ISSPA 87, signal processing, theories, Implementations and Applications. Editor: B. Boashash, Brisbane, Australia 24-28 August, 1987.

Stochastic Analysis of Non-Adaptive Decision Feedback Equalizers*

RODNEY A. KENNEDY, BRIAN D. O. ANDERSON and ROBERT R. BITMEAD
Department of Systems Engineering, Research School of Physical Sciences, Australian National University, G.P.O. Box 4, Canberra, A.C.T., Australia
* Research supported by the Australian Telecommunications and Electronics Research Board (A.T.E.R.D.)

ABSTRACT: The performance of decision feedback equalizers (DFEs) working on a communication channel with correct taps is studied. In the case of noiseless channels we show that the space of channel parameters can accordingly be partitioned into a finite number of sets. The error recovery performance of a DFE is the same for all DFEs within one such set, and is determinable. We also discuss some tight bounds for recovery time, which are particularly important when the number of equalizer taps is not small. We argue that near minimum phase character for the channel does not necessarily guarantee short recovery time. For noisy channels we indicate that certain error probability upper bounds derived in the literature are tight.

1 INTRODUCTION

Our aim is to present some error recovery properties of decision feedback equalizers (DFEs) in terms of the parameters describing the communication channel. We analyze two cases: (i) noiseless channels and (ii) channels with additive noise. We show that in the case of noiseless channels the non−linearity of the decision function in the DFE receiver has the effect of partitioning the space of channel parameters into a finite number of sets. Further, we show that by examining this partition one is led naturally to classify some of the important non−adaptive properties of noiseless DFEs namely: (i) the error recovery time statistics, (ii) the input data sequences which result in arbitrarily long recovery times, and (iii) the identification of channels which are inappropriate for the use of DFEs. Also with the inclusion of noise into the analysis we are able to give some tight upper bounds on the error probability statistics.

Error recovery and error propagation effects of DFEs have been the subject of several papers and our work forms a natural extension of previous ideas. Duttweiler, Mazo and Messerschmitt [1] determine an upper bound on the steady state error probability of the DFE in terms of the probability of error in the absence of past decision errors. However, questions regarding tightness of these bounds were left open. More recently, O'Reilly and de Oliveira Duarte [2,3] extended the techniques and results in [1] by developing a procedure which gives upper and lower bounds on the steady state error statistics and recovery time statistics for a given channel. A different approach to the stochastic analysis of DFEs was given by Cantoni and Butler [4,5] who gave a bound on the expected error recovery time. They include a discussion on the input sequences which result in poor DFE recovery performance. These references (as will we) all use ideas based on finite state Markov processses (see also [6]).

Our contribution (based on [7]) is to analyze the DFE in terms of the noiseless (high signal to noise ratio) communication channel, and to show precisely how the channel parameters affect the stochastic dynamics of DFEs. We provide exact calculations for the classes of channel of greatest theoretical and practical interest. We emphasize that the stochastic analysis of DFEs can be conceptually reduced to the understanding of the one−to−one correspondence between disjoint polytopes in the space of channel parameters and a set of finite state Markov processes (or more generally a set of state transition diagrams). Our results extend, generalize and clarify the important contributions in [4]. The non−trivial but close connection of our results (and those in [4,5]) to the results in [1−3], and the inclusion of channel noise into the analysis, will be given in outline and we refer the reader to [7−8] for the details. We work exclusively using 2−level input sequences. The ideas could easily be generalized to handle M−level inputs (see e.g., [2−5]).

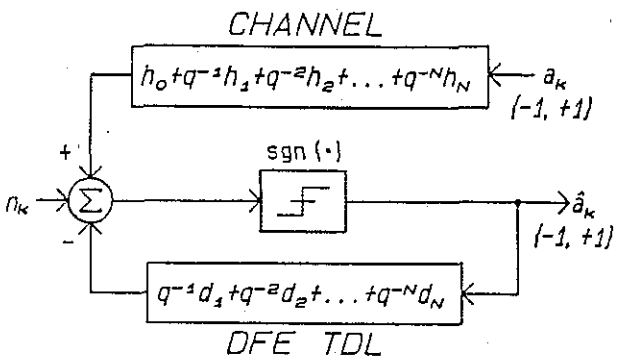

FIGURE 1: Channel and DFE Model.

We preview the contents of the remainder of this paper by section. Section 2 contains our definitions and a development of finite state Markov processes used to analyze the stochastic dynamics of DFEs. The N = 2 tap DFE is treated section 3. By extending the results and concepts found in section 3, we are able to treat in section 4 the general N ≥ 2 tap DFE. The noisy channel case is treated in section 5. The conclusions are given in section 6.

2 PROBLEM FORMULATIONS

2.1. System Definitions

The system under consideration is given in Fig.1. It shows the communication channel modelled as a finite impulse response filter driven by a white, zero mean, binary data stream $\{a_k\}$. This channel will be represented by a cursor $h_0 > 0$ paired with a tail $H_N = [h_1\ h_2\ \ldots\ h_N]^T \in R^N$ (notation: v^T denotes the transpose of v). A discussion of a wide class of physical channels (defined by those which are minimum phase) which generally consist of a dominant cursor followed, but not preceded, by a series of echoes (the tail), is given by Clarke [9]. He advocates the use of DFEs for the equalization of these channels.

The effect of the channel tail is to introduce intersymbol interference (ISI) which corrupts the information carried with the cursor h_0. Additive channel noise n_k may also contaminate the received signal. The receiver structure attempts to remove the introduced ISI by modelling the channel tail with a tapped delay line which is represented by the vector of weights $D_N = [d_1\ d_2\ \ldots\ d_N]^T \in R^N$. Using estimates of the data $\hat{a}_k$, rather than the actual data a_k, the channel ISI is reconstructed and cancelled at the receiver input.

The data estimates $\hat{a}_k$ are generated by a signum function sgn(·) which produces −1 for negative arguments

and +1 otherwise. The fundamental decision equation representing the output of the DFE is then given by,

$$\hat{a}_k = \mathrm{sgn}(h_0 a_k + \sum_{i=1}^{N} h_i a_{k-i} - \sum_{i=1}^{N} d_i \hat{a}_{k-i} + n_k). \quad (2.1)$$

This equation represents an idealization since it assumes that the length of the DFE tapped delay line equals the length of the channel tail (both N).

In section 5 we treat the case where channel noise n_k in (2.1) is significant. Until then we will take $n_k = 0$. We assume the feedback tap weights have the correct values, i.e., $D_N = H_N$, and that the channel is effectively noiseless. Therefore (2.1) becomes,

$$\hat{a}_k = \mathrm{sgn}(h_0 a_k + \sum_{i=1}^{N} h_i (a_{k-i} - \hat{a}_{k-i})). \quad (2.2)$$

Error recovery becomes an object of concern when previous decisions are incorrect since this increases the likelihood of further errors because the ISI is incorrectly cancelled and the eye closes. This effect, termed error propagation, [1] may result in unacceptably high error rates.

Further, many DFE applications are adaptive and involve the adjustment of D_N on−line in response to errors. It is important that the error recovery properties and timescales of the correctly tuned DFE be understood before a sensible analysis of an adaptive DFE can proceed.

2.2. Finite State Markov Processes

Our analysis of the recovery statistics of DFEs uses the theory of finite state Markov processes (FSMPs) as have [1−8]. In modelling the DFE, we can assign a Markov state to the 2N−vector of past data (channel states) and past decisions (DFE tapped delay line states) as follows,

$$X_k \triangleq [a_{k-1} \cdots a_{k-N}\ \hat{a}_{k-1} \cdots \hat{a}_{k-N}]^T \in R^{2N}. \quad (2.3)$$

Each component can take on two values; therefore, we have 4^N different Markov state vectors (2.3) which we refer to as *atomic states*. The complete set of atomic states will be denoted by Ω. The Markovian property arises since the input binary data stream, which takes the values −1 and +1 with equal probability, is *white*, i.e., $\{a_k\}$ is a sequence of i.i.d., random variables.

We will show that by aggregating atomic states one obtains the FSMPs appearing in [1−6]. The advantage of the 4^N atomic state Markov process is twofold: (i) in the analysis of pathological input sequences (section 3.2), and (ii) without modification it may be used to analyze the case when $D_N \neq H_N$ (e.g., with adaptive DFEs). The advantage of aggregated FSMPs is reduced complexity.

We introduce an ordering of the atomic states which will be useful in section 3. Any logical scheme can be used for this purpose and we choose the following rule,

$$i \triangleq \tfrac{1}{2}\,([2^{2N-1}\ 2^{2N-2} \ldots 4\ 2\ 1].X_k + 2^{2N}-1), \quad (2.4)$$

e.g., with N = 2, $X_k = [-1\ +1\ +1\ +1]^T$ gets coded as i = 7. We use the symbol <i> to denote the ith atomic state in this ordering.

It is both convenient and computationally advantageous to aggregate the atomic states X_k into Markov states forming a new process. Such an aggregation forms a FSMP (capable of exactly modelling the transient properties of the atomic FSMP) if and only if the following property holds (in all pairs of aggregated states U and V):− If the probability is p for one particular atomic state in an aggregated state U to transit to the set of atomic states in aggregated state V, then all atomic states in U transit to V with the same probability p. In the DFE problem that we consider, such an aggregation is possible since the DFE tapped delay line weights correspond precisely to the channel tail. Using this observation, one can derive a FSMP where the states are

$$E_k = [e_{k-1}\ e_{k-2} \cdots e_{k-N}]^T \in R^N, \quad (2.5)$$

where $e_{k-i} \triangleq a_{k-i} - \hat{a}_{k-i}$, i = 1,2,..,N. These vectors will be referred to as E−states (short for error states). Each component of this vector can take a value in {−2,0,+2}, hence there are a total of 3^N E−states. This "error model" was used in [1−5] to model the stochastic dynamics of the DFE. For our purposes we introduce a yet more efficient structure which may be derived from the error model. We now show that we can roughly halve the number of states in the error model by exploiting the symmetry in the input symbol probability density.

The new aggregation pairs off the two E−states E_k and $-E_k$ defining a P−state (short for paired state). If $E_k = 0$, then this forms a P−state by itself. This new FSMP has $(3^N+1)/2$ P−states, which means for N = 2 we can model the DFE dynamics exactly with 5 states rather than 9 or 16. Proof that the resulting process is Markovian follows from 2 properties:

Property 1: By *assumption* the input takes values +1 and −1 with equal probability.

Property 2: State E_k transits to state E_{k+1} if and only if state $-E_k$ transits to state $-E_{k+1}$ when the data a_k is of opposite sign.

The proof of property 2 is straightforward (see [7]).

A salient feature of both the error system and the P−state system is that the aggregation has created a single *absorbing* state corresponding to the *closed set* of 2^N atomic states. These atomic states satisfy $\hat{a}_{k-i} = a_{k-i}$, i = 1,2,..,N, and they form a closed set since the only transitions from one such atomic state are to another atomic state in the set. We refer to the absorbing state as A (for absorbing) and it has the simple interpretation that if the system is in state A then we have the last N decisions correct, and hence all future decisions will be correct. Naturally, once in A the system stays in A (in the absence of noise disturbances) and we say the DFE has *recovered*. We will denote the set of atomic states excluding A, i.e., the complement of A, by $\Omega \setminus A$.

We model an initial error condition, e.g., a single noise induced error, in the DFE as an *initial* error distribution π_0 across the atomic states. Without loss of generality, we take the time of this disturbance as k = 0. For time k > 0 we assume the only subsequent decision errors are internally generated by the DFE through error propagation.

In a natural way, a distribution across the atomic states induces a distribution across the aggregated states $\bar{\pi}_0$, through a matrix B,

$$\bar{\pi}_0 = B\,\pi_0, \quad (2.6)$$

where $b_{ij} = 1$ if <j> belongs to aggregated state i, and 0 otherwise. When working with an aggregated state model we will denote that part of the induced distribution excluding set A by $\bar{\pi}^*_0$, i.e., the partial distribution across $\Omega \setminus A$. Therefore, we partition $\bar{\pi}_0$ as follows,

$$\bar{\pi}_0 = \begin{bmatrix} \bar{\pi}^*_0 \\ \sigma \end{bmatrix}. \quad (2.7)$$

Letting $\|.\|_1$ denote the l_1−norm, we have $\sigma = 1 - \|\bar{\pi}^*_0\|_1 \in R$ to ensure $\bar{\pi}_0$ is a probability vector, noting also that we have implicitly ordered the aggregated states with A last.

The main restriction with aggregating atomic states is that we lose information by having our observations based on coarser objects, e.g., P−states rather than atomic states. However, in studying DFE recovery we are mainly concerned with observations concerning A and its complement $\Omega \setminus A$.

3 TWO TAP CASE N = 2

3.1. 29 Classes of Channel

The decision equation, governing the performance of the N = 2 DFE is given from (2.2) by

$$\hat{a}_k = \mathrm{sgn}(h_0 a_k + h_1 e_{k-1} + h_2 e_{k-2});\ h_0 > 0. \quad (3.1)$$

The space of channel parameters (or equivalently the space of tap parameters) will be called H−space, corresponding to R^2, containing the particular channel tail denoted by $H_2 = [h_1\ h_2]^T$. (The cursor h_0 will be carried along implicitly.)

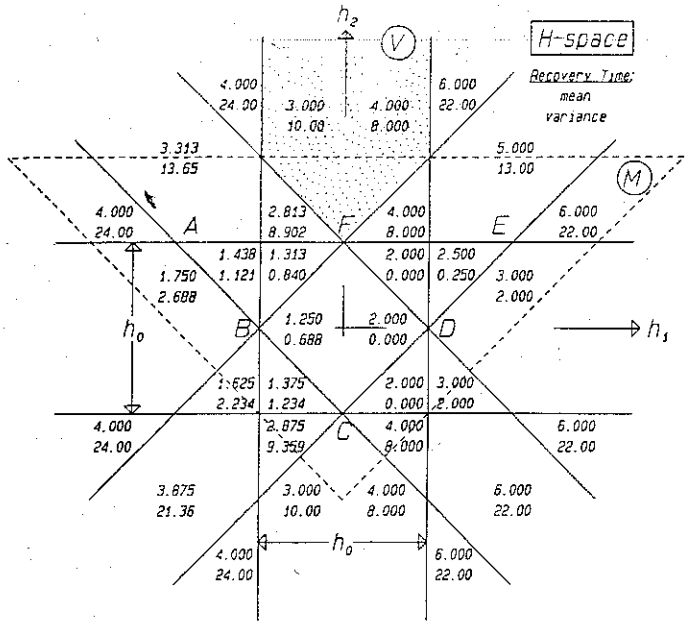

FIGURE 2: The Channel Parameter Space for N=2. The 29 Polytopes are bounded by the solid lines, the Minimum Phase Region by a dashed triangle. Numbers within polytopes give Mean (top) and Variance (bottom) of the Recovery Time. Uniform Distribution values on the left, Single Noise Induced Error Distribution on the right.

The set of lines in H−space given by

$$\{H_2:\ h_1 t_1 + h_2 t_2 = h_0\}, \qquad (3.2)$$

where $(t_1,t_2) \in \{\{-2,0,+2\} \times \{-2,0,+2\}\} \setminus (0,0)$, define the thresholds of the sgn(·) function in (3.1), and as such represent the boundaries between regions in H−space where the performance of the DFE is fixed. So the DFE performance does not depend continuously on H_2 but varies in a *quantized* fashion. These regions are the intersections of half−planes with boundaries given in (3.2) and therefore represent polytopes in $\mathbb{R}^2$.

The 8 boundaries of (3.2) partition H−space into the 29 polytopes of Fig.2. Hence there are only 29 classes of channel to be considered, up to an arbitrary scaling factor (note how the cursor h_0 is incorporated in Fig.2).
Remarkably, *two DFEs working on two different channels within the same polytope, with the same data sequence* $\{a_k\}$ *and initial conditions, have indistinguishable output sequences* $\{\hat{a}_k\}$. (However, with the inclusion of noise into the analysis, the position of the channel vector relative to the polytope boundary affects the DFE behaviour.)

3.2. Pathological Sequences

As noted in [4], there may exist "pathological" input data sequences $\{a_k\}$, depending on the polytope, for which by *definition* the DFE never recovers, i.e., the absorbing group of atomic states A is never reached. In this situation some decision errors must be made after any fixed but arbitrary time. However, from a stochastic viewpoint, it was demonstrated in [4] that these sequences have zero probability in the sense that the probability of the input sequence remaining pathological decreases to zero with time.

Our aim is to classify the pathological sequences in terms of the 29 polytopes for N = 2. For brevity, our demonstration will be for one polytope, labelled V and shown shaded in Fig.2, which is given by

$$V = \{H_2:\ \{h_0 > 2h_1\} \cap \{h_0 > -2h_1\} \cap \{h_0 < 2h_1+2h_2\} \cap \{h_0 < 2h_2-2h_1\}\}. \qquad (3.3)$$

The remaining 28 polytopes (and those for N > 2) may be treated in a similar fashion.

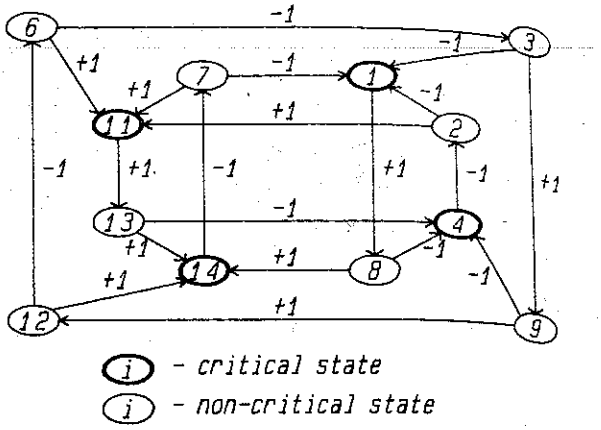

FIGURE 3: Pathological Sequence State Transition Diagram for Polytope V.

We adopt the atomic FSMP consisting of 16 states $\{<0>,\ldots,<15>\}$ for our classification for polytope V. In Fig.3 we have shown polytope V's (restricted) state transition diagram with the set $A \triangleq \{<0>,<5>,<10>,<15>\}$, corresponding to recovery (i.e., error free transmission), removed. Such a diagram encapsulates the complete ensemble of pathological input data sequences (by definition). We note that there is an infinite number of periodic and aperiodic pathological input data sequences associated with polytope V, e.g., we have the periodic sequence $\{+1,+1,-1,-1,+1,+1,..\}$ corresponding to the state transitions $<1> \to <8> \to <14> \to <7> \to <1> \to <8>$,.. etc.

In Fig.3 we also delineate the critical and non−critical states of the polytope V. When the system is in a non−critical state the input data is essentially "don't care" and the system cannot recover in one step. This defines a purely stochastic component of a pathological sequence which is the source of potential aperiodicity. In a critical state, however, we have only a probability of $\frac{1}{2}$ of remaining pathological. For the remaining 28 polytopes we may have either: (i) no pathological sequences; (ii) only periodic pathological sequences; or (iii) as in polytope V, both periodic and aperiodic pathological sequences. In the next section we will determine to which of the 3 groups indicated above that each of the 29 polytopes belong.

3.3. Finite Recovery Time Polytopes

In this subsection we give necessary and sufficient conditions for a polytope to have no pathological sequences or equivalently a bounded recovery time. In [4] it was shown that a necessary condition on the channel parameters for there to be no pathological sequences (a form of "stability", see [4]) is that they lie in the triangle ACE (Fig.2). In [7], the following were demonstrated:

Proposition 1: A necessary and sufficient condition on the channel parameters H_2 for the DFE to have a *bounded* recovery time is $H_2 \in$ triangle ACE (Fig.2).

Proposition 2: (i) Triangle ACE defines the union of 5 polytopes which have an empty set of pathological sequences.
(ii) The 6 polytopes $\in \{H_2:\ \{|h_2| < \frac{1}{2}h_0\} \setminus \{\text{triangle ACE}\}\}$ have only pathological sequences which are periodic.
(iii) The 18 polytopes $\in \{H_2:\ \{|h_2| > \frac{1}{2}h_0\}\}$, e.g., polytope V, have both periodic and aperiodic pathological sequences.

In Fig.2 we have superimposed a region labelled M defining the class of (second order) minimum phase channels (zeros inside the unit circle). It is interesting to note that triangle ACE is contained wholly within M. We will return to the relationship between our channel classification and minimum phase channels in section 4.3.

3.4. Mean Error Recovery Time

Pathological sequences represent the worst case situation since arbitrarily long recovery times are possible. However,

as pointed out in [4], pathological sequences have a special structure which makes their occurrence amongst the ensemble $\{a_k\}$ events of zero probability. It therefore makes sense to focus on statistical recovery properties of DFEs.

The specific aims of this section are to demonstrate how the 29 polytopes (i.e., the 29 classes of channel) affect the mean and variance of the recovery time and to give general formulae for the recovery statistics valid for arbitrary N.

Each of the 29 polytopes defines a different probability transition matrix P which describes the stochastic dynamics of the DFE for that class of channel. The FSMP we choose for our following N = 2 example, is the P−state system consisting of only 5 aggregated states.

The probability transition matrix P (aggregated states) and the associated induced initial error distribution $\bar{\pi}_0$ can always be represented in the following generic form,

$$P = \begin{bmatrix} Q & 0 \\ \underline{r}^T & 1 \end{bmatrix}; \quad \bar{\pi}_0 = \begin{bmatrix} \bar{\pi}_0^* \\ \sigma \end{bmatrix}, \tag{3.4}$$

where we order the aggregated states with A last, and Q is a submatrix, describing transitions between aggregated states belonging to $\Omega \setminus A$, which has eigenvalues satisfying $|\lambda_i(Q)| < 1$. (That the unity eigenvalue of P is simple for our problem is a non−trivial property, see [7]).

Determining the recovery time is precisely a study of transitions made to the absorbing state A. From the generic form (3.4), define $T \triangleq (I - Q)^{-1}$ (non−singular). Then the mean and variance of the absorption (recovery) time ρ can readily be evaluated to yield a standard result in FSMPs. Let $E[\cdot]$ denote the expectation operator, then from [10, pp.89−99], the mean $\mu(\pi_0) \triangleq E[\rho]$ and the variance $\sigma^2(\pi_0) \triangleq E[\rho^2] - \mu(\pi_0)^2$ (in our notation), can be written

$$\mu(\pi_0) = \underset{\sim}{1}^T T.\bar{\pi}_0^*, \tag{3.5a}$$

$$\sigma^2(\pi_0) = \underset{\sim}{1}^T(2T - I)T.\bar{\pi}_0^* - \mu(\pi_0)^2, \tag{3.5b}$$

where $\bar{\pi}_0^*$ is determined from π_0 by using (2.6) and (2.7), and where 1 denotes the vector of 1's.

We now proceed to use (3.5a) and (3.5b) in calculating the first two moments of the recovery time statistics for each of the 29 polytopes, where we assume an initial error distribution corresponding to (i) a single noise induced error π_0^n, and (ii) a uniform error distribution across the 16 atomic states, $\pi_0^u \triangleq 1/16.1$. These are the two cases of greatest practical interest. The first models corruptive channel noise as a rare event (a high signal to noise ratio approximation). The second models a startup of the DFE where the N tap states $\hat{a}_{k-i}$, i = 1,2,..,N take arbitrary initial values (independent of $\{a_k\}$), e.g., during a reset.

We give an outline of the calculation for polytope V defined in (3.3) and shown in Fig.2 (the remaining 28 polytopes can be handled similarly). The atomic states form the following 5 P−states: $P_1 \triangleq \{<3>,<12>\}$, $P_2 \triangleq \{<2>, <7>,<8>,<13>\}$, $P_3 \triangleq \{<6>,<9>\}$, $P_4 \triangleq \{<1>,<4>, <11>,<14>\}$ and $P_5 \triangleq A \triangleq \{<0>,<5>,<10>,<15>\}$. The probability transition matrix (given the above ordering of P−states) for polytope V along with the noise induced error distribution $\bar{\pi}_0^n$ and the uniform error distribution $\bar{\pi}_0^u$ are

$$P = \begin{bmatrix} 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}, \quad \bar{\pi}_0^n = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \bar{\pi}_0^u = \begin{bmatrix} \frac{1}{8} \\ \frac{1}{4} \\ \frac{1}{8} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}$$

Substituting for Q and $\bar{\pi}_0^*$ using (3.5a,b) we get $\mu_k = 4$ and $\sigma_k^2 = 8$ for the first distribution, and $\mu_k = 3$ and $\sigma_k^2 = 10$ for the second. The remaining 28 polytopes can be handled similarly and Fig.2 summarises the results for both distributions. (Due to a symmetry in the recovery time statistics for both distributions with respect to a sign change in h_1, we have saved space by giving the statistics for π_0^n on the right side of the Fig.2 and those for π_0^u on the left.) We make the some remarks based on Fig.2:

Remarks: (i) The inner l_1−ball BCDF (radius $\frac{1}{2}h_0$) has the best recovery statistics of the 29 polytopes for the two distributions (a channel without ISI belongs to this polytope.)

(ii) For the uniform distribution (left hand entries), the maximum mean recovery time of $\mu_k = 4$ is associated with 8 symmetrically placed polytopes. We generalize this observation in section 4.2. (Note, some of the channels in these polytopes are minimum phase.)

In section 4 we will have more to say on the polytopes where the slowest and quickest recovery times have been observed, for the general N ≥ 2 case.

4 GENERAL N ≥ 2 NUMBER OF TAPS

4.1. Finite Class of Channels

Here we generalise the results of section 3.1. The decision equation (2.2) can be written

$$\hat{a}_k = \mathrm{sgn}(h_0 a_k + H_N^T E_k); \quad h_0 > 0. \tag{4.1}$$

The 3^N-1 hyperplanes which partition H−space into a finite number of polytopes are given by the thresholds of the sgn(·) function in (4.1), as follows,

$$\{H_N: H_N^T T_N = h_0\}, \tag{4.2}$$

where $T_N = [t_1\ t_2\ \ldots\ t_N]^T \in R^N$ such that $(t_1,t_2,..,t_N) \in \{\{-2,0,+2\} \times \{-2,0,+2\} \times \ldots \times \{-2,0,+2\}\} \setminus (0,0,..,0)$. The word *polytope* is reserved in this paper for the smallest regions in R^N generated by the boundaries given in (4.2).

It is a difficult problem to determine the exact number of polytopes which result in this partition of R^N. The number of polytopes is bounded below by $2^N.N!$ (see Proposition 3). The size of Z_N for N ≥ 3 prohibits a complete classification of channels. Hence it is important to identify those polytopes which represent the extremes in DFE behaviour and whether these extremes are physically relevant.

In the next subsection we examine precisely those restricted classes of channel which represent the best and worst in terms of the DFE mean recovery time performance.

4.2. Mean Recovery Time Bounds

Cantoni and Butler [4] address the problem of bounding the mean recovery time and, by using the theory of success runs, they demonstrated that an upper bound was (in their notation) $\mu(\frac{1}{2},N) = 2(2^N-1)$. We show that this is the *tight* bound when we have the worst case class of channel (to be defined) and a single noise induced error. The aims of this section are to give both the tight upper and lower bounds for the mean recovery time and to determine some of the corresponding polytopes, i.e., classes of channel which realise these bounds. As in the previous section we will consider the two (same) initial error distributions of greatest practical interest.

We begin with some definitions before deriving our main result (Proposition 3). Let P_B denote the probability of a correct decision *common* to all atomic states belonging to some set B, e.g., if we set B = A (the absorbing group of states) we have $P_A = 1$ (valid for *all* polytopes). We are interested in two cases:

(i) The polytopes where $P_{\Omega \setminus A} = 1$, i.e., before recovery the DFE always makes the correct decision.

(ii) The polytopes where $P_{\Omega \setminus A} = \frac{1}{2}$, i.e., before recovery the decision will be correct or incorrect with equal probability.

Proposition 3: (a) There exists only one polytope where $P_{\Omega \setminus A} = 1$. (b) There exists at least $2^N.N!$ polytopes where $P_{\Omega \setminus A} = \frac{1}{2}$.

Proof: (a) Fact: Only polytope $\{H_N: h_0 > 2\|H_N\|_1\}$ satisfies $P_{\Omega \setminus A} = 1$. (b) If $a_k = \mathrm{sgn}\ H_N^T E_k$ then $\hat{a}_k = a_k$, see (4.1). Therefore, $P_{\Omega \setminus A} = \frac{1}{2}$ if and only if we make an incorrect decision with $a_k = -\mathrm{sgn}\ H_N^T E_k$ for all $E_k \neq 0$. Construct the following polytope,

$$\{H_N: \{h_1 > \tfrac{1}{2}h_0\} \cap \{h_2 > \tfrac{1}{2}h_0 + h_1\} \cap \ldots \cap \{h_N > \tfrac{1}{2}h_0 + \sum_{i=1}^{N-1} h_i\}\} \tag{4.3}$$

Then $|H_N^T E_k| > h_0$ for any channel in this polytope, for

all $E_k \neq 0$. Therefore when $a_k = -\text{sgn } H_N{}^T E_k$ we necessarily have $\hat{a}_k \neq a_k$. This gives the desired result. Further, there are precisely $2^N.N!$ polytopes of this form obtained by 2^N sign changes and N! permutations of the channel vector H_N. (If $N > 2$ there are additional polytopes such that $P_{\Omega \setminus A} = \frac{1}{2}$.) □

We adopt the terminology "best case class of channel" and "worst case class of channel", respectively, for the two cases considered in Proposition 3. Clearly the terminology is justified in the former case since decision errors (in the absence of noise) are never made. In the latter case, we claim the recovery time is maximized when $P_{\Omega \setminus A} = \frac{1}{2}$. This fact can be deduced from an important and intuitively reasonable result in [5] whose proof needs a lengthy formal argument. In our notation this result takes the form,

$$\Pr(X_k \epsilon A \mid P_{\Omega \setminus A} = \tfrac{1}{2},\ \pi_0) \leq \Pr(X_k \epsilon A \mid \pi_0, n_k)$$

In words, the slowest (conceivable) recovering DFE system is precisely one where the probability of a correct decision for all atomic states before recovery is $\frac{1}{2}$. This is true for all arbitrary initial (error) distributions π_0 and whether or not noise n_k is present.

We now evaluate the mean recovery times for the two classes of channel considered in Proposition 3 and thereby obtain lower and upper bounds on the mean recovery times for all channels. Clearly for $P_{\Omega \setminus A} = 1$ channels, recovery is deterministic and therefore the recovery time is bounded above by N.

The calculation for the $P_{\Omega \setminus A} = \frac{1}{2}$ channels (the upper bounds) will be done for the two important practical cases, π_0^u and π_0^n. We use the theory in [4] to determine the upper bound when we start from the single noise induced error distribution π_0^n. From such a distribution we need to make N consecutive correct decisions each of probability $\frac{1}{2}$ to recover. However this mean recovery time was computed in [4] by using the theory of success runs and is given by

$$\mu(\pi_0^n) = 2(2^N - 1). \tag{4.4}$$

(This bound is also derived in [3].) However, our calculation is for an *explicit* channel, e.g., any channel satisfying (4.4), and not for a hypothetical one [4]. The bound is thus tight.

The second initial error distribution of interest is the uniform atomic distribution π_0^u. In this case the mean recovery time can be shown to be $\mu(\pi_0^u) = 2(2^N - 1) - N$, by an elementary but tedious calculation. A more extensive treatment of calculations involving mean recovery times, their variances and error propagation effects may be found in [2,3]. Our contribution to this style of analysis is to demonstrate that channels exist which result in behaviour best described as pathological. However, the question as to the physical significance of these channels needs to be addressed. This is the subject of the next subsection.

4.3 Minimum Phase Channels

We make some remarks regarding whether $P_{\Omega \setminus A} = \frac{1}{2}$ polytopes contain minimum phase channels. Sensible models for physical channels should be causal and (nearly) minimum phase [9]. Note in (4.3) we have a region whose channel impulse responses increase at least exponentially. It is extremely dubious whether such a channel would exist in practice or that a DFE would be contemplated for its equalization. Channels of the form (4.3) appear to be (nearly) *maximum* phase. However, (4.3) represents only one of a class of $2^N.N!$ polytopes. Another, more interesting, polytope in this class is represented by

$$\{H_N : \{h_N > \tfrac{1}{2}h_0\} \cap \{h_{N-1} > \tfrac{1}{2}h_0 + h_N\} \cap \ldots \cap \{h_1 > \tfrac{1}{2}h_0 + \sum_{i=1}^{N-1} h_i\}\}. \tag{4.5}$$

When N = 2 this polytope intersects the triangular minimum phase region (see section 3.3 and Fig.2). We conclude in this case *there exist minimum phase channels with the worst possible error recovery performance.* For N > 2 we have been able to demonstrate that polytope (4.5) contains channels with only one zero outside the unit circle. However, we have been unable to determine whether there are any minimum phase channels satisfying $P_{\Omega \setminus A} = \frac{1}{2}$.

In the next section we show that various upper bounds on the error probability which have appeared in the literature are tight in a certain sense.

5 TIGHT ERROR PROBABILITY BOUNDS FOR NOISY CHANNELS

5.1. Preliminary

The DFE equation (2.1) may be written compactly,

$$\hat{a}_k = \text{sgn}(h_0 a_k + S_k + n_k), \tag{5.1a}$$

$$= \text{sgn}(h_0 a_k + R_k), \tag{5.1b}$$

where we have introduced the shorthand $S_k \triangleq H_N{}^T E_k$ representing the residual ISI, and $R_k \triangleq S_k + n_k$ representing the residual ISI plus noise.

The channel noise n_k is assumed: (i) to be zero mean, (ii) to have finite variance $\sigma_n{}^2 < \infty$, (iii) to be independent of the data a_k and the residual ISI S_k, and (iv) to be white.

We make two simple but fundamental observations regarding (5.1b):

(B1) If $a_k = \text{sgn } R_k$ then $\hat{a}_k = a_k$.

(B2) If $a_k = -\text{sgn } R_k$ then $\hat{a}_k \neq a_k$ if and only if $|R_k| > h_0$.

Most of our subsequent results centre on these two results.

5.2. Global Bound

In [1] it was established that the probability of error is always bounded above by $\frac{1}{2}$. We can see this by using Bayes' Rule. The probability of a decision error is given by

$$\Pr(\hat{a}_k \neq a_k) = \Pr(\hat{a}_k \neq a_k \mid a_k = \text{sgn} R_k).\Pr(a_k = \text{sgn } R_k) + \Pr(\hat{a}_k \neq a_k \mid a_k = -\text{sgn} R_k).\Pr(a_k = -\text{sgn} R_k) \tag{5.2}$$

Using observations (B1) and (B2), and the assumptions: (i) the data takes binary values with equal probability, and (ii) R_k is independent of a_k, then (5.2) reduces to

$$\Pr(\hat{a}_k \neq a_k) = \tfrac{1}{2}.\Pr(\hat{a}_k \neq a_k \mid a_k = -\text{sgn } R_k), \tag{5.3a}$$

$$= \tfrac{1}{2}.\Pr(|R_k| > h_0), \tag{5.3b}$$

and clearly this expression is bounded above by the *global bound* $\frac{1}{2}$. In essence, this translates to a statement that having a DFE as a channel equalizer is generally better than not having one at all *but not always.*

5.3. General Bound

The second bound derived in [1] takes the form

$$P_E \leq \frac{\epsilon 2^N}{2\epsilon(2^N - 1) + 1}, \tag{5.4}$$

where P_E is the probability of an error under stationarity, N is the number of taps, and ϵ is the probability of error in the absence of past decision errors, *viz,*

$$\epsilon \triangleq \Pr(\hat{a}_k \neq a_k \mid E_k = 0) = \tfrac{1}{2} \Pr(|n_k| > h_0). \tag{5.5}$$

Our approach to demonstrating the tightness of (5.4) is simply to construct a channel and noise density which realizes the value of the bound. For simplicity and clarity we assume the channel noise magnitude can be bounded above by B_U as follows,

$$|n_k| \leq B_U < \infty. \tag{5.6}$$

(This requirement can be relaxed and Chebyshev's Inequality invoked to demonstrate the same bound in (5.4) works for unbounded noise with finite variance $\sigma_n{}^2 < \infty$. The analysis given in [8] is typical of modification in style required to treat the more general case.)

Consider the following construction. An arbitrary selection of H_N and T_N (the latter defined in (4.2)) will generally lead to

$$H_N^T T_N \neq 0 \quad \forall\, T_N \neq 0. \tag{5.7}$$

(If not we can perturb one h_i such that (5.7) holds.) Then it is possible to scale H_N to H_N^* (if necessary), for a given h_0 and B_U, such that

$$\min_{(T_N \neq 0)} \left| H_N^{*T} T_N \right| > B_U + h_0. \tag{5.8}$$

Then the channel $\{h_0, H_N^*\}$ has the property that $|R_k| \triangleq |H_N^{*T} E_N + n_k| > h_0$ whenever $E_k \neq 0$, i.e., $\Pr(\hat{a}_k \neq a_k | E_k \neq 0) = \frac{1}{2}$ for all k. Note this is *identical* to the behaviour of a (noiseless) $P_{\Omega \setminus A} = \frac{1}{2}$ polytope channel when $E_k \neq 0$ (see section 4.2). Otherwise in the case $E_k = 0$ the conditional error probability is simply given by ε (5.5). This completes our construction and we now give an example. Let $N = 2$ and $|n_k| < 1.2$, i.e., $B_U = 1.2$. Then the channel $\{h_0 = 1,\ H_2 = [1.15\ 2.26]^T\}$ satisfies (5.8), noting that subject to $E_k \neq 0$, we have $|R_k|_{\min} = 2 \times 2.26 - 2 \times 1.15 - 1.2 = 1.02$.

We now show that the constructed channels have error probabilities given by the right hand side of (5.4). When the system is in a state such that $E_k = 0$ we can either: (i) continue normal error-free operation, or (ii) make a noise induced error (see (5.5)). Then the mean waiting time before a decision error is made is

$$t_1 \triangleq \varepsilon^{-1}, \tag{5.9}$$

from elementary considerations. With a decision error the system transits to a state where the most recent decision is in error. Then the mean time spent during recovery is *by construction* the same as for the noiseless case (4.4), i.e.,

$$t_2 \triangleq 2(2^N - 1). \tag{5.10}$$

Given the Markovian nature of a DFE it is not hard to show that under stationarity, the probability the system will be found in a state where $E_k = 0$ is given by $t_1/(t_1+t_2)$, and the probability the system is recovering (the complementary event) is $t_2/(t_1+t_2)$ (see [8]). Therefore, the *stationary* probability of error P_E for our *constructed* channels is simply given by Bayes' rule,

$$\begin{aligned} P_E &= \Pr(\hat{a}_k \neq a_k | E_k = 0).\Pr(E_k = 0) + \\ &\qquad \Pr(\hat{a}_k \neq a_k | E_k \neq 0).\Pr(E_k \neq 0) \\ &= \varepsilon \frac{t_1}{t_1+t_2} + \tfrac{1}{2} \frac{t_2}{t_1+t_2} = \frac{\varepsilon 2^N}{2\varepsilon(2^N-1)}. \end{aligned} \tag{5.11}$$

Expression (5.11) is precisely the bound (5.4) derived in [1]. Hence we have established that this bound is tight, being achieved by certain noisy channels satisfying (5.6) and (5.8).

Remarks: (i) A straightforward modification to the analysis also yields the same (supremum) bound given the assumption of unbounded channel noise with finite variance.

(ii) Expression (5.11) shows the connection between the worst case recovery time bound (4.4) derived in [4], and the error probability upper bound (5.4) derived in [1].

5.4. Tight, High Signal to Noise Ratio, Upper Bound

As before let $H_N \triangleq [h_1\ h_2 \ldots h_N]^T \in R^N$ and h_0 denote respectively the N-tap tail and cursor of a channel. We partition the tail H_N according to $H_n \triangleq [h_1\ h_2 \ldots h_n]^T \in R^n$ and $H_d \triangleq [h_{n+1}\ h_{n+2} \ldots h_N]^T \in R^{N-n}$. With an l_1-norm overbound on H_d given by $\|H_d\|_1 < \frac{1}{2}h_0$. Duttweiler *et al* [1] were able to demonstrate that asymptotically as $\sigma_n^2 \to 0$

$$P_E \leq \varepsilon . 2^n, \tag{5.12}$$

subject to very mild constraints on the shape of the impulse response tail and a gaussian noise assumption. The demonstration in [1] is also valid if we let $N \to \infty$. With finite N we need no constraint on the shape of the tail. In [8] it is demonstrated that even if the noise is not gaussian and imposing only that the noise variance is finite that the upper bound (5.12) is asymptotically tight. (This bound is asymptotically tight in the sense that certain noisy channels realize the value of the bound as the noise variance decreases to zero.) This solves an open question raised in the conclusions of [1] regarding tightness. The interested reader should consult [8] for the details of the proof.

6 CONCLUSIONS

We have shown that the class of all noiseless channels of a given tap length N can be classified into a *finite* number of subclasses identified with polytopes in R^N. With each polytope we associate a unique FSMP from which one may determine error recovery statistics. In section 3 which examined the N = 2 tap DFE, we showed how to classify the complete set of pathological input sequences, i.e., those input sequences which lead to arbitrarily long recovery times, for a given polytope. We also gave necessary and sufficient conditions for a polytope to have a bounded recovery time. Section 4 gave some results for the $N \geq 2$ tap DFE. In particular, we identified polytopes which represent the best and worst extremes in DFE recovery behaviour (Proposition 3) and thus gave tight bounds on the expected recovery times. (The interested reader should note that an asymptotic formula which gave the tight bound on the probability of a DFE failing to recover in a given time has also been developed, see [7] for details).

Our results demonstrate that DFEs are practical only on restricted classes of channel, else, e.g., with a N = 64 tap DFE running at a sampling rate of 1kHz the mean (error) recovery time can be as high as 10^{10} years, see (4.4). The imposition of just a minimum phase condition on the class of channels does not appear strong enough to guarantee satisfactory DFE operation (see section 4.3).

For DFEs operating on noisy channels we have outlined the proofs that various upper bounds derived in [1] are tight in a particular sense and this resolves open questions raised in the conclusions of [1]. The full proofs are given in [8].

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. Thy. vol. IT-20 July (1974) pp 490-497.

[2] J.J. O'Reilly and A.M. de Oliveira Duarte, 'Error propagation in decision feedback receivers', IEE Proc. F, Commun., Radar and Signal Process., 1985, 132, (7), pp. 561-566.

[3] A.M. de Oliveira Duarte and J.J. O'Reilly, 'Simplified technique for bounding error statistics for DFB receivers', IEE Proc. F, Commun., Radar and Signal Process., 1985, 132, (7), pp. 567-575.

[4] A. Cantoni and P. Butler, 'Stability of Decision Feedback Inverses' IEEE Trans. Commun. vol. COM-24 No.9 Sept (1976) pp 970-977.

[5] A. Cantoni and P. Butler, 'Behaviour of Decision Feedback Inverses with Deterministic and Random Inputs' Univ. Newcastle, Newcastle, Australia, Tech. Rep. EE7505, Mar. 1975.

[6] P. Monsen, 'Adaptive Equalization of a Slow Fading Channel' IEEE Trans. Commun. vol. COM-22 No.8 Aug (1974) pp 1064-1075.

[7] R.A. Kennedy and B.D.O. Anderson, 'Recovery TImes of Decision Feedback Equalizers on Noiseless Channels', IEEE Trans. Commun. (to appear.)

[8] R.A. Kennedy, B.D.O. Anderson and R.R. Bitmead, 'Tight Bounds on the Error Probabilities of Decision Feedback Equalizers' IEEE Trans. Commun. (to appear.)

[9] B.R. Clarke, 'The Time-Domain Response of Minimum Phase Networks' IEEE Trans. Circuits and Systems, vol. CAS-32 No.11 Nov (1985) pp 1187-1189.

[10] D.L. Isaacson and R.W. Madsen 'Markov Chains Theory and Applications', John Wiley & Sons Inc, New York (1976).