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

Sheng Chen

Department of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ
United Kingdom
sqc@ecs.soton.ac.uk

Presented at 2nd IMA Conference on Mathematics in Communications
16-18 December 2002 Lancaster University

PAM Channel Model

- Received signal model

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

where h_i are the CIR taps with n_h the length of the CIR, the AWGN $n(k)$ has variance σ_n^2, and $s(k)$ takes the value from the symbol set

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

- The generic DFE uses the information present in

$$\mathbf{y}(k) = [y(k) \cdots y(k-m+1)]^T \quad \hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1) \cdots \hat{s}(k-d-n_b)]^T$$

to produce an estimate $\hat{s}(k-d)$ of $s(k-d)$.

Observation Vector

- Choose: the decision delay $d = n_h - 1$, feedforward order $m = n_h$ and feedback order $n_b = n_h - 1 \Rightarrow$ sufficient to guarantee a linear separability for different signal classes.

- The observation vector $\mathbf{y}(k)$ can be expressed as

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

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

$$\mathbf{H}_1 = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n_h-1} \\ 0 & h_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & h_1 \\ 0 & \cdots & 0 & h_0 \end{bmatrix} \quad \mathbf{H}_2 = \begin{bmatrix} 0 & \cdots & 0 \\ h_{n_h-1} & \ddots & \vdots \\ \vdots & \ddots & 0 \\ h_1 & \cdots & h_{n_h-1} \end{bmatrix}$$

Feedback as Space Translation

- Assume correct past decisions, we have

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

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

$$\mathbf{r}(k) = \mathbf{y}(k) - \mathbf{H}_2 \ \hat{\mathbf{s}}_b(k)$$

$\Rightarrow$ We can consider the DFE in the space $\mathbf{r}(k)$, instead of the space $\mathbf{y}(k)$.

- Define the set of the signal states in this new space as

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

which can be partitioned into M subsets conditioned on $s(k-d)$:

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

Bayesian Decision Feedback Equaliser

The Bayesian DFE involves:

- computing the M decision variables for $1 \le i \le M$

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

- making the decision according to

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

Investigating the performance of the the Bayesian DFE under the condition of SER lower than 10^{-6} is very difficult if not impossible, using a conventional Monte Carlo simulation.

Importance Sampling Simulation

- Importance sampling aims to reduce the variance of an error rate estimator $\Rightarrow$ achieve a given precision from shorter simulation runs.
- The basic idea is that certain input variable values have more impact on the error probability. If these "important" values are emphasized by sampling more frequently, the estimator variance can be reduced.
- Fundamental issue: choice of the biased distribution to encourage the important regions of the simulation input variables.
- Mean translation approach: the distribution is moved toward the error region, usually by shifting the density to a decision boundary.
- An asymptotically efficient estimator requires a number of simulation trials which grows less than exponentially fast as the error rate tends to zero $\Rightarrow$ realistic to attempt extremely low error rate simulation.

Symmetric Structure of Subset States

- *Lemma 1*: For $1 \le i \le M-1$, $R^{(i+1)}$ is a translation of $R^{(i)}$:

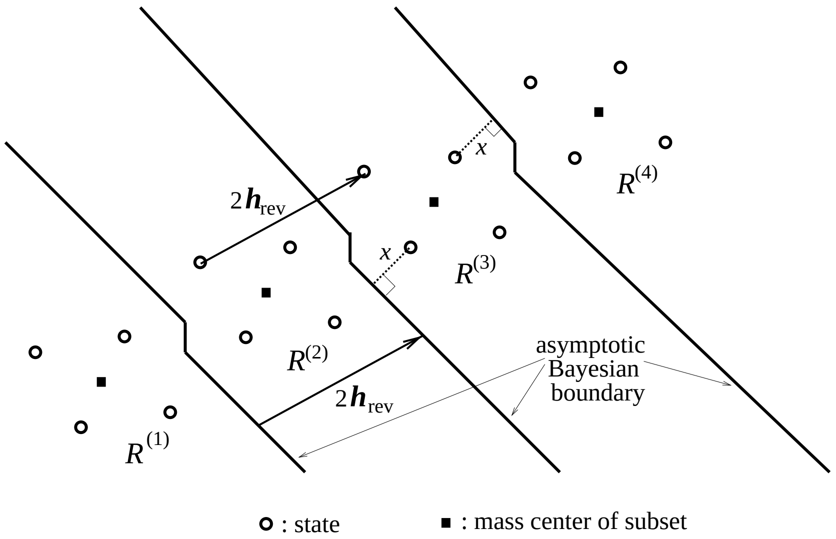
$$R^{(i+1)} = R^{(i)} + 2\mathbf{h}_{\text{rev}} \quad \text{with} \quad \mathbf{h}_{\text{rev}} = [h_{n_h-1} \cdots h_1 \ h_0]^T$$

- Asymptotically as the SNR tends to infinity, the decision boundary $\mathcal{B}_{i+1}$ for separating $R^{(i+1)}$ and $R^{(i+2)}$ is a shift of $\mathcal{B}_i$ by an amount $2\mathbf{h}_{\text{rev}}$.
- Thus, we only need to consider two neighbouring subsets, for example, $R^{(M/2)}$ and $R^{((M/2)+1)}$ (the classes of $s_{M/2} = -1$ and $s_{(M/2)+1} = 1$).
- *Definition 1*: A pair of opposite-class states $(\mathbf{r}^{(+)} \in R^{((M/2)+1)}, \mathbf{r}^{(-)} \in R^{(M/2)}$ is said to be *Gabriel neighbours* if $\forall \mathbf{r}_j \in R^{(M/2)} \bigcup R^{((M/2)+1)}$, $\mathbf{r}_j \ne \mathbf{r}^{(+)}$ and $\mathbf{r}_j \ne \mathbf{r}^{(-)}$:

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

where $\mathbf{r}_0 = (\mathbf{r}^{(+)} + \mathbf{r}^{(-)})/2$.

Illustration of Symmetric and Shifting Properties



Asymptotic Bayesian Decision Boundary

- *Lemma 2*: Asymptotically, $\mathcal{B}_{M/2}$ for separating $R^{(M/2)}$ and $R^{((M/2)+1)}$ is piecewise linear and made up of L hyperplanes. Each hyperplane is defined by a pair of Gabriel neighbours $(\mathbf{r}^{(+)}, \mathbf{r}^{(-)})$.

- A necessary condition for a point $\mathbf{r}_B \in \mathcal{B}_{M/2}$

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

  where $\mathbf{x}^{\perp}$ denotes an arbitrary vector in the subspace orthogonal to $\mathbf{x}$.

- The sufficient conditions for $\mathbf{r}_B \in \mathcal{B}_{M/2}$

$$\begin{gathered}\|\mathbf{r}_B - \mathbf{r}^{(+)}\|^2 < \|\mathbf{r}_B - \mathbf{r}_l\|^2, \ \forall \mathbf{r}_l \in R^{((M/2)+1)}, \ \mathbf{r}_l \neq \mathbf{r}^{(+)} \\ \|\mathbf{r}_B - \mathbf{r}^{(-)}\|^2 < \|\mathbf{r}_B - \mathbf{r}_j\|^2, \ \forall \mathbf{r}_j \in R^{(M/2)}, \ \mathbf{r}_j \neq \mathbf{r}^{(-)} \\ \|\mathbf{r}_B - \mathbf{r}^{(+)}\|^2 = \|\mathbf{r}_B - \mathbf{r}^{(-)}\|^2\end{gathered}$$

Identifying Gabriel Neighbour Pairs

Simple algorithm to select the set of Gabriel neighbour pairs $\{\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)}\}_{l=1}^{L}$:

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

Symbol Error Rate of M-PAM Bayesian DFE

- The aim is to estimate the lower-bound SER of the Bayesian DFE

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

- Define a "binary" Bayesian DFE on $R^{(M/2)}$ and $R^{((M/2)+1)}$:

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

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

- If the error rate of this "binary" Bayesian DFE is P_e, it can be shown:

$$P_E = \gamma P_e \quad \text{with} \quad \gamma = 2(M-1)/M$$

An IS Simulation Technique

- An IS simulation technique evaluate P_e as follows:

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

- $I_E(\mathbf{r}(k)) = 1$ 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 $\mathbf{r}_j \in R^{((M/2)+1)}$, and $N_s = N_f/M$ is the number of states in $R^{((M/2)+1)}$

- The sample $\mathbf{r}_j(k)$ is generated using the simulation density:

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

  where L_j is the number of the bias vectors $\mathbf{c}_{l,j} = -\mathbf{r}_j + \mathbf{v}_{l,j}$ for $\mathbf{r}_j \in R^{((M/2)+1)}$, $p_{l,j} \geq 0$ for $1 \leq l \leq L_j$, and $\sum_{l=1}^{L_j} p_{l,j} = 1$.

Importance Sampling Gain

- The IS gain is defined as the ratio of the numbers of trials required for the same estimate variance using the conventional Monte Carlo and IS methods.
- An estimate of the IS gain for $\hat{P}_e$ is

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

with

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

- The IS simulated P_E is simply $\hat{P}_E = \gamma \hat{P}_e$, and the estimated IS gain for $\hat{P}_e$ will be used as the estimated IS gain for $\hat{P}_E$.

Constructing IS Simulation Density

- Each pair $(\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)})$ in the set of Gabriel neighbours defines a hyperplane

$$H_l(\mathbf{r}) = \mathbf{w}_l^T \mathbf{r} + b_l = 0$$

that is part of the asymptotic decision boundary $\mathcal{B}_{M/2}$, with

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

- H_l is *canonical* with $H_l(\mathbf{r}_l^{(+)}) = 1$ and $H_l(\mathbf{r}_l^{(-)}) = -1$.
- *Definition 2*: A state $\mathbf{r}_j^{(-)} \in R^{(M/2)}$ is *sufficiently separable* by H_l if $\mathbf{w}_l^T \mathbf{r}_j^{(-)} + b_l \le -1$. Similarly, $\mathbf{r}_j^{(+)} \in R^{((M/2)+1)}$ is sufficiently separable by H_l if $\mathbf{w}_l^T \mathbf{r}_j^{(+)} + b_l \ge 1$.

- *Definition 3*: H_l is *reachable* from $\mathbf{r}_j^{(+)} \in R^{((M/2)+1)}$ if the projection of $\mathbf{r}_j^{(+)}$ onto H_l is on $\mathcal{B}_{M/2}$.
- For each $\mathbf{r}_j^{(-)} \in R^{(M/2)}$, its separability index for H_l is $\alpha_{l,j}^{(-)} = 1$ if $\mathbf{r}_j^{(-)}$ is sufficiently separable by H_l; otherwise $\alpha_{l,j}^{(-)} = 0$.

  The separability index $\alpha_{l,j}^{(+)}$ for $\mathbf{r}_j^{(+)} \in R^{((M/2)+1)}$ is similarly defined.
- The reachability of H_l from $\mathbf{r}_j^{(+)} \in R^{((M/2)+1)}$ is tested by computing

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

If $\mathbf{v}_{l,j} = \mathbf{r}_j^{(+)} + \mathbf{c}_{l,j} \in \mathcal{B}_{M/2}$ (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 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})

- For each $\mathbf{r}_j^{(+)} \in R^{((M/2)+1)}$, select those hyperplanes that can sufficiently separate $\mathbf{r}_j^{(+)}$ and are reachable from $\mathbf{r}_j^{(+)}$ or define the integer set:

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

- The convex region $\mathcal{R}_j^{(+)}$ covering $\mathbf{r}_j^{(+)}$ is the intersection of all the half-spaces $\mathcal{H}_q^{(+)} = \{\mathbf{r} : H_q(\mathbf{r}) \ge 0\}$ with $q \in G_j^{(+)}$.

  In fact, a subset of the hyperplanes defined by $G_j^{(+)}$ is sufficient, provided that every opposite-class state in $R^{(M/2)}$ can sufficiently be separated by at least one hyperplane in the subset.

Remarks on the Construction

If the above construction of the simulation density can be done, that is, $G_j^{(+)}$ exists for each $\mathbf{r}_j^{(+)} \in R^{((M/2)+1)}$, then

1. the error region $\mathcal{E}$ satisfies

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

   with the half-spaces $\mathcal{H}_q^{(-)} = \{\mathbf{r} : H_q(\mathbf{r}) < 0\}$;

2. all the hyperplanes defined in $G_j^{(+)}$ are reachable from $\mathbf{r}_j^{(+)}$; and

3. at least one of $\{\mathbf{v}_{q,j}\}$ is the so-called minimum rate point.

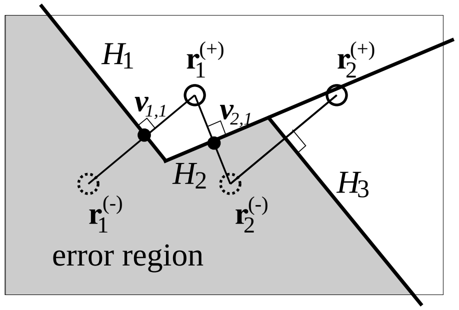
The simulation density constructed with the bias vectors $\{\mathbf{c}_{q,j}\}$, $q \in G_j^{(+)}$, for all j will guarantee asymptotic efficiency.

Illustration of Simulation Density Construction

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.

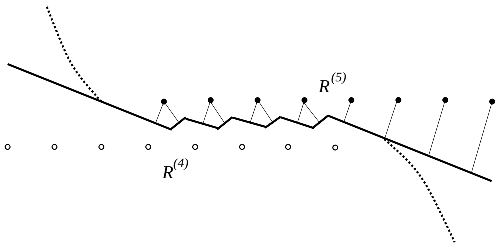
Separability and Reachability Table

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



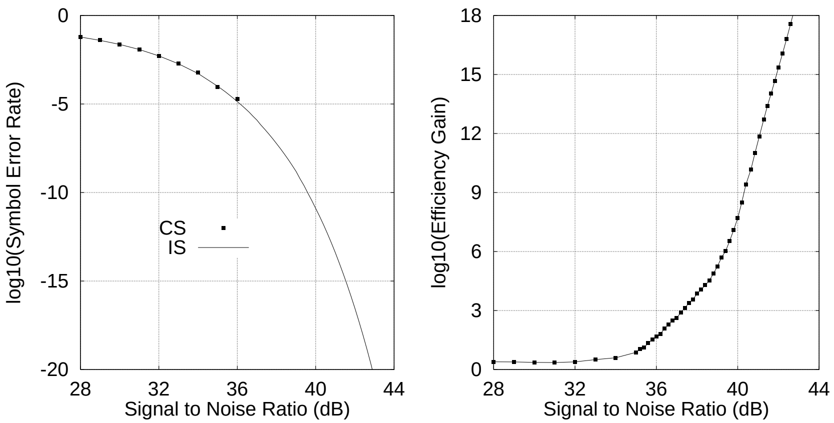
Simulation Example 1

- Channel $\mathbf{h} = [0.3\ 1.0]^T$ with 8-PAM symbols, $m = 2$, $d = 1$ and $n_b = 1$.
- R had 64 states. Nine pairs of Gabriel neighbours were selected from $R^{(4)}$ and $R^{(5)}$, and an AE simulation density was constructed:



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.

- Uniform probability for the bias vectors ($p_{l,j} = 1/L_j$). For each SNR, $N_k = 10^5$. Thus, the total samples used for a given SNR were 8×10^5.



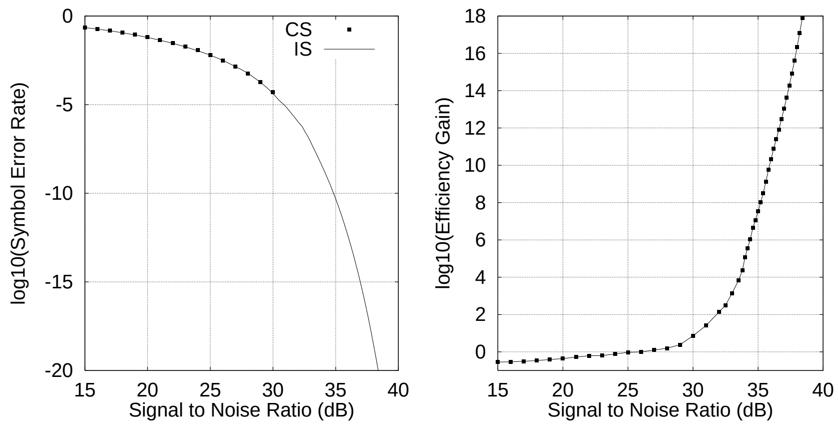
(a) (b)

The lower-bound SERs (a) and the estimated IS gain (b) of the Bayesian DFE for $\mathbf{h} = [0.3\ 1.0]^T$ with 8-PAM symbols.

CS: conventional simulation, IS: importance sampling simulation.

Simulation Example 2

- Channel $\mathbf{h} = [0.3 \ 1.0 \ \ -0.3]^T$ with 8-PAM symbols, $m = 3$, $d = 2$ and $n_b = 2$.
- 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.
- The bias vectors were selected with uniform probability in the simulation.
- For each SNR, $N_k = 10^4$ samples were used for each state in $\mathcal{R}^{(5)}$, resulting in a total of 6.4×10^5 samples for a given SNR.



(a) (b)

The lower-bound SERs (a) and the estimated IS gain (b) of the Bayesian DFE for $\mathbf{h} = [0.3 \ 1.0 \ \ -0.3]^T$ with 8-PAM symbols.

CS: conventional simulation, IS: importance sampling simulation.