AN IMPORTANCE SAMPLING SIMULATION METHOD FOR BAYESIAN DECISION FEEDBACK EQUALIZERS

S. Chen and L. Hanzo

Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.
sqc@ecs.soton.ac.uk

ABSTRACT

An importance sampling (IS) simulation technique is presented for evaluating the lower-bound bit error rate (BER) of the Bayesian decision feedback equalizer (DFE) under the assumption of correct decisions being fed back. A design procedure is developed, which chooses appropriate bias vectors for the simulation density to ensure asymptotic efficiency of the IS simulation.

1. INTRODUCTION

For the class of equalizers based on symbol-by-symbol decision with decision feedback, the Bayesian DFE [1]–[4] is known to provide the best performance. Due to its complexity, performance analysis of the Bayesian DFE is usually based on conventional Monte Carlo simulation, which is computationally very costly even for low signal to noise ratio (SNR) conditions. Iltis [5] developed a randomized bias technique for the IS simulation of Bayesian equalizers. Although it can only guarantee asymptotic efficiency, as defined in [6], for certain channels, this IS simulation technique provides a valuable method in assessing the performance of the Bayesian equalizer.

We apply the IS simulation technique to evaluate the lower-bound BER of the Bayesian DFE. By viewing decision feedback as a geometric translation, the Bayesian DFE is "converted" to the Bayesian equalizer in the translated space [7], with a desired property that the subsets of opposite-class channel states are always linearly separable. It can further be shown that the asymptotic decision boundary is piecewise linear. A design procedure is developed, which determines the set of hyperplanes that form the asymptotic Bayesian decision boundary and constructs the convex regions associated with individual states by intersecting hyperplanes that are reachable from the states concerned. This provides the appropriate bias vectors for the simulation density to ensure asymptotic efficiency.

2. THE BAYESIAN DFE

We will assume that the channel is real-valued and the received signal sample is given by:

$$y(k) = \sum_{i=0}^{n_a-1} a_i s(k-i) + e(k)\,, \tag{1}$$

where n_a is the channel impulse response (CIR) length, a_i are the channel taps, the Gaussian white noise $e(k)$ has zero mean and variance σ_e^2, and the transmitted symbol sequence $\{s(k)\}$ takes values from the set $\{\pm 1\}$. A DFE uses the observation vector $\mathbf{y}(k) = [y(k) \cdots y(k-m+1)]^T$ and the past detected symbol vector $\hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1) \cdots \hat{s}(k-d-n)]^T$ to produce an estimate $\hat{s}(k-d)$ of $s(k-d)$. Without the loss of generality, the decision delay of $d = n_a - 1$, feedforward order of $m = n_a$ and feedback order of $n = n_a - 1$ are chosen, as this choice is sufficient to guarantee the linear separability (see lemma 1 below).

The received signal vector can be expressed as: $\mathbf{y}(k) = F_1\ \mathbf{s}_f(k) + F_2\ \mathbf{s}_b(k) + \mathbf{e}(k)$, where $\mathbf{s}_f(k) = [s(k) \cdots s(k-d)]^T$, $\mathbf{s}_b(k) = [s(k-d-1) \cdots s(k-d-n)]^T$, and

$$F_1 = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n_a-1} \\ 0 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1 \\ 0 & \cdots & 0 & a_0 \end{bmatrix} \tag{2}$$

$$F_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{n_a-1} & 0 & \ddots & \vdots \\ a_{n_a-2} & a_{n_a-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \cdots & a_{n_a-2} & a_{n_a-1} \end{bmatrix} \tag{3}$$

are the $m \times (d+1)$ and $m \times n$ CIR matrices, respectively. Under the assumption of correct decision feedback, we have

$\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k)$, and the decision feedback translates the original signal space $\mathbf{y}(k)$ into a new space $\mathbf{r}(k)$:

$$\mathbf{r}(k) \stackrel{\triangle}{=} \mathbf{y}(k) - F_2\, \hat{\mathbf{s}}_b(k)\,. \tag{4}$$

Let the $N_f = 2^{d+1}$ sequences of $\mathbf{s}_f(k)$ be $\mathbf{s}_{fj}$, $1 \le j \le N_f$. The set of the noiseless channel states in the translated space is defined as $R \stackrel{\triangle}{=} \{\mathbf{r}_j = F_1\, \mathbf{s}_{fj},\ 1 \le j \le N_f\}$, which can be partitioned into the two subsets conditioned on $s(k-d)$:

$$R^{(\pm)} \stackrel{\triangle}{=} \{\mathbf{r}_j \in R : s(k-d) = \pm 1\}\,. \tag{5}$$

Lemma 1 $R^{(+)}$ and $R^{(-)}$ are linearly separable.

Proof: Choose the weights of a hyperplane $H(\mathbf{r}) = \mathbf{w}^T\mathbf{r} = 0$ to be: $\mathbf{w}^T = \left[0 \cdots 0\ \frac{1}{a_0}\right]$. For any $\mathbf{r}^{(+)} \in R^{(+)}$ and $\mathbf{r}^{(-)} \in R^{(-)}$, we have $\mathbf{w}^T\mathbf{r}^{(+)} = 1 > 0$ and $\mathbf{w}^T\mathbf{r}^{(-)} = -1 < 0$.

Although it is always possible to construct a single hyperplane to correctly separate opposite-class states for the DFE, the optimal decision boundary in general cannot be realized by one hyperplane.

Proposition 1 The asymptotic decision boundary ∂E of the Bayesian DFE for large SNR is piecewise linear and made up of a set of L hyperplanes. Each of these hyperplanes is defined by a pair of *dominant* opposite-class states ($\mathbf{r}_l^{(+)} \in R^{(+)}, \mathbf{r}_l^{(-)} \in R^{(-)}$), such that the hyperplane is orthogonal to the line connecting the pair of dominant states and passes through the midpoint of the line.

Proof: See [5]. As $\sigma_e^2 \to 0$, a necessary condition for a point $\mathbf{r} \in \partial E$ is

$$\mathbf{r} = \frac{\mathbf{r}_l^{(+)} + \mathbf{r}_l^{(-)}}{2} + \left[\frac{\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}}{2}\right]^{\perp}, \tag{6}$$

where $\mathbf{x}^{\perp}$ denotes an arbitrary vector in the subspace orthogonal to $\mathbf{x}$, $\mathbf{r}_l^{(+)}$ and $\mathbf{r}_l^{(-)}$ are a pair of dominant states; and the sufficient conditions for $\mathbf{r} \in \partial E$ are

$$\|\mathbf{r} - \mathbf{r}_l^{(+)}\|^2 < \|\mathbf{r} - \mathbf{r}_i\|^2,\ \forall \mathbf{r}_i \in R^{(+)},\ \mathbf{r}_i \ne \mathbf{r}_l^{(+)}, \tag{7}$$

$$\|\mathbf{r} - \mathbf{r}_l^{(-)}\|^2 < \|\mathbf{r} - \mathbf{r}_j\|^2,\ \forall \mathbf{r}_j \in R^{(-)},\ \mathbf{r}_j \ne \mathbf{r}_l^{(-)}, \tag{8}$$

$$\|\mathbf{r} - \mathbf{r}_l^{(+)}\|^2 = \|\mathbf{r} - \mathbf{r}_l^{(-)}\|^2\,. \tag{9}$$

Proposition 1 follows as a direct consequence.

The set of all the dominant state pairs $\{\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)}\}_{l=1}^{L}$ can easily be determined using an algorithm given in [5],[8].

3. IS SIMULATION METHOD

Since the Bayesian DFE is reduced to the Bayesian equalizer in the translated space, the IS simulation technique of [5] can be extended to evaluate its lower-bound BER, which is given by:

$$\hat{P}_e = \frac{1}{N_s}\frac{1}{N_k}\sum_{i=1}^{N_s}\sum_{k=1}^{N_k} I_E(\mathbf{r}_i(k))\frac{p(\mathbf{r}_i(k)|\mathbf{r}_i)}{p^*(\mathbf{r}_i(k)|\mathbf{r}_i)}\,, \tag{10}$$

where $I_E(\mathbf{r}(k)) = 1$ if $\mathbf{r}(k)$ causes an error, and $I_E(\mathbf{r}(k)) = 0$ otherwise; $p(\mathbf{r}_i(k)|\mathbf{r}_i)$ is the true conditional density given $\mathbf{r}_i \in R^{(+)}$, and $N_s = 2^d$ is the number of states in $R^{(+)}$; the sample $\mathbf{r}_i(k)$ is generated using the simulation density $p^*(\mathbf{r}_i(k)|\mathbf{r}_i)$ chosen to be

$$p^*(\mathbf{r}_i(k)|\mathbf{r}_i) = \sum_{j=1}^{L_i} p_{ji}\frac{1}{(2\pi\sigma_e^2)^{\frac{m}{2}}}\exp\left(-\frac{\|\mathbf{r}_i(k) - \mathbf{v}_{ji}\|^2}{2\sigma_e^2}\right)\,. \tag{11}$$

In the simulation density (11), L_i is the number of the bias vectors $\mathbf{c}_{ji} = -\mathbf{r}_i + \mathbf{v}_{ji}$ for $\mathbf{r}_i \in R^{(+)}$, $p_{ji} \ge 0$ for $1 \le j \le L_i$, and $\sum_{j=1}^{L_i} p_{ji} = 1$. An estimate of the IS gain, 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 in [5]. To achieve asymptotic efficiency, $\{\mathbf{c}_{ji}\}$ must meet certain conditions [6]. We present the following procedure of constructing $p^*(\mathbf{r}_i(k)|\mathbf{r}_i)$ to meet these conditions.

Each of the L dominant state pairs $\{\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)}\}$ defines a hyperplane $H_l(\mathbf{r}) = \mathbf{w}_l^T\mathbf{r} + b_l = 0$. The weight vector $\mathbf{w}_l$ and bias b_l of the hyperplane are given by:

$$\mathbf{w}_l = \frac{2\left(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\right)}{\|\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\|^2}\,, \tag{12}$$

$$b_l = -\frac{(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)})^T(\mathbf{r}_l^{(+)} + \mathbf{r}_l^{(-)})}{\|\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\|^2}\,. \tag{13}$$

Notice that the theory of support vector machines [9],[10] has been applied to determine the hyperplane H_l with ($\mathbf{r}_l^{(+)}$, $\mathbf{r}_l^{(-)}$) as its two support vectors, and the hyperplane H_l is a *canonical* hyperplane having the property $H_l(\mathbf{r}_l^{(+)}) = 1$ and $H_l(\mathbf{r}_l^{(-)}) = -1$.

A state $\mathbf{r}_i \in R$ is said to be *sufficiently separable* by the hyperplane H_l, if H_l can separate $\mathbf{r}_i$ correctly with $|\mathbf{w}_l^T\mathbf{r}_i + b_l| \ge 1$. Thus, if $\mathbf{w}_l^T\mathbf{r}_i^{(+)} + b_l \ge 1$ for $\mathbf{r}_i^{(+)} \in R^{(+)}$, $\mathbf{r}_i^{(+)}$ is sufficiently separable by H_l and a separability index $h_{li}^{(+)}$ is set to 1; otherwise $h_{li}^{(+)} = 0$. Similarly, if $\mathbf{r}_i^{(-)} \in R^{(-)}$

satisfies $\mathbf{w}_l^T \mathbf{r}_i^{(-)} + b_l \leq -1$, it is sufficiently separable by H_l and $h_{li}^{(-)} = 1$; otherwise $h_{li}^{(-)} = 0$. The reachability of H_l from $\mathbf{r}_i^{(+)} \in R^{(+)}$ can be tested by computing

$$\mathbf{c}_{li} = -0.5 \left(\mathbf{w}_l^T \mathbf{r}_i^{(+)} + b_l \right) \left(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)} \right) . \qquad (14)$$

If $\mathbf{v}_{li} = \mathbf{r}_i^{(+)} + \mathbf{c}_{li} \in \partial E$, H_l is said to be reachable from $\mathbf{r}_i^{(+)}$ ($\mathbf{c}_{li}$ is then a bias vector), and the reachability index is $\gamma_{li} = 1$; otherwise $\gamma_{li} = 0$. The process produces the following separability and reachability table:

	$\mathbf{r}_1^{(-)}$	$\cdots$	$\mathbf{r}_{N_s}^{(-)}$	$\mathbf{r}_1^{(+)}$	$\cdots$	$\mathbf{r}_{N_s}^{(+)}$
H_1	$h_{11}^{(-)}$	$\cdots$	$h_{1N_s}^{(-)}$	$h_{11}^{(+)}$ (γ_{11})	$\cdots$	$h_{1N_s}^{(+)}$ (γ_{1N_s})
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
H_L	$h_{L1}^{(-)}$	$\cdots$	$h_{LN_s}^{(-)}$	$h_{L1}^{(+)}$ (γ_{L1})	$\cdots$	$h_{LN_s}^{(+)}$ (γ_{LN_s})

In order to construct a convex region $\mathcal{R}_i^{(+)}$ for $\mathbf{r}_i^{(+)} \in R^{(+)}$, we select those hyperplanes that can *sufficiently* separate $\mathbf{r}_i^{(+)}$ and that are reachable from $\mathbf{r}_i^{(+)}$ with the aid of the above table. This yields the following integer set:

$$G_i^{(+)} \triangleq \{ j : \ h_{ji}^{(+)} = 1 \text{ and } \gamma_{ji} = 1 \} . \qquad (15)$$

Then $\mathcal{R}_i^{(+)}$ is the intersection of all the half-spaces $\mathcal{H}_j^{(+)} \triangleq \{\mathbf{r} : H_j(\mathbf{r}) \geq 0\}$ with $j \in G_i^{(+)}$. In fact, it is not necessary to use every hyperplanes defined in $G_i^{(+)}$ to construct $\mathcal{R}_i^{(+)}$. A subset of these hyperplanes will be sufficient, provided that every opposite-class state in $R^{(-)}$ can sufficiently be separated by at least one hyperplane in the subset. If such a $G_i^{(+)}$ exists for each $\mathbf{r}_i^{(+)}$, the simulation density constructed with the bias vectors $\{\mathbf{c}_{ji}\}$, $j \in G_i^{(+)}$, will achieve asymptotic efficiency, since all the hyperplanes defined in $G_i^{(+)}$ are reachable from $\mathbf{r}_i^{(+)}$ and obviously at least one of $\{\mathbf{v}_{ji}\}$ is the minimum rate point (as defined in [6]), and the error region E satisfies

$$E \subset \overline{\mathcal{R}_i^{(+)}} \triangleq \bigcup_{j \in G_i^{(+)}} \mathcal{H}_j^{(-)} \qquad (16)$$

with the half-spaces $\mathcal{H}_j^{(-)} \triangleq \{\mathbf{r} : H_j(\mathbf{r}) < 0\}$.

For the 2-tap channel $\mathbf{a} = [a_0 \ a_1]^T$, it is straightforward to verify that the simulation density for the Bayesian DFE can always be constructed to satisfy the conditions for asymptotic efficiency. This is in contrast to the case of the Bayesian equalizer where, for the 2-tap channel, asymptotic efficiency is not always guaranteed [5]. We believe that asymptotic efficiency of the IS simulation for the Bayesian DFE can generally be ensured, although a rigorous proof is still under consideration. This may be because of the linear separability and because of the associated property of a much more sparse state distribution due to the decision feedback. We have tested a variety of channels, and no counter example has been found.

4. SIMULATION EXAMPLE

The IS technique for the Bayesian DFE was simulated using the 3-tap CIR defined by:

$$\mathbf{a} = [-0.8 \ \ 1.0 \ \ -0.5]^T . \qquad (17)$$

The bias vectors were generated using the procedure described in the previous section. As in [5], the bias vectors were selected with uniform probability in the simulation. For all the cases, 10^5 iterations were employed at each SNR, averaging over all the possible states in $R^{(+)}$. Since the channel had a length of $n_a = 3$, the DFE structure was specified by $m = 3$, $d = 2$ and $n = 2$. The asymptotic decision boundary consisted of 5 hyperplanes. Table 1 gives the separability and reachability table for this channel.

	$R^{(-)}$				$R^{(+)}$			
H_1	1	1	0	1	0	0	1 (1)	0
H_2	1	0	1	1	1 (1)	1 (1)	0	1 (1)
H_3	1	1	1	1	0	1 (1)	0	0
H_4	0	1	0	0	1 (1)	0	1 (0)	1 (1)
H_5	0	0	1	0	1 (1)	1 (1)	1 (1)	1 (1)

Table 1: The separability and reachability table for the CIR of $\mathbf{a} = [-0.8 \ \ 1.0 \ \ -0.5]^T$. The DFE structure is defined by $m = 3$, $d = 2$ and $n = 2$. $R^{(\pm)} = \{\mathbf{r}_1^{(\pm)}, \mathbf{r}_2^{(\pm)}, \mathbf{r}_3^{(\pm)}, \mathbf{r}_4^{(\pm)}\}$.

The states $\mathbf{r}_1^{(+)}$ and $\mathbf{r}_4^{(+)}$ require the two hyperplanes H_2 and H_4 to separate them from all the opposite-class states, and H_2 and H_4 are reachable from both states. Thus, there are two bias vectors for $\mathbf{r}_1^{(+)}$ and $\mathbf{r}_4^{(+)}$, respectively, and $E \subset \mathcal{H}_2^{(-)} \bigcup \mathcal{H}_4^{(-)}$. The state $\mathbf{r}_2^{(+)}$ is separated from $R^{(-)}$ by the single reachable hyperplane H_3. The state $\mathbf{r}_3^{(+)}$ is separated from $R^{(-)}$ by the two reachable hyperplanes H_1 and H_5, and $E \subset \mathcal{H}_1^{(-)} \bigcup \mathcal{H}_5^{(-)}$. Asymptotic efficiency of the IS simulation is therefore guaranteed for this example. Fig. 1 shows the lower-bound BERs obtained using the IS and conventional simulation methods, respectively. It can be seen that the conventional Monte Carlo results for low SNR conditions agreed with those of the IS simulation. The estimated IS gains, depicted in Fig. 2, indicate that exponential IS gains were obtained with increasing SNRs.

5. CONCLUSIONS

We have extended the randomized bias technique for IS simulation of [5] to evaluate the lower-bound BER of the Bayesian DFE. A design procedure has been presented for constructing the simulation density that meets the asymptotic efficiency conditions. Although asymptotic efficiency for the general channel has not rigorously been proven, we are unable to find a counter example suggesting that the asymptotic efficiency conditions are not met. The more difficult problem of how to derive an upper-bound BER of the Bayesian DFE, taking into account error propagation, remains an open question and is still under investigation.

6. REFERENCES

[1] K. Abend and B.D. Fritchman, "Statistic detection for communication channels with intersymbol interference," *Proc. IEEE*, Vol.58, No.5, pp.779–785, 1970.

[2] D. Williamson, R.A. Kennedy and G.W. Pulford, "Block decision feedback equalization," *IEEE Trans. Communications*, Vol.40, No.2, pp.255-264, 1992.

[3] S. Chen, B. Mulgrew and S. McLaughlin, "Adaptive Bayesian equaliser with decision feedback," *IEEE Trans. Signal Processing*, Vol.41, No.9, pp.2918–2927, 1993.

[4] S. Chen, S. McLaughlin, B. Mulgrew and P.M. Grant, "Adaptive Bayesian decision feedback equaliser for dispersive mobile radio channels," *IEEE Trans. Communications*, Vol.43, No.5, pp.1937–1946, 1995.

[5] R.A. Iltis, "A randomized bias technique for the importance sampling simulation of Bayesian equalizers," *IEEE Trans. Communications*, Vol.43, No.2/3/4, pp.1107–1115, 1995.

[6] J.S. Sadowsky and J.A. Bucklew, "On large deviations theory and asymptotically efficient Monte Carlo estimation," *IEEE Trans. Information Theory*, Vol.36, No.3, pp.579–588, 1990.

[7] S. Chen, B. Mulgrew, E.S. Chng and G. Gibson, "Space translation properties and the minimum-BER linear-combiner DFE," *IEE Proc. Communications*, Vol.145, No.5, pp.316–322, 1998.

[8] E.S. Chng, B. Mulgrew, S. Chen and G. Gibson, "Optimum lag and subset selection for radial basis function equaliser," in *Proc. 5th IEEE Workshop Neural Networks for Signal Processing* (Cambridge, USA), Aug.31-Sept.2, 1995, pp.593–602.

[9] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[10] V. Vapnik, "The support vector method of function estimation," in *Nonlinear Modeling: Advanced Black-Box Techniques*, J.A.K. Suykens and J. Vandewalle, Eds. Boston, MA: Kluwer, 1998, pp.55–85.

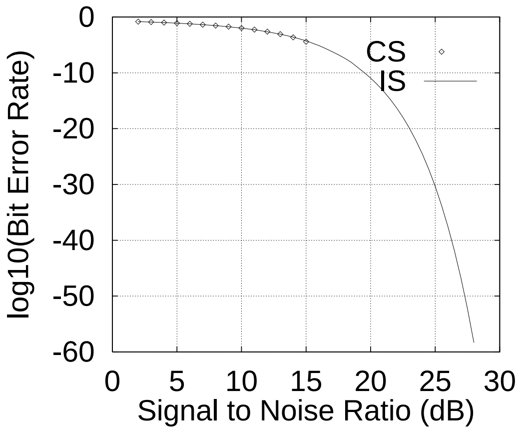

Figure 1: The lower-bound BERs of the Bayesian DFE for the CIR of $\mathbf{a} = [-0.8\ \ 1.0\ \ -0.5]^T$ using conventional sampling (CS) and importance sampling (IS) simulation. The DFE structure is defined by $m = 3$, $d = 2$ and $n = 2$.

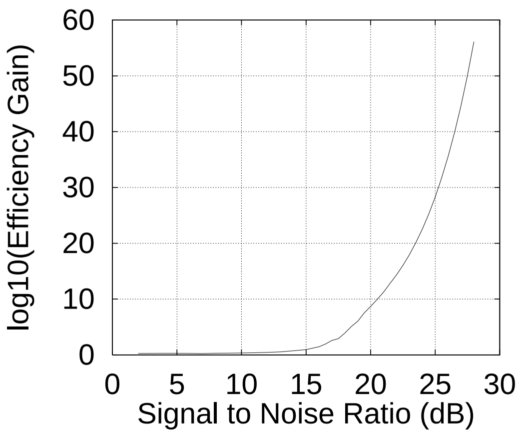

Figure 2: The IS gain of the Bayesian DFE for the CIR of $\mathbf{a} = [-0.8\ \ 1.0\ \ -0.5]^T$. The DFE structure is defined by $m = 3$, $d = 2$ and $n = 2$.