Importance Sampling Simulation for Evaluating Lower-Bound Symbol Error Rate of the Bayesian DFE with Multi-Level Signalling Schemes

S. Chen

Abstract

For the class of equalizers that employs a symbol-decision finite-memory structure with decision feedback, the optimal solution is known to be the Bayesian decision feedback equalizer (DFE). The complexity of the Bayesian DFE however increases exponentially with the length of the channel impulse response (CIR) and the size of the symbol constellation. Conventional Monte Carlo simulation for evaluating the symbol error rate (SER) of the Bayesian DFE becomes impossible for high channel signal to noise ratio (SNR) conditions. It has been noted that the optimal Bayesian decision boundary separating any two neighbouring signal classes is asymptotically piecewise linear and consists of several hyperplanes, when the SNR tends to infinity. This asymptotic property can be exploited for efficient simulation of the Bayesian DFE. An importance sampling (IS) simulation technique is presented based on this asymptotic property for evaluating the lower-bound SER of the Bayesian DFE with a multi-level pulse amplitude modulation (M-PAM) scheme, 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 (AE) of the IS simulation.

Keywords

Bayesian decision feedback equalizer, asymptotic decision boundary, Monte Carlo simulation, importance sampling, symbol error rate.

I. Introduction

Equalization technique plays an ever-increasing role in combating distortion and interference in communication links [1],[2] and high-density data storage systems [3],[4]. For the class of equalizers based on a symbol-by-symbol decision with decision feedback, the maximum *a posteriori* probability equalizer with decision feedback or Bayesian DFE [5]-[8] is known to provide the best performance. The complexity of this optimal Bayesian solution, however, increases exponentially with the CIR length and the size of symbol constellation. Furthermore, due to its complicated structure, performance analysis of the Bayesian DFE is usually based on conventional Monte Carlo simulation, which is computationally costly even for modest SNR conditions. To obtain a reliable SER estimate, at least 100 errors should occur during a simulation. Thus, for an SER level of 10^{-6}, at least 10^8 data samples are needed. Investigating the Bayesian DFE under SER performance better than 10^{-6} is very difficult if not impossible, using a conventional Monte Carlo simulation.

Within the context of communication systems, IS refers to a simulation technique which aims to reduce the variance of the error rate estimator. By reducing the variance of error rate estimator, IS can achieve a given precision from shorter simulation runs, compared with a conventional Monte Carlo

SP EDICS: 3-PERF Performance Analysis, Optimization, and Limits.
The author is with the Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, U.K. (Tel./Fax: +44 (0)23 8095 6660/4508; E-mail: sqc@ecs.soton.ac.uk.)

simulation. For an excellent review of IS techniques, see [9]. The basic idea behind IS is that certain values of the input random variables in a simulation have more impact on the error probability being estimated than others. If these "important" values are emphasized by sampling more frequently, the estimator variance can be reduced. The fundamental issue in IS simulation is then the choice of the biased distribution, which encourages the important regions of the input variables. One of the most effective IS techniques is the mean translation approach [10]–[14], where the distribution is moved toward the error region. This is usually corresponding to shifting the density to a decision boundary. It is highly desired that a chosen IS technique is asymptotically efficient. For a precise definition of AE, see for example, [11]. Loosely speaking, an AE estimator requires a number of simulation trials which grows less than exponentially fast as the error rate tends to zero. Thus, when AE estimators are available, it is realistic to attempt extremely low error probability simulation.

Application of a mean-translation based IS technique to practical simulation systems is by no means a straightforward and easy task. For the binary phase shift keying (BPSK) modulation scheme, Iltis [15] developed a randomized bias technique for the IS simulation of Bayesian equalizers without decision feedback. The simulation density in Iltis' scheme consists of a sum of Gaussian distributions with the bias vector being chosen from a fixed set in a random manner. Although it can only guarantee asymptotic efficiency for certain channels, this IS simulation technique provides a valuable method in assessing the performance of the Bayesian equalizer. This IS simulation technique was extended to evaluate the lower-bound (assuming correct decision feedback) bit error rate of the Bayesian DFE with the BPSK scheme [16],[17]. This paper considers an IS simulation for evaluating the lower-bound SER of the Bayesian DFE with M-PAM symbols. Based on a geometric translation property for the M subsets of noise-free channel states, the asymptotic Bayesian decision boundary for separating any two neighbouring signal classes can be deduced [18]. Furthermore, by exploiting a symmetric distribution within each subset of channel states, the SER of the Bayesian DFE for the M-PAM symbol constellation is shown to be a scaled error rate of the equivalent "binary" Bayesian DFE evaluated on any two neighbouring signal subsets. These two properties enable an extension of the IS simulation technique for the binary Bayesian DFE [16],[17] to the general M-PAM case.

II. The Bayesian decision feedback equalizer

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

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

where h_i are the CIR taps, n_h is the CIR length, the Gaussian white noise $n(k)$ has zero mean and variance σ_n^2, and the M-PAM symbol $s(k)$ takes the value from the symbol set:

$$\mathcal{S} \triangleq \{s_i = 2i - M - 1,\ 1 \le i \le M\}\,. \tag{2}$$

The channel SNR is defined as

$$\mathrm{SNR} \triangleq \left(\sum_{i=0}^{n_h-1} h_i^2\right) \sigma_s^2/\sigma_n^2 \tag{3}$$

where σ_s^2 is the symbol variance. The generic DFE uses the information present in the noisy observation vector $\mathbf{y}(k) = [y(k)\ y(k-1) \cdots y(k-m+1)]^T$ and the past detected symbol vector $\hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1) \cdots \hat{s}(k-d-n_b)]^T$ to produce an estimate $\hat{s}(k-d)$ of $s(k-d)$, where d, m and n_b are the decision delay, the feedforward and feedback orders, respectively. The choice of $d = n_h - 1$, $m = n_h$ and $n_b = n_h - 1$ will be used, as this choice is sufficient to guarantee a desired linear separability for different signal classes [19]. With this choice, the observation vector $\mathbf{y}(k)$ can be expressed as [8],[19]:

$$\mathbf{y}(k) = \mathbf{H}_1\ \mathbf{s}_f(k) + \mathbf{H}_2\ \mathbf{s}_b(k) + \mathbf{n}(k) \tag{4}$$

where $\mathbf{s}_f(k) = [s(k) \cdots s(k-d)]^T$, $\mathbf{s}_b(k) = [s(k-d-1) \cdots s(k-d-n_b)]^T$, $\mathbf{n}(k) = [n(k) \cdots n(k-m+1)]^T$, and

$$\mathbf{H}_1 = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n_h-1} \\ 0 & h_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & h_1 \\ 0 & \cdots & 0 & h_0 \end{bmatrix} \tag{5}$$

$$\mathbf{H}_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ h_{n_h-1} & 0 & \ddots & \vdots \\ h_{n_h-2} & h_{n_h-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ h_1 & \cdots & h_{n_h-2} & h_{n_h-1} \end{bmatrix} \tag{6}$$

are the $m \times (d+1)$ and $m \times n_b$ CIR matrices, respectively.

Assuming correct past decisions, we have $\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k)$ and

$$\mathbf{y}(k) = \mathbf{H}_1\ \mathbf{s}_f(k) + \mathbf{H}_2\ \hat{\mathbf{s}}_b(k) + \mathbf{n}(k)\,. \tag{7}$$

Thus the decision feedback translate the original observation space $\mathbf{y}(k)$ into a new space $\mathbf{r}(k)$:

$$\mathbf{r}(k) \triangleq \mathbf{y}(k) - \mathbf{H}_2\ \hat{\mathbf{s}}_b(k)\,. \tag{8}$$

There are $N_f = M^{d+1}$ possible values or sequences of $\mathbf{s}_f(k)$, denoted as $\mathbf{s}_{f,j}$, $1 \le j \le N_f$. The set of the noiseless channel states in the translated signal space is then defined by

$$R \triangleq \{\mathbf{r}_j = \mathbf{H}_1\ \mathbf{s}_{f,j},\ 1 \le j \le N_f\}\,. \tag{9}$$

The channel state set R can be partitioned into M subsets conditioned on the value of $s(k-d)$:

$$R^{(i)} \triangleq \{\mathbf{r}_j \in R: \ s(k-d) = s_i\}, \ 1 \le i \le M \,. \tag{10}$$

The optimal Bayesian DFE [8] can now be summarized. The M decision variables are given by

$$\rho_i(\mathbf{r}(k)) = \sum_{\mathbf{r}_j \in R^{(i)}} \exp\left(-\frac{\|\mathbf{r}(k) - \mathbf{r}_j\|^2}{2\sigma_n^2}\right), \ 1 \le i \le M \tag{11}$$

and the minimum-error-probability decision is defined by

$$\hat{s}(k-d) = s_{i^*} \quad \text{with} \quad i^* = \arg\max_{1 \le i \le M} \{\rho_i(\mathbf{r}(k))\} \,. \tag{12}$$

A. Symmetric structure of subset states and asymptotic Bayesian decision boundary

In [18], a geometric translation property has been established, relating any two "neighbouring" subsets of channel states. This property is re-iterated here in lemma 1.

Lemma 1: For $1 \le i \le M-1$, the subset $R^{(i+1)}$ is a translation of $R^{(i)}$ by the amount $2\mathbf{h}_{\text{rev}}$:

$$R^{(i+1)} = R^{(i)} + 2\mathbf{h}_{\text{rev}} \tag{13}$$

where $\mathbf{h}_{\text{rev}} = [h_{n_h-1} \cdots h_1 \ h_0]^T$. Furthermore, $R^{(i)}$ and $R^{(i+1)}$ are linearly separable.

It is obvious that $R^{(1)}$ has one neighbour $R^{(2)}$, $R^{(M)}$ has one neighbour $R^{(M-1)}$, and $R^{(i)}$, $2 \le i \le M-1$, has two neighbours $R^{(i-1)}$ and $R^{(i+1)}$. This shifting property implies that asymptotically the decision boundary $\mathcal{B}_{i+1}$ for separating $R^{(i+1)}$ and $R^{(i+2)}$ is a shift of $\mathcal{B}_i$ for separating $R^{(i)}$ and $R^{(i+1)}$ by an amount $2\mathbf{h}_{\text{rev}}$. Thus the construction of the asymptotic Bayesian decision boundary for the binary Bayesian DFE [20] can readily be applied for the construction of the asymptotic decision boundary for separating any two neighbouring signal classes [18]. For the completeness, the relevant results given in [18] are summarized. Without the loss of generality, consider the two neighbouring subsets $R^{\left(\frac{M}{2}\right)}$ and $R^{\left(\frac{M}{2}+1\right)}$, which corresponds to the two classes $s_{\frac{M}{2}} = -1$ and $s_{\frac{M}{2}+1} = 1$. First define the concept of Gabriel neighbour states.

Definition 1: A pair of opposite-class channel states $(\mathbf{r}^{(+)} \in R^{\left(\frac{M}{2}+1\right)}, \mathbf{r}^{(-)} \in R^{\left(\frac{M}{2}\right)})$ is said to be a Gabriel neighbour pair if $\forall \mathbf{r}_j \in R^{\left(\frac{M}{2}\right)} \bigcup R^{\left(\frac{M}{2}+1\right)}$, $\mathbf{r}_j \ne \mathbf{r}^{(+)}$ and $\mathbf{r}_j \ne \mathbf{r}^{(-)}$:

$$\|\mathbf{r}_j - \mathbf{r}_0\|^2 > \|\mathbf{r}^{(+)} - \mathbf{r}_0\|^2 \tag{14}$$

where $\bigcup$ denotes the union operator and

$$\mathbf{r}_0 = \frac{\mathbf{r}^{(+)} + \mathbf{r}^{(-)}}{2} \,. \tag{15}$$

The following lemma describes the optimal decision boundary $\mathcal{B}_{\frac{M}{2}}$ that separates $R^{\left(\frac{M}{2}\right)}$ and $R^{\left(\frac{M}{2}+1\right)}$ in the asymptotic case of $\sigma_n^2 \rightarrow 0$.

Lemma 2: Asymptotically, the optimal decision boundary $\mathcal{B}_{\frac{M}{2}}$ separating $R^{\left(\frac{M}{2}\right)}$ and $R^{\left(\frac{M}{2}+1\right)}$ is piecewise linear and made up of a set of L hyperplanes. Each of these hyperplanes is defined by a pair of Gabriel neighbour states, the hyperplane is orthogonal to the line connecting the Gabriel neighbour pair and passes through the midpoint of the line.

Consequently, a necessary condition for a point $\mathbf{r}_B \in \mathcal{B}_{\frac{M}{2}}$ is

$$\mathbf{r}_B = \frac{\mathbf{r}^{(+)} + \mathbf{r}^{(-)}}{2} + \left[\frac{\mathbf{r}^{(+)} - \mathbf{r}^{(-)}}{2}\right]^{\perp} \tag{16}$$

where $\mathbf{x}^{\perp}$ denotes an arbitrary vector in the subspace orthogonal to $\mathbf{x}$, $\mathbf{r}^{(+)}$ and $\mathbf{r}^{(-)}$ are a pair of Gabriel neighbour states; and the sufficient conditions for $\mathbf{r}_B \in \mathcal{B}_{\frac{M}{2}}$ are

$$\|\mathbf{r}_B - \mathbf{r}^{(+)}\|^2 < \|\mathbf{r}_B - \mathbf{r}_l\|^2, \ \forall \mathbf{r}_l \in R^{\left(\frac{M}{2}+1\right)}, \ \mathbf{r}_l \neq \mathbf{r}^{(+)}, \tag{17}$$

$$\|\mathbf{r}_B - \mathbf{r}^{(-)}\|^2 < \|\mathbf{r}_B - \mathbf{r}_j\|^2, \ \forall \mathbf{r}_j \in R^{\left(\frac{M}{2}\right)}, \ \mathbf{r}_j \neq \mathbf{r}^{(-)}, \tag{18}$$

$$\|\mathbf{r}_B - \mathbf{r}^{(+)}\|^2 = \|\mathbf{r}_B - \mathbf{r}^{(-)}\|^2 . \tag{19}$$

Based on these necessary and sufficient conditions, a simple algorithm can be used to select the set of all the L Gabriel neighbour pairs $\{\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)}\}_{l=1}^{L}$, as in the binary case [15],[20]. For the completeness, the algorithm is summarized:

$L = 0$;
FOR $\mathbf{r}_q^{(+)} \in R^{\left(\frac{M}{2}+1\right)}$
  FOR $\mathbf{r}_j^{(-)} \in R^{\left(\frac{M}{2}\right)}$
    $\mathbf{x} = \frac{\mathbf{r}_q^{(+)} + \mathbf{r}_j^{(-)}}{2}$; $\eta = \|\mathbf{r}_q^{(+)} - \mathbf{x}\|^2$;
    IF $\left(\|\mathbf{r}_l^{(+)} - \mathbf{x}\|^2 > \eta, \forall \mathbf{r}_l^{(+)} \in R^{\left(\frac{M}{2}+1\right)}, l \neq q\right)$ AND
     $\left(\|\mathbf{r}_l^{(-)} - \mathbf{x}\|^2 > \eta, \forall \mathbf{r}_l^{(-)} \in R^{\left(\frac{M}{2}\right)}, l \neq j\right)$
      $L \mathrel{+}= 1$;
      $R_{\text{Gabriel}} \leftarrow (\mathbf{r}_L^{(+)}, \mathbf{r}_L^{(-)}) \triangleq (\mathbf{r}_q^{(+)}, \mathbf{r}_j^{(-)})$;
    END IF
  NEXT $\mathbf{r}_j^{(-)}$
NEXT $\mathbf{r}_q^{(+)}$

The number of Gabriel neighbour pairs L depends on the CIR and the size of the symbol constellation, and is automatically determined in the above algorithm.

A useful property regarding the distribution of a subset $R^{(i)}$ should be emphasized. Due to the symmetric distribution of the symbol constellation $\mathcal{S}$, defined in (2), the states of $R^{(i)}$ are distributed *symmetrically* around the mass center $\mathbf{m}_i$ of $R^{(i)}$. In particular, if a point $\mathbf{r}_j \in R^{(i)}$ has a distance x to the decision boundary $\mathcal{B}_{i-1}$, then there is another point $\mathbf{r}_l \in R^{(i)}$ with the same distance to the other decision boundary $\mathcal{B}_i$. This symmetric distribution property together with the shifting property are illustrated in Fig. 1.

B. SER of the Bayesian DFE with M-PAM symbols

Although there exists no close-form expression for the SER of the Bayesian DFE with M-PAM symbols, the calculation of the theoretic lower-bound SER for the Bayesian DFE

$$P_E \stackrel{\triangle}{=} \text{Prob}\{\hat{s}(k-d) \neq s(k-d)\} \tag{20}$$

can be simplified by utilizing the above-mentioned properties. Consider the conditional error probability given $s(k-d) = s_i$ with $1 < i < M$. Denote this conditional error probability as $P_{E|s_i}$. The decision region $\mathcal{D}_i$ for $\hat{s}(k-d) = s_i$ is defined by the two decision boundaries $\mathcal{B}_{i-1}$ and $\mathcal{B}_i$. Error occurs when $\mathbf{r}(k) \notin \mathcal{D}_i$, that is, when the noise makes the observation either crossing over $\mathcal{B}_{i-1}$ or over $\mathcal{B}_i$. Because of the symmetric distribution of $R^{(i)}$, probability for $\mathbf{r}(k)$ crossing over $\mathcal{B}_{i-1}$ is equal to that of over $\mathcal{B}_i$. Denote this "one-side" error probability as $P_{E|s_i|1}$. Then $P_{E|s_i} = 2P_{E|s_i|1}$, $1 < i < M$. The cases of $i = 1$ and M are special, as the decision regions $\mathcal{D}_1$ and $\mathcal{D}_M$ are half spaces, each defined by a single decision boundary. Thus $P_{E|s_1} = P_{E|s_1|1}$ and $P_{E|s_M} = P_{E|s_M|1}$. Since all these one-side conditional error probabilities $P_{E|s_i|1}$, $1 \leq i \leq M$, are equal, the error probability or SER of the Bayesian DFE is simply

$$P_E = \frac{2(M-1)}{M} P_{E|s_i|1} \, . \tag{21}$$

Now consider a "binary" Bayesian DFE defined on $R^{\left(\frac{M}{2}\right)}$ and $R^{\left(\frac{M}{2}+1\right)}$ with the decision function given by

$$f_b(\mathbf{r}(k)) = \sum_{\mathbf{r}_j \in R^{\left(\frac{M}{2}+1\right)}} \exp\left(-\frac{\|\mathbf{r}(k)-\mathbf{r}_j\|^2}{2\sigma_n^2}\right) - \sum_{\mathbf{r}_l \in R^{\left(\frac{M}{2}\right)}} \exp\left(-\frac{\|\mathbf{r}(k)-\mathbf{r}_l\|^2}{2\sigma_n^2}\right) \tag{22}$$

and the decision rule defined by

$$\hat{s}(k-d) = \begin{cases} 1, & \text{sgn}(f_b(\mathbf{r}(k))) \geq 0 \, , \\ -1, & \text{sgn}(f_b(\mathbf{r}(k))) < 0 \, . \end{cases} \tag{23}$$

But the error probability of this "binary" Bayesian DFE, denoted as P_e, is equal to $P_{E|s_i|1}$. Thus the SER of the Bayesian DFE for M-PAM symbols is the scaled error probability of the equivalent binary Bayesian DFE, defined in eqns. (22) and (23), by a factor $\frac{2(M-1)}{M}$.

Remark 1: Strictly speaking, P_E equals to $\frac{2(M-1)}{M} P_e$ only at the asymptotic case. If the SNR is not sufficiently large, the decision boundary defined by $\{\mathbf{r} : f_b(\mathbf{r}) = 0\}$ may non longer coincide with $\mathcal{B}_{\frac{M}{2}}$.

More fundamentally, the realization of the optimal decision boundary $\mathcal{B}_i$ by the multiple-hyperplane decision boundary as described in lemma 2 is accurate only for sufficiently large SNR.

III. IS simulation for the Bayesian DFE with M-PAM symbols

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 the two neighbouring subsets $R^{\left(\frac{M}{2}\right)}$ and $R^{\left(\frac{M}{2}+1\right)}$. The IS simulation technique [16],[17] can readily be used to evaluate P_e as follows:

$$\hat{P}_e = \frac{1}{N_s}\frac{1}{N_k}\sum_{j=1}^{N_s}\sum_{k=1}^{N_k} I_E(\mathbf{r}_j(k))\frac{p(\mathbf{r}_j(k)|\mathbf{r}_j)}{p^*(\mathbf{r}_j(k)|\mathbf{r}_j)} \tag{24}$$

where the error indicator function $I_E(\mathbf{r}(k)) = 1$ if $\mathbf{r}(k)$ causes an error, and $I_E(\mathbf{r}(k)) = 0$ otherwise; $p(\mathbf{r}_j(k)|\mathbf{r}_j)$ is the true conditional density given $\mathbf{r}_j \in R^{\left(\frac{M}{2}+1\right)}$, which is Gaussian with mean $\mathbf{r}_j$ and covariance $\sigma_n^2\mathbf{I}_m$, $\mathbf{I}_m$ is the $m \times m$ identity matrix, and $N_s = M^d = N_f/M$ is the number of states in $R^{\left(\frac{M}{2}+1\right)}$; N_k is the number of samples used for each signal pattern $\mathbf{r}_j$, and the sample $\mathbf{r}_j(k)$ is generated using the simulation density $p^*(\mathbf{r}_j(k)|\mathbf{r}_j)$ chosen to be

$$p^*(\mathbf{r}_j(k)|\mathbf{r}_j) = \sum_{l=1}^{L_j} p_{l,j}\frac{1}{(2\pi\sigma_n^2)^{\frac{m}{2}}}\exp\left(-\frac{\|\mathbf{r}_j(k)-\mathbf{v}_{l,j}\|^2}{2\sigma_n^2}\right). \tag{25}$$

In the simulation density (25), L_j is the number of the bias vectors $\mathbf{c}_{l,j} = -\mathbf{r}_j + \mathbf{v}_{l,j}$ for $\mathbf{r}_j \in R^{\left(\frac{M}{2}+1\right)}$, $p_{l,j} \geq 0$ for $1 \leq l \leq L_j$, and $\sum_{l=1}^{L_j} p_{l,j} = 1$. 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 [11],[15]:

$$\Gamma = \frac{\hat{P}_e(1-\hat{P}_e)}{\hat{\eta}-\hat{P}_e^2} \tag{26}$$

with

$$\hat{\eta} = \frac{1}{N_s}\frac{1}{N_k}\sum_{j=1}^{N_s}\sum_{k=1}^{N_k} I_E(\mathbf{r}_j(k))\left(\frac{p(\mathbf{r}_j(k)|\mathbf{r}_j)}{p^*(\mathbf{r}_j(k)|\mathbf{r}_j)}\right)^2. \tag{27}$$

The IS simulated P_E is simply

$$\hat{P}_E = \frac{2(M-1)}{M}\hat{P}_e\,. \tag{28}$$

The estimated IS gain for $\hat{P}_e$ will be used as the estimated IS gain for $\hat{P}_E$.

A. The construction of the IS simulation density

To achieve AE, the bias vectors $\{\mathbf{c}_{l,j}\}$ must meet certain conditions [11]. A design procedure is presented for constructing the simulation density $p^*(\mathbf{r}_j(k)|\mathbf{r}_j)$ to meet these conditions. Let $R_{\text{Gabriel}} = \{\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)}\}_{l=1}^{L}$ be the set of Gabriel neighbour pairs selected from $R^{\left(\frac{M}{2}\right)}$ and $R^{\left(\frac{M}{2}+1\right)}$. Each Gabriel neighbour pair $(\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)})$ defines a hyperplane $H_l(\mathbf{r}) = \mathbf{w}_l^T \mathbf{r} + b_l = 0$ that is part of the asymptotic decision boundary $\mathcal{B}_{\frac{M}{2}}$. The weight vector $\mathbf{w}_l$ and bias b_l of the hyperplane are given by:

$$\mathbf{w}_l = \frac{2\left(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\right)}{\|\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\|^2}, \quad b_l = -\frac{(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)})^T(\mathbf{r}_l^{(+)} + \mathbf{r}_l^{(-)})}{\|\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\|^2}. \tag{29}$$

Notice that the theory of support vector machines [21], [22] has been applied to determine H_l with ($\mathbf{r}_l^{(+)}$, $\mathbf{r}_l^{(-)}$) as its two support vectors, and H_l is a *canonical* hyperplane having the property $H_l(\mathbf{r}_l^{(+)}) = 1$ and $H_l(\mathbf{r}_l^{(-)}) = -1$. The following two definitions are useful in the construction of the simulation density.

Definition 2: A state $\mathbf{r}_j^{(-)} \in R^{\left(\frac{M}{2}\right)}$ is said to be *sufficiently separable* by the hyperplane H_l if $\mathbf{w}_l^T \mathbf{r}_j^{(-)} + b_l \le -1$. Similarly, a state $\mathbf{r}_j^{(+)} \in R^{\left(\frac{M}{2}+1\right)}$ is said to be *sufficiently separable* by H_l if $\mathbf{w}_l^T \mathbf{r}_j^{(+)} + b_l \ge 1$.

Definition 3: The hyperplane H_l is *reachable* from $\mathbf{r}_j^{(+)} \in R^{\left(\frac{M}{2}+1\right)}$ if the projection of $\mathbf{r}_j^{(+)}$ onto H_l is on the asymptotic decision boundary $\mathcal{B}_{\frac{M}{2}}$.

For each $\mathbf{r}_j^{(-)} \in R^{\left(\frac{M}{2}\right)}$, its separability index for H_l is $\alpha_{l,j}^{(-)} = 1$ if $\mathbf{r}_j^{(-)}$ is sufficiently separable by H_l; otherwise $\alpha_{l,j}^{(-)} = 0$. Similarly, $\alpha_{l,j}^{(+)} = 1$ if $\mathbf{r}_j^{(+)} \in R^{\left(\frac{M}{2}+1\right)}$ is sufficiently separable by H_l, and $\alpha_{l,j}^{(+)} = 0$ otherwise. The reachability of H_l from $\mathbf{r}_j^{(+)} \in R^{\left(\frac{M}{2}+1\right)}$ can be tested by computing

$$\mathbf{c}_{l,j} = -0.5\left(\mathbf{w}_l^T \mathbf{r}_j^{(+)} + b_l\right)\left(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\right). \tag{30}$$

If $\mathbf{v}_{l,j} = \mathbf{r}_j^{(+)} + \mathbf{c}_{l,j} \in \mathcal{B}_{\frac{M}{2}}$ (i.e. $f_b(\mathbf{v}_{l,j}) = 0$), H_l is reachable from $\mathbf{r}_j^{(+)}$ ($\mathbf{c}_{l,j}$ is then a bias vector), and the reachability index is $\gamma_{l,j} = 1$; otherwise $\gamma_{l,j} = 0$. The whole 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	$\alpha_{1,1}^{(-)}$	$\cdots$	$\alpha_{1,N_s}^{(-)}$	$\alpha_{1,1}^{(+)}$ $(\gamma_{1,1})$	$\cdots$	$\alpha_{1,N_s}^{(+)}$ (γ_{1,N_s})
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
H_L	$\alpha_{L,1}^{(-)}$	$\cdots$	$\alpha_{L,N_s}^{(-)}$	$\alpha_{L,1}^{(+)}$ $(\gamma_{L,1})$	$\cdots$	$\alpha_{L,N_s}^{(+)}$ (γ_{L,N_s})

In order to construct a convex region $\mathcal{R}_j^{(+)}$ covering a $\mathbf{r}_j^{(+)} \in R^{\left(\frac{M}{2}+1\right)}$, first select those hyperplanes that can sufficiently separate $\mathbf{r}_j^{(+)}$ and that are reachable from $\mathbf{r}_j^{(+)}$ with the aid of the above table. This yields the following integer set:

$$G_j^{(+)} \triangleq \{q : \ \alpha_{q,j}^{(+)} = 1 \text{ and } \gamma_{q,j} = 1\}. \tag{31}$$

Then $\mathcal{R}_j^{(+)}$ is the intersection of all the half-spaces $\mathcal{H}_q^{(+)} \triangleq \{\mathbf{r} : H_q(\mathbf{r}) \ge 0\}$ with $q \in G_j^{(+)}$. In fact, it is not necessary to use every hyperplanes defined in $G_j^{(+)}$ to construct $\mathcal{R}_j^{(+)}$. A subset of these hyperplanes will be sufficient, provided that every opposite-class state in $R^{\left(\frac{M}{2}\right)}$ can sufficiently be separated by at least one hyperplane in the subset. If this can be done, the error region $\mathcal{E}$ satisfies

$$\mathcal{E} \subset \overline{\mathcal{R}_j^{(+)}} \triangleq \bigcup_{q \in G_j^{(+)}} \mathcal{H}_q^{(-)} \tag{32}$$

with the half-spaces $\mathcal{H}_q^{(-)} \triangleq \{\mathbf{r} : H_q(\mathbf{r}) < 0\}$. Obviously, all the hyperplanes defined in $G_j^{(+)}$ are reachable from $\mathbf{r}_j^{(+)}$ and at least one of $\{\mathbf{v}_{q,j}\}$ is the minimum rate point (as defined in [11]). Notice that these are sufficient conditions that a set of bias vectors (related to a $\mathbf{r}_j^{(+)}$) must be met to achieve AE [11]. If such a $G_j^{(+)}$ exists for each $\mathbf{r}_j^{(+)} \in R^{\left(\frac{M}{2}+1\right)}$, the simulation density constructed with the bias vectors $\{\mathbf{c}_{q,j}\}$, $q \in G_j^{(+)}$, for all j will guarantee AE.

An illustrative example with $M = 2$ and $n_h = 2$ is depicted in Fig. 2. In this example, $R^{(1)} = \{\mathbf{r}_1^{(-)}, \mathbf{r}_2^{(-)}\}$ and $R^{(2)} = \{\mathbf{r}_1^{(+)}, \mathbf{r}_2^{(+)}\}$. It is obvious that $(\mathbf{r}_1^{(+)}, \mathbf{r}_1^{(-)})$ is a Gabriel neighbour pair, as all the other states satisfy the condition (14) given $\mathbf{r}_0 = (\mathbf{r}_1^{(+)} + \mathbf{r}_1^{(-)})/2$. Similarly $(\mathbf{r}_1^{(+)}, \mathbf{r}_2^{(-)})$ and $(\mathbf{r}_2^{(+)}, \mathbf{r}_2^{(-)})$ are Gabriel neighbour pairs. Thus the asymptotic decision boundary is formed from the corresponding three hyperplanes H_1, H_2 and H_3. Obviously, $\mathbf{r}_1^{(-)}$ is sufficiently separable by H_1 while $\mathbf{r}_2^{(-)}$ is not, as $H_1(\mathbf{r}_1^{(-)}) = -1$ and $H_1(\mathbf{r}_2^{(-)}) > 0$. Both $\mathbf{r}_1^{(+)}$ and $\mathbf{r}_2^{(+)}$ are sufficiently separable by H_1, as $H_1(\mathbf{r}_1^{(+)}) = 1$ and $H_1(\mathbf{r}_2^{(+)}) > 1$, but H_1 is not reachable from $\mathbf{r}_2^{(+)}$, as the projection of $\mathbf{r}_2^{(+)}$ onto H_1 is not at the asymptotic decision boundary. Continuing this process for the other two hyperplanes leads to the separability and reachability table given in Table I. As $\mathbf{r}_1^{(+)}$ is sufficiently separated from the opposite-class states by the two reachable hyperplanes H_1 and H_2, there are two bias vectors $\mathbf{c}_{1,1}$ and $\mathbf{c}_{2,1}$ for $\mathbf{r}_1^{(+)}$ with $\mathbf{v}_{2,1}$ being the minimum rate point, and the error region is covered by the half space formed from H_1 and H_2. Since $\mathbf{r}_2^{(+)}$ is sufficiently separable by a single reachable hyperplane H_3, there is one bias vector $\mathbf{c}_{1,2}$ for $\mathbf{r}_2^{(+)}$ with $\mathbf{v}_{1,2}$ being the minimum rate point, and the error region is covered by the half space formed from H_3. For this example, the constructed IS simulation density achieves AE.

Remark 2: The construction procedure for the IS simulation density discussed above, if can be done, will guarantee the AE of IS simulation. Strictly speaking, however, AE can only be guaranteed at the asymptotic case. As pointed out in Remark 1, if the SNR is too small, the multiple-hyperplane decision boundary may deviate from the true optimal decision boundary $\mathcal{B}_{\frac{M}{2}}$. Shifting the density to the asymptotic decision boundary is then not "optimal". This is the main source for a relatively small IS gain when SNR is small, as can be observed in the simulation results.

Remark 3: Since it is assumed that correct decisions are fed back, the IS simulation procedure considered here provides a *lower-bound* SER for the Bayesian DFE. In practice, it is more useful to provide some *upper-bound* SER and to take into account error propagation caused by incorrect decisions being fed back. However, due to its highly complicated structure, derivation of an upper-bound SER for the Bayesian DFE will be extremely difficult if not impossible. In the lack of any upper-bound, the lower-bound SER is the only means that can be used to evaluate potential performance of the Bayesian DFE.

B. Numerical examples

Example 1. The IS technique was simulated for the Bayesian DFE with 4-PAM symbols using the 3-tap CIR defined by $\mathbf{h} = [0.4\ 1.0\ 0.6]^T$. The DFE structure was specified by $m = 3$, $d = 2$ and $n_b = 2$. The channel state set R had $N_f = 64$ states. Five pairs of Gabriel neighbour states were found from the subsets $R^{(2)}$ and $R^{(3)}$, giving rise to 5 separating hyperplanes. The separability and reachability table for this example is listed in Table II, from which the required bias vectors were generated. For this example, it is straightforward to verify that the constructed simulation density achieves AE. An inspection of Table II shows that the states $\mathbf{r}_1^{(+)}$ to $\mathbf{r}_5^{(+)}$ in $R^{(3)}$ are sufficiently separated from the opposite-class $R^{(2)}$ by the two reachable hyperplanes H_1 and H_2, $\mathbf{r}_6^{(+)}$ to $\mathbf{r}_9^{(+)}$ are sufficiently separated from $R^{(2)}$ by the two reachable hyperplanes H_3 and H_4, and $\mathbf{r}_{10}^{(+)}$ to $\mathbf{r}_{16}^{(+)}$ are sufficiently separated from $R^{(2)}$ by the hyperplane H_5.

As in [15], the bias vectors were selected with uniform probability in the simulation with $p_{l,j} = \frac{1}{L_j}$ for $1 \leq l \leq L_j$, that is, no attempt was made to optimize the probabilities $p_{l,j}$ in (25). For each SNR, 10^5 iterations were employed, averaging over all the possible states in $\mathcal{R}^{(3)}$. Thus, the total samples used for a given SNR were 1.6×10^6. Fig. 3 (a) shows the lower-bound SERs obtained using the IS and conventional sampling (CS) simulation methods, respectively. It can be seen that the conventional Monte Carlo results for low SNR conditions based directly on the Bayesian DFE of (11) and (12) agreed with those of the IS simulation. The estimated IS gains, depicted in Fig. 3 (b), indicate that exponential IS gains were obtained with increasing SNRs. It can be seen that for small SNR conditions, the IS gain is relatively small for the reason given in Remark 2. For example, given SNR=25 dB, the IS gain was a modest value of $\Gamma = 3.5$. It should be emphasized that a IS simulation is really needed at very low SER or high SNR situations. Under such conditions, the proposed IS simulation technique is extremely efficient. For example, given SNR=30 dB, the SER of the Bayesian DFE with correct symbols being fed back evaluated by the IS technique was approximately 10^{-9} with an estimated IS gain of $\Gamma = 1,326,783$. The CS method could not work under the same SNR condition and it would require approximately 2.1×10^{12} samples to achieve a similar estimation variance.

Example 2. A 2-tap channel $\mathbf{h} = [0.3\ 1.0]^T$ with 8-PAM symbols was simulated, and the Bayesian

DFE structure was defined by $m = 2$, $d = 1$ and $n_b = 1$. The channel state set R had $N_f = 64$ states. Nine pairs of Gabriel neighbour states were selected from the subsets $R^{(4)}$ and $R^{(5)}$, and Table III lists the separability and reachability table for this example. It is straightforward to verify that the constructed simulation density achieves AE. The state $\mathbf{r}_1^{(+)}$ is sufficiently separable from the opposite-class $R^{(4)}$ by the two reachable hyperplanes H_1 and H_2, $\mathbf{r}_2^{(+)}$ is sufficiently separable from $R^{(4)}$ by the two reachable hyperplanes H_3 and H_4, $\mathbf{r}_3^{(+)}$ is separable by the two reachable hyperplanes H_5 and H_6, $\mathbf{r}_4^{(+)}$ is separable by the two reachable hyperplanes H_7 and H_8, and $\mathbf{r}_5^{(+)}$ to $\mathbf{r}_8^{(+)}$ are separable from $R^{(4)}$ by the single reachable hyperplane H_9. As this is a two-dimensional example, the graphic illustration of the simulation density construction can be made and is shown in Fig. 4. Notice the difference between the true optimal decision boundary $\mathcal{B}_4$ under a low SNR condition and the asymptotic decision boundary. This explains the relatively small IS gain in the simulation for low SNR conditions.

Again the bias vectors were selected with uniform probability in the simulation. For each SNR, 10^5 samples were used for each state in $\mathcal{R}^{(5)}$, resulting in a total of 8×10^5 samples for a given SNR. Fig. 5 (a) depicts the lower-bound SERs obtained using the IS and CS simulation methods, respectively. Again, the conventional Monte Carlo results for low SNR conditions agreed with those of the IS simulation. It can be seen from Fig. 5 (b) that exponential IS gains were obtained with increasing SNRs.

IV. Conclusions

A randomized bias technique for 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 and consists of several hyperplanes. Furthermore, it has been shown that asymptotically the SER of the Bayesian DFE for the M-PAM symbol constellation is a scaled error rate of the equivalent binary Bayesian DFE evaluated on any two neighbouring signal subsets. Although asymptotic efficiency of the proposed IS simulation method for the general channel has not rigorously been proved, a design procedure has been presented for constructing the simulation density that meets the asymptotic efficiency conditions. The SER evaluated is under the assumption that correct symbols are fed back, and error propagation is not taken into account. Nevertheless, the method provides a practical means of evaluating the potential performance for the Bayesian DFE under high SNR conditions.

References

[1] S.U.H. Qureshi, "Adaptive equalization," *Proc. IEEE*, Vol.73, No.9, pp.1349–1387, 1985.
[2] J.G. Proakis, *Digital Communications*. 3rd edition, New York: McGraw-Hill, 1995.
[3] J. Moon, "The role of SP in data-storage systems," *IEEE Signal Processing Magazine*, Vol.15, No.4, pp.54–72, 1998.
[4] J.G. Proakis, "Equalization techniques for high-density magnetic recording," *IEEE Signal Processing Magazine*, Vol.15, No.4, pp.73–82, 1998.
[5] K. Abend and B.D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, Vol.58, No.5, pp.779–785, 1970.
[6] D. Williamson, R.A. Kennedy and G.W. Pulford, "Block decision feedback equalization," *IEEE Trans. Communications*, Vol.40, No.2, pp.255-264, 1992.
[7] 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.
[8] S. Chen, S. McLaughlin, B. Mulgrew and P.M. Grant, "Bayesian decision feedback equaliser for overcoming co-channel interference," *IEE Proc. Communications*, Vol.143, No.4, pp.219–225, 1996.
[9] P.J. Smith, M. Shafi and H. Gao, "Quick simulation: a review of importance sampling techniques in communications systems," *IEEE J. Selected Areas in Communications*, Vol.15, No.4, pp.597–613, 1997.
[10] D. Lu and K. Yao,"Improved importance sampling techniques for efficient simulation of digital communication systems," *IEEE J. Selected Areas in Communications*, Vol.6, No.1, pp.67–75, 1988.
[11] 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.
[12] G.C. Orsak and B. Aazhang, "Efficient importance sampling techniques for simulation of multiuser communication systems," *IEEE Trans. Communications*, Vol.40, No.6, pp.1111-1118, 1992.
[13] H.J. Schlebusch, "On the asymptotic efficiency of importance sampling techniques," *IEEE Trans. Information Theory*, Vol.39, pp.710–715, 1993.
[14] J.C. Chen, D. Lu, J.S. Sadowsky and K. Yao, "On importance sampling in digital communications — Part I: fundamentals," *IEEE J. Selected Areas in Communications*, Vol.11, No.3, pp.289–299, 1993.
[15] 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.
[16] S. Chen, "Importance sampling simulation for evaluating the lower-bound BER of the Bayesian DFE," *IEEE Trans. Communications*, to appear, 2002.
[17] S. Chen and L. Hanzo, "An importance sampling simulation method for Bayesian decision feedback equalizers," in *Proc. 5th Int. Conf. Mathematics in Signal Processing* (University of Warwick, U.K.), Dec.18-20, 2000, 4 pages.
[18] S. Chen, L. Hanzo and B. Mulgrew, "Decision feedback equalization using multiple-hyperplane partitioning for detecting ISI-corrupted M-ary PAM signals," *IEEE Trans. Communications*, Vol.49, No.5, pp.760–764, 2001.
[19] 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.
[20] S. Chen, B. Mulgrew and L. Hanzo, "Asymptotic Bayesian decision feedback equalizer using a set of hyperplanes," *IEEE Trans. Signal Processing*, Vol.48, No.12, pp.3493–3500, 2000.
[21] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[22] S. Chen, S. Gunn and C.J. Harris, "Decision feedback equalizer design using support vector machines," *IEE Proc. Vision, Image and Signal Processing*, Vol.147, No.3, pp.213–219, 2000.

Sheng Chen (SM'97) received the Ph.D. degree in control engineering from the City University, London, U.K., in 1986. From 1986 to 1999, he held various research and academic appointments at the Universities of Sheffield, Edinburgh, and Portsmouth, U.K.. Since 1999, he has been with the Department of Electronics and Computer Science, the University of Southampton, U.K., where he currently holds a post of Readership in Communications Signal Processing. His research interests include modeling and identification of nonlinear systems, adaptive nonlinear signal processing, artificial neural network research, finite-precision digital controller design, evolutionary computation methods, and optimization. He has published over 170 research papers.

TABLE I

SEPARABILITY AND REACHABILITY TABLE FOR THE EXAMPLE GIVEN IN FIG. 2.

hyperplane	$\mathbf{r}_1^{(-)}$	$\mathbf{r}_2^{(-)}$	$\mathbf{r}_1^{(+)}$	$\mathbf{r}_2^{(+)}$
H_1	1	0	1 (1)	1 (0)
H_2	0	1	1 (1)	0
H_3	1	1	0	1 (1)

TABLE II

SEPARABILITY AND REACHABILITY TABLE FOR THE CHANNEL $\mathbf{h} = [0.4\ 1.0\ 0.6]^T$ WITH 4-PAM SYMBOLS

	H_1	H_2	H_3	H_4	H_5
	1	0	1	0	1
	1	0	1	0	1
	1	0	1	0	1
	1	1	1	0	1
	1	0	1	0	1
	1	0	1	0	1
	1	1	1	0	1
$R^{(2)}$	1	1	1	1	1
	0	1	1	0	1
	0	1	1	0	1
	0	1	1	1	1
	0	1	1	1	1
	0	1	0	1	1
	0	1	0	1	1
	0	1	0	1	1
	0	1	0	1	1
	1 (1)	1 (1)	0	1 (1)	0
	1 (1)	1 (1)	0	1 (1)	0
	1 (1)	1 (1)	0	1 (1)	0
	1 (1)	1 (1)	0	1 (0)	0
	1 (1)	1 (1)	1 (1)	1 (1)	0
	1 (0)	1 (1)	1 (1)	1 (1)	0
	1 (0)	0	1 (1)	1 (1)	0
$R^{(3)}$	1 (0)	0	1 (1)	1 (1)	0
	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
	1 (1)	0	1 (0)	1 (1)	1 (1)
	1 (1)	0	1 (0)	0	1 (1)
	1 (1)	0	1 (0)	0	1 (1)
	1 (1)	0	1 (1)	1 (1)	1 (1)
	1 (1)	0	1 (1)	0	1 (1)
	1 (1)	0	1 (1)	0	1 (1)
	1 (1)	0	1 (1)	0	1 (1)

TABLE III

Separability and reachability table for the channel $\mathbf{h} = [0.3\ 1.0]^T$ with 8-PAM symbols

	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9
$R^{(4)}$	1	0	1	0	1	0	1	0	1
	1	0	1	0	1	0	1	0	1
	1	0	1	0	1	0	1	0	1
	1	0	1	0	1	0	1	0	1
	0	1	1	0	1	0	1	0	1
	0	1	0	1	1	0	1	0	1
	0	1	0	1	0	1	1	0	1
	0	1	0	1	0	1	0	1	1
$R^{(5)}$	1 (1)	1 (1)	0	1 (0)	0	1 (1)	0	1 (1)	0
	1 (1)	0	1 (1)	1 (1)	0	1 (0)	0	1 (1)	0
	1 (0)	0	1 (1)	0	1 (1)	1 (1)	0	1 (0)	0
	1 (1)	0	1 (0)	0	1 (1)	0	1 (1)	1 (1)	0
	1 (1)	0	1 (1)	0	1 (0)	0	1 (1)	0	1 (1)
	1 (1)	0	1 (1)	0	1 (1)	0	1 (1)	0	1 (1)
	1 (1)	0	1 (1)	0	1 (1)	0	1 (1)	0	1 (1)
	1 (1)	0	1 (1)	0	1 (1)	0	1 (1)	0	1 (1)

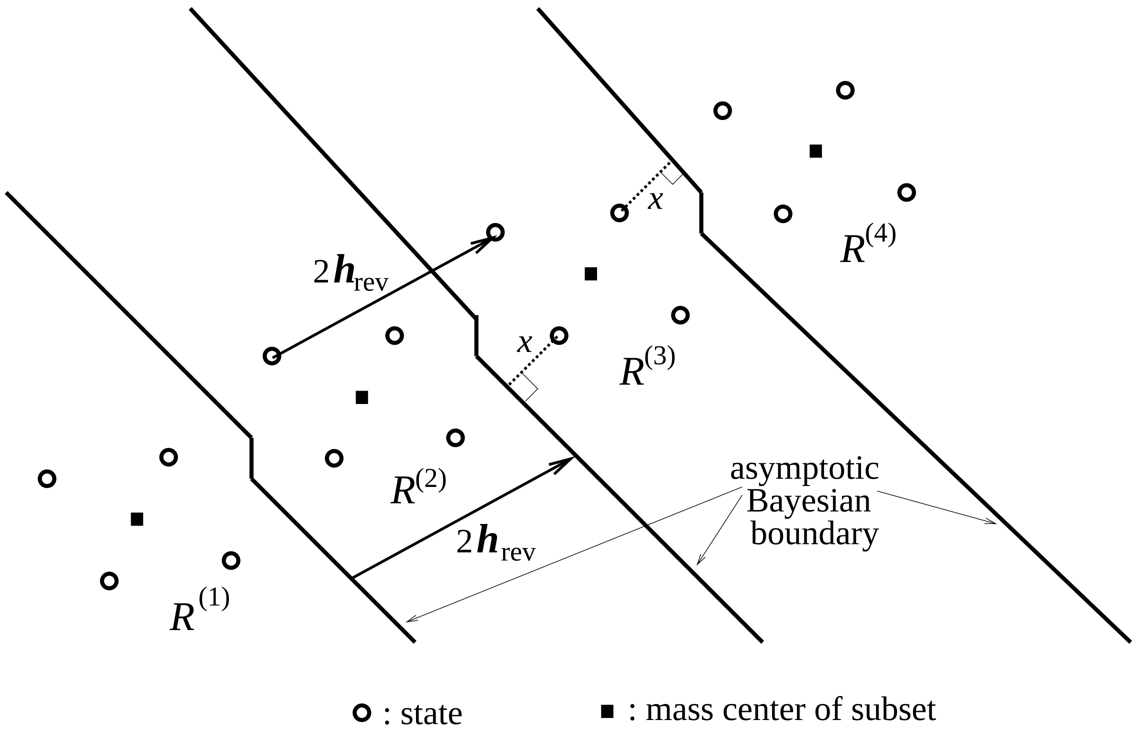


Fig. 1. Illustration of symmetric and shifting properties of subset states.

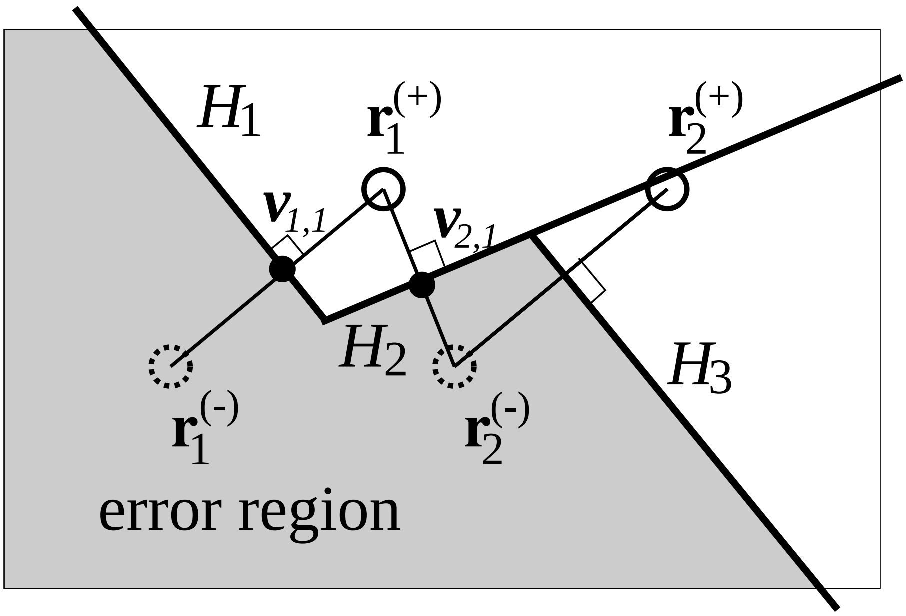
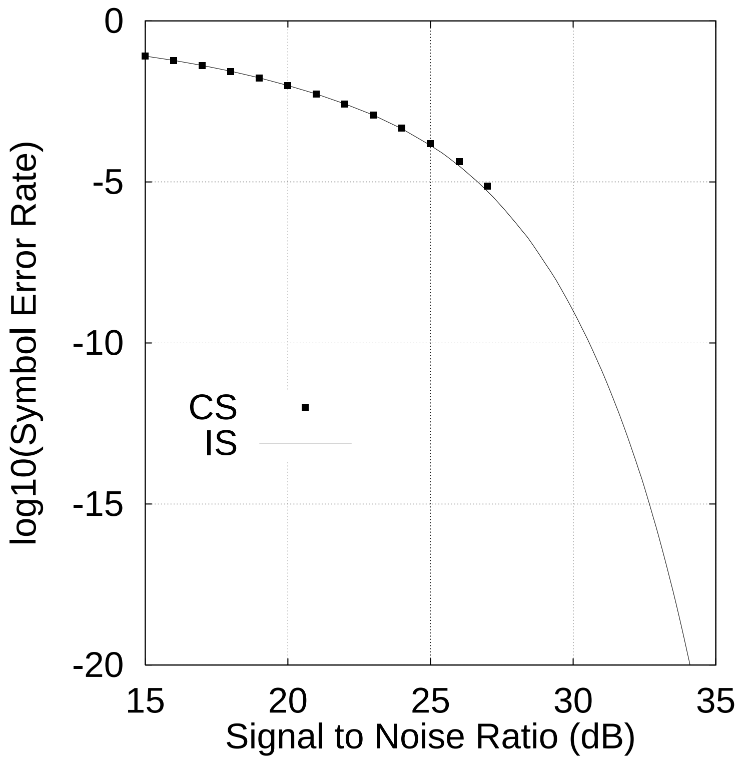
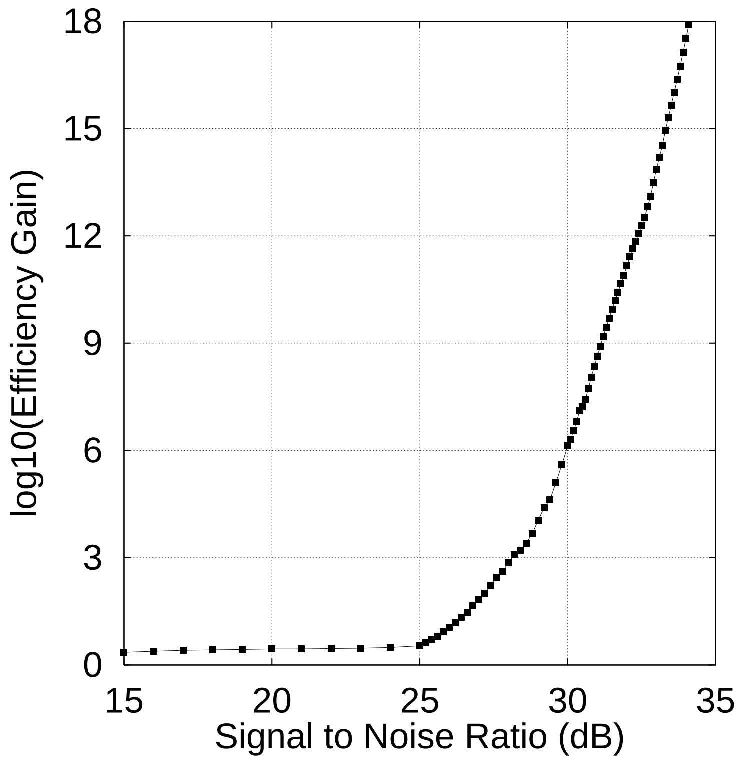


Fig. 2. An illustration of the simulation density construction for the case of binary ($M = 2$) symbols with a two-tap channel. In this example, there are three Gabriel neighbour pairs $(\mathbf{r}_1^{(+)}, \mathbf{r}_1^{(-)})$, $(\mathbf{r}_1^{(+)}, \mathbf{r}_2^{(-)})$ and $(\mathbf{r}_2^{(+)}, \mathbf{r}_2^{(-)})$. The asymptotic decision boundary is formed from the three corresponding hyperplanes H_1, H_2 and H_3. The separability and reachability table for this example is given in Table I.



(a)



(b)

Fig. 3. The lower-bound SERs (a) and the estimated IS gain (b) of the Bayesian DFE for the CIR $\mathbf{h} = [0.4\ 1.0\ 0.6]^T$ with 4-PAM symbols, using conventional sampling (CS) and importance sampling (IS) simulation.

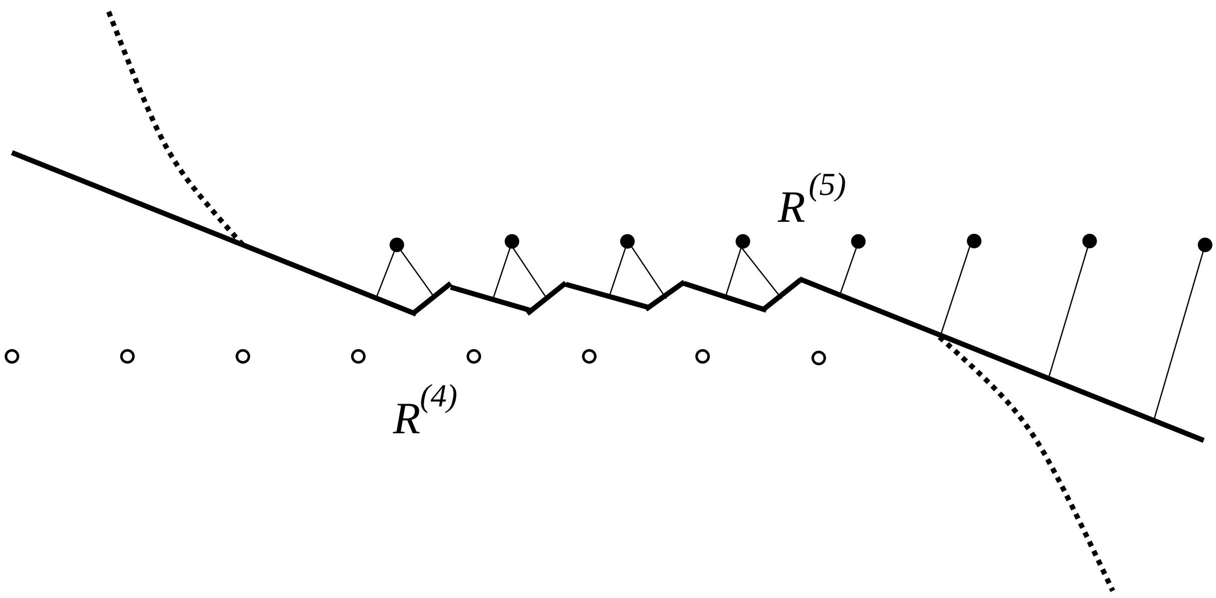
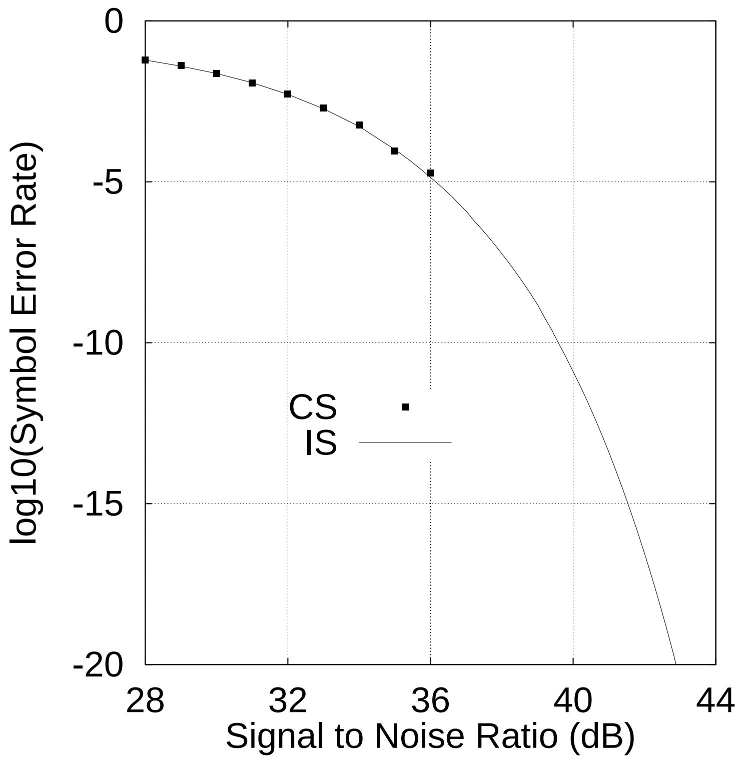
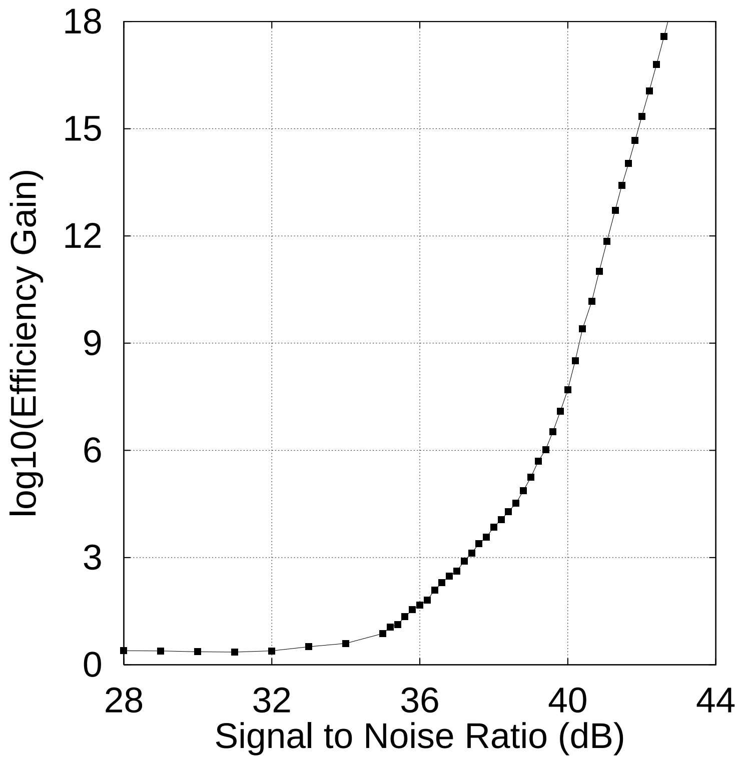


Fig. 4. Illustration of the simulation density construction for the channel $\mathbf{h} = [0.3\ 1.0]^T$ with 8-PAM symbols. Thick solid curve indicates the asymptotic decision boundary, thick dashed curve the true optimal decision boundary for small SNR, and thin lines indicates the bias vectors used in the simulation density.



(a)



(b)

Fig. 5. The lower-bound SERs (a) and the estimated IS gain (b) of the Bayesian DFE for the CIR $\mathbf{h} = [0.3\ 1.0]^T$ with 8-PAM symbols, using conventional sampling (CS) and importance sampling (IS) simulation.