Importance Sampling Simulation for Evaluating the Lower-Bound BER of the Bayesian DFE

Sheng Chen

***Abstract*—An importance sampling (IS) simulation technique, originally derived by Iltis for Bayesian equalizers, is extended to evaluate the lower-bound bit error rate of the Bayesian decision feedback equalizer (under the assumption of correct decisions being fed back. Using a geometric translation approach, it is shown that the two subsets of opposite-class channel states are always linearly separable. A design procedure is presented, which chooses appropriate bias vectors for the simulation density to ensure asymptotic efficiency of the IS simulation.**

***Index Terms*—Asymptotic decision boundary, Bayesian decision feedback equalizer, importance sampling, Monte Carlo simulation.**

I. Introduction

AMONG the equalizers with symbol-decision structure and decision feedback, the maximum *a posteriori* probability or Bayesian decision feedback equalizer (DFE) [1]–[3] is known to provide the best performance. Due to its complexity, the performance of the Bayesian DFE is usually simulated using the conventional Monte Carlo approach, which is computationally very costly even for modest SNR conditions. Iltis [4] developed a randomized bias technique for the importance sampling (IS) simulation of Bayesian equalizers. Although asymptotic efficiency of this IS simulation technique can only be guaranteed for certain channels, it provides a valuable method in assessing the performance of the Bayesian equalizer.

We extend this IS simulation technique to evaluate the lower-bound bit error rate (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 [5], with a desired property that the subsets of opposite-class channel states are always linearly separable. 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 of the IS simulation as defined in [6].

Paper approved by M. Z. Win, the Editor for Equalization and Diversity of the IEEE Communications Society. Manuscript received August 20, 1998; revised August 20, 1999, and April 27, 2001.

The author is with the Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, U.K. (e-mail: sqc@ecs.soton.ac.uk).

Publisher Item Identifier S 0090-6778(02)01371-5.

II. Space Translation and Linear Separability

For the notational simplicity, we will assume the real-valued channel modeled to be

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

where n_a is the channel length, a_i the channel taps, the Gaussian white noise $e(k)$ has zero mean and variance σ_e^2, and the 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 $s(k-d)$. The integers d, m, and n are the decision delay and the feedforward and feedback orders, respectively. Without the loss of generality, $d = n_a - 1$, $m = n_a$, and $n = n_a - 1$ are chosen, as this choice is sufficient to guarantee the linear separability (cf. Lemma 1).

The channel observation vector can be expressed as $\mathbf{y}(k) = F_1\,\mathbf{s}_f(k) + F_2\,\mathbf{s}_b(k) + \mathbf{e}(k)$, where $\mathbf{s}_f(k) = [s(k) \cdots s(k-d)]^T$, $\mathbf{s}_b(k) = [s(k-d-1) \cdots s(k-d-n)]^T$, the $m \times (d+1)$ matrix F_1 and $m \times n$ matrix F_2 are defined by

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

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

Under the assumption of correct decision feedback, $\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k)$ and the decision feedback translates the original space $\mathbf{y}(k)$ into a new space $\mathbf{r}(k)$ as

$$\mathbf{r}(k) \triangleq \mathbf{y}(k) - F_2\,\hat{\mathbf{s}}_b(k). \tag{3}$$

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, $R \triangleq \{\mathbf{r}_j = F_1\,\mathbf{s}_{fj},\ 1 \le j \le N_f\}$, can be partitioned into the two subsets conditioned on $s(k-d)$ as

$$R^{(\pm)} \triangleq \{\mathbf{r}_j \in R : s(k-d) = \pm 1\}. \tag{4}$$

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

The proof is straightforward. Choose the weights of a hyperplane $H(\mathbf{r}) = \mathbf{w}^T \mathbf{r} = 0$ to be $\mathbf{w}^T = [0 \cdots 0\,(1/a_0)]$. For any

$\mathbf{r}^{(+)} \in R^{(+)}$ and $\mathbf{r}^{(-)} \in R^{(-)}$, we have $\mathbf{w}^T\mathbf{r}^{(+)} = 1 > 0$ and $\mathbf{w}^T\mathbf{r}^{(-)} = -1 < 0$.

Lemma 1 states that it is always possible to construct a single hyperplane to correctly separate opposite-class states for the DFE, although the optimal decision boundary in general cannot be realized by one hyperplane. In fact, the asymptotic decision boundary ∂E of the Bayesian DFE for large SNR consists of L hyperplanes. Each of these hyperplanes is defined by a pair of "dominant" opposite-class states $(\mathbf{r}_l^{(+)} \in R^{(+)}, \mathbf{r}_l^{(-)} \in R^{(-)})$ called Gabriel neighbors and the hyperplane is orthogonal to the line connecting the pair of Gabriel neighbors. The proposition 1 in [4] and the simple algorithm in [7] show how these L Gabriel neighbor pairs $\{\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)}\}$ can be found.

III. IS Simulation for the Bayesian DFE

Utilizing the geometric translation, the Bayesian DFE can be summarized as [8]

$$\hat{s}(k-d) = \mathrm{sgn}(f_B(\mathbf{r}(k))) \text{ with}$$
$$f_B(\mathbf{r}(k)) = \sum_{\mathbf{r}_j \in R^{(+)}} e^{\frac{-\|\mathbf{r}(k)-\mathbf{r}_j\|^2}{2\sigma_e^2}} - \sum_{\mathbf{r}_l \in R^{(-)}} e^{\frac{-\|\mathbf{r}(k)-\mathbf{r}_l\|^2}{2\sigma_e^2}} \quad (5)$$

where it has been assumed that channel states are equiprobable. Since the Bayesian DFE is reduced to the Bayesian equalizer in the translated space, the IS simulation technique of [4] can be extended to evaluate its lower-bound BER as follows:

$$\hat{P}_e = \frac{1}{N_s}\frac{1}{N_k}\sum_{i=1}^{N_s}\sum_{k=1}^{N_k} I_E(\mathbf{r}_i(k))\frac{p(\mathbf{r}_i(k)|\mathbf{r}_i)}{p^*(\mathbf{r}_i(k)|\mathbf{r}_i)} \quad (6)$$

where the indicator function $I_E(\mathbf{r}(k)) = 1$ if $\mathbf{r}(k)$ causes an error by (5) 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

$$p^*(\mathbf{r}_i(k)|\mathbf{r}_i) = \sum_{j=1}^{L_i} p_{ji}\frac{1}{(2\pi\sigma_e^2)^{\frac{m}{2}}} e^{\frac{-\|\mathbf{r}_i(k)-\mathbf{v}_{ji}\|^2}{2\sigma_e^2}}. \quad (7)$$

In the simulation density (7), L_i is the number of the bias vectors $\mathbf{c}_{ji} = -\mathbf{r}_i + \mathbf{v}_{ji}$ for $\mathbf{r}_i \in R^{(+)}$, $p_{ji} \geq 0$, 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 [4]. To achieve asymptotic efficiency, $\{\mathbf{c}_{ji}\}$ must meet certain conditions [6]. The following procedure of constructing $p^*(\mathbf{r}_i(k)|\mathbf{r}_i)$ shows how these conditions are met.

Each of the L Gabriel neighbor pairs $\{\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)}\}$ defines a hyperplane $H_l(\mathbf{r}) = \mathbf{w}_l^T\mathbf{r} + b_l = 0$, with the weight vector $\mathbf{w}_l$ and bias b_l given by

$$\mathbf{w}_l = \frac{2\left(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\right)}{\|\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\|^2}$$
$$b_l = -\frac{(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)})^T(\mathbf{r}_l^{(+)} + \mathbf{r}_l^{(-)})}{\|\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\|^2}. \quad (8)$$

Notice that the hyperplane defined by (8) is a *canonical* hyperplane with $(\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)})$ as its two support vectors and having the property $H_l(\mathbf{r}_l^{(+)}) = 1$ and $H_l(\mathbf{r}_l^{(-)}) = -1$ [9], [10].

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 $|\mathbf{w}_l^T\mathbf{r}_i + b_l| \geq 1$. Thus, if $\mathbf{w}_l^T\mathbf{r}_i^{(+)} + b_l \geq 1$ 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

$$\mathbf{c}_{li} = -0.5\left(\mathbf{w}_l^T\mathbf{r}_i^{(+)} + b_l\right)\left(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\right). \quad (9)$$

If $\mathbf{v}_{li} = \mathbf{r}_i^{(+)} + \mathbf{c}_{li} \in \partial E$, H_l is reachable from $\mathbf{r}_i^{(+)}$ ($\mathbf{c}_{li}$ is then a bias vector) and the reachability index $\gamma_{li} = 1$, otherwise $\gamma_{li} = 0$. The process produces the separability and reachability table shown at the bottom of the page.

To construct a convex region $\mathcal{R}_i^{(+)}$ associated with $\mathbf{r}_i^{(+)} \in R^{(+)}$, select those hyperplanes that can *sufficiently* separate $\mathbf{r}_i^{(+)}$ and are reachable from $\mathbf{r}_i^{(+)}$ and denote

$$G_i^{(+)} \triangleq \{j : h_{ji}^{(+)} = 1 \text{ and } \gamma_{ji} = 1\}. \quad (10)$$

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 hyperplane defined in $G_i^{(+)}$ to construct $\mathcal{R}_i^{(+)}$. A subset of these hyperplanes will be enough, 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 $\{\mathbf{c}_{ji}\}$,

	$\boldsymbol{r}_1^{(-)}$	$\dots$	$\boldsymbol{r}_{N_s}^{(-)}$	$\boldsymbol{r}_1^{(+)}$	$\dots$	$\boldsymbol{r}_{N_s}^{(+)}$
H_1	$h_{11}^{(-)}$	$\dots$	$h_{1N_s}^{(-)}$	$h_{11}^{(+)}(\gamma_{11})$	$\dots$	$h_{1N_s}^{(+)}(\gamma_{1N_s})$
$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
H_L	$h_{L1}^{(-)}$	$\dots$	$h_{LN_s}^{(-)}$	$h_{L1}^{(+)}(\gamma_{L1})$	$\dots$	$h_{LN_s}^{(+)}(\gamma_{LN_s})$

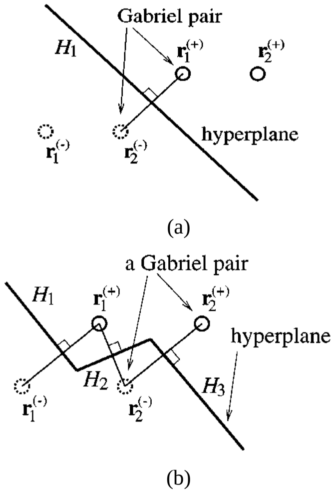


Fig. 1. The two typical cases of the asymptotic Bayesian decision boundary for channel $\mathbf{a} = [a_0\ a_1]^T$. The DFE structure is defined by $m = 2$, $d = 1$, and $n = 1$.

$j \in G_i^{(+)}$, 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 and the error region

$$E \subset \overline{\mathcal{R}_i^{(+)}} \triangleq \bigcup_{j \in G_i^{(+)}} \mathcal{H}_j^{(-)} \tag{11}$$

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

For the two-tap channel $\mathbf{a} = [a_0\ a_1]^T$, the existence of $G_i^{(+)}$ is guaranteed. This is because for the two-tap channel there exist only two scenarios as illustrated in Fig. 1. Case (a) is trivial. In case (b), there are three Gabriel neighbor pairs and the asymptotic decision boundary is made up of three hyperplanes. H_1 and H_2 are reachable from $\mathbf{r}_1^{(+)}$, $E \subset \mathcal{H}_1^{(-)} \bigcup \mathcal{H}_2^{(-)}$ and one of $\{\mathbf{v}_{11}, \mathbf{v}_{21}\}$ is the minimum rate point. Similarly, H_3 is reachable from $\mathbf{r}_2^{(+)}$ and the error region is fully contained in the half-space $\mathcal{H}_3^{(-)}$. Thus, for the two-tap channel, 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 for the two-tap channel [4], where asymptotic efficiency is not always guaranteed. Without a rigorous proof, we believe that asymptotic efficiency of the IS simulation for the Bayesian DFE can generally be ensured. This is because of the linear separability and a much sparse state distribution due to decision feedback. We have tested a variety of channels and no counter example has been found. A rigorous proof of asymptotic efficiency is still under investigation.

IV. Simulation Results

The IS technique for the Bayesian DFE was simulated using two channels defined by

$$\left.\begin{array}{ll} \text{Channel 1}: & \mathbf{a} = [0.4\ 0.7\ 0.4]^T \\ \text{Channel 2}: & \mathbf{a} = [0.35\ 0.8\ 1.0\ 0.8]^T. \end{array}\right\} \tag{12}$$

The bias vectors were generated using the procedure described in the previous section. As in [4], the bias vectors were selected with uniform probability in the simulation, i.e., $p_{ji} = 1/L_i$, $1 \le j \le L_i$. For all the cases, 10^5 iterations at each SNR were run, averaging over all the possible states in $R^{(+)}$.

Channel 1 had a length $n_a = 3$ and, therefore, the DFE structure was specified by $m = 3$, $d = 2$, and $n = 2$. The asymptotic decision boundary consisted of five hyperplanes. Table I gives the separability and reachability table for this channel. The state $\mathbf{r}_1^{(+)}$ requires the two hyperplanes H_1 and H_2 to separate it from all the opposite-class states and H_1 and H_2 are reachable from $\mathbf{r}_1^{(+)}$. Thus, there are two bias vector $\mathbf{c}_{11}$ and $\mathbf{c}_{21}$ and $E \subset \mathcal{H}_1^{(-)} \bigcup \mathcal{H}_2^{(-)}$. The states $\mathbf{r}_2^{(+)}$ and $\mathbf{r}_3^{(+)}$ are separated from $R^{(-)}$ by the two reachable hyperplanes H_3 and H_4 and $E \subset \mathcal{H}_3^{(-)} \bigcup \mathcal{H}_4^{(-)}$. The state $\mathbf{r}_4^{(+)}$ is separated from $R^{(-)}$ by the single reachable hyperplane H_5. Asymptotic efficiency of the IS simulation is therefore guaranteed for this example. Fig. 2(a) shows the lower-bound BERs obtained using the IS and conventional simulation methods, respectively. It can be seen that the conventional Monte Carlo simulation results for low-SNR conditions 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 SNR.

TABLE I
THE SEPARABILITY AND REACHABILITY TABLE FOR CHANNEL $\mathbf{a} = [0.4\ 0.7\ 0.4]^T$. THE DFE STRUCTURE IS DEFINED BY $m = 3$, $d = 2$, AND $n = 2$. $R^{(\pm)} = \{\mathbf{r}_1^{(\pm)}, \mathbf{r}_2^{(\pm)}, \mathbf{r}_3^{(\pm)}, \mathbf{r}_4^{(\pm)}\}$

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

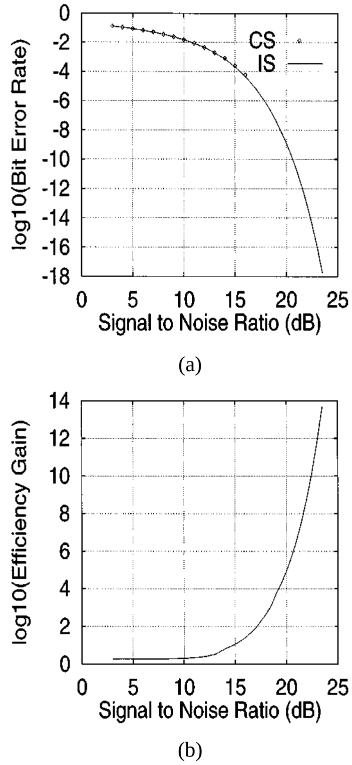
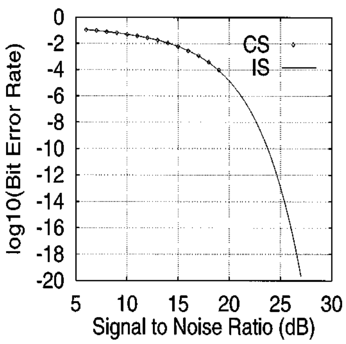


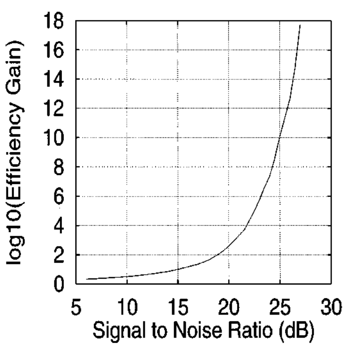
Fig. 2. (a) The lower-bound BERs and (b) IS gain of the Bayesian DFE for channel $\mathbf{a} = [0.4\ 0.7\ 0.4]^T$. The DFE structure is defined by $m = 3$, $d = 2$, and $n = 2$.

TABLE II
THE SEPARABILITY AND REACHABILITY TABLE FOR CHANNEL $\mathbf{a} = [0.35\ 0.8\ 1.0\ 0.8]^T$. THE DFE STRUCTURE IS DEFINED BY $m = 4$, $d = 3$ AND $n = 3$. $R^{(\pm)} = \{\mathbf{r}_1^{(\pm)}, \mathbf{r}_2^{(\pm)}, \mathbf{r}_3^{(\pm)}, \mathbf{r}_4^{(\pm)}, \mathbf{r}_5^{(\pm)}, \mathbf{r}_6^{(\pm)}, \mathbf{r}_7^{(\pm)}, \mathbf{r}_8^{(\pm)}\}$

	$R^{(-)}$								$R^{(+)}$							
H_1	1	1	1	1	1	1	1	1	1 (1)	0	1 (1)	1 (1)	0	0	1 (1)	0
H_2	1	1	1	0	1	1	0	0	0	1 (1)	1 (1)	1 (1)	1 (1)	1 (0)	1 (1)	1 (1)
H_3	1	1	0	1	1	1	1	1	1 (1)	1 (1)	1 (1)	1 (1)	0	0	1 (1)	1 (1)
H_4	1	1	1	1	1	1	0	0	0	0	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
H_5	1	1	0	0	1	1	1	1	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	0	1 (1)	1 (1)
H_6	1	1	1	1	1	1	1	0	0	0	1 (1)	1 (1)	0	1 (1)	1 (1)	1 (1)
H_7	0	1	0	0	1	1	0	1	1 (0)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)



(a)



(b)

Fig. 3. (a) The lower-bound BERs and (b) IS gain of the Bayesian DFE for channel $\mathbf{a} = [0.35\ 0.8\ 1.0\ 0.8]^T$. The DFE structure is defined by $m = 4$, $d = 3$, and $n = 3$.

As channel 2 had a length $n_a = 4$, the DFE structure was specified by $m = 4$, $d = 3$, and $n = 3$. The asymptotic decision boundary was made up of seven hyperplanes. Table II shows the separability and reachability table for this channel. From Table II, a simulation density with asymptotic efficiency was obtained. Fig. 3 depicts the lower-bound BERs obtained using the IS and conventional Monte Carlo methods as well as the estimated IS gains. It can be seen that the results of the conventional Monte Carlo simulation for low SNRs agreed with those of the IS simulation and exponential IS gains were obtained with increasing SNR.

V. CONCLUSION

We have extended the randomized bias technique for IS simulation 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. For the two-tap channel, asymptotic efficiency is guaranteed when using the IS simulation technique to estimate the lower-bound BER of the Bayesian DFE. Although asymptotic efficiency for the general channel has not rigorously been proved, 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.

REFERENCES

[1] D. Williamson, R. A. Kennedy, and G. W. Pulford, "Block decision feedback equalization," *IEEE Trans. Commun.*, vol. 40, pp. 255–264, Feb. 1992.

[2] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian equaliser with decision feedback," *IEEE Trans. Signal Processing*, vol. 41, pp. 2918–2927, Sept. 1993.

[3] S. Chen, S. McLaughlin, B. Mulgrew, and P. M. Grant, "Adaptive Bayesian decision feedback equaliser for dispersive mobile radio channels," *IEEE Trans. Commun.*, vol. 43, pp. 1937–1946, May 1995.

[4] R. A. Iltis, "A randomized bias technique for the importance sampling simulation of Bayesian equalizers," *IEEE Trans. Commun.*, vol. 43, pp. 1107–1115, Feb./Mar./Apr. 1995.

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

[6] J. S. Sadowsky and J. A. Bucklew, "On large deviations theory and asymptotically efficient Monte Carlo estimation," *IEEE Trans. Inform. Theory*, vol. 36, pp. 579–588, 1990.

[7] 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, MA, 1995, pp. 593–602.

[8] S. Chen, S. McLaughlin, B. Mulgrew, and P. M. Grant, "Bayesian decision feedback equaliser for overcoming co-channel interference," *Proc. IEECommun.*, vol. 143, no. 4, pp. 219–225, 1996.

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

[10] S. Chen, S. Gunn, and C. J. Harris, "Decision feedback equalizer design using support vector machines," *Proc. IEE Vision, Image Signal Processing*, vol. 147, no. 3, pp. 213–219, 2000.