

Importance Sampling for the Efficient Simulation of Adaptive Systems in Frequency Nonselective Slow Rayleigh Fading

W. A. Al-Qaq
J. K. Townsend

Center for Communications and Signal Processing
Department of Electrical and Computer Engineering
North Carolina State University

TR-94/4
April 1994

Submitted to the *IEEE Global Conference on Communications, GLOBECOM '94*
Technical Area: Modeling and Simulation Techniques

Importance Sampling for the Efficient Simulation of Adaptive Systems in Frequency Nonselective Slow Rayleigh Fading¹

Wael A. Al-Qaq ²
J. Keith Townsend ³

Center for Communications and Signal Processing,
Department of Electrical & Computer Engineering,
North Carolina State University, Raleigh, NC 27695-7914
Tel: (919)515-7353
Fax: (919)515-5523

Abstract

Importance sampling (IS) is recognized as an efficient technique in reducing the simulation run time needed to estimate low bit error rates (BER's) in digital communication systems. However, IS applications presented in the literature thus far have been primarily limited to systems with additive white Gaussian noise (AWGN).

In this paper, we present an IS stochastic technique for the efficient simulation of adaptive systems which employ diversity in the presence of frequency nonselective slow Rayleigh fading and AWGN. After accounting for the overhead of the optimization algorithm, average speed-up factors of up to 6 orders of magnitude (over conventional Monte Carlo (MC)) were attained for error probabilities as low as 10^{-11} .

¹This work was supported in part by the Center for Communications & Signal Processing, North Carolina State University.

²W. A. Al-Qaq is an IBM Graduate Fellow.

³Corresponding Author: J. Keith Townsend, Tel: 919-515-7353, Fax: 919-515-5523

1 Introduction

As the demand for wireless data communications increases, so will the need for low bit error rate (BER) wireless links. Since the wireless channel is frequently characterized by Rayleigh fading and AWGN, adaptive reception as well as diversity combining schemes [1, 2, 3] are techniques useful for mitigating time variations in the channel and of achieving lower BER's for data applications. Unfortunately, this added complexity makes closed form analysis of the BER infeasible and renders MC simulation as the primary substitute for performance evaluation. However, utilizing conventional MC for low BER estimation can in itself be a prohibitive task due to long run times.

As an alternative, MC based IS techniques are frequently applied to significantly reduce simulation run time for a given estimator precision. This substantial reduction, however, is normally accompanied by the difficulty of specifying an efficient IS scheme and the corresponding optimal parameter settings. For the diversity system considered in here, the cumulative memory behavior of the adaptive receiver, the decision-directed phenomenon, and the nonlinearity of the adaptive algorithm render analytical optimization techniques [4, 5, 6] as ineffective. In addition, it is unclear if and how numerical optimization (large deviations) techniques [7, 8, 9, 10] are applicable in this case.

The main contribution of this paper is to introduce a stochastic IS methodology for the efficient simulation of systems characterized by diversity and adaptive receivers in the presence of nonselective slow Rayleigh fading and AWGN. This IS stochastic gradient descent (SGD) algorithm, which we first presented in [11], is utilized in here to determine the near-optimal IS parameters that characterize the dominant fading process. The fading model assumed in this paper obeys a first order Markovian process [1, 12, 13, 14]. Using a simple illustrative example, we show that for this case, the IS simulation technique is mostly efficient when the statistics of the fading model obey a modified Markov chain distribution [7, 8, 15].

A fourth order diversity system was simulated at three different time instants and using the corresponding optimal IS parameters. In addition, two signaling formats were considered, BPSK and QPSK. After accounting for the overhead of the optimization algorithm, average

speed-up factors of up to 6 orders of magnitude (over conventional Monte Carlo (MC)) were attained for error probabilities as low as 10^{-11} .

2 System Description

Consider a carrier-modulated real-valued signal $S(t)$ transmitted over L independent diversity channels, where

$$S(t) = \text{Re} \left\{ \sum_{k=0}^{\infty} \mathbf{d}(k)g(t - kT) \exp(jw_c t) \right\} \quad (1)$$

where $\sum_{k=0}^{\infty} \mathbf{d}(k)g(t - kT)$ is the complex lowpass envelope with $\mathbf{d}(k)$ being the k th data symbol (real or complex), and $g(t)$ being the impulse response of the transmit filter. w_c is the carrier frequency in rad/sec. Each channel introduces a frequency nonselective and slow Rayleigh fading in addition to AWGN. The equivalent lowpass and complex time-varying impulse response of each channel is given by [16]

$$h_i(\tau, t) = c_i(t)\delta(\tau), \quad i = 1, \dots, L \quad (2)$$

where for $i = 1, \dots, L$, $c_i(t)$ is a complex-valued Gaussian random process. The slow fading process is assumed to be constant for the duration of one symbol period [1, 12, 16]. Employing this piecewise constant approximation for the fading process, the received signal over the i th channel can be expressed as [1, 16]

$$x_i(t) = \text{Re} \left\{ \left[\sum_{k=0}^{\infty} \mathbf{c}_i(k)\mathbf{d}(k)g(t - kT) + n_i(t) \right] \exp(jw_c t) \right\}, \quad i = 1, \dots, L \quad (3)$$

where $[\sum_{k=0}^{\infty} \mathbf{c}_i(k)\mathbf{d}(k)g(t - kT) + n_i(t)]$ represents the received complex lowpass envelope. The fading gain of the i th ($1 \leq i \leq L$) fading channel $\mathbf{c}_i(k)$ is a complex Gaussian random variable (CGRV) with $E\{\mathbf{c}_i(k)\} = 0$ and $E\{|\mathbf{c}_i(k)|^2\} = 2\sigma^2$. For $i \neq j$, $\mathbf{c}_i(k)$ and $\mathbf{c}_j(k)$ are independent processes. In this paper, the complex fading process $\mathbf{c}_i(k) = \mathbf{a}_i(k) + j\mathbf{b}_i(k)$ has an even Doppler power spectrum [17] which implies that $\{\mathbf{a}_i(k)\}$ and $\{\mathbf{b}_i(k)\}$ are independent Gaussian processes with a zero mean and an identical autocorrelation function $R(l)$, where

$$R(l) = E\{\mathbf{a}_i(k)\mathbf{a}_i(k+l)\} = E\{\mathbf{b}_i(k)\mathbf{b}_i(k+l)\}, \quad i = 1, \dots, L \quad (4)$$

The fading model assumed in this paper obeys a first order Markovian process [1, 12, 13, 14], namely

$$R(l) = \sigma^2 \rho^{|l|}, \quad 0 < \rho \leq 1 \quad (5)$$

The correlation parameter ρ is a measure of the rate of the fading channel fluctuations. This parameter is primarily determined by the product of the symbol period T and the fading bandwidth f_d (i.e., the 3dB cutoff frequency of the Doppler power spectrum). In the case of a first order Butterworth Doppler power spectrum [18], the correlation parameter is given by

$$\rho = \exp(-f_d T) \quad (6)$$

where for a slow Rayleigh fading channel, we have $0 < f_d T \ll 1$. For example, a symbol rate of 50 KHz and a fading bandwidth of 10 Hz would correspond to $\rho = 0.9998$. Other expressions for ρ can be found in [18] for several commonly encountered Doppler power spectra.

The autocorrelation function in (5) yields the following state equation

$$\mathbf{c}_i(k+1) = \rho \mathbf{c}_i(k) + \mathbf{w}_i(k), \quad i = 1, \dots, L \quad (7)$$

where for $i = 1, \dots, L$, $\{\mathbf{w}_i(k)\}$ is a sequence of iid zero-mean CGRV's with $E\{|\mathbf{w}_i(k)|^2\} = 2\sigma^2(1 - \rho^2)$, and for $i \neq j$, $\{\mathbf{w}_i(k)\}$ and $\{\mathbf{w}_j(k)\}$ are independent processes.

The complex additive noise at the i th antenna is given by $n_i(t)$, where each $n_i(t)$ is a complex AWGN process, with $E\{n_i(t)\} = 0$ and $E\{n_i(t)n_i^H(t + \tau)\} = N_o\delta(\tau)$ (H denotes the complex conjugate), with $n_i(t)$ and $n_j(t)$ being independent processes for $i \neq j$.

The received signal is demodulated by a local carrier $\exp(-j\omega_c t)$, match-filtered by $g(T - t)$, and sampled every T seconds. Thus, the resulting i th output (input to the i th equalizer) during the k th symbol period is given by

$$\mathbf{x}_i(k) = \mathbf{c}_i(k)\mathbf{d}(k) + \mathbf{n}_i(k), \quad i = 1, \dots, L \quad (8)$$

In the above equation, it is assumed that the transmit filter is normalized (i.e., $\int_0^T |g(t)|^2 dt = 1$), therefore, $E\{|\mathbf{n}_i(k)|^2\} = N_o$. A block diagram of the the digital communication system described above is depicted in Fig. 1.

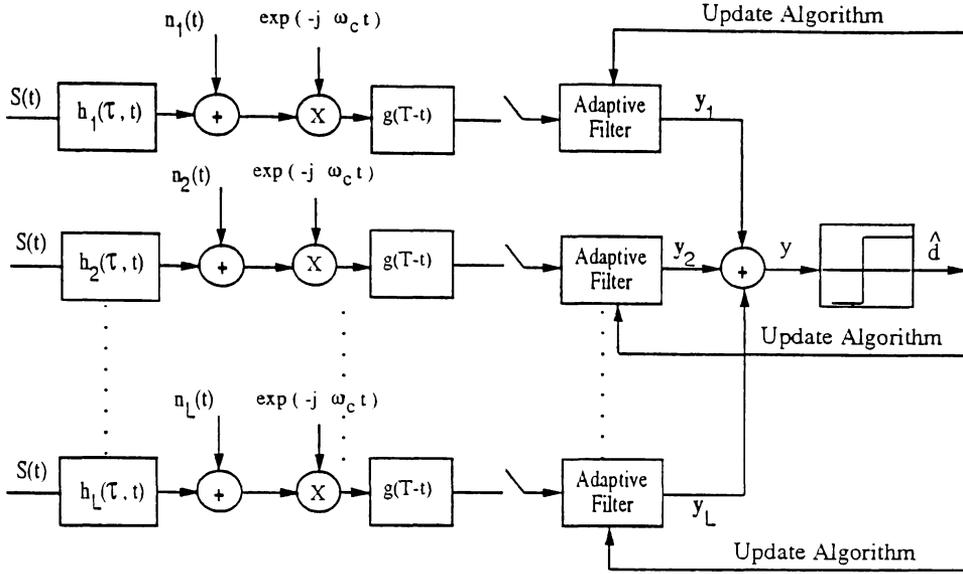


Figure 1: A block diagram of a digital communication system with diversity and adaptive filtering in the presence of nonselective slow Rayleigh fading and AWGN.

A partially coherent reception technique was employed in [1, 12] using a decision-directed adaptive Kalman filter. This receiver assumes full knowledge of the statistics of the incoming signal. In this paper, a decision-directed recursive least square [19, 20] adaptive receiver is utilized to compensate for the fading process on the i th diversity. This partially coherent algorithm does not assume any prior knowledge of the statistics of the input signal, and is therefore more practical to implement. Define the random data vector at instant k as

$$\mathbf{D}(k) = [\mathbf{d}(k), \mathbf{d}(k-1), \dots, \mathbf{d}(0)],$$

the i th AWGN vector at the output of the receive matched-filter at instant k as

$$\mathbf{N}_i(k) = [\mathbf{n}_i(k), \mathbf{n}_i(k-1), \dots, \mathbf{n}_i(0)] \quad 1 \leq i \leq L,$$

and the corresponding vector $\mathbf{N}(k)$ of the iid random vectors $\{\mathbf{N}_i(k)\}_{i=1}^L$ as

$$\mathbf{N}(k) = [\mathbf{N}_1(k), \mathbf{N}_2(k), \dots, \mathbf{N}_L(k)].$$

Also, define the i th complex Gaussian random fading vector at instant k as

$$\mathbf{C}_i(k) = \mathbf{A}_i(k) + \mathbf{B}_i(k) = [\mathbf{c}_i(k), \mathbf{c}_i(k-1), \dots, \mathbf{c}_i(0)] \quad 1 \leq i \leq L,$$

where

$$\mathbf{A}_i(k) = [\mathbf{a}_i(k), \mathbf{a}_i(k-1), \dots, \mathbf{a}_i(0)]$$

and

$$\mathbf{B}_i(k) = [\mathbf{b}_i(k), \mathbf{b}_i(k-1), \dots, \mathbf{b}_i(0)],$$

and let $\mathbf{C}(k)$ be the corresponding vector of the iid random vectors $\{\mathbf{C}_i(k)\}_{i=1}^L$, namely

$$\mathbf{C}(k) = \mathbf{A}(k) + \mathbf{B}(k) = [\mathbf{C}_1(k), \mathbf{C}_2(k), \dots, \mathbf{C}_L(k)] \quad (9)$$

where $\mathbf{A}(k)$ and $\mathbf{B}(k)$ are defined in a fashion similar to $\mathbf{C}(k)$. In addition, define the received random vector at instant k as

$$\mathbf{X}_i(k) = [\mathbf{x}_i(k), \mathbf{x}_i(k-1), \dots, \mathbf{x}_i(0)] \quad 1 \leq i \leq L,$$

and the corresponding vector $\mathbf{X}(k)$ of the iid random vectors $\{\mathbf{X}_i(k)\}_{i=1}^L$ as

$$\mathbf{X}(k) = [\mathbf{X}_1(k), \mathbf{X}_2(k), \dots, \mathbf{X}_L(k)] \quad (10)$$

The output of the i th adaptive receiver for the k th transmitted symbol $\mathbf{d}(k)$ ($k \geq k_o$) is given by [19, 21]

$$\begin{aligned} \mathbf{y}_i(k) &= g(\mathbf{X}_i(k)) \\ &= \mathbf{x}_i(k) \left\{ \frac{\sum_{l=0}^{k_o-1} \lambda^{k-l-1} \mathbf{d}(l) \mathbf{x}_i^H(l) + \sum_{l=k_o}^{k-1} \lambda^{k-l-1} \hat{\mathbf{d}}(l) \mathbf{x}_i^H(l)}{\sum_{l=0}^{k-1} \lambda^{k-l-1} |\mathbf{x}_i(l)|^2 + \delta \lambda^{k-1}} \right\}, \quad i = 1, \dots, L \end{aligned} \quad (11)$$

where k_o is the number of symbols used to train the equalization algorithm (training sequence), and δ is a small positive number. $0 < \lambda \leq 1$ represents the forgetting factor of the algorithm. The purpose of this factor is to weight the most recent symbols more heavily and thus allow the equalizer to track time variations in the channel. The input to the decision device is given by

$$\begin{aligned} \mathbf{y}(k) &= \sum_{i=1}^L \mathbf{y}_i(k) \\ &= \sum_{i=1}^L g(\mathbf{X}_i(k)) = G(\mathbf{X}(k)) \end{aligned} \quad (12)$$

At the end of the training period, a decision-directed adaptive algorithm is employed whereby decisions made on the output sequence $\{\mathbf{y}(k)\}$ (i.e., $\{\hat{\mathbf{d}}(k)\}$) are used to replace the actual transmitted symbol sequence $\{\mathbf{d}(k)\}$. This will effectively aid the adaptive receiver in tracking channel variations, but may increasingly hinder the tracking capability with time due error propagation. This error propagation effect will be demonstrated later on when considering a fourth order diversity example.

3 IS Formulation

The adaptive system considered in the previous section clearly possesses a cumulative memory behavior as evident from (11). This behavior is quite common to all adaptive algorithms where the decision made at each time instant k utilizes a weighted version of all the past received samples up to time zero. In addition to this behavior, the decision-directed phenomenon, the strong correlation present in the input signal, and the non-linearity of the adaptive algorithm make closed form analysis of the BER infeasible. As an alternative, MC simulation is commonly applied to estimate the BER performance.

Since the statistical distribution of the complex random vector $\mathbf{X}(k)$ is dependent on the correlation parameter ρ and the power in the fading process $2\sigma^2$, it would be appropriate to denote its corresponding probability density function (pdf) as $f_{\mathbf{X}(k)}(X(k), \Theta)$, where

$$\Theta \triangleq [\rho, \sigma] \quad (13)$$

and

$$\begin{aligned} f_{\mathbf{X}(k)}(X(k), \Theta) &= f_{\mathbf{C}(k)}(C(k), \Theta) f_{\mathbf{N}(k)}(N(k)) f_{\mathbf{D}(k)}(D(k)) \\ &= f_{\mathbf{A}(k)}(A(k), \Theta) f_{\mathbf{B}(k)}(B(k), \Theta) f_{\mathbf{N}(k)}(N(k)) f_{\mathbf{D}(k)}(D(k)) \end{aligned} \quad (14)$$

For the fading model assumed in (5), the distribution of the i th independent random vectors $\mathbf{A}_i(k)$ and $\mathbf{B}_i(k)$ are identical with the i th marginal pdf $f_{\mathbf{A}_i(k)}(A_i(k), \Theta)$ being given by

$$f_{\mathbf{A}_i(k)}(A_i(k), \Theta) = f_{\mathbf{a}_i(0)}(a_i(0), \Theta) \prod_{j=0}^{k-1} f_{\mathbf{a}_i(j+1)|\mathbf{a}_i(j)}(a_i(j+1), \Theta|a_i(j))$$

$$\begin{aligned}
 &= \frac{1}{(2\pi\sigma^2)^{(k+1)/2}(1-\rho^2)^{k/2}} \exp\left\{-\frac{a_i^2(0)}{2\sigma^2}\right\} \times \\
 &\quad \exp\left\{-\frac{\sum_{j=0}^{k-1}(a_i(j+1) - \rho a_i(j))^2}{2(1-\rho^2)\sigma^2}\right\}, \quad i = 1, \dots, L \quad (15)
 \end{aligned}$$

Observe that the notation $f_{\mathbf{X}(k)}(X(k), \Theta)$ was only chosen to explicitly show the dependence on Θ since for most practical cases and cases where IS is applicable (i.e., high signal to noise ratio (SNR)), the slow Rayleigh fading process is the dominant factor. Therefore, we consider biasing only the fading processes $A(k)$ and $B(k)$ leaving the AWGN random process $N(k)$ unbiased. In addition, the parameter vector Θ is related to the biasing scheme which we adopt and address in the next section.

Let the expectation with respect to the marginal pdf $f_{\mathbf{X}(k)}(X(k), \Theta)$ be denoted as $E_{k,\Theta}\{\cdot\}$. The subscript k is used to emphasize the *time-dependent* nature of the expectation. In addition, let $I(G(\mathbf{X}(k)))$ be equal to 1 when the detected symbol is in error and zero otherwise, where $G(\mathbf{X}(k))$ is given by (12). The probability of making a decision error at some time instant k is given by the following instantaneous BER expression

$$P(k, \Theta) = E_{k,\Theta}\{1_{\Omega(k)}\} \quad (16)$$

where $1_{\Omega(k)}$ is the indicator function of the set $\Omega(k) = \{\mathbf{X}(k) : I(G(\mathbf{X}(k))) = 1\}$.

To apply IS, note that (16) can be written as

$$P(k, \Theta) = E_{k,\Theta^*(k)}\{1_{\Omega(k)}w_{\mathbf{X}(k)}(\mathbf{X}(k), \Theta, \Theta^*(k))\} \quad (17)$$

where $E_{k,\Theta^*(k)}\{\cdot\}$ denotes the expectation operator with respect to the IS pdf $f_{\mathbf{X}(k)}(X(k), \Theta^*(k))$ with the time-dependent parameter vector

$$\Theta^*(k) = [\rho^*(k), \sigma^*(k)] \quad (18)$$

and

$$\begin{aligned}
 w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k)) &= w_{\mathbf{A}(k)}(A(k), \Theta, \Theta^*(k))w_{\mathbf{B}(k)}(B(k), \Theta, \Theta^*(k)) \\
 &= \frac{f_{\mathbf{A}(k)}(A(k), \Theta) f_{\mathbf{B}(k)}(B(k), \Theta)}{f_{\mathbf{A}(k)}(A(k), \Theta^*(k)) f_{\mathbf{B}(k)}(B(k), \Theta^*(k))} \quad (19)
 \end{aligned}$$

Observe that the random vectors $\mathbf{A}(k)$ and $\mathbf{B}(k)$ are iid with

$$f_{\mathbf{A}(k)}(A(k), \Theta) = \prod_{i=1}^L f_{\mathbf{A}_i(k)}(A_i(k), \Theta) \quad (20)$$

and

$$f_{\mathbf{A}(k)}(A(k), \Theta^*(k)) = \prod_{i=1}^L f_{\mathbf{A}_i(k)}(A_i(k), \Theta^*(k)) \quad (21)$$

Upon incorporating the preceding two equations into (19), we get

$$w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k)) = \prod_{i=1}^L w_{\mathbf{A}_i(k)}(A_i(k), \Theta, \Theta^*(k)) w_{\mathbf{B}_i(k)}(B_i(k), \Theta, \Theta^*(k)) \quad (22)$$

where $w_{\mathbf{A}_i(k)}(A_i(k), \Theta, \Theta^*(k))$ and $w_{\mathbf{B}_i(k)}(B_i(k), \Theta, \Theta^*(k))$ have an identical distribution, with

$$w_{\mathbf{A}_i(k)}(A_i(k), \Theta, \Theta^*(k)) = \frac{f_{\mathbf{A}_i(k)}(A_i(k), \Theta)}{f_{\mathbf{A}_i(k)}(A_i(k), \Theta^*(k))} \quad (23)$$

A MC-based IS estimator of (17) is given by

$$\hat{P}(k, \Theta) = \frac{1}{N_X} \sum_{i=1}^{N_X} I(G(X(k, i))) w_{\mathbf{X}(k)}(X(k, i), \Theta, \Theta^*(k)) \quad (24)$$

It is straight forward to show that the estimator in (24) is an *unbiased* estimator of $P(k, \Theta)$ (i.e., $E_{k, \Theta^*(k)}\{\hat{P}(k, \Theta)\} = P(k, \Theta)$). The variance of the IS estimator in (24) is given

$$\begin{aligned} V\{\hat{P}(k, \Theta)\} &= \mathcal{V}(k, \Theta, \Theta^*(k)) \\ &= \frac{1}{N_X} V_{k, \Theta^*(k)} \left\{ 1_{\Omega(k)} w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k)) \right\} \end{aligned} \quad (25)$$

where $\mathcal{V} : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow [0, \infty)$, and $V_{k, \Theta^*(k)}\{\cdot\}$ denotes the time-dependent variance operator with respect to the simulation pdf $f_{\mathbf{X}(k)}(X(k), \Theta^*(k))$. The empirical precision of the estimator in (24) may be found by using the sample variance estimator

$$\hat{V}\{\hat{P}(k, \Theta)\} = \frac{1}{N_X^2} \sum_{i=1}^{N_X} I(G(X(k, i)))^2 w_{\mathbf{X}(k)}^2(X(k, i), \Theta, \Theta^*(k)) - \frac{\hat{P}^2(k, \Theta)}{N_X} \quad (26)$$

For a given relative precision $\alpha_o > 0$, the simulation is terminated when the condition $\sqrt{\hat{V}\{\hat{P}(k, \Theta)\}}/\hat{P}(k, \Theta) \leq \alpha_o$ is satisfied.

3.1 The IS Biasing Scheme

Two important issues are worth addressing at this point. First, note that for $i = 1, \dots, L$, the in-phase and quadrature Gaussian processes $\mathbf{A}_i(k)$ and $\mathbf{B}_i(k)$ are identically modified according to the parameter vector $\Theta^*(k)$. This is the case since $\mathbf{A}_i(k)$ and $\mathbf{B}_i(k)$ are equally weighted in the real and imaginary parts of $\mathbf{y}_i(k)$ in (11). In addition, each of the random processes $\{\mathbf{A}_i(k)\}_{i=1}^L$ and $\{\mathbf{B}_i(k)\}_{i=1}^L$ must be equally biased as well since according to (12), the output $\mathbf{y}(k)$ is the sum of the equally weighted iid outputs $\{\mathbf{y}_i(k)\}_{i=1}^L$. This implies that the modification of the underlying density function $f_{\mathbf{X}(k)}(X(k), \Theta)$ is entirely governed by the parameter vector $\Theta^*(k)$.

Second, observe that altering the parameter vector Θ implies that the adopted IS simulation technique is performed under a modified Markov chain distribution [15, 8, 7]. At a given time instant k , this technique effectively transforms the original unbiased autocorrelation function $R(l)$ of (5) into

$$R^*(l, k) = \sigma^{*2}(k) \rho^{*|l|}(k), \quad 0 < \rho^*(k) < 1 \quad (27)$$

In the simple case of a memoryless nonlinearity followed by an integrate and dump receiver, and when the input samples have a Markov chain distribution, it was shown [7] that the above simulation technique is more efficient than translating the mean [4, 10] or scaling the variance [5, 6] of the iid random variables $\{\mathbf{w}_i(k)\}$ in (7). The ineffectiveness of the of the mean translation and variance scaling techniques is primarily attributed to the fact that the random process $\mathbf{X}(k)$ has a zero mean and $X(k) = 0 \in \Omega(k)$.

This can be illustrated by Considering the system discussed in the previous section with a first order diversity and binary signaling ($\mathbf{d}(k) = \pm 1$). Furthermore, consider the time instant $k = 1$, and let the $SNR = \infty$ (i.e., $\mathbf{N}_1(1) = \mathbf{0}$) along with the quadrature random vector $\mathbf{B}_1(1) = \mathbf{0}$. Conditioning on $d(1) = +1$ and using Eq. (11) yields

$$P(1, \Theta) = \Pr\{\mathbf{a}_1(0)\mathbf{a}_1(1) < 0\},$$

where $\mathbf{A}_1(1) = [\mathbf{a}_1(1), \mathbf{a}_1(0)] \sim N(\mathbf{0}, \Sigma_{\mathbf{A}_1(1)})$, and

$$\Sigma_{\mathbf{A}_1(1)} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

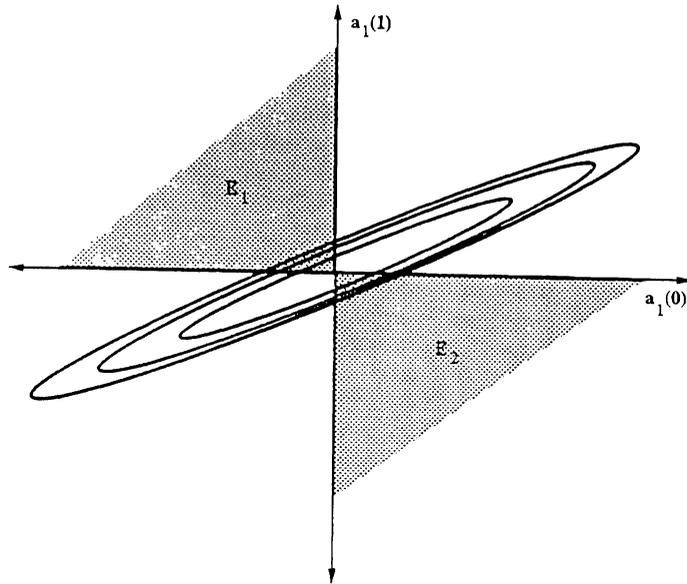


Figure 2: A contour plot of the original pdf, $f_{A_1(1)}(A_1(1), \Theta)$ and the corresponding error region.

A plot of the contours of $f_{A_1(1)}(A_1(1), \Theta)$ and the error region ($E_1 \cup E_2$) for $\Theta = [0.99, 1]$ is shown in Fig. 2. The probability of error in this case is equal to $P(1, \Theta) = 4.51 \times 10^{-2}$. Note that as the value of $\rho \rightarrow 1$, the contours area over the region $E_1 \cup E_2$ will gradually vanish and as a result, $P(1, \Theta) \rightarrow 0$. This should be the case since it can be shown that

$$P(1, \Theta) = \frac{1}{\pi} \tan^{-1} \left[\frac{\sqrt{1 - \rho^2}}{\rho} \right]$$

Clearly, $\lim_{\rho \rightarrow 1} P(1, \Theta) = 0$.

Even though translating the contours diagonally into E_1 or E_2 will increase the raw (i.e., unweighted) error count, the variance of the IS weight function over the entire error region ($E_1 \cup E_2$) will rather increase since the modified pdf can only concentrate its mass on either E_1 or E_2 but not both. Uniform variance scaling (i.e., biasing σ^2 only) will, on the other hand, proportionally scale the contours in all dimensions and the net increase in the raw error count will be zero. This conclusion can also be deduced from the fact that $P(1, \Theta)$, as given above, is independent of σ .

The difficulty encountered with the above two schemes can be surmounted by modifying the parameter vector Θ . such an approach will help provide the two ingredients necessary to

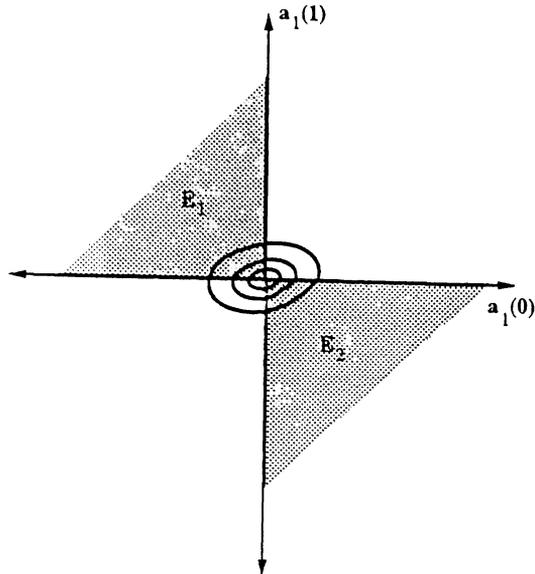


Figure 3: A contour plot of the modified pdf, $f_{\mathbf{A}_1(1)}(\mathbf{A}_1(1), \Theta_{opt}^*(1))$, $\Theta_{opt}^*(1) = [0.4674, 0.0211]$.

realize a good IS biasing scheme, namely increasing the raw error count and evenly distribute the weight function over the error region $E_1 \cup E_2$.

The IS variance (Eq. (25)) for this simple case can be attained in a closed form and minimized to yield $\Theta_{opt}^*(1)$. Instead, the search for $\Theta_{opt}^*(1)$ was conducted stochastically using the SGD algorithm, which we first presented in [11], with Θ being the starting point. The search yielded $\Theta^*(1) \approx \Theta_{opt}^*(1) = [0.4674, 0.0211]$, and a plot of the contours of the modified pdf is shown in Fig. 3.

4 Optimizing the IS Distribution

In order to attain a significant reduction in the number of detector decisions N_X needed to estimate $P(k, \Theta)$ with a given accuracy, the optimal parameter setting $\Theta_{opt}^*(k)$ which minimizes the variance in (25) must be determined. Since a closed form expression of (25) is not available, analytical optimization techniques [4, 5, 6] are rendered ineffective. In addition, as we illustrated earlier, having a zero mean random process $\mathbf{X}(k)$ and $X(k) = \mathbf{0} \in \Omega(k)$ coupled with the presence of a decision-directed phenomenon and the complexity of the

adaptive algorithm make it unclear how numerical optimization (large deviations) techniques [10, 7, 8, 9, 10] can be modified to accommodate the current application.

As an alternative, the IS SGD algorithm which we first presented in [11, 22] is utilized to search for the optimal IS parameter vector $\Theta_{opt}^*(k)$ in a two dimensional space. In summary, the SGD algorithm is based on the fact that for any vector $\Theta_{opt}^*(k)$ to be a local or global minimum of $\mathcal{V}(k, \Theta, \Theta^*(k))$ in (25), the following equation must hold

$$\nabla_{\Theta^*(k)} \mathcal{V}(k, \Theta, \Theta^*(k))|_{\Theta^*(k)=\Theta_{opt}^*(k)} = \mathbf{0} \quad (28)$$

Thus, the search for the optimal IS setting is conducted by the SGD algorithm in a descending fashion according to

$$\Theta^*(k, n+1) = \Theta^*(k, n) - \beta(n) \widehat{\nabla}_{\Theta^*(k)} \mathcal{V}(k, \Theta, \Theta^*(k))|_{\Theta^*(k)=\Theta^*(k, n)} \quad (29)$$

where $\beta(n)$ is the step size taken at the n th iteration, and $\widehat{\nabla}_{\Theta^*(k)} \mathcal{V}(k, \Theta, \Theta^*(k))$ is an *unbiased* estimate of the left-hand side of (28). It was shown [11], that such an estimate can be constructed using the following estimator

$$\begin{aligned} \widehat{\nabla}_{\Theta^*(k)} \mathcal{V}(\Theta, \Theta^*(k)) &= \frac{1}{N_X^2} \sum_{j=1}^{N_X} \text{I}(G(X(k, i))) w_{\mathbf{X}(k)}(X(i, k), \Theta, \Theta^*(k)) \times \\ &\quad \nabla_{\Theta^*(k)} w_{\mathbf{X}(k)}(X(k, i), \Theta, \Theta^*(k)) \end{aligned} \quad (30)$$

In order to implement the above estimate, we need to compute the gradient of the IS weight function.

$$\nabla_{\Theta^*(k)} w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k)) = \left[\frac{\partial w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k))}{\partial \rho^*(k)}, \frac{\partial w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k))}{\partial \sigma^*(k)} \right] \quad (31)$$

Introducing the following definitions will help simplify the expression of the above gradient. Let for $i = 1, \dots, L$,

$$\begin{aligned} R_1(A_i(k), \Theta^*(k)) &\triangleq \frac{\rho^*(k) \sigma^{*2}(k) \sum_{j=0}^{k-1} (a_i(j+1) - \rho^*(k) a_i(j))^2}{(1 - \rho^{*2}(k))^2 \sigma^{*4}(k)} - \\ &\quad \frac{(1 - \rho^{*2}(k)) \sigma^{*2}(k) \sum_{j=0}^{k-1} (a_i(j+1) a_i(j) - \rho^*(k) a_i(j)^2)}{(1 - \rho^{*2}(k))^2 \sigma^{*4}(k)} \end{aligned} \quad (32)$$

$$R_2(A_i(k), \Theta^*(k)) \triangleq \frac{\sum_{j=0}^{k-1} (a_i(j+1) - \rho^*(k)a_i(j))^2}{(1 - \rho^{*2}(k))\sigma^{*3}(k)} \quad (33)$$

with $R_1(B_i(k), \Theta^*(k))$ and $R_2(B_i(k), \Theta^*(k))$ being defined in a fashion similar to (32) and (33) respectively. After taking the derivative of both sides in (22) and utilizing the above definitions, it can be shown that

$$\frac{\partial w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k))}{\partial \rho^*(k)} = \left\{ \frac{-2kL\rho^*(k)}{(1 - \rho^{*2}(k))} + \sum_{i=1}^L (R_1(A_i(k), \Theta^*(k)) + R_1(B_i(k), \Theta^*(k))) \right\} \times w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k)) \quad (34)$$

and

$$\begin{aligned} \frac{\partial w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k))}{\partial \sigma^*(k)} &= \left\{ \frac{2L(k+1)}{\sigma^*(k)} \right\} w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k)) - \\ &\sum_{i=1}^L \left[\frac{(a_i^2(0) + b_i^2(0))}{\sigma^{*3}(k)} \right] w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k)) - \\ &\sum_{i=1}^L [R_2(A_i(k), \Theta^*(k)) + R_2(B_i(k), \Theta^*(k))] \times \\ &w_{\mathbf{X}(k)}(X(k), \Theta, \Theta^*(k)) \end{aligned} \quad (35)$$

Observe that during the n th iteration of the SGD algorithm in (29), the IS estimators of all the three quantities $P(k, \Theta)$, $V\{\hat{P}(k, \Theta)\}$, and $\nabla_{\Theta^*(k)} V\{\hat{P}(k, \Theta)\}$ are obtained by sampling from the simulation pdf at the n th iteration, namely $f_{\mathbf{X}(k)}(X(k), \Theta^*(k, n))$. In addition to yielding an optimal estimator of $P(k, \Theta)$ as $n \rightarrow \infty$, this approach will generally provide suboptimal estimates of $V\{\hat{P}(k, \Theta)\}$ and $\nabla_{\Theta} V\{\hat{P}\}$, which are sufficiently accurate to successfully perform the SGD algorithm.

Let the SNR per diversity be defined as

$$\gamma = \frac{E\{|\mathbf{d}(k)\mathbf{c}_i(k)|^2\}}{E\{\mathbf{n}_i(k)\}} = \frac{E\{|\mathbf{d}(k)|^2\} 2\sigma^2}{N_o}, \quad i = 1, \dots, L. \quad (36)$$

For a given diversity order $1 \leq i \leq L$, a SNR, and a time instant $k \geq k_o$, the SGD algorithm is applied for some value of Θ such that $\hat{P}(k, \Theta) \sim 10^{-2}$. this technique will help circumvent the difficulty of specifying an initial starting point by simply setting $\Theta^*(k, 1) = \Theta$. With this range of values for $\hat{P}(k, \Theta)$ and the above starting point, a manageable number of decisions

($N_X \approx 1000$) can be sufficient to accurately compute $\widehat{\nabla}_{\Theta^*(k)} V\{\hat{P}(k, \Theta)\}|_{\Theta^*(k)=\Theta^*(k,1)}$. The optimal parameter vector Θ_{opt}^* determined for a high $\hat{P}(k, \Theta)$ is then used to choose a starting point $\Theta^*(1)$ at higher values of $P(k, \Theta)$. This ‘‘extrapolation’’ technique was shown to be very effective in a variety of practical digital communication and queing systems applications [11, 22] and for highly nonlinear systems with large dimensionality [23]. with this technique applied efficiently, the overhead (in number of decisions) involved in determining $\Theta_{opt}^*(k)$ will be insignificant compared to the savings in number of decisions N_X needed to accurately estimate a low $P(k, \Theta)$.

As an example, consider a second, and a fourth order diversity systems with BPSK signaling and a per-diversity SNR of $\gamma = 30$ dB. The equalizer was trained using a fixed training sequence of 12 symbols and the BER was measured 10 symbols after training (i.e., $P(22, \Theta)$). Due to the absence of intersymbol interference (ISI) and the symmetry of the signal constellation about the origin, the probability of error is insensitive to the choice of the data sequence after the training period. This effectively results in an optimal IS setting that is sequence-independent, as was empirically verified. In each case, Θ was chosen to attain a high $P(22, \Theta)$. Since the BER being considered is high, the improvement will not be significant, and the optimal vector is identified by the convergence of the SGD algorithm.

The search for the near-optimal vector $\Theta_{opt}^*(22)$ was executed iteratively using the SGD algorithm with $N_X = 1000$ decisions per estimate of $\nabla_{\Theta^*(k)} V\{\hat{P}(k, \Theta)\}$ per iteration. The step size $\beta(n)$ at the n th iteration was chosen according to

$$\beta(n) = \frac{\Delta}{\|V\{\hat{P}(k, \Theta)\}|_{\Theta^*(k)=\Theta^*(k,n)}\|} \quad (37)$$

On the n th iteration, this step size will result in a maximum incremental or decremental change of Δ to the components of the parameter vector $\Theta^*(k, n)$. As observed in [11, 23], selecting a small Δ will result in a slow convergence but a higher accuracy in locating the optimal setting as opposed to a larger Δ that yields a faster convergence rate but may cause some deviation from the the correct path of the gradient descent. Typical values of Δ may range from 0.0001 to 0.005. The forgetting factor of the adaptive algorithm λ was set equal

Diversity (L)	Θ	$\Theta_{opt}^*(22)$	$\hat{P}(22, \Theta)$	Raw count
1	[0.992, 0.7071]	[0.9637, 0.3295]	4.34×10^{-2}	17.3%
2	[0.98, 0.7071]	[0.9602, 0.504]	2.78×10^{-2}	8.75%
4	[0.96, 0.7071]	[0.947, 0.6042]	1.19×10^{-2}	2.7%

Table 1: The optimal IS parameters for $k = 22$. The signaling format is BPSK and $\gamma = 30$ dB.

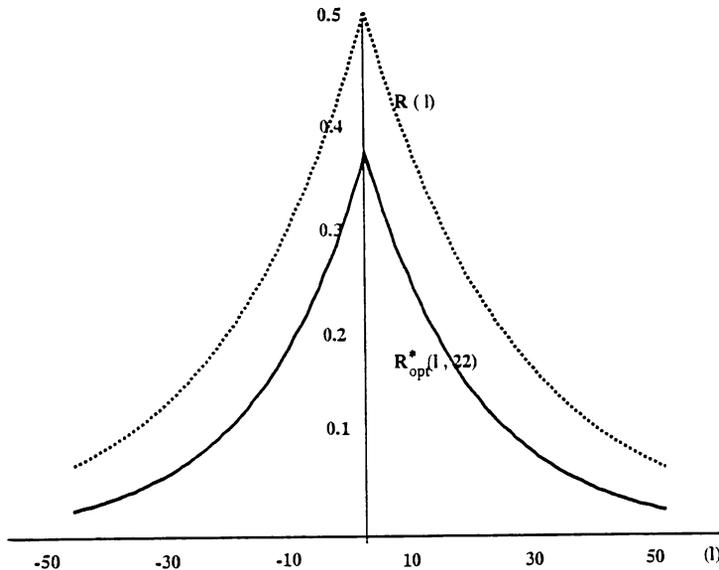


Figure 4: A plot of the original and optimal IS autocorrelation functions for BPSK. $L = 4$, $\Theta = [0.96, 0.7071]$, and $\gamma = 30$ dB.

to ρ , and $\delta = 4 \times 10^{-5}$. The results are shown in Table 1. $\Theta^*(k) \approx \Theta_{opt}^*(k, n)$. For each case considered in Table 1, the optimal IS autocorrelation function $R_{opt}^*(l, k)$ yields a less correlated fading signal, as compared to $R(l)$, with a reduced mean power (i.e., less SNR). This effectively translates into an increase in the Doppler frequency bandwidth (i.e., faster fading) and a reduction in the energy of the fading power spectrum. A plot of the original and modified autocorrelation functions for the case $L = 4$ is shown in Fig. 4.

Two important observations can be deduced from Table 1. First, the raw error count at the optimal IS setting decreases as the diversity order increases. Moreover, it was experimentally observed that the error count corresponding to the different diversities in Table 1 is roughly maintained at the same level for different values of ρ and k .

Second, note that the optimal parameter vector approximately satisfies the following

equation

$$\sigma^2(1 - \rho^2) \approx \sigma^{*2}(k)(1 - \rho^{*2}(k)) \quad (38)$$

where $k = 22$. In fact, as our empirical results will show later, the above equation holds for other time instants as well, and its accuracy increases as the correlation parameter ρ approaches unity. Clearly, the above equation can be exploited to transform a 2-D search into a 1-D search, however, we only chose to utilize the above equation in identifying a good starting point prior to conducting the SGD search.

Thus, for a given SNR, a correlation function $R(l) = \sigma^2 \rho^{|l|}$, a time instant $k_1 \geq k_o$, and a diversity order L , the near-optimal parameter setting $\Theta_{opt}^*(k)$, or equivalently, the near-optimal autocorrelation function $R_{opt}^*(l, k_1)$ is determined as follows:

- Choose a $\rho' < \rho$ and a $\sigma' = \sigma$ such that $\hat{P}(k_1, \Theta') \approx 10^{-2}$.
- Using the IS SGD algorithm, determine the near-optimal setting for the system with $\Theta' = [\rho', \sigma']$,
- Choose a starting point $\Theta^*(k_1, 1)$ such that

$$\Theta^*(k_1, 1) = [\rho_{opt}^{*'}(k_1), \sigma^*(k_1, 1)] \quad (39)$$

where

$$\sigma^2(1 - \rho^2) = \sigma^{*2}(k_1, 1)(1 - \rho_{opt}^{*2}(k_1)) \quad (40)$$

- Perform a 2-dimensional search using the SGD algorithm to locate $\Theta_{opt}^*(k_1)$

The optimal setting $\Theta_{opt}^*(k_1)$ can then be used as the starting point of the SGD algorithm to locate the optimal parameters at a time instant $k_2 > k_1$, and so on.

4.1 A Fourth Order Diversity Example

Consider a fourth order diversity system ($L = 4$). The simulation algorithm discussed in the previous section was applied to simulate the BER at the time instants, $k_1 = 22$, $k_2 = 32$, and $k_3 = 42$ (i.e., 10, 20, and 30 symbols after training). The power in the fading signal

k	$\Theta_{opt}^*(k)$	$\hat{P}(k, \Theta)$	$\hat{V}\{\hat{P}(k, \Theta)\}$	$S_p(k)$	Raw Count
22	[0.9708, 0.0417]	1.894×10^{-11}	1.852×10^{-22}	8.236×10^7	2.28%
32	[0.9753, 0.0455]	4.567×10^{-11}	2.372×10^{-21}	1.53×10^7	2.9%
42	[0.9802, 0.0503]	1.621×10^{-10}	2.135×10^{-19}	1.6×10^5	2.82%

Table 2: The optimal IS parameters and the corresponding estimated probabilities, variances, and speed-up factors. $\Theta = [0.9999, 0.7071]$ and $\gamma = 30$ dB, and the signaling format is BPSK.

k	$\Theta_{opt}^*(k)$	$\hat{P}(k, \Theta)$	$\hat{V}\{\hat{P}(k, \Theta)\}$	$S_p(k)$	Raw Count
22	[0.9873, 0.0631]	2.74×10^{-9}	4.97×10^{-17}	1.82×10^4	2.96%
32	[0.9910, 0.0745]	5.64×10^{-9}	3.46×10^{-17}	1.169×10^5	2.7%
42	[0.9921, 0.0792]	1.725×10^{-8}	1.56×10^{-15}	4.1×10^3	3.11%

Table 3: The optimal IS parameters and the corresponding estimated probabilities, variances, and speed-up factors. $\Theta = [0.9999, 0.7071]$, $\gamma = 30$ dB, and the signaling format is QPSK.

was always normalized to unity ($E\{|\mathbf{c}_i(k)|^2\} = 1$). The symbol rate was set to 50 KHz, the fading bandwidth f_D was assumed to be 2.2 Hz which corresponds to $\rho = .9999$, and a SNR of $\gamma = 30$ dB was considered. In addition, two signaling formats were considered, BPSK and QPSK. In each case, the forgetting factor of the adaptive algorithm λ was set equal to the correlation parameter ρ and $\delta = 4 \times 10^{-5}$. The starting point for the first entry in Table 2 was $\Theta^*(22, 1) = [.9476, .0313]$ which was determined using Table 1 and Eq. (40). In addition, for the given Θ , γ , and k , the near-optimal IS parameters determined for the BPSK case were also used as a starting point for the QPSK case. In each case, the step size was chosen according to (37) and the values of Δ ranged from 0.0003 to 0.001. Another interesting case to consider is when $\gamma = \infty$ since it represents the irreducible BER performance. This case was simulated for QPSK and $\Theta = [0.9999, .7071]$, and the results are shown in Table 4.

The estimates of the variance $V\{\hat{P}(k, \Theta)\}$ and the instantaneous BER $P(k, \Theta)$ were computed using an ensemble of $N_E = 50$ estimates of $N_X = 1000$ decisions per estimate. The time-dependent speed-up factor $S_p(k)$, corresponding to $\hat{V}\{\hat{P}(k, \Theta)\}$ and the estimator in (24), was calculated according to

$$S_p(k) = \frac{N_{MC}}{N_X N_E} \quad (41)$$

k	$\Theta_{opt}^*(k)$	$\hat{P}(k, \Theta)$	$\hat{V}\{\hat{P}(k, \Theta)\}$	$S_p(k)$	Raw Count
22	[0.9879, 0.0635]	1.89×10^{-9}	2.15×10^{-17}	3.17×10^4	1.17%
32	[0.9910, 0.0745]	3.82×10^{-9}	2.31×10^{-17}	1.15×10^5	1.56%
42	[0.9922, 0.0804]	1.3368×10^{-8}	8.205×10^{-16}	6.134×10^3	2.09%

Table 4: The optimal IS parameters and the corresponding estimated probabilities, variances, and speed-up factors. $\Theta = [0.9999, 0.7071]$, $\gamma = \infty$, and the signaling format is QPSK.

where N_{MC} is the conventional MC number of decisions required to attain the same accuracy as our IS scheme. N_{MC} was computed based on a 95% confidence interval [24]. The overhead (in number of decisions) was not included in the computations of the speed-up factor. The reduction in speed-up factors due to overhead ranged from 1 to 2.5 orders of magnitude. This reduction is clearly dependent on the choice of Δ .

Several interesting observations can be made from the results in Tables 2, 3, and 4. First note the increase in $\hat{P}(k, \Theta)$ as time (k) increases. This increase is due to the error propagation effect of the decision-directed algorithm (i.e., feeding back erroneous decisions while updating the adaptive algorithm). As time increases, $P(k, \Theta)$ will also increase until it eventually reaches an intolerable rate, and transmitting a new training sequence would become necessary to improve the adaptation algorithm. This signaling technique which periodically interleaves training and data sequences is frequently used by adaptive algorithms [2, 25]. In addition, observe that the increase in $P(k, \Theta)$ is also accompanied by a decrease in $S_p(k)$ as would have been expected. Any further decrease in $S_p(k)$ might also be attributed to an increase in the memory of the algorithm as k increases.

Another two important observations pertaining to the optimal IS setting $\Theta_{opt}^*(k)$ are the decrease in the bias as k increases, and how well $\Theta_{opt}^*(k)$ satisfies Eq. (38). A plot of the optimal trajectory $(\rho_{opt}^*(k), \sigma_{opt}^*(k))$ in Table 2 vs. time is shown in Fig. 5.

5 Conclusion

In this paper, we presented a stochastic IS methodology for the efficient simulation of adaptive systems in the presence of frequency nonselective slow Rayleigh fading and AWGN.

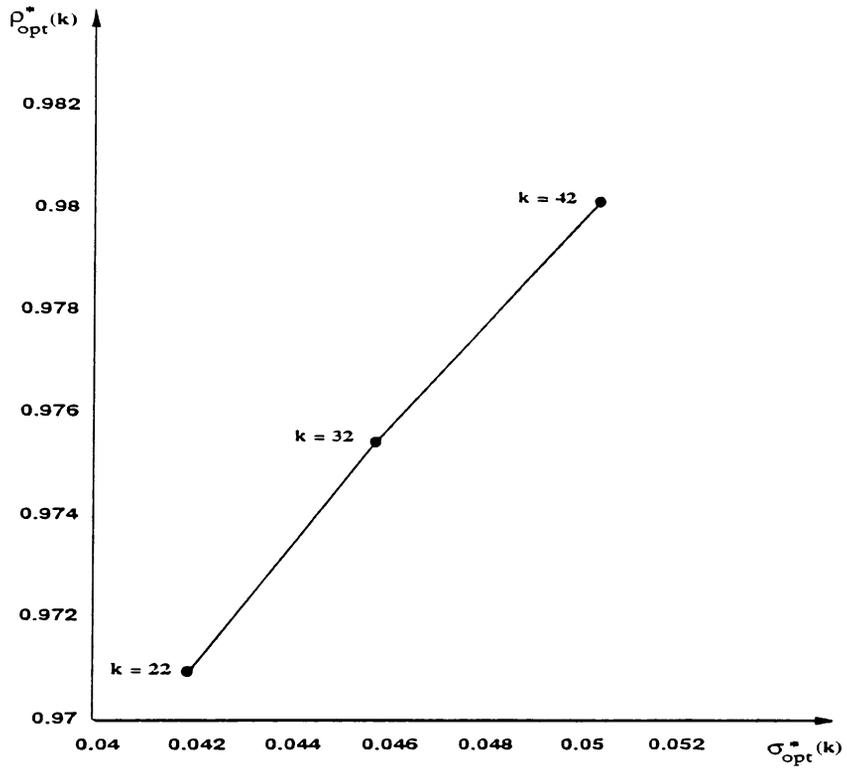


Figure 5: A plot of the optimal trajectory $(\rho_{opt}^*(k), \sigma_{opt}^*(k))$. $L = 4$, $\Theta = [0.9999, 0.7071]$, and $\gamma = 30$ dB.

Using a first order Markovian model for the fading channel, the search for the optimally modified Markov distribution was accomplished by the robust SGD algorithm. Extending this technique to accommodate adaptive algorithms other than the RLS algorithm is straight forward. Speed-up factors of up to 6 orders of magnitude over conventional MC were achieved for error probabilities as low as 10^{-11} after accounting for overhead of the SGD algorithm.

References

- [1] Pooi Yuen Kam. Adaptive Diversity Reception Over a Slow Nonselective Fading Channel. *IEEE Trans. Commun.*, 35(5):572–574, May 1987.
- [2] N. W. K. Lo, D. D. Falconer, and A. U. H. Sheikh. Adaptive Equalization and Diversity Combining for Mobile Radio Using Interpolated Channel Estimates. *IEEE Trans. Veh. Technol.*, 40(3):636–645, Aug. 1991.
- [3] P. Balaban and J. Salz. Dual diversity Combining and Equalization in Digital Cellular Mobile Radio. *IEEE Trans. Veh. Technol.*, 40(2):342–354, May 1991.
- [4] D. Lu and K. Yao. Improved Importance Sampling Technique for Efficient Simulation of Digital Communication Systems. *IEEE J. Select. Areas Commun.*, 6(1), Jan. 1988.
- [5] K. S. Shanmugan and P. Balaban. A Modified Monte-Carlo Simulation Technique for the Evaluation of Error Rate in Digital Communication Systems. *IEEE Trans. Commun.*, COM-28(11):1916–1924, Nov. 1980.
- [6] B. R. Davis. An Improved Importance Sampling Method for Digital Communication System Simulations. *IEEE Trans. Commun.*, COM-34(7):715–719, Jul. 1986.
- [7] R. K. Bahr and J. A. Bucklew. Quick Simulation of Detector Error Probabilities in the Presence of Memory and Nonlinearity. *IEEE Trans. on Commun.*, 41:1610–1617, Nov. 1993.

- [8] J. A. Bucklew, P. Ney, and J. S. Sadowsky. Monte Carlo Simulation and Large Deviations Theory for Uniformly Recurrent Markov Chains. *J. Appl. Prob.*, 27:44–59, Mar. 1990.
- [9] J. S. Sadowsky and J. A. Bucklew. On Large Deviation Theory and Asymptotically Efficient Monte Carlo Estimation. *IEEE Trans. Inform. Theory*, IT-36(3):579–588, May 1990.
- [10] J-C. Chen, D. Lu, J. S. Sadowsky, and K. Yao. On Importance Sampling in Digital Communications — Part I: Fundamentals. *IEEE J. Select. Areas in Commun.*, 11(3):289–299, Apr. 1993.
- [11] W. Al-Qaq, M. Devetsikiotis, and J. K. Townsend. Simulation of Digital Communication Systems Using a Stochastically Optimized Importance Sampling Technique. In *Proc. IEEE Global Telecom. Conf., GLOBECOM '93*, Houston, Dec. 1993.
- [12] P. Y. Kam. Optimal Detection of Digital Data Over the Nonselective Rayleigh Fading Channel with Diversity Reception. *IEEE Trans. Commun.*, 39(2):214–219, Feb. 1991.
- [13] Pooi Yuen Kam. Optimal Detection of Digital Data Over the Nonselective Rayleigh Fading Channel with Diversity Reception. *IEEE Trans. Commun.*, 39(2):214–219, Feb. 1991.
- [14] M. Stojanovic, J. G. Proakis, and J. Catipovic. Analysis of the Performance of a Decision Feedback Equalizer on Fading Multipath Channels in the Presence of Channel Estimation Errors. In *Proc. of CISS*, Chicago, June 1993.
- [15] J. A. Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation*. New York: John Wiley & Sons, 1990.
- [16] John G. Proakis. *Digital Communications*. New York: McGraw-Hill, 1989.
- [17] Ed. W. C. Jakes, Jr. *Microwave Mobile Communications*. New York: Wiley, 1974.

- [18] Pooi Yuen Kam. Bit Error Probabilities of MDPSK Over the Nonselective Rayleigh Fading Channel with Diversity Reception. *IEEE Trans. Commun.*, 39(2):220–224, Feb. 1991.
- [19] S. S. Haykin. *Adaptive Filter Theory*. Englewood Cliffs, New Jersey: Prentice-Hall, 1986.
- [20] E. Eleftheriou and D. D. Falconer. Tracking Properties and Steady-State Performance of RLS Adaptive Filter Algorithms. *IEEE Trans. Acoust., Speech, Signal Processing*, 34(5):1097–1109, Oct. 1986.
- [21] John G. Proakis and Dimitris G. Manolakis. *Digital Signal Processing*. New York: Macmillan, 1988.
- [22] M. Devetsikiotis, W. Al-Qaq, J. A. Freebersyser, and J. K. Townsend. Stochastic Gradient Techniques for the Efficient Simulation of High-Speed Networks Using Importance Sampling. In *Proc. IEEE Global Telecom. Conf., GLOBECOM '93*, Houston, Dec. 1993.
- [23] W. Al-Qaq and J. K. Townsend. A Stochastic Importance Sampling Methodology for the Efficient Simulation of Digital Communication Systems with RLS Adaptive Equalizers. To be presented at *IEEE Int. Conf. Commun., ICC '94*.
- [24] M. C. Jeruchim. Techniques for Estimating the Bit Error Rate in the Simulation of Digital Communication Systems. *IEEE J. Select. Areas Commun.*, SAC-2(1):153–170, Jan. 1984.
- [25] Pooi Yuen Kam and Cho Huak Teh. Reception of PSK Signals Over Fading Channels Via Quadrature Amplitude Estimation. *IEEE Trans. Commun.*, 31(8):1024–1027, Aug. 1983.