Tight Bounds on the Error Probabilities of Decision Feedback Equalizers

RODNEY A. KENNEDY, STUDENT MEMBER, IEEE, BRIAN D. O. ANDERSON, FELLOW, IEEE, AND ROBERT R. BITMEAD, SENIOR MEMBER, IEEE

Abstract—**A decision feedback equalizer (DFE) with correct tap weights operating on a noisy channel is considered. We show how the results concerning a noiseless channel can be extended to yield tight bounds on the stationary error probability performance for the noisy case. The effect of noise on DFE performance is classified according to the noise distribution and the channel parameters.**

I. Introduction

In this paper, we study the range of performance that one can expect from a decision feedback equalizer (DFE) as we vary across the class of noisy finite impulse response (FIR) channels used for binary transmission. We will be interested in finding a subclass of noisy channels which *realize* the worst possible error probability performance of the DFE. This analysis extends the techniques in [1] (see also [2], [3]) which treats DFE's operating on *noiseless* channels. In the noiseless case, it is meaningless to study steady-state error probabilities because with probability one, the DFE converges to a closed set of *error-free* states [1]. This implies that the stationary error probability is zero. This situation changes with the introduction of channel noise (of sufficient amplitude), and we may observe unacceptably high error rates, reflected in a high error probability, in the DFE due to an error propagation mechanism [4].

In [4], Duttweiler *et al.* derive upper bounds on the stationary error probability for classes of noisy channels subject to various constraints. However, these bounds *appear* conservative and it seems natural to question the *tightness* of the bounds. Indeed, this very question is raised in the conclusions of [4]. We repose and solve this problem.

A significant advance on the work of Duttweiler *et al.* [4] has appeared recently in the DFE literature. O'Reilly and de Oliveira Duarte [5], [6] develop a procedure which gives upper and lower bounds on various error statistics for a *given* channel. (Their techniques are applicable to multilevel data sequences and correlated noise.) It is clear that this procedure produces an upper bound on the stationary error probability which cannot exceed that given in [4]. However, like [4], these results give no indication about what range of error probabilities one might expect, apart from what may be indicated by specific examples.

Our contribution to this style of analysis is not to extend the techniques and results found in [5], [6], but rather to contrive noisy channels which realize the upper bounds in [4], thereby settling the open questions regarding tightness. (These bounds are realized by manipulating the channel parameters typically in the presence of small noise, rather than by taking the limit as the noise variance increases [2].) It is not our intention to suggest that such contrived channels will or do arise in practice (although it is not clear that they do not). Rather, the merit of our results rests in showing the need for imposing stronger hypotheses in characterizing the channel parameters for *practical* systems. This would enable tighter bounds on the error probability to be derived, thereby better reflecting the DFE performance to be expected in practice. However, we argue that imposing a minimum phase property on the (high signal-to-noise ratio) channels does not appear strong enough to guarantee improvements on the error rate bounds in [4]. Another consequence of our analysis is the suggestion that by adding dither to a DFE, one may improve its error performance. Both these latter points seem, at least initially, to be counterintuitive.

Finally, the explicit use of the recovery time bound derived by Cantoni and Butler [2], [3] (shown tight in [1]) to give a straightforward proof of an error probability bound in [4], illuminates the nontrivial but close connection between the two important early contributions to the analysis of the error propagation mechanism in DFE's.

The ideas in this paper are based on two-level input sequences. Extending the results to M-level inputs appears straightforward (see [2], [3], [5], [6]).

The remainder of this paper is set out as follows. Section II outlines the notation developed in [1] which also is adopted here. We also present our assumptions regarding the channel noise. In Section III, we treat a special situation where we assume bounded channel noise to demonstrate the tightness of a bound given in [4]. (Extending our calculation to the unbounded but finite variance noise case is straightforward and not presented in detail.) High signal-to-noise ratio channels which satisfy an l_1-norm overbound property of the tail are considered in Section IV. Section V quantifies the effect noise has on DFE performance relative to the noiseless case. Finally, Section VI summarizes the main results and considers a natural extension to the work.

II. Definitions, Assumptions, and Some Notation

Our notation forms an extension of that found in [1]. However, to keep this treatment self-contained, we give a *minimal* review of definitions. Reference [1], treating the noiseless case for which this paper forms the sequel, should be consulted to obtain a more complete (and sympathetic) review of the notation, general principles of DFE operation, and the use of finite state Markov processes (FSMP's).

The communication channel is modeled as an FIR filter with cursor h_0 and tail $H_N \triangleq [h_1\ h_2 \cdots h_N]^T \in \mathbb{R}^N$. The DFE tapped delay line (see Fig. 1) is represented $D_N \triangleq [d_1\ d_2 \cdots d_N]^T \in \mathbb{R}^N$. The data a_k are assumed white and take values in $\{-1, +1\}$ with equal probability (k is the time index). We denote the binary decision or data estimate by $\hat{a}_k$. The kth decision error is defined $e_k \triangleq a_k - \hat{a}_k \in \{-2, 0, +2\}$. In our analysis, we will have cause to use the past decision error vector

$$E_k \triangleq [e_{k-1}\ e_{k-2} \cdots e_{k-N}]^T \in \mathbb{R}^N. \tag{2.1}$$

Paper approved by the Editor for Digital Communications of the IEEE Communications Society. Manuscript received September 25, 1986; revised March 13, 1987. This work was supported by the Australian Telecommunications and Electronics Research Board.

The authors are with the Department of Systems Engineering, Research School of Physical Sciences, Australian National University, Canberra, A.C.T., 2601, Australia.

IEEE Log Number 8716585.

0090-6778/87/1000-1022$01.00

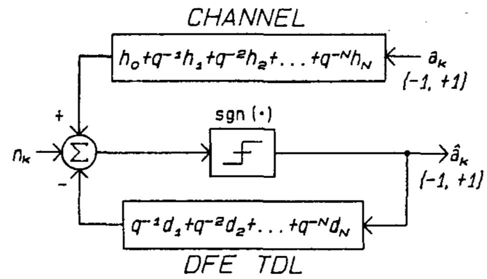


Fig. 1. Channel and DFE model.

Now, under ideal conditions, $D_N = H_N$ and we assume this. With this last condition, E_k forms a state [1]. More generally, when $D_N \neq H_N$ (which we do not treat in this paper), E_k is not a state, but the following vector is:

$$x_k \triangleq [a_{k-1} \cdots a_{k-N}\, \hat{a}_{k-1} \cdots \hat{a}_{k-N}]^T \in \mathbb{R}^{2N}. \quad (2.2)$$

Naturally, x_k contains all the information E_k does, but not vice versa. Each possible value taken by x_k will be referred to as an *atomic state*, and the complete collection of 4^N atomic states will be denoted Ω. Note, there is no need conceptually to differentiate between x_k and E_k when we refer to the state of the system (Fig. 1). We will, however, generally be referring to the atomic states x_k to keep the treatment compatible with [1].

We will be carrying over the finite state Markov process (FSMP) modeling from [1] noting the set of 4^N x_k's form the states. As in [1], we consider lumping the atomic states into aggregated states in such a way to preserve the Markov property. While in principle such a lumping is not necessary, it does reduce the computational burden [1]. Finally, we let n_k denote the channel noise (see Fig. 1). Assumptions regarding the channel noise are outlined below.

The DFE decision equation which is central to our analysis is given by (see Fig. 1)

$$\hat{a}_k = \text{sgn}\left(h_0 a_k + \sum_{i=1}^{N} h_i a_{k-i} - \sum_{i=1}^{N} d_i \hat{a}_{k-i} + n_k\right); \qquad h_0 > 0.$$

assuming correct tap weights $D_N = H_N$, the above equation reduces to

$$\hat{a}_k = \text{sgn}\left(h_0 a_k + \sum_{i=1}^{N} h_i e_{k-i} + n_k\right); \qquad h_0 > 0 \quad (2.3)$$

which may be compactly written

$$\hat{a}_k = \text{sgn}\,(h_0 a_k + H_N^T E_k + n_k); \qquad h_0 > 0 \quad (2.4a)$$

$$= \text{sgn}\,(h_0 a_k + S_k + n_k) \quad (2.4b)$$

$$= \text{sgn}\,(h_0 a_k + R_k) \quad (2.4c)$$

where we have introduced the shorthand $S_k \triangleq H_N^T E_k$ (notation: v^T denotes the transpose of v) representing the residual intersymbol interference (ISI), and $R_k \triangleq S_k + n_k$ representing the residual ISI plus noise.

The channel noise n_k is assumed: 1) to be zero mean, 2) to have finite variance $\sigma_n^2 < \infty$, 3) to be independent of the data a_k and the residual ISI S_k, and 4) to be white. Assumptions 3) and 4) are particularly important because they ensure one can use FSMP modeling for a noisy DFE as was done for the noiseless case [1] (see also [2]–[6]).

We make two simple but fundamental observations regarding (2.4c).

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 center on these two results.

III. Error Probability Bounds

A. Global Bound

In [4], it was established that the probability of error is always bounded above by 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*. We rederive this result because its proof will be useful in later sections. Using Bayes' theorem, the probability of a decision error is given by

$$\Pr(\hat{a}_k \neq a_k) = \Pr(\hat{a}_k \neq a_k \mid a_k = \text{sgn}\, R_k) \cdot \Pr(a_k = \text{sgn}\, R_k)$$
$$+ \Pr(\hat{a}_k \neq a_k \mid a_k = -\text{sgn}\, R_k) \cdot \Pr(a_k = -\text{sgn}\, R_k). \quad (3.1)$$

Using observations B1) and B2), and the assumptions: 1) the data take binary values with equal probability, and 2) R_k is independent of a_k, then (3.1) reduces to

$$\Pr(\hat{a}_k \neq a_k) = \frac{1}{2} \cdot \Pr(\hat{a}_k \neq a_k \mid a_k = -\text{sgn}\, R_k) \quad (3.2a)$$

$$= \frac{1}{2} \cdot \Pr(|R_k| > h_0) \quad (3.2b)$$

and clearly this expression is bounded above by the *global bound* 1/2. In Section III-C, we construct a set of channels and a noise density which realizes the pathological value 1/2. (This bound is realized *without* the assumption that the signal-to-noise ratio is vanishingly small, in which case the error probability can be made arbitrarily close to 1/2 [2].)

B. General Bound

The second bound derived in [4] takes the form

$$P_E \leq \frac{\epsilon 2^N}{2\epsilon(2^N - 1) + 1} \quad (3.3)$$

where P_E is the probability of an error under stationarity, ϵ is the probability of error in the absence of past decision errors, and N is the number of taps. Our approach to demonstrating the tightness of (3.3) 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:

$$|n_k| \leq B_U < \infty. \quad (3.4)$$

(This requirement can be relaxed and Chebyshev's inequality invoked to demonstrate that the same bound in (3.3) works for unbounded noise with finite variance $\sigma_n^2 < \infty$. The analysis given in Section IV is typical of modification in style required to treat the more general case.)

The system we consider is given in Fig. 2, which we shall now explain. (The state labeling and the transition probabilities marked on the arrows joining states need to be separately described. We will show subsequently that there exists a set of channels and noise densities for which Fig. 2 represents an *exact* stochastic description under suitable interpretation.) We adopt the "recovery distance" aggregated FSMP states given in [6]. (This set of aggregated states was also used to treat the noiseless channel case [1, sect. IV.E].) In review, aggregated state 0 comprises the 2^N set of atomic states, each of which satisfy $E_k = 0$. (In [1], this set was called the set $\mathcal{A}$.) Further, an atomic state x_k belongs to the aggregated state $i > 0$ if the DFE requires *i consecutive* correct decisions to reach aggregated state 0. Despite the fact that aggregated state 0 is no

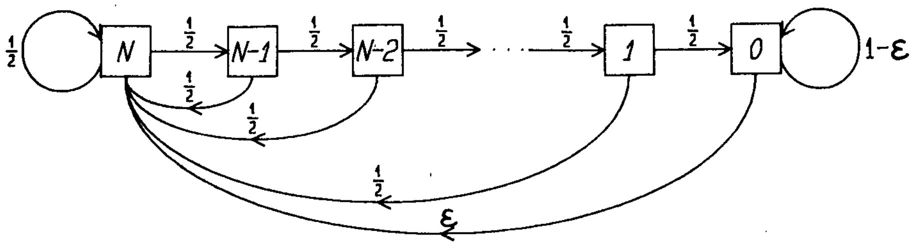


Fig. 2. Finite state Markov process of a class of channels with bounded noise.

longer absorbing, as it was for the noiseless case, we will continue to refer to the DFE as having *recovered* when it is in aggregated state 0.

We now describe the transition probability assignments shown in Fig. 2. When in aggregated state 0 (where $E_k = 0$), only noise can cause a decision error. With such an error, the system transits to state N with probability ϵ defined by

$$\epsilon \triangleq \Pr(\hat{a}_k \neq a_k | E_k = 0)$$
$$= \frac{1}{2} \Pr(|n_k| > h_0) \tag{3.5}$$

where we have used (3.2b) with $R_k = n_k$. Note, varying the channel does not change ϵ. Therefore, this error probability is the same for all atomic states in aggregated state 0.

Next we consider the transitions emanating from the aggregated states $N, N-1, \cdots, 1$ in Fig. 2. It is clear from the definition of recovery distance that if the DFE makes an error, it transits to aggregated state N. Otherwise, with each consecutive correct decision, the index defining the aggregated state, i.e., the recovery distance, decrements. Now the transition probabilities shown in Fig. 2 imply that for some *hypothetical* channel, correct decisions occur with probability 1/2 and incorrect decisions also with probability 1/2 (in the presence of bounded noise). We realize this hypothetical channel by constructing a class of channels with the desired property.

Consider the DFE error probability *during recovery* (where $E_k \neq 0$). Then

$$\Pr(\hat{a}_k \neq a_k | E_k \neq 0) = \frac{1}{2} \cdot \Pr(|R_k| > h_0 | E_k \neq 0) \tag{3.6}$$

derived in analogy to (3.2b), except conditioned on the DFE not having recovered. To make this last expression equal 1/2, we simply need to ensure that R_k, considered as a random variable, has a *conditional* probability density which is zero in the interval $[-h_0, h_0]$. Now since the noise is assumed bounded above by B_U (3.4), it is sufficient that the random variable S_k has zero probability density in the interval $[-h_0 - B_U, h_0 + B_U]$. This in turn ensures that while the system is in the aggregated states where $E_k \neq 0$, the decision in (2.4a) is based solely on R_k, and therefore $\hat{a}_k$ is completely uncorrelated with a_k. We can guarantee S_k has the desired property by a suitable choice of H_N.

We assert that the nonempty region in the channel parameter space given by

$$\Big\{H_N : \{2h_N > h_0 + B_U\} \cap \{2h_{N-1} > 2h_N + h_0 + B_U\}$$
$$\cap \cdots \cap \{2h_1 > 2 \sum_{i=2}^{N} h_i + h_0 + B_U\}\Big\} \tag{3.7}$$

defines a proper subset of the channels which have Fig. 2 as their FSMP representation. We now show this. The formula (3.7) simply ensures the parameters $h_1 > h_2 > \cdots > h_N$ are spaced sufficiently far apart that $|H_N^T E_k| > h_0 + |n_k|$ for all $E_k \neq 0$. This implies, via the triangle inequality, that $|H_N^T E_k + n_k| > h_0$, i.e., $|R_k| > h_0$. Then using (3.6), we conclude that for all channels satisfying (3.7), $\Pr(\hat{a}_k \neq a_k | E_k \neq 0) = 1/2$. Therefore, we have constructed a class of channels (3.7) where: 1) *the statistical behavior of the DFE before recovery is the same whether or not bounded channel noise is present*, and 2) Fig. 2 is an *exact* stochastic description under suitable interpretation. We will now show how this property and our previous results for the noiseless case provide an elementary derivation and reinterpretation of the bound (3.3) derived in [4]. In light of (3.5), we consider only the nondegenerate case where $\Pr(|n_k| > h_0) > 0$ which ensures aggregated state 0 is nonabsorbing, i.e., $\epsilon > 0$.

First, we give an example of our construction. 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 (3.7), noting that subject to $E_k \neq 0$, we have $|R_k|_{\min} = 2 \times 2.26 - 2 \times 1.15 - 1.2 = 1.02$.

Now consider Fig. 2. When the system is in aggregated state 0, we can either: 1) continue normal error-free operation, or 2) make a noise-induced error. Once in aggregated state 0, the mean waiting time before a decision error is made is

$$t_1 \triangleq \epsilon^{-1} \tag{3.8}$$

from elementary considerations. With a decision error, the system transits to aggregated state N. The mean time spent during recovery when we start in aggregated state N is, by construction, the same as for the noiseless case [1, eq. (4.6)], i.e.,

$$t_2 \triangleq 2(2^N - 1). \tag{3.9}$$

Now since the arrangement in Fig. 2 depicts an FSMP, the time it takes to transit from aggregated state N to 0 is independent of the transitions which led the system to state N. Similarly for the case starting in 0 transiting to state N. Therefore, under stationarity, the probability the system will be found in state 0 is given by $t_1/(t_1 + t_2)$, and the probability the system is recovering (the complement event) is $t_2/(t_1 + t_2)$. Therefore, the *stationary* probability of error P_E for our *constructed* channels is simply given by Bayes' theorem:

$$P_E \triangleq \Pr(\hat{a}_k \neq a_k | \text{stationarity})$$
$$= \Pr(\hat{a}_k \neq a_k | E_k = 0) \cdot \Pr(E_k = 0)$$
$$+ \Pr(\hat{a}_k \neq a_k | E_k \neq 0) \cdot \Pr(E_k \neq 0)$$
$$= \epsilon \frac{t_1}{t_1 + t_2} + \frac{1}{2} \frac{t_2}{t_1 + t_2} \tag{3.10a}$$
$$= \frac{\epsilon 2^N}{2\epsilon(2^N - 1) + 1}. \tag{3.10b}$$

Expression (3.10b) is precisely the bound (3.3) derived in [4].

Hence, we have established that this bound is tight, being achieved by certain noisy channels satisfying (3.4) and (3.7).

Remarks:

1) A straightforward modification to the analysis also yields the same (supremum) bound given the assumption of unbounded channel noise with finite variance. In this case, we select B_U (as a parameter rather than a bound) sufficiently large in the expression for the channel (3.7) to make P_E arbitrarily close to (3.10b). However, with larger B_U in (3.7), the resulting channels become more contrived and less likely to appear in practice.

2) In the high signal-to-noise ratio limit, i.e., as $\sigma_n^2 \to 0$, the channels that are analogous to (3.7), which yield a stationary error probability P_E that is arbitrarily close to the bound (3.3) are the "$P_{\Omega \setminus A} = 1/2$ polytopes" considered in [1]. (We give no formal proof of this, but simply note that as $B_U \to 0$ in (3.7), we obtain a region which corresponds to a $P_{\Omega \setminus A} = 1/2$ polytope in [1, eq. (4.4)].) Hence, questions regarding the *physical relevance* of these high signal-to-noise ratio channels coincide with the questions raised in [1] concerning noiseless minimum phase channels. (The interested reader should consult [1, sect. IV-D].)

3) Expression (3.10a) shows the connection between the worst case recovery time bound (3.9) derived in [2] and the error probability upper bound (3.3) derived in [4].

C. Range of Realizable Error Probabilities

In this subsection, we contrive noise densities (distributions) which realize all values of ϵ in the interval [0, 1/2], and hence from (3.10b), all values of P_E in the interval [0, 1/2]. Consider noise n_k which can be bounded above and below as follows:

$$h_0 < B_L \le |n_k| \le B_U. \tag{3.11}$$

Then the presence of the lower bound implies $\epsilon = 1/2$ from (3.5). By Section III-B, the presence of the upper bound ensures that a channel can be constructed (3.7) such that (3.10b) is realized. With $\epsilon = 1/2$ in (3.10b), we get $P_E = 1/2$, a most pathological situation. Constructing further noise densities yielding the remaining values in [0, 1/2] for P_E is straightforward, but not of practical interest.

IV. Asymptotically Tight Error Probability Bound

A. Preliminary

As before, let $H_N \triangleq [h_1\ h_2 \cdots h_N]^T \in \mathbb{R}^N$ and h_0 denote, respectively, the N-tap tail and cursor of a channel. We partition the tail H_N according to $H_n \triangleq [h_1\ h_2 \cdots h_n]^T \in \mathbb{R}^n$ and $H_d \triangleq [h_{n+1}\ h_{n+2} \cdots h_N]^T \in \mathbb{R}^{N-n}$. With an l_1-norm overbound on H_d given by

$$\|H_d\|_1 < \frac{1}{2} \cdot h_0 \tag{4.1}$$

Duttweiler *et al.* [4] were able to demonstrate that asymptotically as $\sigma_n^2 \to 0$

$$P_E \le \epsilon \cdot 2^n \tag{4.2}$$

subject to mild constraints on the shape of the impulse response tail and a Gaussian noise assumption. The demonstration in [4] is also valid if we let $N \to \infty$, which we will not consider here for brevity. With finite N, we need no constraint on the shape of the tail. We also drop the Gaussian noise assumption and require only that the noise variance is finite to demonstrate that the upper bound (4.2) 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.

In the following sections, we construct a class of channels whose stationary error probability is bounded below by an expression which approaches (4.2). Given (4.2) is an upper bound [4], this implies (4.2) forms the asymptotic error probability for this class.

The bound (4.2) is important because it shows that by imposing conditions on the shape of the channel, viz. (4.1), corresponding to physical reality, less pessimistic error probability bounds than (3.10b) are possible.

B. Construction of a Candidate Class of Channels

The N-tap channel decision equation (2.3) is algebraically equivalent to

$$\hat{a}_k = \operatorname{sgn}\left(\tilde{h}_0(k)a_k + \sum_{i=1}^{n} h_i e_{k-i} + n_k\right); \qquad n<N \tag{4.3a}$$

where

$$\tilde{h}_0(k) \triangleq h_0 + a_k \sum_{i=n+1}^{N} h_i e_{k-i} \in [h_0 - D,\ h_0 + D] \tag{4.3b}$$

and

$$D \triangleq 2 \sum_{i=n+1}^{N} |h_i| = 2\|H_d\|_1. \tag{4.3c}$$

With constraint (4.1), we simply have

$$D < h_0 \tag{4.4}$$

and then (4.3a) can be *interpreted* as an equation for an n-tap channel with a random, time-varying cursor bounded to *strictly positive* values, i.e., $\tilde{h}_0(k) > 0$. We note that, as mentioned in [4], condition (4.1) corresponds to the eye being open (i.e., the residual ISI satisfies $|S_k| < h_0$) after the first n taps are discarded. Therefore, in the absence of noise n_k, recovery is guaranteed if we make n rather than N consecutive correct decisions.

It is now straightforward to construct an N-tap channel H_N (for a given h_0) such that Fig. 3 describes the transition probabilities between the aggregated states 0 through N *when channel noise n_k is absent* (these aggregated states are the same as those in Fig. 2). (We will subsequently demonstrate that for all such channels, $P_E \to \epsilon \cdot 2^n$ asymptotically for small noise.) Note from Fig. 3 that the probability of a correct decision is one for aggregated states 0 through $N - n$ because the eye is open after n or more consecutive correct decisions have been made. When the DFE is in an atomic state within any of the aggregated states $N, N - 1, \cdots, N - n + 1$, we claim we can construct a suitable H_n such that the probability of error is precisely 1/2 (in the absence of noise). If so, then Fig. 3 forms a valid FSMP representation of this *noiseless* channel.

We begin our construction by defining a region in the space of the *first n* channel parameters in terms of a positive quantity $\psi > 0$:

$$\Xi(\psi) \triangleq \left\{H_n : \{2h_n > \psi\} \cap \{2h_{n-1} > 2h_n + \psi\} \cap \cdots \cap \{2h_1 > 2\sum_{i=2}^{n} h_i + \psi\}\right\}. \tag{4.5}$$

Then we assert that a sufficient condition on the N-tap channel H_N subject to (4.1) to have the desired behavior in states N through $N - n + 1$ is the following condition on its first n parameters:

$$H_n \in \Xi(h_0 + D). \tag{4.6}$$

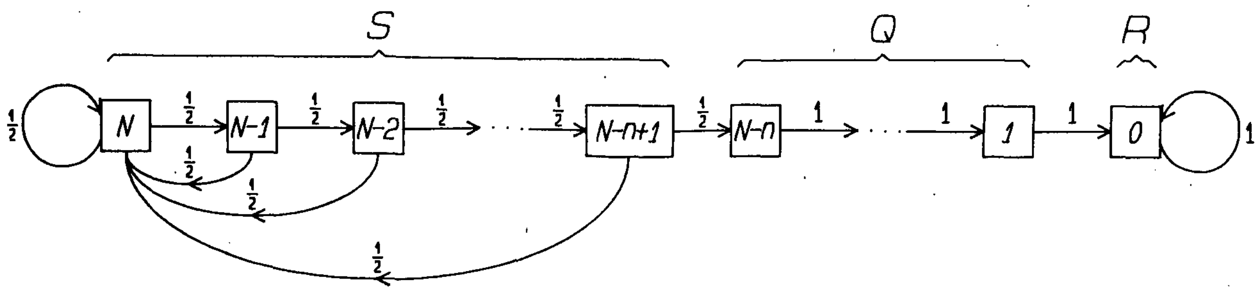


Fig. 3. Finite state Markov process of a class of noiseless channels.

To see this, first consider a (truncated) noiseless channel H_n with cursor $h_0 + D$. Then (4.6) is identical to a sufficient condition that a noiseless channel has exactly an error probability of 1/2 when it has made less than n consecutive correct decisions [1, Proposition IV.1(b)]. We now demonstrate that the channel pair $\{\tilde{h}_0(k), H_n\}$, which has the same fixed tail H_n but the time-varying cursor of (4.3b), has this behavior also. From the definitions (4.5) and (4.6), it is clear that for all $H_n \in \Xi(h_0 + D) \Rightarrow H_n \in \Xi(\tilde{h}_0(k))$, i.e.,

$$\Xi(h_0+D) \subset \Xi(\tilde{h}_0(k)). \tag{4.7}$$

Due to the algebraic equivalence of $\{\tilde{h}_0(k), H_n\}$ to $\{h_0, H_N\}$ expressed through (2.3) and (4.3a), we have determined a subset of the class of channels H_N, subject to (4.1) and (4.6), which have the FSMP representation of Fig. 3. (For a more detailed account of this style of construction of noiseless channels, see [1].)

We now give an example of our construction. Let $N = 7$, $n = 3$, and $h_0 = 1$. We pick any vector in $\mathbb{R}^4$ whose l_1 norm satisfies (4.1), say $H_d = [-0.3\ 0.1\ 0.05\ -0.01]^T$. Here $\|H_d\|_1 = 0.46$, and therefore $D = 0.92$ from (4.3c). Next we select any $H_3 \in \Xi(1.92)$ according to (4.5) and (4.6). A suitable choice is, say, $H_3 = [3.95\ 2.00\ 0.97]^T$. Then we claim that the channel $\{h_0 = 1, H_7 = [3.95\ 2.00\ 0.97\ -0.3\ 0.1\ 0.05\ -0.01]^T\}$ in the presence of noise n_k will have an error probability $P_E \to 8\epsilon$ (4.2) as $\sigma_n^2 \to 0$ where $\epsilon \to 0$ is given by (3.5).

We now *include channel noise* n_k in the analysis of our constructed channel. We note that the lumping of atomic states into aggregated states corresponding to "recovery distance" [6] no longer yields an FSMP (for our constructed channel) when noise is present. To see this, observe that each atomic state within a single aggregation will have different noise thresholds, implying different error probabilities. This destroys the Markov property because knowledge of the recovery distance is insufficient to deduce exact error probabilities. Fortunately, in deriving our error probability bound, we do not need to use a (noisy) FSMP. However, the aggregation of atomic states into recovery distance states will be useful in deriving approximate bounds as was done in [6].

We define a further coarser partition of the set of atomic states Ω into three sets. Set S (slow phase of recovery) is defined as the set of atomic states whose recovery distance is greater than $N - n$. Set Q (quick phase of recovery) is those whose recovery distance is between (and including) n and 1. Set R (recovery) is those corresponding to recovery. These three sets which delineate three phases of DFE recovery behavior on our constructed channels (in the high signal-to-noise ratio limit) are shown in Fig. 3.

C. Stationary Probabilities

To facilitate a simple and systematic derivation of the desired asymptotic formula, we introduce some shorthand. The probability of error *under stationarity* given that the present atomic state, whatever it is, lies in a subset $X \in \Omega$ will be denoted

$$P_{E|X} \triangleq \Pr(\hat{a}_k \neq a_k | X). \tag{4.8}$$

Similarly, the probability that the present state x_k lies in X (again assuming stationarity) is denoted

$$\rho_X \triangleq \Pr(x_k \in X). \tag{4.9}$$

Based on these definitions and noting the sets S, Q, and R partition Ω, we have *under stationarity* and using Bayes' theorem

$$\begin{aligned} P_E &= P_{E|R} \cdot \rho_R + P_{E|Q} \cdot \rho_Q + P_{E|S} \cdot \rho_S \\ &= \epsilon \cdot \rho_R + P_{E|Q} \cdot \rho_Q + P_{E|S} \cdot \rho_S. \end{aligned} \tag{4.10}$$

By bounding the components of the right-hand side of (4.10), we will be able to determine a lower bound on the stationary error probability P_E for our constructed channels.

D. Lower Bound on Error Probability for Constructed Channel Class

To prove the upper bound (4.2) of Duttweiler *et al.* is tight, it is sufficient to find a lower bound for our constructed channel which approaches (4.2) in the high signal-to-noise ratio limit, i.e., as $\sigma_n^2 \to 0$. From (4.10), we have

$$\begin{aligned} P_E &\geq \epsilon \cdot \rho_R + P_{E|S} \cdot \rho_S \\ &\geq \epsilon \cdot \boldsymbol{\rho}_R + \boldsymbol{P}_{E|S} \cdot \boldsymbol{\rho}_S \end{aligned} \tag{4.11}$$

where $\boldsymbol{\rho}_R$, $\boldsymbol{P}_{E|S}$, $\boldsymbol{\rho}_S$ (to be determined) are lower bounds on ρ_R, $P_{E|S}$, ρ_S, respectively.

Consider first the calculation for $\boldsymbol{\rho}_S$ in (4.11). Let ρ_i denote the probability that the DFE will be found in (recovery distance) aggregated state i under stationarity. Then by definition (see Fig. 3),

$$\rho_S = \rho_N + \rho_{N-1} + \cdots + \rho_{N-n+1} \tag{4.12a}$$

and

$$\rho_Q = \rho_{N-n} + \rho_{N-n-1} + \cdots + \rho_1. \tag{4.12b}$$

Now with every decision error, the DFE transits to an atomic state whose recovery distance is N (by definition). This shows $\rho_N = P_E$. Further, by Section III-A, the least (imaginable) probable transition from aggregated state N to $N - 1$ or $N - 1$ to $N - 2$, etc., occurs with probability 1/2. This shows $\rho_i \geq P_E(1/2)^{N-i}$. Using (4.12a), we may then define

$$\begin{aligned} \boldsymbol{\rho}_S &\triangleq P_E + \frac{1}{2} P_E + \cdots + \left(\frac{1}{2}\right)^{n-1} P_E \\ &= 2P_E(1 - 2^{-n}). \end{aligned} \tag{4.13}$$

We now compute a suitable $\boldsymbol{\rho}_R$ in (4.11). First, we give a very conservative upper bound on $\rho_S + \rho_Q$ using (4.12a) and

(4.12b). Clearly, $\rho_i \leq P_E$ for $i = N, N-1, \cdots, 1$ because aggregated state $i \in \{1, 2, \cdots, N-1\}$ can only be reached from aggregated state $i+1$. Therefore,

$$\rho_S + \rho_Q \leq \sum_{i=1}^{N} P_E$$

and this implies (noting $\rho_S + \rho_Q + \rho_R = 1$) that a suitable (conservative) ρ_R is given by

$$\rho_R \triangleq 1 - NP_E. \tag{4.14}$$

The calculation for $P_{E|S}$ in (4.11) can proceed by invoking Chebyshev's inequality. Consider some particular $x_k \in S$; then from (3.2b), we have

$$\Pr(\hat{a}_k \neq a_k | x_k) = \frac{1}{2} \cdot \Pr(|R_k| > h_0 | x_k); \qquad x_k \in S.$$

Now for all $x_k \in S$, we have $|S_k| - h_0 > 0$ by construction (4.5). Define

$$K_S \triangleq \min_{x_k \in S} \{|S_k| - h_0\} > 0,$$

i.e., in set S, the eye is always closed by at least K_S in the absence of noise. Therefore, when $|n_k| = |R_k - S_k| < K_S$, the eye remains closed, and we deduce (outline only)

$$\Pr(\hat{a}_k \neq a_k | x_k) \geq \frac{1}{2} \cdot \Pr(|R_k - S_k| < K_S | x_k); \qquad \forall x_k \in S.$$

Therefore, applying Chebyshev's inequality, we obtain

$$P_{E|S} \triangleq \Pr(\hat{a}_k \neq a_k | S) \geq \frac{1}{2} \cdot \left(1 - \frac{\sigma_n^2}{K_S^2}\right),$$

i.e., anticipating the high signal-to-noise ratio limit and using a signed order notation,

$$P_{E|S} \triangleq \frac{1}{2} - O(\sigma_n^2). \tag{4.15}$$

Clearly, by letting $\sigma_n^2 \to 0$, we obtain the noiseless result.

A similar calculation involving Chebyshev's inequality, for $M = Q$ or R where $h_0 - |S_k| > 0$ (eye open without noise), yields

$$P_{E|M} \leq \frac{1}{2} \cdot \frac{\sigma_n^2}{K_M^2}$$

where

$$K_M \triangleq \min_{x_k \in M} \{h_0 - |S_k|\} > 0.$$

Hence, if $M = R$ where $S_k = 0$ ($E_k = 0$) for all atomic states, we have $K_R = h_0$ which leads to

$$\epsilon \leq \frac{1}{2} \cdot \frac{\sigma_n^2}{h_0^2}. \tag{4.16}$$

This result says that: 1) the probability of error *after recovery* is bounded above by the reciprocal of twice the S/N ratio, and 2) $\epsilon \leq O(\sigma_n^2)$.

We can now compute the lower bound on P_E (4.11) by substituting for (4.13), (4.14), (4.15), and (4.16).

$$\begin{aligned} P_E &\geq \epsilon \cdot \rho_R + P_{E|S} \cdot \rho_S \\ &= \epsilon \cdot (1 - NP_E) + \left(\frac{1}{2} - O(\sigma_n^2)\right) \cdot 2P_E(1 - 2^{-n}) \\ &\geq \epsilon + P_E - 2^{-n}P_E - P_E \cdot O(\sigma_n^2) \\ &\geq \epsilon \cdot 2^n - O(\sigma_n^4). \end{aligned} \tag{4.17}$$

This completes our demonstration that the class of channels which satisfy (4.1) and (4.6) have a high signal-to-noise ratio limiting stationary error probability P_E which is arbitrarily close to the upper bound derived in [4], which is thus shown to be tight. (These channels are also precisely those whose stochastic dynamics in the absence of noise are described by the FSMP in Fig. 3.)

Remarks:

1) The asymptotic upper bound derived in [4] is for Gaussian noise n_k. This bound also holds for non-Gaussian noise distributions with finite variance. To show this, an analogous bounding procedure to that given in this section can be used.

2) The introduction of channel noise n_k into the analysis of DFE's involves only $O(\sigma_n^2)$ modifications relative to the noiseless case. This justifies the practical relevance of studying noiseless DFE's as was done in [1] when the signal-to-noise ratio is high.

3) The proof is invalid if $N \to \infty$. To incorporate this case, we need to modify (4.14) by imposing a further constraint on the channel tail; see [4].

V. The Effect of Noise on Error Recovery

Based on our results, we classify the effect of channel noise on the error recovery performance of the DFE relative to the noiseless case [1]. We distinguish three cases.

1) No Effect on DFE Recovery: We have constructed a channel where, before recovery, (bounded) noise is inconsequential to DFE recovery; see (3.4) and (3.7). After recovery, it is straightforward to construct (bounded) noise densities which realize any value of ϵ and hence P_E in the interval [0, 1/2]; see (3.1).

2) Noise Worsens DFE Recovery: This might be regarded as the expected or standard case. As an example, let the channel satisfy $2\|H_N\|_1 < h_0$, i.e., the eye is always open in the absence of noise. Then let $\Pr(|n_k| > h_0) > 0$. Clearly, with noise, the eye can close.

3) Noise Improves DFE Recovery: An example of this case was obtained in Section IV where, by construction for the noiseless case, $P_{E|S} = 1/2$, but with noise, $P_{E|S} \leq 1/2 - O(\sigma_n^2)$; see (4.15).

The third case is, of course, counterintuitive and the most interesting. It is a nontrivial problem to determine which classes of channels result in improved (transient) error recovery for certain noise distributions. Also, it is not clear whether these channels can be expected in practice.

Remark:

1) It is also possible to contrive situations where for a particular channel parameter/channel noise distribution combination, not only is the (transient) error recovery performance improved, as indicated by (4.15), but also the stationary error probability P_E is reduced. However, it is dubious whether such channels will appear in practice.

VI. Conclusions

A. Summary

In this paper, we have demonstrated that the most pathological conceivable DFE behavior is realized by some channel parameter/channel noise combination. By purely constructive methods, the bounds derived by Duttweiler *et al.* [4] were shown to be tight, thereby settling the open question raised in the conclusions of [4]. Further, we demonstrated that the presence of noise either: 1) has no effect on, 2) worsens, or 3) improves the DFE error recovery performance relative to the noiseless case. It seems a nontrivial exercise to classify explicitly those channels which benefit from the presence of channel noise. (Naturally, the level of benefit depends also on the noise distribution.)

B. Extensions

The broad classification given in Section V suggests a number of questions. For example: What is the optimal (possibly nonunique, possibly degenerate) channel noise distribution which minimizes the stationary error probability of a nonadaptive DFE on a given channel?

An application would be to improve the error recovery rate and the overall error probability by deliberately adding dither in a high signal-to-noise ratio DFE receiver, a technique well known in nonlinear control systems design.

C. Discussion

In contriving channels and noise densities which display pathological behavior, it is not our intention to suggest they reflect behavior to be observed in practice. (On the contrary, we would not expect or hope this to be the case.) More importantly, our results indicate the manner in which stronger hypotheses need to be imposed on the channel and noise, to tighten further the bounds on P_E and thereby reflect better the DFE performance to be expected in practice.

We have argued that results concerning the behavior of high signal-to-noise ratio DFE's can be suitably approximated by the noiseless DFE given in [1]. In [1], we presented evidence that a minimum phase constraint on the channel parameters was insufficient to guarantee satisfactory recovery behavior. A similar conclusion thus carries over to the noisy case where it appears certain minimum phase channels may give unacceptably high error rates through error propagation, even though as a class they generally appear attractive; see [7].

References

[1] R. A. Kennedy and B. D. O. Anderson, "Recovery times of decision feedback equalizers on noiseless channels," *IEEE Trans. Commun.*, this issue, pp. 1012–1021.

[2] A. Cantoni and P. Butler, "Stability of decision feedback inverses," *IEEE Trans. Commun.*, vol. COM-24, pp. 970–977, Sept. 1976.

[3] ——, "Behaviour of decision feedback inverses with deterministic and random inputs," Univ. Newcastle, Newcastle, Australia, Tech. Rep. EE7505, Mar. 1975.

[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] J. J. O'Reilly and A. M. de Oliveira Duarte, "Error propagation in decision feedback receivers," *Proc. IEE F, Commun., Radar, Signal Processing*, vol. 132, no. 7, pp. 561–566, 1985.

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

[7] B. R. Clarke, "The time-domain response of minimum phase networks," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 1187–1189, Nov. 1985.

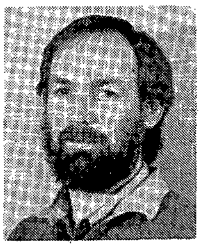
★

Rodney A. Kennedy (S'86), for a photograph and biography, see this issue, p. 1021.

★

Brian D. O. Anderson (S'62–M'66–SM'74–F'75), for a photograph and biography, see this issue, p. 1021.

★



Robert R. Bitmead (S'76–M'79–SM'86) was born in Sydney, Australia, in 1954. He received the B.Sc. degree in applied mathematics from the University of Sydney in 1976 and the M.E. and Ph.D. degrees in electrical engineering from the University of Newcastle, Australia, in 1977 and 1979, respectively.

His current position is Fellow in the Department of Systems Engineering, Australian National University, Canberra, which he has held since late 1982. For three years prior to his move to Canberra, he was a Lecturer in Electrical and Electronic Engineering at James Cook University of North Queensland, Townsville, Australia. In 1984 he held a visiting appointment in the School of Electrical Engineering, Cornell University, Ithaca, NY.

Dr. Bitmead was an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL from 1984 to 1986, and is currently Editor for Adaptive Control Theory of the *International Journal of Adaptive Control and Signal Processing*.