Performance Evaluation of Bayesian Decision Feedback Equalizer with $M$-PAM Symbols Using Importance Sampling Simulation

S. Chen
Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.

Abstract—An importance sampling (IS) simulation method is presented for evaluating the lower-bound symbol error rate (SER) of the Bayesian decision feedback equalizer (DFE) with $M$-PAM symbols, under the assumption of correct decision feedback. By exploiting an asymptotic property of the Bayesian DFE, a design procedure is developed, which chooses appropriate bias vectors for the simulation density to ensure asymptotic efficiency (AE) of the IS simulation.

I. INTRODUCTION

As the complexity of the Bayesian DFE [1]-[3] increases exponentially with the channel impulse response (CIR) length $n_h$ and the symbol size $M$, SER evaluation under high signal-to-noise-ratio (SNR) conditions becomes impossible using a conventional Monte Carlo simulation. This paper considers SER evaluation of the Bayesian DFE using an IS simulation method. The idea of IS is that certain values of the input variables have more impact on the error rate than others and, by sampling these “important” values more frequently, the estimator variance can be reduced [4]. The issue is then how to choose a biased distribution to encourage the important regions of the input variables. One of the most effective IS techniques is the mean translation approach [5]-[7], where the distribution is moved toward the error region.

For binary symbols, Ittis [8] developed a randomized bias technique for the IS simulation of Bayesian equalizers. This IS simulation method was extended to evaluate the lower-bound (assuming correct decision feedback) bit error rate of the Bayesian DFE with binary symbols [9],[10]. For the $M$-PAM case, the asymptotic Bayesian decision boundary for separating any two neighbouring signal classes can be deduced [11]. By exploiting a symmetric distribution within each signal subset, the SER of the Bayesian DFE for the $M$-PAM case can be shown to be a scaled error rate of the equivalent “binary” Bayesian DFE evaluated on two neighbouring signal subsets. These two properties enable an extension of the IS simulation technique to the $M$-PAM case.

II. THE BAYESIAN DECISION FEEDBACK EQUALIZER

Consider the real-valued channel generates the received signal samples of:

$$y(k) = \sum_{i=0}^{n_h-1} h_i s(k-i) + n(k)$$  \hspace{1cm} (1)

where $h_i$ are the CIR taps, the Gaussian white noise $n(k)$ has zero mean and variance $\sigma_n^2$, and $s(k)$ takes the value from the symbol set $S = \{s_i = 2i - M - 1, 1 \leq i \leq M\}$. The DFE uses the observed vector $y(k) = [y(k) \cdots y(k-m+1)]^T$ and the past detected symbol vector $\hat{s}_i(k) = [\hat{s}(k-d-1) \cdots \hat{s}(k-d-n_h)]^T$ to produce an estimate $\hat{s}(k-d)$ of $s(k-d)$, where $d$, $m$ and $n_h$ are the decision delay, feedforward and feedback orders, respectively. The choice of $d = n_h - 1$, $m = n_h$ and $n_h = n_h - 1$ will be used, as this choice guarantees a linear separability for different signal classes [12]. Let $s_f(k) = [s(k) \cdots s(k-d)]^T$ and $\hat{s}_i(k) = [\hat{s}(k-d-1) \cdots \hat{s}(k-d-n_h)]^T$. Express $y(k)$ as

$$y(k) = H_1 s_f(k) + H_2 \hat{s}_i(k) + n(k)$$  \hspace{1cm} (2)

where

$$H_1 = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n_h-1} \\ 0 & h_0 & \cdots & \vdots \\ \vdots & \ddots & \ddots & h_1 \\ 0 & 0 & \cdots & h_0 \end{bmatrix}$$  \hspace{1cm} (3)

$$H_2 = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \\ h_{n_h-1} & \cdots & h_{n_h-1} \end{bmatrix}$$  \hspace{1cm} (4)

are the $m \times (d+1)$ and $m \times n_h$ CIR matrices, respectively. Assuming correct past decisions, we have

$$y(k) = H_1 s_f(k) + H_2 \hat{s}_i(k) + n(k).$$  \hspace{1cm} (5)

Thus the decision feedback translate $y(k)$ into a new space:

$$r(k) = y(k) - H_2 \hat{s}_i(k).$$  \hspace{1cm} (6)

Let the $N_f = M^{d+1}$ combinations of $s_f(k)$ be $s_{f,j}, 1 \leq j \leq N_f$. The set of the noiseless channel states, namely, $\tilde{R} = \{r_j = H_1 s_{f,j}, 1 \leq j \leq N_f\}$ can be partitioned into $M$ subsets conditioned on $s(k-d)$:

$$R^{(i)} = \{r_j \in \tilde{R}: s(k-d) = s_i\}, 1 \leq i \leq M.$$  \hspace{1cm} (7)

The optimal Bayesian DFE [3] involves computing the $M$ decision variables for $1 \leq i \leq M$:

$$p_i(r(k)) = \sum_{r_j \in R^{(i)}} \exp \left( -\frac{||r(k) - r_j||^2}{2\sigma_n^2} \right)$$  \hspace{1cm} (8)
and making the decision according to
\[ \hat{s}(k - d) = s_{i*}, \quad \text{with} \quad i* = \arg \max_{1 \leq i \leq M} \{ \rho_i(r(k)) \}. \quad (9) \]

A geometric translation property [11] is reiterated here. For \( 1 \leq i \leq M - 1 \), \( R^{(i+1)} \) is a translation of \( R^{(i)} \) by the amount \( 2\h_{\text{rev}} \): \[ R^{(i+1)} = R^{(i)} + 2\h_{\text{rev}}, \] where \( \h_{\text{rev}} = [h_{M-1} \cdots h_1 h_0]^T \). This shifting property implies that, asymptotically when the SNR tends to infinity, the decision boundary \( B_{i+1} \) for separating \( R^{(i+1)} \) and \( R^{(i+2)} \) is a shift of \( B_i \) by an amount \( 2\h_{\text{rev}} \). Without the loss of generality, consider the two neighbouring subsets \( R^{(M/2)} \) and \( R^{(M/2)+1} \). A pair of opposite-class channel states \( (r^{(+)}, r^{(-)}) \in R^{(M/2)} \) is said to be Gabriel neighbours if \( \forall r_j \in R^{(M/2)} \cup R^{(M/2)+1}, r_j \neq r^{(+)}, r_j \neq r^{(-)} \):
\[ \|r_j - r_0^+\|^2 > \|r_j^+ - r_0\|^2 \quad (10) \]
where \( r_0 = (r^{(+)}, r^{(-)}/2) \). The following lemma [11] describes the asymptotic decision boundary \( B_{M/2} \)

**Lemma 1:** Asymptotically, the optimal decision boundary \( B_{M/2} \) separating \( R^{(M/2)} \) and \( R^{(M/2)+1} \) is piecewise linear and made up of \( L \) hyperplanes. Each of these hyperplanes is defined by a pair of Gabriel neighbours, the hyperplane is orthogonal to the line connecting the pair of Gabriel neighbours and passes through the midpoint of the line.

Consequently, a necessary condition for \( r_B \in B_{M/2} \) is
\[ r_B = \frac{r^{(+)}}{2} + \left[ \frac{r^{(+)}, r^{(-)}}{2} \right]^\perp \quad (11) \]
where \( x^\perp \) denotes an arbitrary vector in the subspace orthogonal to \( x \), \( r^{(+)}, r^{(-)} \) are a pair of Gabriel neighbours; and the sufficient conditions for \( r_B \in B_{M/2} \) are
\[ \|r_B - r^{(+)}/2\|^2 < \|r_B - r_0\|^2, \quad \forall r_i \in R^{(M/2)+1}, r_i \neq r^{(+)}, \quad (12) \]
\[ \|r_B - r^{(-)}\|^2 < \|r_B - r_0\|^2, \quad \forall r_j \in R^{(M/2)}, r_j \neq r^{(-)}, \quad (13) \]
\[ \|r_B - r^{(+)}\|^2 = \|r_B - r^{(-)}\|^2. \quad (14) \]

A simple algorithm can be used to select the set of all \( L \) pairs of Gabriel neighbours \( \{r_i^+, r_i^-\}_{i=1}^{L} \) [8,11].

Due to the symmetric distribution of \( S \), the states of \( R^{(i)} \) are distributed symmetrically around the mass center of \( R^{(i)} \). In particular, if a point \( r_j \in R^{(i)} \) has a distance \( x \) to the decision boundary \( B_{i-1} \), then there is another point \( r_l \in R^{(i)} \) with the same distance to the other decision boundary \( B_i \). Now consider the lower-bound SER for the Bayesian DFE
\[ P_E = \text{Prob}\{\hat{s}(k - d) \neq s(k - d)\} \quad (15) \]
First create a “binary” Bayesian DFE defined on \( R^{(M/2)} \) and \( R^{(M/2)+1} \) with the decision function given by
\[ f_i(r(k)) = \sum_{r_j \in R^{(M/2)+1}} \exp \left( -\frac{\|r(k) - r_j\|^2}{2\sigma_n^2} \right) - \sum_{r_i \in R^{(M/2)}} \exp \left( -\frac{\|r(k) - r_i\|^2}{2\sigma_n^2} \right) \quad (16) \]
and the decision rule defined by
\[ \hat{s}(k - d) = \begin{cases} 1, & \text{sgn}(f_i(r(k))) \geq 0, \\ -1, & \text{sgn}(f_i(r(k))) < 0. \end{cases} \quad (17) \]
Denote the error probability of this “binary” Bayesian DFE as \( P_e \). Taking into account of the shifting and symmetric properties discussed previously, it is straightforward to verify that \( P_E = \gamma P_e \), with \( \gamma = 2(M - 1)/M \).

### III. IS SIMULATION FOR THE M-PAM CASE

To evaluate the SER, \( P_E, \) of the Bayesian DFE with \( M \)-PAM symbols, it is only needed to evaluate the error probability, \( P_e, \) of the equivalent binary Bayesian DFE defined on \( R^{(M/2)} \) and \( R^{(M/2)+1} \). The IS simulation technique [9,10] can readily be used to evaluate \( P_e \) as follows:
\[ \hat{P}_e = 1 - \frac{1}{N_i \cdot \sum_{j=1}^{N_i} \sum_{k=1}^{N_e} I_E(r_j(k)) \cdot \frac{p(r_j(k)|r_j)}{p^*(r_j(k)|r_j)}} \quad (18) \]
where \( I_E(r_j(k)) = 1 \) if \( r(k) \) causes an error, and \( I_E(r_j(k)) = 0 \) otherwise; \( p(r_j(k)|r_j) \) is the true conditional density given \( r_j \in R^{(M/2)+1} \), and \( N_i = M^d = N_j/M \) is the number of states in \( R^{(M/2)+1} \); the sample \( r_j(k) \) is generated using the simulation density \( p^*(r_j(k)|r_j) \) chosen to be
\[ p^*(r_j(k)|r_j) = \sum_{j=1}^{L_j} \frac{p_{i,j}}{(2\pi\sigma_n^2)^d} \exp \left( -\frac{\|r_j(k) - v_{i,j}\|^2}{2\sigma_n^2} \right) \quad (19) \]
In the simulation density (19), \( L_j \) is the number of the bias vectors \( c_{i,j} = -r_j + v_{i,j} \) for \( r_j \in R^{(M/2)+1} \), \( p_{i,j} \geq 0 \) for \( 1 \leq l \leq L_j \), and \( \sum_{j=1}^{L_j} p_{i,j} = 1 \). An estimate of the IS gain for \( \hat{P}_e \), which is defined as the ratio of the numbers of trials required for the same estimate variance using the Monte Carlo and IS methods, is given as [6,8]:
\[ \Gamma = \frac{\hat{P}_e(1 - \hat{P}_e)}{\hat{P}_e - \hat{P}^2_e} \quad (20) \]
with
\[ \hat{P}_e = \frac{1}{N_s \cdot N_k \cdot \sum_{j=1}^{N_j} \sum_{k=1}^{N_s} I_E(r_j(k)) \left( \frac{p(r_j(k)|r_j)}{p^*(r_j(k)|r_j)} \right)^2. \quad (21) \]
The IS simulated \( P_E \) is simply \( \hat{P}_E = \gamma \hat{P}_e \), and the estimated IS gain for \( P_e \) will be used as the estimated IS gain for \( \hat{P}_E \).
A design procedure is given for constructing the simulation density $P_r(x|k) \mid p_j$ that meets the conditions for AE [6]. Let $\{r_i^{(+)}, r_i^{(-)}\}_{i=1}^L$ be the set of Gabriel neighbours selected from $R^{(M/2)}$ and $R^{(M/2)+1}$. Each pair $(r_i^{(+)}, r_i^{(-)})$ defines a hyperplane $H_i(r) = w_i^T r + b_i = 0$ that is part of the asymptotic decision boundary $B_M/2$, with

$$w_i = 2 \frac{(r_i^{(+)}/r_i^{(-)})}{\|r_i^{(+)}/r_i^{(-)}\|^2}, \quad b_i = -\frac{(r_i^{(+)}/r_i^{(-)})^T(r_i^{(+)}/r_i^{(-)})}{\|r_i^{(+)}/r_i^{(-)}\|^2}. \quad (22)$$

Note that $H_i$ is a canonical hyperplane with $H_i(r_i^{(+)}/r_i^{(-)}) = 1$ and $H_i(r_i^{(-)}) = -1$. A state $r_i^{(-)} \in R^{(M/2)}$ is sufficiently separable by the hyperplane $H_i$ if $w_i^T r_i^{(-)} + b_i \leq -1$. Similarly, $r_i^{(+)}/r_i^{(-)} \in R^{(M/2)+1}$ is sufficiently separable by $H_i$ if $w_i^T r_i^{(-)} + b_i \geq 1$. The hyperplane $H_i$ is reachable from $r_i^{(+)}/r_i^{(-)} \in R^{(M/2)+1}$ if the projection of $r_i^{(+)}$ onto $H_i$ is on $B_M/2$. For each $r_i^{(-)} \in R^{(M/2)}$, its separability index for $H_i$ is $\alpha_i^{(-)} = 1$ if $r_i^{(-)}$ is sufficiently separable by $H_i$; otherwise $\alpha_i^{(-)} = 0$. The separability index $\alpha_i^{(+)}$ for $r_i^{(+)}/r_i^{(-)} \in R^{(M/2)+1}$ is similarly defined. The reachability of $H_i$ from $r_i^{(+)}/r_i^{(-)} \in R^{(M/2)+1}$ can be tested by computing

$$c_{i,j} = -0.5 \left(w_i^T r_j^{(+)}/b_i \right) \left(r_i^{(+)}/r_i^{(-)} \right). \quad (24)$$

If $v_i,j = r_i^{(+)}/c_{i,j} \in B_M/2$ (i.e. $f_i(v_{i,j}) = 0$), $H_i$ is reachable from $r_i^{(+)}/c_{i,j}$ is then a bias vector), and the reachability index is $\gamma_{i,j} = 1$; otherwise $\gamma_{i,j} = 0$. The whole process produces the following separability and reachability table:

<table>
<thead>
<tr>
<th>$r_i^{(-)}$</th>
<th>$r_i^{(+)}/r_i^{(-)}$</th>
<th>$r_i^{(-)}$</th>
<th>$r_i^{(+)}/r_i^{(-)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{i,1}$</td>
<td>$a_{i,1}/N_i$</td>
<td>$a_{i,1}$</td>
<td>$a_{i,1}/N_i$</td>
</tr>
<tr>
<td>$a_{i,2}$</td>
<td>$a_{i,2}/N_i$</td>
<td>$a_{i,2}$</td>
<td>$a_{i,2}/N_i$</td>
</tr>
<tr>
<td>$a_{i,N_i}$</td>
<td>$a_{i,N_i}/N_i$</td>
<td>$a_{i,N_i}$</td>
<td>$a_{i,N_i}/N_i$</td>
</tr>
</tbody>
</table>

For each $r_j^{(+)}/r_j^{(-)} \in R^{(M/2)+1}$, select those hyperplanes that can sufficiently separate $r_j^{(+)}/r_j^{(-)}$ and are reachable from $r_j^{(+)}/r_j^{(-)}$ with the aid of the above table. This yields the integer set:

$$G_j^{(+)}/r_j^{(-)} = \{ q : \alpha_q^{(+)} = 1 \text{ and } \gamma_{q,j} = 1 \}. \quad (25)$$

The convex region $R_j^{(+)}/r_j^{(-)}$ covering $r_j^{(+)}/r_j^{(-)}$ is the intersection of all the half-spaces $H_q^{(+)}/r_q^{(-)} = \{ r : H_q(r) \geq 0 \}$ with $q \in G_j^{(+)}/r_j^{(-)}$. In fact, it is not necessary to use every hyperplane defined in $G_j^{(+)}/r_j^{(-)}$ to form $R_j^{(+)}/r_j^{(-)}$. A subset of these hyperplanes will be sufficient, provided that every opposite-class state in $R^{(M/2)}$ can sufficiently be separated by at least one hyperplane in the subset. If this can be done, the error region $E$ satisfies

$$E \subset R_j^{(+)}/r_j^{(-)} = \bigcup_{q \in G_j^{(+)}/r_j^{(-)}} H_q^{(+)}/r_q^{(-)} \quad (26)$$

with the half-spaces $H_q^{(+)}/r_q^{(-)} = \{ r : H_q(r) \leq 0 \}$. Obviously, all the hyperplanes defined in $G_j^{(+)}/r_j^{(-)}$ are reachable from $r_j^{(+)}/r_j^{(-)}$ and at least one of $\{ v_{i,j} \}$ is the minimum rate point (as defined in [6]). If $G_j^{(+)}/r_j^{(-)}$ exists for each $r_j^{(+)}/r_j^{(-)} \in R^{(M/2)+1}$, the simulation density constructed with the bias vectors $\{ v_{i,j} \}$, $q \in G_j^{(+)}/r_j^{(-)}$, for all $j$ will guarantee AE.

IV. SIMULATION EXAMPLES

Example 1. A 2-tap channel $h = [0.3 1.0]^T$ with 8-PAM symbols was simulated, given $m = 2$, $d = 1$ and $n_b = 1$. The set $R$ had 64 states. Nine pairs of Gabriel neighbours were selected from $R^{(4)}$ and $R^{(5)}$, leading to the separability and reachability table from which an AE simulation density was constructed. The simulation density construction is illustrated in Fig. 1. The bias vectors were selected with uniform probability in the simulation ($p_{i,j} = 1/L_{i,j}$). For each SNR, 10$^7$ iterations were used for each state in $R^{(5)}$. Thus, the total samples used for a given SNR were $8 \times 10^5$. Fig. 2 (a) shows the lower-bound SERs obtained using the IS and conventional sampling (CS) simulations, respectively. It can be seen that the conventional Monte Carlo results for low SNR conditions based directly on the Bayesian DFE of (8) and (9) agreed with those of the IS simulation. The estimated IS gains, depicted in Fig. 2 (b), indicate that exponential IS gains were obtained with increasing SNRs.

Example 2. A 3-tap channel $h = [0.3 1.0 - 0.3]^T$ with 8-PAM symbols was tested, given $m = 3$, $d = 2$ and $n_b = 2$. The set $R$ had 512 states. Nineteen pairs of Gabriel neighbours were found from $R^{(4)}$ and $R^{(5)}$, leading to the separability and reachability table from which an AE simulation density was constructed. Again the bias vectors were selected with uniform probability in the simulation. For each SNR, 10$^4$ samples were used for each state in $R^{(5)}$, resulting in a total of $6.4 \times 10^6$ samples for a given SNR. Fig. 3 (a) depicts the lower-bound SERs obtained using the IS and CS simulations, respectively. Again, the conventional Monte Carlo results for low SNR conditions agreed with those of the
Fig. 2. The lower-bound SERs (a) and the estimated IS gain (b) of the Bayesian DFE for $h = [0.3\ 1.0]^T$ with 8-PAM symbols.

IS simulation. It can be seen from Fig. 3 (b) that exponential IS gains were obtained with increasing SNRs.

V. CONCLUSIONS

An IS simulation has been extended to evaluate the lower-bound SER of the Bayesian DFE with $M$-PAM symbols. It has been noted that the Bayesian decision boundary separating any two neighbouring signal classes is asymptotically piecewise linear. Furthermore, the SER of the Bayesian DFE for the $M$-PAM case is a scaled error rate of the equivalent binary Bayesian DFE evaluated on any two neighbouring signal subsets. A design procedure has been presented for constructing the simulation density that meets the asymptotic efficiency conditions.

REFERENCES