E_{k-d}). For convenience we introduce the notation (with reference to (2.3))

$$\begin{aligned} & f_d^L(y_k, \cdots, y_{k-d}; \widehat{u}_{k-d-1}, \cdots, \widehat{u}_{k-d-L})|_{n_k = \cdots = n_{k-d} = 0} \\ \triangleq \ & g_d^L(u_k, \cdots, u_{k-d-L}; \widehat{u}_{k-d-1}, \cdots, \widehat{u}_{k-d-L}). \end{aligned} \tag{3.2}$$

We aim to determine a tight upper bound on the probability of an incorrect decision at time k from an arbitrary nonzero initial error state. This measure of performance will be seen to be central to our analysis of the particular neural net structure to follow. With this in mind, we define an extended state $\overline{X}_k$ ($\overline{E}_k$ in the tuned case) having the same form as X_k but in which the decisions which appear in the initial state may take the additional value zero. Now each element $\overline{e}_{k-i}$ of the extended error state $\overline{E}_k$ may take one of 5 possible values: $0, \pm 1, \pm 2$, so there are a total of 5^L $\overline{E}_k$ states comprising the set $\overline{\mathcal{E}}$. Of course, a DFE with an initial state in $\overline{\mathcal{E}}$ reverts to a DFE with state in $\mathcal{E}$ after L time units because the binary decisions being fed back will displace the initial conditions. The concept of an extended state, together with the error probability bound, will allow us to gauge the effect of omitting the feedback (or recursive) part (*i.e.*, decisions) in the recursive representation for the DFE, thus obtaining a *feedforward* neural net structure generated by

$$\begin{aligned} \widetilde{u}_k \ &= \ g_d^L(u_k, \cdots, u_{k-d}, u_{k-d-1}, \cdots, u_{k-d-L}; \\ & \qquad \widetilde{u}_{k-d-1} := 0, \cdots, \widetilde{u}_{k-d-L} := 0), \end{aligned} \tag{3.3}$$

where $\widetilde{u}_k$ is binary, and by the notation $\widetilde{u}_{k-d-j} := 0$, we mean that any feedback paths in the recursive neural net (3.2) are to be deleted. Note that the same $\widetilde{u}_k$ as (3.3) would be generated by a DFE (with feedback), started in an abnormal initial state $\overline{E}_{k-d}$ with fictitious past decisions $\widehat{u}_{k-d-1} = 0, \cdots, \widehat{u}_{k-d-L} = 0$ and fed with the sequence of inputs $u_{k-d}, \cdots, u_k$. Thus, the feedforward neural net in (3.3) 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 6.

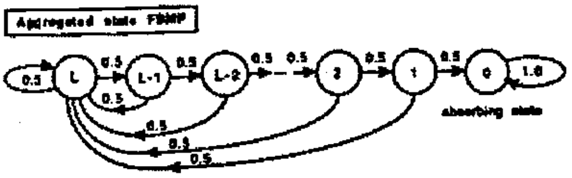

Figure 3. FSMP of a worst case channel

4 Worst Case Channels

Extending the development found in [4,7], we now determine a class of channels on which the DFE has the worst possible performance (in terms of error propagation) from a possibly zero initial condition (*i.e.*, a state in $\overline{\mathcal{E}}$ but not in $\mathcal{E}$). This will allow us to bound the performance of the DFE on an arbitrary FIR channel. Any channel ($h_0 = 1, H'$) satisfying

$$\min_{\overline{E}_k \neq 0} | H'\overline{E}_k | > 1 \tag{4.1}$$

has the property that $Pr(\widehat{u}_k \neq u_k) = \frac{1}{2}$ for any nonzero extended error state $\overline{E}_k$. This follows since the inputs are equiprobable and therefore u_k has a probability of $\frac{1}{2}$ of having the same sign as $H'\overline{E}_k$ (noting that $\widehat{u}_k = \text{sgn}(u_k + H'\overline{E}_k)$). Channels satisfying (4.1) will be termed "worst case" channels. We claim that the expected error recovery time [2] is maximised for such channels, which form a subset of the worst case class in [2] since we are allowing E_k to have the additional values ± 1. Any channel (with $h_0 = 1$) whose parameters belong to the set

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

will also fall into the worst case category. This is because the h_j have been spaced so far apart that no linear combination with coefficients in the set $\{0, \pm 1, \pm 2\}$ has magnitude less than 1. (The same is true of any channel that can be obtained from this set by changes of sign and/or permutation of parameters.)

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

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

demonstrating that [1 1.2 3.5] is a worst case channel. That is, no $L = 2$ channel may have an expected recovery time which exceeds that of the above channel [2].

5 Worst Case Finite State Markov Process

In what follows we may assume the DFE is operating on a worst case channel so as to obtain a tight bound on performance. The order of our FSMP model can be reduced by aggregating states, provided that states to be aggregated have identical transition probabilities [10]. Because all transitions outside the absorbing state have probability $\frac{1}{2}$, we choose to aggregate the $\overline{E}_k$ states according to the following rule [7]:

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

From the shift register property (3.1) applied to $\overline{E}_k$ (by allowing $\widehat{u}_k = 0$), it is clear that at most L successive correct decisions are needed to force an arbitrary nonzero $\overline{E}_k$ state to the absorbing state. Hence there are $L + 1$ aggregated states c_k in the new FSMP. From a given state $c_k = q$ ($q \neq 0$) there is a probability of $\frac{1}{2}$ (for equiprobable inputs) of transiting either to state $c_{k+1} = q - 1$ (with a correct decision) or to state $c_{k+1} = L$ (with an incorrect decision) at the next time instant (see Fig. 3). Subsequently the transition probability matrix can be partitioned as

[2] Expected time to reach the absorbing state

$$P \stackrel{\Delta}{=} (p_{ij}) = \left[\begin{array}{c|c} Q & 0_L \\ \hline r' & 1 \end{array}\right] \in R^{(L+1)\times(L+1)} \quad (5.1)$$

where

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

and

$$Q = \left[\begin{array}{ccc|c} \frac{1}{2} & \cdots & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2}I_{L-1} & & 0_{L-1} \end{array}\right] \in R^{L\times L} \quad (5.2)$$

$$r' = \left[0 \cdots 0 \ \frac{1}{2}\right] \in R^{1\times L}.$$

Since $\epsilon_k = 0$ is the unique absorbing class (containing only $E_k = \underline{0}$), the eigenvalues of Q are less than one (in magnitude) [9]. 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 $L+1-i$, i.e.,

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

This state distribution vector evolves according to

$$\pi_{k+1} = P\pi_k. \quad (5.3)$$

Now suppose the initial error state $\overline{E}_0 \in \overline{\mathcal{E}}$ induces the distribution π_0 at time $k=0$. We can compute the probability of the system failing to reach the absorbing state ϵ_{k+1} at time $k+1$, while operating on a worst case channel, by

$$p_k(\pi_0) \stackrel{\Delta}{=} Pr(\epsilon_{k+1} \neq 0 \mid \pi_0) = \sum_{i=1}^{L} Pr(\epsilon_{k+1} = i \mid \pi_0). \quad (5.4)$$

In other words, $p_k(\pi_0)$ is the sum of the first L components of the vector π_k. If we partition π_k conformably with (5.1) as

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

and make repeated use of (5.3), we have from (5.4)

$$p_k(\pi_0) = [\underbrace{1 \cdots 1\ 0}_{L+1}]\pi_k = [\underbrace{1 \cdots 1}_{L}]Q^{k+1}\overline{\pi}_0,$$

in which $\overline{\pi}_0$ is the initial distribution across nonzero aggregated error states ϵ_k. Applying the power method to the matrix Q, we obtain the upper bound stated in the following theorem (see [4] for a full proof):

Theorem 1 (Tuned DFE): *Consider a tuned DFE operating on a noiseless worst case FIR channel of order L, initially in nonzero error state $\overline{E}_0 \in \overline{\mathcal{E}}$ at time 0. If $\overline{E}_0$ induces the aggregated state distribution π_0, the probability of not reaching the absorbing state $\epsilon_k = 0$ at time k is given by*

$$p_k(\pi_0) \stackrel{\Delta}{=} Pr(\epsilon_k \neq 0 \mid \pi_0) = \alpha_1\lambda_1^k(1 + o(2^{-k})),$$

where $\alpha_1 = \widehat{w}_1'\overline{\pi}_0 \in (0,1]$,

$$\lambda_1 \stackrel{\Delta}{=} \max_{1\leq i\leq L} \lambda_i(Q) \in (0,1),$$

is the unique dominant eigenvalue of Q (5.2), and $\widehat{w}_1 = w_1/\|w_1\| \in R^L$ where w_1 is the eigenvector of Q corresponding to λ_1, given by

$$w_1 = \left[1, \sum_{j=1}^{L-1}\mu^j, \sum_{j=1}^{L-2}\mu^j, \cdots, \mu+\mu^2, \mu\right]',$$

with $\mu = (2\lambda_1)^{-1}$.
□

Since the calculations assumed a worst case channel, the above bound indicates the highest degree to which an initial error can influence subsequent decisions, due to error propagation alone, on any channel. The bound may be conservative for general channels in that $\lambda_1 \approx 1$ for worst case channels. We list λ_1 for varyious channel lengths (L) in Table 1. For practical channels (*e.g.*, with decaying impulse responses) the functional form of the bound is still valid (with correspondingly smaller λ_1), but in general it is not possible to aggregate the FSMP model to obtain the Lth order description, so the 5^L order nonaggregated FSMP must be used.

$L = dim(Q)$	2	3	4	5	6
λ_1	0.8090	0.9196	0.9638	0.9830	0.9918

Table 1. Dominant eigenvalue of Q

6 Error Probability Bound for Neural Net

In the last section we saw that the probability of an incorrect decision in the DFE, due to some initial decision error, becomes exponentially less likely with time. Equivalently we can say that after a sufficient time, the effect of an initial error is negligibly small. Because of the structural link between the DFE and the recursive neural net (3.3), this "settling time" will be shown to be an indicator of the depth of recursion d (or number of layers) needed to obtain a *feedforward* neural net that is a good approximation (in the sense of error probability) to the original DFE. We imply by "feedforward" that the recursive representation in section 2 is truncated after d steps and there is no feedback of past decisions.

At each time istant k, the feedforward neural net (3.3) is equivalent (in the sense of producing the same output from a given sequence of inputs) to a (tuned noiseless) DFE that has been initialised to the nonzero error state

$$\overline{E}_{k-d} = (u_{k-d-1},\cdots,u_{k-d-L})' \quad (6.1)$$

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

We can reformulate (3.3) in a way which reflects more lucidly the internal structure of the neural net. We denote by γ_k^i $(1 \leq i \leq d+1)$ the internal binary decision generated by the ith layer of the net, used in computing the eventual output $\widetilde{u}_k$ at time k. These preliminary decisions are obtained iteratively as follows:

$$\gamma_k^{i+1} = \begin{cases} \mathrm{sgn}(u_{k-d+i} + H'\overline{E}_{k-d+i}) & \text{if } 0 \leq i \leq d \\ 0 & \text{otherwise} \end{cases}$$

$$\overline{E}_{k-d+i} = (u_{k-d+i-1} - \gamma_k^i,\cdots,u_{k-d+i-L} - \gamma_k^{i+1-L})' \in \overline{\mathcal{E}}, \quad (6.2)$$

and $\widetilde{u}_k = \gamma_k^{d+1}$. *Note that we have assigned $\gamma_k^i = 0$ for $i < 1$ to match* the initial conditions and produce the same $\widetilde{u}_k$ as in (3.3). Thus $\widetilde{u}_k$ is the output of a FSMP with initial state $\overline{E}_{k-d}$ driven by the input sequence $\{u_{k-d},\cdots,u_k\}$, and passing through the sequence of states $\overline{E}_{k-d+i}$ $(i = 0,\cdots,d)$. In what follows, we take $(1, H')$ as a worst case channel and aggregate the states according to the same rule as before.

We can calculate the probability that the neural net 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 (6.2)) at the $(d+1)$st (output) layer of the neural net. Thus

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

Now, the ϵ_k states have the transition diagram (Fig. 2) corresponding to the choice of a worst case channel. Hence in the absorbing state $\epsilon_k = 0$ the probability of a correct decision is 1, whereas for all states $(\epsilon_k \neq 0)$ outside the absorbing state this probability is $\frac{1}{2}$. With a view to applying Theorem 1 we let $Pr(\epsilon_k \neq 0|\pi_{k-d}) = p_k(\pi_{k-d})$. We now have

$$\begin{aligned} Pr(\widetilde{u}_k = u_k|\pi_{k-d}) &= \frac{1}{2}p_k(\pi_{k-d}) + 1 - p_k(\pi_{k-d}) \\ &= 1 - \frac{1}{2}p_k(\pi_{k-d}). \end{aligned} \quad (6.3)$$

However, unlike in the DFE, the "initial" distribution π_{k-d} (for each k) cannot be chosen arbitrarily. In fact all "initial" error states $\overline{E}_{k-d}$ for the neural net (3.3) will belong to the aggregated state $\epsilon_{k-d} = L$. To see this, recall the shift register property for the extended states (6.2). *Clearly we would need a sequence of L correct decisions* $\gamma_k^{i+1} = u_{k-d+i}$ $(i = 0,\cdots,L-1)$ to attain the absorbing state $\overline{E}_{k-d+L-1} = \underline{0}$. Therefore, the particular "initial" distribution we seek is

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

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

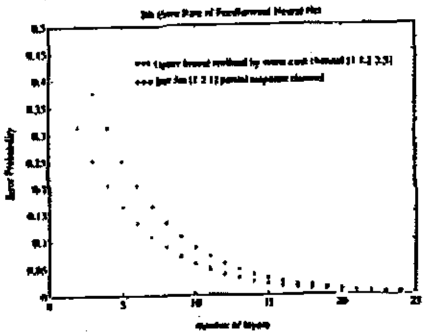

Figure 4. Error probabilities for 2nd-order channel

We now drop the conditioning on the left hand side of (6.3) since only one π_{k-d} is possible in the neural net case. Applying Theorem 1 to evaluate $p_k(\pi_0)$ with $\pi_0 \triangleq \pi^*_{k-d}$, yields the asymptotic formula for the error probability which we state below.

Theorem 2 (Feedforward Neural Net): *The decision error probability $Pr(\tilde{u}_k \neq u_k)$ for the $d+1$-layer feedforward neural net generated by (3.3), with worst case channel parameters, is given by*

$$Pr(\tilde{u}_k \neq u_k) = \frac{1}{2} p_k(\pi^*_{k-d}) = \frac{1}{2} \sigma_1^* \lambda_1^d (1 + o(2^{-d})),$$

where

$$\sigma_1^* = \Big(1 + \sum_{j=1}^{L-1} \big(\sum_{i=1}^{j} (2\lambda_1)^{-i}\big)^2\Big)^{-\frac{1}{2}},$$

and λ_1 is the (unique) dominant eigenvalue of Q (5.2).
□

The reasoning carries over to the feedforward neural net because of the definition of the extended error state $\overline{E}_k$, which allows for the disconnection of the feedback part of the recursive structure in (3.2). The above bound is tight in the sense that certain channels that realise the bound exist, but on practical channels, fewer layers would be required to obtain the same error probability.

In Figure 4 we have plotted the upper bound resulting from the aggregated FSMP realised by a worst case 2nd-order channel. On the same graph, the decision error probability (simulated on 10^5 points) for a feedforward neural net operating on the 2nd-order partial response channel [1 2 1], as a function of the number of layers $(d+1)$, is displayed.

7 Quality of Approximation

The possibility arises of exact representation of a DFE by a nonrecursive neural net. This is indeed the case when the error propagation events have a guaranteed finite length. For certain classes of channels (known as stable channels), for example those which satisfy for all $\theta \in [0, 2\pi)$ the frequency domain condition

$$h_0/2 + \sum_{k=1}^{L} h_k \cos(k\theta) > 0,$$

it is known that the DFE has a deterministic, bounded error recovery time (from a nonzero $\mathcal{E}$ state) [6]. If we call this time D, then the feedforward neural net derived from the DFE for this stable channel will reproduce the output sequence $\{\tilde{u}_k\}$ exactly (from the same sequence of inputs) if the number of layers is chosen to be $\geq D + L$. The reason for this is that the decision $\tilde{u}_k$ is independent of $\tilde{u}_{k-D-j}$ for $j > 0$ due to the finite extent of error propagation on stable channels. Hence (including a further L layers to regain the $\mathcal{E}$ state space) a feedforward neural net with $D+L$ or more layers loses no information relative to deciding $\tilde{u}_k$ by neglecting these past decisions, bearing in mind that we are only treating the noiseless case.

We have tabulated in Table 2 the simulated (over 10^6 points) noiseless bit error rates (BER) for the neural net as a function of the number of layers $(d+1)$ while operating on the following 10th-order channels: (i) an exponential impulse response channel with coefficients $h_k = (0.8)^k \cos(k\pi/6)$, and (ii) a passive channel with coefficients [1.0 −0.27 −0.18 −0.31 0.27 0.09 −0.05 0.06 −0.08 0.084 0.01].

d	1	2	3	4
BER (i)	0.04129	0.02003	0.01287	0.00000
BER (ii)	0.01070	0.00100	0.00000	0.00000

Table 2. Error rate of neural net on two 10th order channels

It is clear, in this example, that the representation becomes exact (in the sense of reproducing exactly the input sequence in the absence of noise) when the number of layers is chosen to exceed the maximum duration of error events (known to be finite on both the above channels).

8 Conclusions

The problem of existence of a feedforward multilayer neural network, that approximates a nonadaptive DFE on a given channel with desired accuracy, has been addressed. The computation of a tight bound on the error probability of such a net, as a function of the number of layers, was obtained by using an extended state-space and aggregating the finite state Markov process for a worst case channel. This bound was evaluated for a simple example and the neural net simulated on various channels. It was explained that for channels on which the DFE has finite error recovery (e.g., so-called "passive" channels), a sufficient condition for the corresponding neural net to be an exact functional representation of the DFE is that the number of layers exceed the error recovery time by L (the channel order).

The important problem of representing an adaptive DFE by a neural net remains open, as does the application of neural nets in other areas of channel equalisation, to say nothing of the adaptation of such devices.

References

[1] 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.

[2] A. Cantoni, and P. Butler, "Behaviour of Decision Feedback Inverses", *IEEE Trans. Commun.*, vol. COM-24, pp. 1064-1075, Sept. 1976. (Also: Univ. Newcastle Tech. Report EE7505, Mar. 1975.)

[3] R. Hecht-Nielsen, "Theory of the Backpropagation Neural Network", *Proc. Int. Joint Conf. on Neural Networks*, pp. 593-605, 1989.

[4] 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.

[5] R.A. Kennedy, B.D.O. Anderson, and R.R. Bitmead, "Channels leading to rapid error recovery for decision feedback equalizers", *IEEE Trans. Commun.*, vol. COM-37, pp. 1146-1155, Nov. 1989.

[6] 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.

[7] A.M. de Oliveira Duarte, and J.J. O'Reilly, "Simplified Techniques for Bounding Error Statistics for DFB Receivers", *Proc. F, Commun., Radar and Signal Process.*, vol. 132, no. 7, pp. 567-575, 1985.

[8] G.J. Gibson, S. Siu, and C.F.N. Cowan, "Application of Multilayer Perceptrons as Adaptive Channel Equalisers", *IFAC Adaptive Systems in Control and Signal Processing*, Glasgow, 1989.

[9] F.R. Gantmacher, "The Theory of Matrices", vol. 2, Chelsea Publishing Company, N.Y., 1974.

[10] L. Isaacson, and R.W. Madsen, "Markov Chains, Theory and Applications", John Wiley and Sons Inc., 1985.