Gradient Estimation for Stochastic Optimization of Optical Code-Division Multiple-Access Systems: Part I—Generalized Sensitivity Analysis

Narayan B. Mandayam, *Member, IEEE*, and Behnaam Aazhang, *Senior Member, IEEE*

***Abstract*— For optimizing the performance of optical code-division multiple-access (CDMA) systems, there is a need for determining the sensitivity of the bit-error rate (BER) of the system to various system parameters. Asymptotic approximations and bounds, used for system bit-error probabilities, seldom capture the sensitivities of the system performance. In this paper, we develop single-run gradient estimation methods for such optical CDMA systems using a discrete-event dynamic systems (DEDS) approach. Specifically, computer-aided techniques such as infinitesimal perturbation analysis (IPA) and likelihood ratio (LR) methods are used for analyzing the sensitivity of the average BER to a wide class of system parameters. It is shown that the above formulation is equally applicable to time-encoded and frequency-encoded systems. Further, the estimates derived are unbiased, and also optimality of the variance of these estimates is shown via the theory of common random variates and importance sampling techniques.**

***Index Terms*— Discrete-event simulations, frequency-encoded OCDMA, infinitesimal perturbation analysis, likelihood ratio method, time-encoded OCDMA.**

I. INTRODUCTION

OFTEN the performance criterion of optical CDMA systems is specified by means of the average bit-error rate (BER) of a given user and is usually given by

$$\overline{P}_e = E[L(\theta, \Psi)] \tag{1}$$

where θ could be some parameter (possibly vector-valued) specifying the characteristics of some system parameter, for example, the desired user codes, and Ψ, for instance, could characterize the random-valued interferers. For optimum performance of such systems, it is required to address the following optimization problem

$$\min_{\theta} \overline{P}_e = \min_{\theta} E[L(\theta, \Psi)] \tag{2}$$

which may or may not have additional constraints. Under these circumstances, it is imperative to compute the derivatives $(\partial/\partial\theta)\overline{P}_e$. However, the performance measures used for optical CDMA systems are often analytically intractable, that computing the bit-error probability in itself is a hard problem. Owing to this analytical intractability, the optimization of such systems (with respect to any class of parameters) seldom results in analytical solutions. An alternative and very effective approach that can be exploited for addressing this optimization is to use discrete-event simulations that have been used to address design issues in queuing systems and queuing networks [1], [2]. The goal of this paper is to develop efficient infinitesimal perturbation analysis (IPA) [3, p. 17] and likelihood ratio (LR) [4] techniques for evaluating derivatives (gradients) required in addressing problems of the type given in (2).

In particular, we address issues concerning the unbiasedness of the sensitivity estimates as well as variance reduction for these. Since these single-run gradient estimation techniques are based on inferences from a single sample path (or single simulation run), it is often required to use additional variance reduction techniques while estimating the sensitivities. We use the techniques of common random variates [5] and importance sampling [6]–[10] for reducing the variance of these gradient estimation methods. We illustrate the use of these sensitivity analysis techniques by studying the variation of the BER to unequal received powers as well as code parameters [11]. In the companion paper (see [12]), we address optimum detection for such systems, with the goal being the minimization of the bit-error rate as in (2).

The paper is organized as follows. We begin with a brief description of discrete-event dynamic systems (DEDS) and associated sensitivity analysis techniques. In Section III, the optical CDMA system being considered is described in a DEDS-based formulation. Using the above formulation, we illustrate IPA methods for studying the sensitivity of such systems by analyzing BER sensitivity to unequal received power effects. In Section V, the likelihood ratio method discussed for sensitivity analysis is illustrated by studying the sensitivity of these systems to code parameters. We finally present our conclusions in Section VI.

II. PRELIMINARIES

A. Discrete-Event Dynamical Systems (DEDS)

A DEDS can be represented by the pair (θ, Ψ), where θ is a parameter (possibly vector) of the DEDS and Ψ

Manuscript received May 5, 1996; revised October 25, 1996. This work was supported by the Advanced Technology Program of the Texas Higher Education Coordinating Board under Grant 003604-018. This paper was presented in part at the 1993 Conference on Information Sciences and Systems, The Johns Hopkins University, Baltimore, MD, March 1993.

N. B. Mandayam is with the Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08855-0909 USA.

B. Aazhang is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251-1892 USA.

Publisher Item Identifier S 0733-8716(97)02679-6.

is a random vector representing all the randomness in the system. Typically, the components of Ψ are random variables uniformly, independently, and identically distributed on [0, 1), i.e., Ψ is defined on the underlying probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and is a measurable mapping $\Psi\colon\ \Omega \rightarrow [0, 1)^d$ where d could be finite or infinite. The above representation is due to the fact that in computer simulations, the uniform random variable is used as the seed from which all random variables (vectors) are generated. For example, in an optical code-division multiple-access (OCDMA) system using $\{0, 1\}$ optical pulse sequences, θ could be the laser power of a user's transmission and Ψ could be the random vector whose components are the random powers due to interfering users. Each (θ, Ψ) determines a sample path of the system, and a performance measure obtained from such a sample path is denoted as $L(\theta, \Psi)$. For any realization of Ψ, $L(\theta, \Psi)$ is a function of θ, and is called the sample performance function. In most cases, we are interested in the expected value of the performance $J(\theta) = E[L(\theta, \Psi)]$, and sensitivity analysis is concerned with estimating $(\partial/\partial\theta)\, J(\theta)$. In both the PA and LR methods, the above quantity is estimated as $(\partial/\partial\theta)\, \widehat{J(\theta)} = (\partial L(\theta, \Psi)/\partial\theta)$, where the sample derivative calculated from a single sample path is defined as

$$\frac{\partial L(\theta, \Psi)}{\partial\theta} = \lim_{\Delta\theta \rightarrow 0} \frac{L(\theta + \Delta\theta, \Psi) - L(\theta, \Psi)}{\Delta\theta}. \tag{3}$$

The above estimate is said to be unbiased if and only if

$$E\left[\frac{\partial L(\theta, \Psi)}{\partial\theta}\right] = \frac{\partial}{\partial\theta}\, J(\theta) = \frac{\partial}{\partial\theta}\, E[L(\theta, \Psi)]. \tag{4}$$

Thus, the unbiasedness of the sensitivity estimate is equivalent to the interchangeability of the two operators E and $(\partial/\partial\theta)$. This can be satisfied by imposing some conditions on $L(\theta, \Psi)$ that can be specified using the Lebesgue dominated convergence theorem.

B. Perturbation Analysis (PA)

In perturbation analysis [3, p. 17], a sample path (θ, Ψ) (called the nominal sample path) is observed, and then a perturbed path $(\theta + \Delta\theta, \Psi)$ is constructed via a set of perturbation generation and perturbation propagation rules. Then the sensitivity estimate is given by

$$\frac{\partial}{\partial\theta}\, \widehat{J(\theta)} = \frac{L(\theta + \Delta\theta, \Psi) - L(\theta, \Psi)}{\Delta\theta} \tag{5}$$

where the perturbation in $\Delta\theta$ is such that the nominal and perturbed sample paths are "deterministically similar" [3, pp. 38–74]. In terms of discrete-event simulation, this would amount to two different simulation runs, but both having the same initial seed for the random number generator. Thus, this technique precludes using two different sample realizations for estimating the derivative.

C. Likelihood Ratio (LR) Method

If the probability density function of the underlying random process $f(\theta, \Psi)$ is "well behaved" (see the Appendix), then one can write

$$\frac{\partial J(\theta)}{\partial\theta} = E\left[L(\theta, \Psi)\, \frac{\partial \ln f(\theta, \Psi)}{\partial\theta}\right] \tag{6}$$

and the sensitivity analysis estimate in this case is known as the likelihood ratio estimate [4] and is given as

$$\frac{\partial}{\partial\theta}\, \widehat{J(\theta)} = \lim_{\Delta\theta \rightarrow 0} \frac{L(\theta, \Psi)}{\Delta\theta}\, \{\Lambda(\theta, \Delta\theta, \Psi) - 1\} \tag{7}$$

where $\Lambda(\theta, \Delta\theta, \Psi) = (f(\theta + \Delta\theta, \Psi)/f(\theta, \Psi))$ is the likelihood ratio. This estimate is unbiased and can be used when the form of the probability density function governing the underlying randomness in the system is explicitly known. This technique is simpler to implement than IPA since it involves only perturbing the likelihood function, and can be used as an alternative when the underlying probability density function is known. If the form of the probability density function is not known, then it is necessary to use IPA since it requires only the nominal and perturbed sample paths for sensitivity estimation.

III. System Description

The objective of the study reported here is to use discrete-event simulation techniques to analyze sensitivities, and eventually optimize performance of optical CDMA systems. We propose techniques that can be universally applied for optimization of both time-encoded as well as frequency-encoded optical multiple-access systems. In order to achieve this, a discrete-event dynamic system formulation is needed for such systems. We will consider an optical CDMA system (time-encoded) where an optical encoder maps each bit of information into a very high-rate optical sequence, which is then coupled into a single-mode fiber channel. At the receiver end, the optical pulse sequence is correlated to a stored replica of itself (correlation process), and the resulting sequence serves as the intensity that excites photoelectrons at the output of the photodetector. These photoelectrons are compared to a threshold at the comparator for data recovery. Consider a K-user optical CDMA system (see Fig. 1) where the users employ codes from a family of optical pulse sequences of length N_c. If user k is sending bit i in the time interval $[0, T)$, then the intensity of the modulated light will be $\lambda_i^{(k)}(t)$ where

$$\lambda_i^{(k)}(t) = \sum_{n=1}^{N_c} \lambda_i^{(k)}(n)\Pi_{T_c}(t - nT_c), \quad i = 0, 1;\ \text{for } t \in [0, T) \tag{8}$$

and $\Pi_{T_c}(t)$ is a unit rectangular pulse of duration T_c, and $\boldsymbol{\lambda}_i^{(k)} = [\lambda_i^{(k)}(1), \cdots, \lambda_i^{(k)}(N_c)]$ is a signature sequence of length $N_c = T/T_c$ with each $\lambda_i^{(k)}(n) \in \{0, \lambda_k\}$. This gives rise to the following two hypotheses at the receiver of the

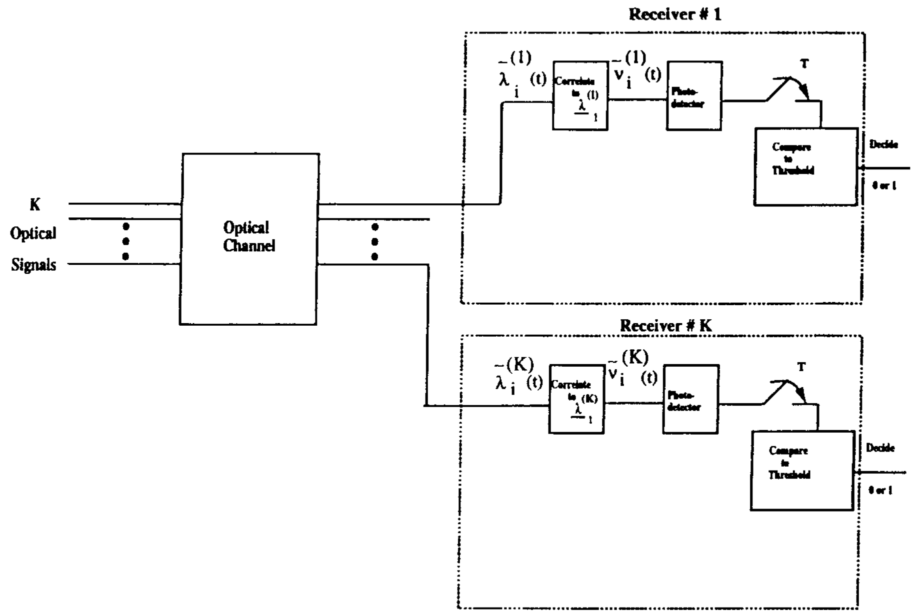


Fig. 1. K-user optical CDMA system with the receiver structure shown for users 1 and K.

desired user (taken to be user 1) in the time interval $[0, T)$ as

$$H_i\colon\ \tilde{\lambda}_i^{(1)}(t) = \lambda_i^{(1)}(t) + \sum_{k=2}^{K} \lambda_b^{(k)}(t), \qquad i = 0, 1 \tag{9}$$

where $\tilde{\lambda}_i^{(1)}(t)$ is the sum of the intensities in the channel under hypothesis H_i, due to the first user and $K-1$ interferers, and in $\lambda_b^{(k)}(t)$ the symbol b denotes the information of the kth user. The receiver corresponding to user 1 has a replica of the signature sequence assigned to user 1, and the light in the channel due to the user 1 and other $K-1$ interferers is correlated with this replicated signature sequence. Without loss of generality, we assume that each user is employing on–off keying, and hence at the first receiver, the intensities are correlated with $\boldsymbol{\lambda}_1^{(1)}$ since $\boldsymbol{\lambda}_0^{(1)} = \mathbf{0}$. For convenience, let $\lambda_1 = 1$, and $\lambda = \langle \boldsymbol{\lambda}_1^{(1)}, \boldsymbol{\lambda}_1^{(1)} \rangle$. Then the output of the correlator due to correlation at the nth chip interval, i.e., the nth observation from the sample path, is given by $\alpha_n + \beta_n$, where α_n is the contribution of the desired user sequence to the (auto) correlation process with $\sum_{n=1}^{N} \alpha_n = \lambda$, and β_n is the contribution of the interfering users to the (cross) correlation and is given by

$$\beta_n = \sum_{k=2}^{K} \Lambda_{k,n}, \qquad n = 1, \cdots, N_c \tag{10}$$

where $\Lambda_{k,n}$ is either λ_k or 0 with probability $p_{k,n}$ and $1-p_{k,n}$, respectively, $\forall k = 2, \cdots, K, \forall n = 1, \cdots, N_c$. The above probabilities depend on the cross-correlation properties of the codes, and also the relative delays of the users. From observing such a sample path, we can compute the sample performance function under hypothesis H_i (i.e., the probability of error from a single sample path) as

$$P_{e,i} = \sum_{r \in \Delta_{1-i}} \mathcal{K}(i, r), \qquad i = 0, 1 \tag{11}$$

where Δ_i is the decision region for the photoelectron count r that hypothesis H_i is true, and the kernel $\mathcal{K}(i, r)$ is given by

$$\mathcal{K}(i, r) = \frac{\left(i\lambda + \sum_{n=1}^{N_c} \beta_n \right)^r}{r!} \cdot \exp\left(-\left(i\lambda + \sum_{n=1}^{N_c} \beta_n \right) \right). \tag{12}$$

The average probability of bit error is given (for binary symmetric hypothesis) by

$$\overline{P}_e = \tfrac{1}{2}\overline{P}_{e,0} + \tfrac{1}{2}\overline{P}_{e,1}$$

where $\overline{P}_{e,i} = E[P_{e,i}]$, $i = 0, 1$.

We observe the output of the correlator at the desired user (see Fig. 2), and this constitutes the nominal sample path. Based on the observation of a single sample path, we first compute the sample performance function, and then, according to appropriate perturbation generation and propagation rules, we compute the sample derivative. Even though the above formulation has been presented for chip-synchronous codes, a similar analysis can be performed for chip-asynchronous codes as well. In this case, the observation is over a period $[0, T)$, which may not correspond to exactly N_c observations over each of the chips. In this case, $\Lambda_{k,n}$ can be either 0, $\tau\lambda_k$, $(T_c - \tau)\lambda_k$, or λ_k with corresponding probability (for a chip delay of τ).

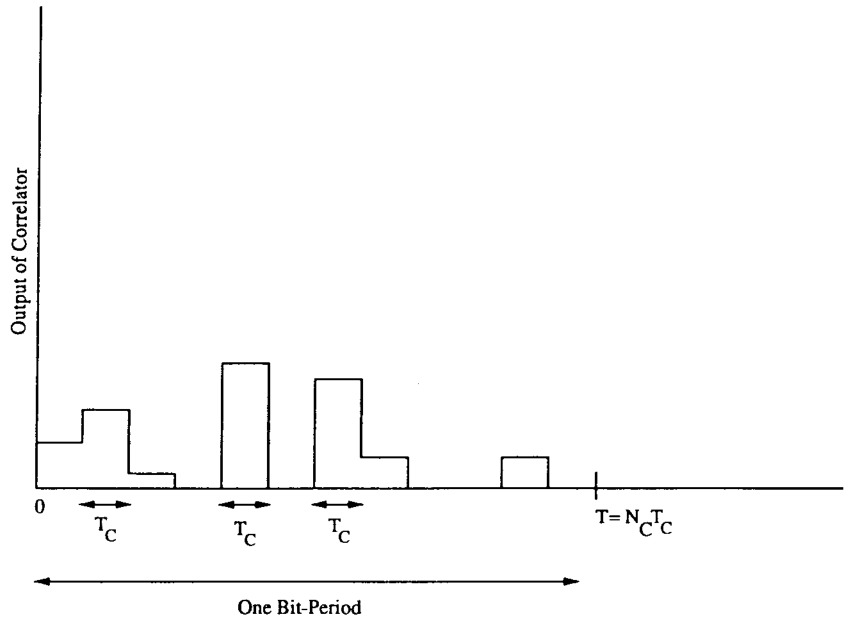


Fig. 2. Correlator output shown for an optical CDMA system over a sample realization of a bit period (for the chip-synchronous case).

For the case of the frequency-encoded system, the above formulation can be developed in an identical manner. In such an optical spectral amplitude CDMA system (see Fig. 3) [13], [14], the optical beam is passed through a grating and focused onto a mask. The grating angularly disperses the spectral content of the beam, so that the signal incident on the mask has spatially decomposed frequency content. The masks are length N_c sequences of open and closed slits of width $W_c = W/N_c$, where each user's optical source has an effective bandwidth W centered at frequency f_0 and a symbol rate of T. The pattern of the open and closed slits determines each user's codes. The result is a spread-spectrum encoding scheme that relies on on–off modulation of spectral bands. The encoded spectrum is then recombined spatially by passing through another grating. At the receiving end, the decoding of the signal is accomplished by passing the incident optical signal through another grating, and detecting the frequency bands individually by a linear array of photodiodes. The observation period at each of these photodetectors corresponds to the bit period, and owing to the superposition principle of Poisson random variables, the photoelectron counts at the outputs of these photodetectors can be combined to form a comprehensive statistic for detection (see [13]). Therefore, (12) can be written identically for the optical spectral amplitude encoded system, but the N_c now correspond to spectral bins. Therefore, without loss of generality, we will consider a time-encoded optical CDMA system in the remainder of the paper, and develop techniques for estimating the BER sensitivity. However, the methods presented here are equally applicable to frequency-encoded optical CDMA systems as well.

IV. IPA for Unequal Received Power Effects

In this section, we illustrate the use of IPA methods by analyzing the sensitivity of the probability of bit error to unequal received power effects. For a K-user optical CDMA system, for the kth user, let $\theta_{k,n}$ be the mean of random variable $\Lambda_{k,n}$. The sensitivity of $\overline{P}_{e,i}$ to $\theta_{k',n}$ for some k' is then estimated as

$$\frac{\widehat{\partial \overline{P}_{e,i}}}{\partial \theta_{k',n}} = \frac{\partial}{\partial \theta_{k',n}} \sum_{r\in\Delta_{1-i}} \mathcal{K}(i, r), \qquad i = 0, 1. \tag{13}$$

This estimate is then unbiased if and only if $E[(\widehat{\partial \overline{P}_{e,i}}/\partial\theta_{k',n})] = (\partial/\partial\theta_{k',n})\, E[P_{e,i}]$.

We will consider the case when the user powers are unequal, and show that the estimate in (13) is unbiased. We will first develop the following preliminaries. Let $(\theta_{k',n}, \Psi)$ denote the nominal sample path, and let $(\theta'_{k',n}, \Psi)$ denote the perturbed sample path where $\theta_{k',n}$ has been perturbed to $\theta'_{k',n} = \theta_{k',n} + \Delta\theta_{k',n}$. Now, $\Lambda_{k,n}$ is given by $\Lambda_{k',n} = (\theta_{k',n}/p_{k',n}) I_{(0,p_{k',n}]}(u_{k',n})$ where $u_{k',n}$ is a uniform random variable on $[0, 1)$, and $I_{(0,a]}(b)$ is 0 or 1 if $b > a$ or $b \le a$, respectively. According to the perturbation generation rule, $\Lambda'_{k',n}$ is given by

$$\Lambda'_{k',n} = \frac{\theta'_{k',n}}{p_{k',n}} I_{(0,p_{k',n}]}(u_{k',n}). \tag{14}$$

We will now state a proposition that establishes the unbiasedness of the sensitivity estimate.

Proposition 1: For the perturbation generation rule given in (14), the sensitivity estimate in (13) is an unbiased estimate.

Proof: The perturbation generation rule given in (14) holds even when the realization is $(\theta'_{k',n}-, \Psi)$. Therefore, $\Lambda'_{k',n}$ is left continuous at $\theta'_{k',n}$ except for the realization when $I_{(0,p_{k',n}]}(u_{k',n})$ is zero, i.e., the kth user is not contributing to the cross correlation at the nth chip. For the same reason, it is also right continuous at $\theta'_{k',n}$. Each term in the sum $\sum_{r\in\Delta_{1-i}} \mathcal{K}(i, r)$ is a continuous function of $\Lambda'_{k',n}$, i.e., for every value of r, $\mathcal{K}(i, r)$ is a continuous function of $\Lambda'_{k',n}$. Using the fact that the composition of continuous functions is continuous and the sum of continuous functions

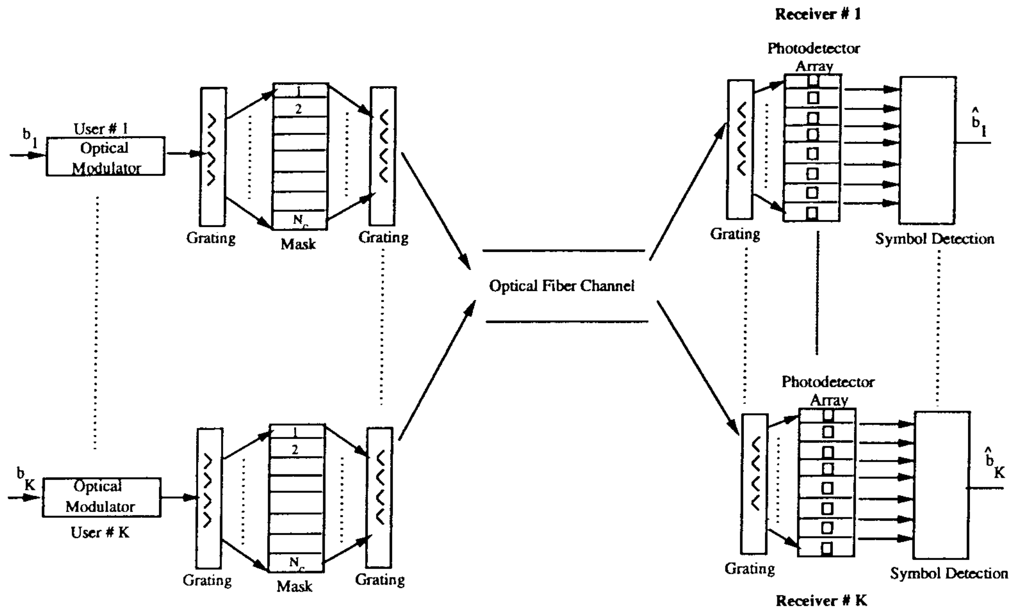


Fig. 3. K-user optical spectral amplitude CDMA system with the encoder and decoder structure shown for users 1 and K.

is continuous, it follows that $\sum_{r\in\Delta_{1-i}} \mathcal{K}(i, r)$ is an almost everywhere continuous function of $\theta_{k',n}$. It is easy to see that $\sum_{r\in\Delta_{1-i}} \mathcal{K}(i, r) \leq 1$, and hence is bounded. The proof now follows by the use of the Lebesgue dominated convergence theorem. □

The sensitivity of the average probability of error to the received power due to the kth user, i.e., $\theta_k = \sum_{n=1}^{N_c} \theta_{k,n}$, can now be computed using (13) and the chain rule. Thus, we see that the perturbation analysis estimate of the sensitivity of the error probability to user powers is unbiased. The issue of variance reduction for these estimates is discussed in the following.

A. Variance Reduction for IPA

In this section, we will show that the IPA estimates obtained for the optical CDMA system as described in the previous sections is the minimum variance estimate. We will use the theory of common random numbers (CRN) [5], which is a widely used variance reduction technique. Assume that X and Y are random variables with known marginal cumulative distribution functions (cdf's) F_X and F_Y, respectively. Moreover, X and Y are generated by the inverse transform method, i.e.,

$$X = F_X^{-1}(U_1) = \inf_b\{F_X(b) \geq U_1\}$$
$$Y = F_Y^{-1}(U_2) = \inf_b\{F_Y(b) \geq U_2\}$$

where U_1 and U_2 are uniformly distributed on [0, 1]. It has been shown that [15]

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2\,\text{cov}(X, Y)$$

is minimized when common random numbers are used, i.e., $U_1 = U_2 = U$. A generalization of the above result was given in [5], which is stated in the following proposition.

Proposition 2: For random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, the vector of common random numbers (VCRN), i.e., $\boldsymbol{U_1} = \boldsymbol{U_2} = \boldsymbol{U}$, is optimal for

$$\min\ \text{var}[g(\mathbf{X}) - h(\mathbf{Y})]$$

given that the marginal distributions are $F_{\boldsymbol{X}}(\boldsymbol{x})$ and $F_{\boldsymbol{Y}}(\boldsymbol{y})$, if the functions $g(\boldsymbol{x})$ and $h(\boldsymbol{y})$, $\boldsymbol{x}, \boldsymbol{y} \in \Re^n$ are monotonic in the same direction for each of the n respective components. Moreover, dependence is permitted only between like components of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

We note that the IPA estimate of sensitivity of the probability of bit error is essentially a VCRN estimate. This is due to the fact that the sample performance functions evaluated from the perturbed and nominal paths have the same initial seed vector.

Proposition 3: For the bit-error probability of an optical CDMA system given in (11), the IPA estimate evaluated from the sample path (at the output of the correlator) is optimal in that it has minimum variance.

Proof: The sample performance function evaluated at the output of the correlator is given as $P_{e,i} = \sum_{r\in\Delta_{1-i}} \mathcal{K}(i, r),\ i = 0, 1$ for both the nominal and the perturbed paths. This is merely the Poisson cdf for each realization, and the IPA estimate is simply the difference of the perturbed and nominal sample performance functions divided by the perturbation. The monotonicity of the two sample performance functions in the same direction follows from the fact that the Poisson cdf is a decreasing function of the mean. Using Proposition 2, the proof is complete. □

B. Numerical Results

We address the issue of sensitivity of the probability of bit error to user power for optical CDMA systems employing direct detection. Infinitesimal perturbation analysis is further used to study the unequal received power effects on single-user and multiuser detection of optical CDMA systems (see

Fig. 4). For such a system, the desired user's power is kept fixed, and the sensitivity of the error probability to the power of the interfering users is estimated as in (13). The slopes $(\partial \overline{P}_{e,i}/\partial \theta_{k'})$ are evaluated at different values of $\theta_{k'}$ by using IPA and averaging over several simulation runs, and then $\overline{P}_{e,i}$ is computed by a straight line interpolation as

$$\overline{P}_{e,i}(m) = \overline{P}_{e,i}(m-1) + \frac{\partial \overline{P}_{e,i}(m)}{\partial \theta_{k'}} \cdot (\theta_{k'}(m) - \theta_{k'}(m-1)) \tag{15}$$

where $\theta_{k'}(1) < \cdots < \theta_{k'}(m-1) < \theta_{k'}(m) < \theta_{k'}(m+1) < \cdots$ are the different points at which $(\partial \overline{P}_{e,i}/\partial \theta_{k'})$ is evaluated. Specifically, we consider the correlation detector [16]–[18], the optimal single-user detector [19], and the optimized single-user detector [19]. The correlation detector is optimal for a single user, while in the optimized single-user detector, the statistics of the interference are assumed to be known, and the detector is optimized for that interference distribution. The optimal single-user detector is one where the probability of error is minimum, i.e., it serves as the lower bound on the performance among all detectors for optical CDMA systems, and is also referred to as the "known-interference" detector. In Fig. 4, the unequal power effects for an optical CDMA system with $K = 15$ users and employing codes of length $N_c = 500$ and weight 50 are shown. The probability of bit error [obtained via (15)] is plotted against the near–far factor ν, which is defined as the ratio of the desired power to the interfering user power assuming all interferers are of equal power. For this example, IPA is used to estimate the derivative required in (15) from a single simulation run. Then, the above estimate is locally averaged over 100 simulation runs at each point of the above interpolation. Note that the results follow closely the simulations of the BER, which are also shown in the figure. For each point, the simulations require several runs approximately on the order of $10/\overline{P}_e$, while IPA requires only a few runs for locally averaging the derivative estimates. This is due to the optimality of the variance of these estimates established in the previous section. The relative merits of the different detectors is similar to that observed in [13]. As expected, the degradation due to unequal power effects is the most severe in the case of the correlation detector, while the optimized single-user detector degrades more gracefully, with the "known-interference" detector being the lower bound on the probability of bit error. Therefore, we see that IPA analysis is easily implemented for such systems, and is independent of the detection scheme. Further, these methods can be uniformly applied to analyze the sensitivity of optical CDMA systems to a variety of system parameters.

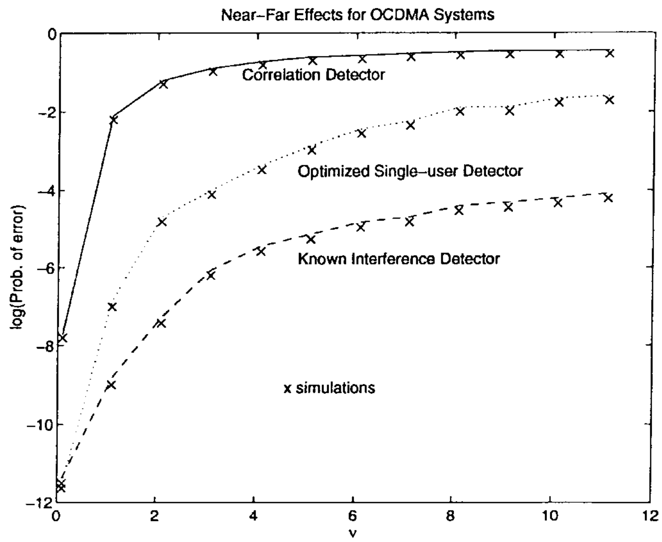


Fig. 4. Unequal received power effects for optical CDMA system. The degradation due to unequal received power levels is shown for the correlation detector, the optimized single-user detector, and the optimal single-user detector (known interference detector). These were plotted using infinitesimal perturbation analysis to estimate the gradient at each point. ν is the near–far factor as defined in [19].

V. Likelihood Ratio Method for Sensitivity to Codes

Having developed IPA for sensitivity analysis of optical CDMA systems, we now consider the other technique (LR) that is commonly used for DEDS. Specifically, we will illustrate the use of LR for studying the sensitivity of the probability of error to interference probabilities and code parameters such as code weights and code lengths. Consider an optical CDMA system where the K users employ $\{N_c, J\}$ optical codes, i.e., a family of sequences of length N_c and weight J. The chip length T_c is given as $T_c = T/N_c$, where T corresponds to the bit period. As before, let the desired user's (taken to be user 1) power be λ. Without loss of generality, let the interferer powers in each chip be either 0 or λ, depending on whether the interfering pulse due to that user is present or not. Since at each chip, there can be any of $\{0, 1, \cdots, K-1\}$ interferers, the probability of bit error of the desired user under hypothesis H_i can be written as

$$\overline{P}_{e,i} = \sum_{r \in \Delta_{1-i}} E_{\boldsymbol{p}}[\mathcal{K}(i, r)], \qquad i = 0, 1 \tag{16}$$

where $\mathcal{K}(i, r)$ is as defined in (12) and Δ_{1-i} is the decision region for H_i. The above equation is exactly the same as before, and we introduce different notation only to emphasize the fact that the parameter vector here is the probability vector $\boldsymbol{p}$. The above notation is also useful in establishing the fact that the LR technique requires knowing the underlying density functions since, in this case, the likelihood is perturbed as opposed to a sample path in IPA (see [3]). The expectation here is taken with respect to the probability vector $\boldsymbol{p}$, i.e., $\beta_n \in \{0, 1, \cdots, K-1\}$ with probabilities $\boldsymbol{p} = (p_0, p_1, \cdots, p_{K-1})$, respectively, $\forall n \in \{1, \cdots, N_c\}$.

We are interested in the sensitivity of $\overline{P}_{e,i}$ with respect to $\boldsymbol{p}$, i.e., the effect of perturbing the interference probabilities on the probability of bit error for the desired user. This quantity can then be directly related to the sensitivity to code parameters via some functional relationship between $\boldsymbol{p}$ and the code parameters like code length and code weight. Note that for an $\{N_c, J\}$ family of optical pulse sequences, there is some mapping $\mathcal{C}$: $\{N_c, J\} \rightarrow \boldsymbol{p}$, and therefore the chain rule of differentiation can be applied to arrive at the sensitivity of $\overline{P}_{e,i}$ to $\{N_c, J\}$. We will now present a theorem that yields an unbiased estimate of the sensitivity of $\overline{P}_{e,i}$ to $\boldsymbol{p}$. The following is similar to that due to Reiman and Weiss [20].

Theorem 1: Let $\boldsymbol{v}$ be a vector with $v_j = 0$ whenever $p_j = 0$, and $\langle \boldsymbol{v}, \mathbf{1}\rangle = \sum_{j=0}^{K-1} v_j = 0$; then

$$\widehat{\nabla_{\boldsymbol{p}} \overline{P}}_{e,i} = \left\langle \frac{\boldsymbol{N}}{\boldsymbol{p}}, \boldsymbol{v} \right\rangle \mathcal{K}(i, r) \tag{17}$$

is an unbiased estimate of the sensitivity of $\overline{P}_{e,i}$ with respect to $\boldsymbol{p}$, where $\boldsymbol{N} = (N_0, N_1, \cdots, N_{K-1})$ is the number of selections of $\{0, 1, \cdots, K-1\}$, respectively, in $[0, T)$, and $(\boldsymbol{N}/\boldsymbol{p})$ denotes componentwise division of the elements of the vectors.

Proof: Without loss of generality, we assume that each p_j is positive since $N_j = 0 = v_j$ when $p_j = 0$. We will show that

$$\lim_{\delta \to 0} \frac{E_{\boldsymbol{p}+\delta \boldsymbol{v}}[\mathcal{K}(i,r)] - E_{\boldsymbol{p}}[\mathcal{K}(i,r)]}{\delta} = E_{\boldsymbol{p}}\left[\left\langle \frac{\boldsymbol{N}}{\boldsymbol{p}}, \boldsymbol{v}\right\rangle \mathcal{K}(i, r)\right] \tag{18}$$

and then the proof follows via linearity of the expectation operator. The likelihood ratio of the two densities $dP_{\boldsymbol{p}+\delta\boldsymbol{v}}$ and $dP_{\boldsymbol{p}}$ can be written as

$$\frac{dP_{\boldsymbol{p}+\delta\boldsymbol{v}}}{dP_{\boldsymbol{p}}} = \left(1 + \frac{\delta v_0}{p_0}\right)^{N_0} \cdots \left(1 + \frac{\delta v_{K-1}}{p_{K-1}}\right)^{N_{K-1}}.$$

Then we can write

$$\frac{E_{\boldsymbol{p}+\delta \boldsymbol{v}}[\mathcal{K}(i,r)] - E_{\boldsymbol{p}}[\mathcal{K}(i,r)]}{\delta} = \int \mathcal{K}(i, r)\left(\left\langle \frac{\boldsymbol{N}}{\boldsymbol{p}}, \boldsymbol{v}\right\rangle + A_\delta\right) dP_{\boldsymbol{p}}$$

where

$$A_\delta = \frac{\prod_{j=0}^{K-1} \left(1 + \frac{\delta v_j}{p_j}\right)^{N_j} - 1}{\delta} - \left\langle \frac{\boldsymbol{N}}{\boldsymbol{p}}, \boldsymbol{v}\right\rangle.$$

We note that $\lim_{\delta \to 0} A_\delta = 0$. As before, $\mathcal{K}(i, r)$ is bounded. A_δ is also bounded because $v_j = 0$ whenever $p_j = 0$, and N_j is bounded for $j = \{0, \cdots, K-1\}$ (i.e., there are only a finite number of users in the system and a finite number of chips N_c). Therefore, $|\mathcal{K}(i, r)(\langle \boldsymbol{N}/\boldsymbol{p}, \boldsymbol{v}\rangle + A_\delta)|$ is bounded, and the proof is complete by the Lebesgue dominated convergence theorem. □

The above theorem yields an unbiased estimate of the sensitivity of the error probability (to the interference probabilities) that can be computed from a single sample path, i.e., $\mathcal{K}(i, r)$ and $\langle(\boldsymbol{N}/\boldsymbol{p}), \boldsymbol{v}\rangle$ can be evaluated by observing a sample path. In the following, we will develop a method for reducing the variance of such estimates.

A. Variance Reduction for Likelihood Ratio Methods

We will consider without loss of generality the case of a CDMA system with a symmetric detection problem (i.e., the error under each hypothesis is identical). The probability of error for such a system is given by

$$\overline{P}_e = \int I_\Delta(r) f(\theta, r) \tag{19}$$

where $I_\Delta(r)$ is the indicator function over the region of an incorrect decision and $f(\theta, r)$ is the probability density function (parameterized by θ) of the underlying randomness in the system. The Monte Carlo estimate of the error probability is given by $\widehat{\overline{P}_e} = (1/M) \sum_{i=1}^{M} I_\Delta(r_i)$. The LR estimate of the sensitivity of $\overline{P}_e$ to the parameter θ is then given by

$$\widehat{\frac{\partial}{\partial\theta}\overline{P}}_e = \frac{1}{M} \sum_{i=1}^{M} I_\Delta(r_i) \frac{\partial}{\partial\theta} \ln f(\theta, r_i) \tag{20}$$

where the samples r_i are generated from $f(\theta, r)$. The above estimator is meaningful only when the probability density function $f(\theta, r)$ satisfies the regularity conditions imposed in (6). Thus, while estimating the error probability, simultaneously one can estimate the sensitivity as well. However, the estimator in (20) suffers from high variance [3, pp. 257–258], and therefore there is a need for variance reduction methods. We use the importance sampling method of variance reduction [6]–[10] and rewrite the estimator in (20) as

$$\widehat{\frac{\partial}{\partial\theta}} P_e^* = \frac{1}{M} \sum_{i=1}^{M} I_\Delta(r_i^*) \frac{\partial}{\partial\theta} \ln f(\theta, r_i^*) \frac{f(\theta, r_i^*)}{f^*(\theta, r_i^*)} \tag{21}$$

where the samples r_i^* are generated from $f^*(\theta, r)$, which serves as the biasing density. Clearly, the estimators in both (20) and (21) are unbiased. The problem at hand is to choose an appropriate $f^*(r, \theta)$ such that the variance of the estimator in (21) is reduced as compared to the variance of that in (20). After some algebraic manipulation, it can be shown that the variance of the standard Monte Carlo estimator in (20) is

$$\operatorname{var}\left(\widehat{\frac{\partial}{\partial\theta}} \overline{P}_e\right) = \frac{1}{M}\left\{E_f\left[I_\Delta(r)\left(\frac{\partial}{\partial\theta} \ln f(\theta, r)\right)^2\right] - \left(\frac{\partial}{\partial\theta}\overline{P}_e\right)^2\right\} \tag{22}$$

and the variance of the estimator in (21) is

$$\operatorname{var}\left(\widehat{\frac{\partial}{\partial\theta}} P_e^*\right) = \frac{1}{M}\left\{E_{f^*}\left[I_\Delta(r)\left(\frac{\partial}{\partial\theta}\frac{f}{f^*} + \frac{f}{f^*}\frac{\partial}{\partial\theta} \cdot \ln f^*(r, \theta)\right)^2\right] - \left(\frac{\partial}{\partial\theta}\overline{P}_e\right)^2\right\} \tag{23}$$

where the expectations in (22) and (23) are taken with respect to f and f^*, respectively. We will now determine a biasing density f^* that minimizes the variance in (23). It can be easily seen that the optimal biasing density that satisfies

$$\min_{f^*} \operatorname{var}\left(\widehat{\frac{\partial}{\partial\theta}} P_e^*\right) \tag{24}$$

is degenerate, i.e., $f_{\text{opt}}^* = ((\partial/\partial\theta) f/(\partial/\partial\theta)\overline{P}_e) I_\Delta(r)$, and thus depends on $(\partial/\partial\theta)\overline{P}_e$. Therefore, we will look for suboptimal solutions that reduce the variance of the estimator in

(23). From (22) and (23), it can be shown that the sufficient condition for reducing the variance of the estimator in (23) with respect to (22) is given by $f < f^*, \forall r$ such that $I_\Delta(r) = 1$. This sufficient condition is exactly the same as that required for the importance sampling method for estimating the probability of error for optical CDMA systems [9]. Thus, we see that importance sampling simulations of both the error probability as well as its sensitivity can be performed simultaneously. We use the importance sampling technique developed for optical CDMA systems [9], [10], where the sub-optimal biasing density was derived through a combination of the above sufficient conditions and an exponential change of measure [21], on the original density. The asymptotic optimality of this method is discussed in the following.

B. Asymptotic Optimality

We first begin by rewriting the estimator in (20) in its finite-difference version as

$$\frac{d\overline{P}_e}{d\theta} = \frac{1}{\Delta\theta}\left[\int I_\Delta(r)\,\frac{dF_{\theta+\Delta\theta}}{dF_\theta}\,dF_\theta - \int I_\Delta(r)\,dF_\theta\right] \quad (25)$$

where $F_{\theta+\Delta\theta}$ and F_θ are the probability measures corresponding to the perturbed and nominal densities, and $(dF_{\theta+\Delta\theta}/dF_\theta)$ is the Radon–Nikodym derivative of the perturbed probability measure with respect to the nominal one. Analogously, we can rewrite the importance sampling estimator in (21) as

$$\frac{dP_e^*}{d\theta} = \frac{1}{\Delta\theta}\left[\int I_\Delta(r)\,\frac{dF_{\theta+\Delta\theta}}{dF^*{}_\theta}\,dF_\theta^* - \int I_\Delta(r)\,\frac{dF_\theta}{dF^*{}_\theta}\,dF_\theta^*\right] \quad (26)$$

where F_θ^* is the probability measure corresponding to the biasing density, and $(dF_{\theta+\Delta\theta}/dF_\theta^*)$ and (dF_θ/dF_θ^*) are the appropriate Radon–Nikodym derivatives. In general, for any two measures $F_{\theta+\Delta\theta}$ and F_θ on the σ field $\mathcal{F}$ of subsets of Ω, $F_{\theta+\Delta\theta}$ is said to be absolutely continuous with respect to F_θ (i.e., $F_{\theta+\Delta\theta} \ll F_\theta$) if and only if $F_\theta(A) = 0$ implies $F_{\theta+\Delta\theta}(A) = 0\,(A \in \mathcal{F})$. Furthermore, if $F_{\theta+\Delta\theta} \ll F_\theta$, then $(dF_{\theta+\Delta\theta}/dF_\theta)$ is well defined. The derivative $(dF_{\theta+\Delta\theta}/dF_\theta)$ is well defined because $F_{\theta+\Delta\theta} \ll F_\theta$ (i.e., the perturbed and the nominal probability densities have the same support). Before addressing the derivatives $(dF_{\theta+\Delta\theta}/dF_\theta^*)$ and (dF_θ/dF_θ^*), we need to first define an exponential change of measure [21, p. 13].

Definition 1: An exponential change of measure from F to F^* is defined as $dF^* = (e^{sx}\,dF/M(s))$, where $s \in \Re$ and $M(s)$ is the corresponding moment-generating function.

The above definition implies that $F_\theta \ll F_\theta^*$, and hence $F_{\theta+\Delta\theta} \ll F_\theta^*$. Therefore, the Radon–Nikodym derivatives in (26) are well defined, and the weights for the importance sampling estimator can be written as a function of $(dF_{\theta+\Delta\theta}/dF_\theta^*) - (dF_\theta/dF_\theta^*)$. From the sufficient conditions established for variance reduction, it follows that minimizing the weights minimizes the variance of the importance sampling estimator.

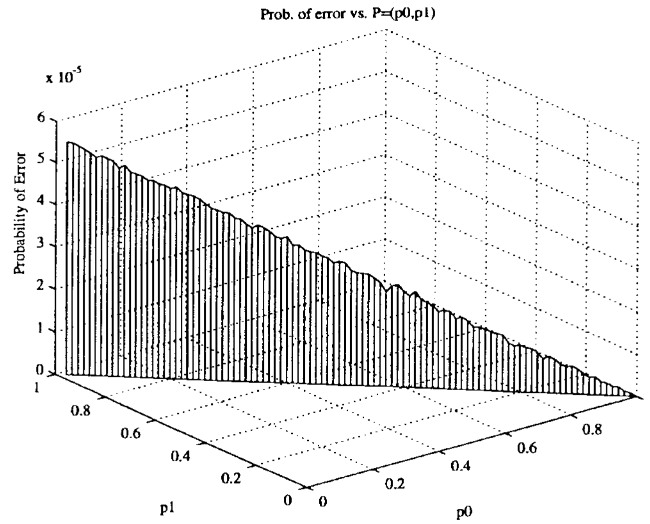


Fig. 5. Probability of interference versus P_e for optical CDMA system. The LR estimates of the variation of the error probability with the probability of no interferer (p_0) and the probability of one interferer (p_1) is shown for a two-user optical CDMA system using the correlation detector and employing codes of length $N_c = 500$.

Theorem 2: Let $A \in \mathcal{F}$ be the set which maps to $I_\Delta(r) = 1$. Then $\lim_{P_e \to 0} F_{\theta+\Delta\theta}(A) \to 0$.

The proof of the above theorem is presented in the Appendix. Since $F_\theta^* \ll F_\theta$ and $F_{\theta+\Delta\theta} \ll F_\theta^*$, we see from the theorem that the weights approach zero as the probability of error approaches zero, i.e., the variance of the importance sampling estimator approaches zero for arbitrarily small values of the error probability. Thus, the importance sampling method using an exponential change of measure is asymptotically optimal for the LR method. This corresponds to the importance sampling method established in [9] and [10] for estimating the probability of bit error for optical CDMA systems, and can be used simultaneously for estimating the sensitivity of the probability of bit error via the LR method.

C. Numerical Results

In this section, we illustrate the use of LR techniques by estimating the sensitivity of bit-error probabilities to parameters such as interference probabilities, code weight J, and code length N_c. The variance of these estimates is reduced using the importance sampling method described in [9]. In Fig. 5, the probability of error is plotted against the probabilities of having one interferer (p_1) and no interferer (p_0) for a two-user optical CDMA system employing codes of length $N_c = 500$ and a correlation detector. The values of $\widehat{\overline{P}_{e,i}}$ are evaluated at different values of p_1 and p_0 which satisfy $p_0 + p_1 = 1$ by using a linear interpolation as in (15). The gradient at each point $(p_0, p_1, \widehat{\overline{P}_{e,i}})$ along $p_0 + p_1 = 1$ is given by the LR estimate in (18). For the above example, the number of trials required to derive the gradient estimates is similar to that required for the importance sampling method described in [9]. It is seen that the probability of error decreases in the direction of increasing p_0. This is very intuitive owing to the fact that the system employs a correlation detector. The issue of greater interest is the sensitivity of the error

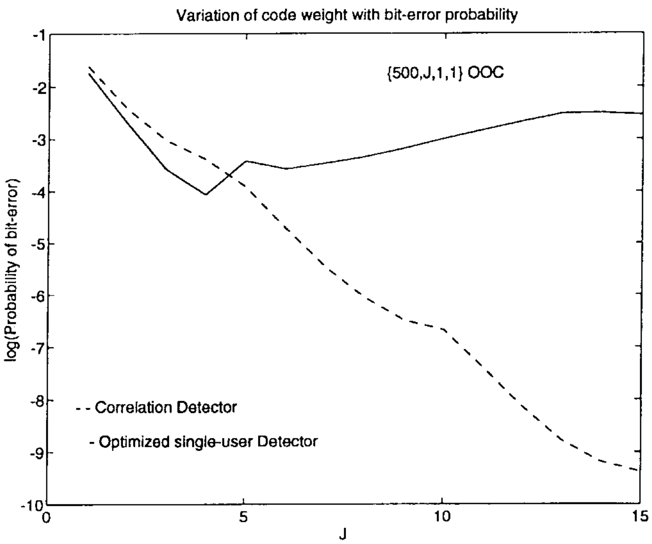


Fig. 6. Code weight versus P_e for optical CDMA system. The LR estimates of the probability of error of the desired user is shown as the code weight J varies for a two-user optical CDMA system using the correlation detector (—) and the optimized single-user detector (–). The system employed OOC's from a $\{500,\, J,\, 1,\, 1\}$ family. For the optimized single-user detector, the received power of the interferer was ten times stronger than that of the desired user, while for the correlation detector, the interferer was ten times weaker.

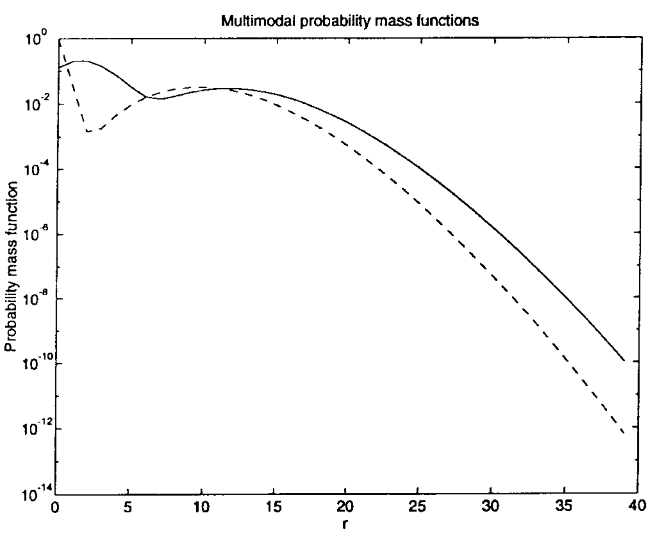


Fig. 7. Probability mass functions for photoelectron count. The probability mass functions for the photoelectron count are shown for a two-user optical CDMA system employing OOC's from a $\{500,\, J,\, 1,\, 1\}$ family for hypotheses H_0 (–) and H_1 (—). The received power of the interferer was ten times stronger than that of the desired user. It is seen that these probability mass functions have multiple modes.

probability to code parameters. For instance, in the case of optical orthogonal codes (OOC) with $\rho_c = 1$ [16], the mapping $\mathcal{C}\colon\ \{N_c,\, J\} \rightarrow \boldsymbol{p}$ is given by $(p_0,\, p_1) = (1 - J^2/N_c^2,\, J^2/2N_c^2)$, and the sensitivity of the error probability to the code parameter J can be evaluated by using the estimate in (18) and chain rule as $\langle \boldsymbol{x},\, \boldsymbol{y}\rangle$ where the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are given by $\boldsymbol{x} = ((\partial \widehat{\boldsymbol{P}}_{\boldsymbol{e},\boldsymbol{i}}/\partial \boldsymbol{p_0}),\, (\partial \widehat{\boldsymbol{P}}_{\boldsymbol{e},\boldsymbol{i}}/\partial \boldsymbol{p_1}))'$ and $\boldsymbol{y} = ((\partial \boldsymbol{p_0}/\partial \boldsymbol{J}),\, (\partial \boldsymbol{p_1}/\partial \boldsymbol{J}))'$. Since J is always integer-valued, we use finite-difference versions of the estimate $(\partial p_i/\partial J)$. In Fig. 6, the bit-error probability is plotted for a two-user optical CDMA system using a correlation detector and employing codes with $\{500,\, J,\, 1,\, 1\}$ against the code weights J. As before, a linear interpolation scheme is used by evaluating the slopes for different values J. The effect of code weight on error probability is found to behave as predicted in [16]. However, this intuitive result does not hold for the case of a system using the optimized single-user detector. As an illustration, we show the results of the likelihood ratio estimation for a two-user system employing codes with $\{500,\, J,\, 1,\, 1\}$ in Fig. 6. The received power of the interferer is ten times stronger than that of the desired user in this example. It is seen that the minimum probability of bit error occurs for $J = 4$, and is not decreasing with increasing code weight. This is due to the fact that the probability mass functions governing the exact photoelectron count are multimodal for this scenario, and hence there exist multiple threshold levels for detection. To illustrate this fact, the probability mass functions for the above example are plotted in Fig. 7. Therefore, similar to IPA, LR techniques described above can be used to derive sensitivities of the BER, and further used to determine code parameters for optimum performance. This method is once again independent of the detection scheme, and can be uniformly applied to analyze the sensitivity of optical CDMA systems to a wide class of parameters.

VI. Conclusion

In this paper, we embarked upon optical CDMA systems from a discrete-event dynamic system point of view. This approach allows us to design and analyze the sensitivity of these systems based upon observations of sample paths, and thus is independent of the constraints imposed on the system due to detection and coding schemes. Further, the sensitivity analysis techniques are equally applicable to time-encoded as well as frequency-encoded optical CDMA systems. In particular, we illustrated the use of infinitesimal perturbation analysis by analyzing the sensitivity of the error probability of such systems to unequal received power effects. Similarly, the likelihood ratio method was illustrated by deriving estimates of the sensitivity of the error probability to code parameters. The techniques were shown to yield unbiased estimates of the sensitivity of the error probability, and the variance of these estimates was shown to be optimal via the theory of common random variates and importance sampling. Thus, the techniques developed in this paper are very robust, and yield a powerful tool for computer-aided design and sensitivity analysis of such systems without encountering problems of bounds, approximations, and asymptotics on performance. In the companion paper [12], IPA-based stochastic gradient algorithms are used for developing a class of optimum detectors that minimize the average bit-error rate.

Appendix

A. Characterization of Probability Density Functions for Likelihood Ratio Method

We now present a theorem that characterizes probability density functions for which the likelihood ratio method of sensitivity estimation is valid. For convenience, we rewrite $P_e = \int I_\Delta(r) f(\theta, r)\, dr$, where the probability density function f is parameterized by some parameter θ. Without loss of generality, consider $\theta \in \Re$. The likelihood ratio estimate of $(dP_e/d\theta)$ is given as

$$\frac{dP_e}{d\theta} = \int I_\Delta(r) f_1(\theta, r)\, dr \tag{27}$$

where $f_1(\theta, r)$ is the partial derivative of f with respect to θ. The characterization of the probability densities for which (27) holds is given by the following theorem.

Theorem 3: If $f_1(\theta, r)$ exists $\forall\, r \in \Delta \equiv \{r\colon I_\Delta(r) = 1\}$, and if there exists a Borel measurable function $h\colon \Delta \to \Re$ such that $|f_1(\theta, r)| \le h(r), \forall\, r, \theta$ where $\int h(r)\, dr < \infty$, then $(dP_e/d\theta)$ exists $\forall\, \theta$ and is given as in (27).

Proof: Let $\theta \in \Theta$. Consider $\theta_0 \in \Theta$, and let $\theta_n \to \theta_0, \theta_n \neq \theta_0$. Then

$$\frac{1}{\theta_n - \theta_0}\left[\int I_\Delta(r) f(\theta_n, r)\, dr - \int I_\Delta(r) f(\theta_0, r)\, dr\right] = \int I_\Delta(r) \frac{[f(\theta_n, r) - f(\theta_0, r)]}{\theta_n - \theta_0}\, dr.$$

By the mean value theorem

$$\frac{[f(\theta_n, r) - f(\theta_0, r)]}{\theta_n - \theta_0} = f_1(\lambda_n, r)$$

for some $\lambda_n = \lambda_n(r)$ between θ_n and θ_0. By the hypothesis, $|f_1(\lambda_n, r)| \le h(r)$, where h is integrable, and the result follows from the dominated convergence theorem (since $([f(\theta_n, r) - f(\theta_0, r)]/\theta_n - \theta_0) \to f_1(\theta_0, r)$, $f_1(\cdot, \theta)$ is Borel measurable for each θ). □

B. Asymptotic Optimality of Exponential Change of Measure

We now prove Theorem 2 for which we will need the following two theorems, the proofs of which can be found in [22, pp. 10–43].

Theorem 4: If μ is countably additive on the σ field $\mathcal{F}$, and if $A_1, A_2, \cdots, \in \mathcal{F}$, $A_n \downarrow A$, and $\mu(A_1)$ is finite, then $\mu(A_n) \to \mu(A)$ as $n \to \infty$.

Theorem 5: Let h be a Borel measurable function such that $\int_\Omega h\, d\mu$ exists. Define $\lambda(A) = \int_A h\, d\mu$, $A \in \mathcal{F}$. Then λ is countably additive on $\mathcal{F}$.

We are now ready to prove Theorem 2, which we shall state here once again for convenience.

Theorem 6: Let $A \in \mathcal{F}$ be the set where $I_\Delta(r) = 1$. Then

$$\lim_{P_e \to 0} F_{\theta + \Delta\theta}(A) \to 0.$$

Proof: Now, $P_e = \int_A dF_\theta$, therefore, $P_e \to 0$ is equivalent to $F_\theta(A) \to 0$. However, $F_\theta^* \ll F_\theta$, implying that $F_\theta^*(A) \to 0$. Since $F_{\theta+\Delta\theta} \ll F_\theta^*$, we can write

$$F_{\theta+\Delta\theta}(A) = \int_A \frac{M(s)}{e^{sx}}\, g\, dF_\theta^*$$

where g is the Radon–Nikodym derivative $(dF_{\theta+\Delta\theta}/dF_\theta^*)$. Further

$$F_{\theta+\Delta\theta}(A) \le M_e \int_A |g|\, dF_\theta^*$$

where M_e is the maximum value that $(M(s)/e^{sx})$ takes over the set A. Therefore, we could write

$$\begin{aligned} F_{\theta+\Delta\theta}(A) &\le M_e \left(\int_{A \cap \{|g| \le n\}} |g|\, dF_\theta^* + \int_{A \cap \{|g| > n\}} |g|\, dF_\theta^* \right) \\ &\le M_e n F_\theta^*(A) + M_e \int_{\{|g| > n\}} |g|\, dF_\theta^*). \end{aligned}$$

We can now use Theorems 4 and 5 to make the second integral arbitrarily small for some large n, say $n \ge N$. Therefore, taking $\int_{\{|g|>n\}} |g|\, dF_\theta^*) < \epsilon/2M_e$ and $F_\theta^*(A) < \epsilon/2NM_e$, we have

$$F_{\theta+\Delta\theta}(A) \le \epsilon$$

and the proof is complete. □

REFERENCES

[1] R. Suri and Y. T. Leung, "Single run optimization of discrete event simulations—An empirical study using the M/M/1 queue," *IIE Trans.*, vol. 21, pp. 35–49, Mar. 1989.

[2] E. K. P. Chong and P. J. Ramadge, "Optimization of queues using an infinitesimal perturbation analysis-based stochastic algorithm with general update times," *SIAM J. Contr. Optimiz.*, vol. 31, May 1993.

[3] Y.-C. Ho and X.-R. Cao, *Perturbation Analysis of Discrete Event Dynamic Systems.* Boston: Kluwer Academic, 1991.

[4] V. Aleksandrov, V. Sysoyev, and V. Shemeneva, "Stochastic optimization," *Eng. Cybern.*, vol. 5, pp. 11–16, 1968.

[5] R. Y. Rubinstein and G. Samorodnitsky, "Variance reduction by the use of common and antithetic random variables," *J. Statist. Computation and Simulation*, vol. 22, pp. 161–180, 1985.

[6] D. Lu and K. Yao, "Improved importance sampling technique for efficient simulation of digital communication systems," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 67–75, Jan. 1988.

[7] G. Orsak and B. Aazhang, "On the theory of importance sampling applied to the analysis of detection systems," *IEEE Trans. Commun.*, vol. 30, pp. 332–339, Apr. 1989.

[8] J. Bucklew, P. Ney, and J. Sadowsky, "Monte Carlo simulation and large deviation theory for uniformly recurrent Markov chains," *J. Appl. Prob.*, vol. 27, pp. 44–59, Mar. 1990.

[9] N. B. Mandayam and B. Aazhang, "Importance sampling for analysis of direct-detection optical communication systems," *IEEE Trans. Commun.*, vol. 43, pp. 229–239, Feb. 1995.

[10] K. Ben Letaief, "Performance analysis of digital lightwave systems using efficient computer simulation techniques," *IEEE Trans. Comm.*, vol. 43, pp. 240–251, Feb. 1995.

[11] N. B. Mandayam and B. Aazhang, "Generalized sensitivity analysis of optical code division multiple access systems," in *1993 Annu. Conf. Inform. Sci. Syst.*, Mar. 1993.

[12] ——, "Gradient estimation for stochastic optimization of optical code-division multiple-access systems: Part II—Adaptive detection," this issue, pp. 742–750.

[13] M. Brandt-Pearce, "High throughput optical code-division-multiple-access communication systems," Ph.D. dissertation, Rice Univ., Houston, TX, 1993.

[14] M. Brandt-Pearce and B. Aazhang, "Optical spectral amplitude code division multiple access system," in *Proc. Int. Symp. Inform. Theory*, San Antonio, TX, Jan. 1993.

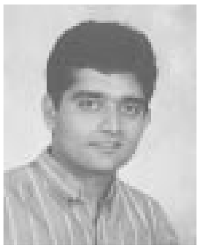
[15] W. Whitt, "Bivariate distributions with given marginals," *Annals Statist.*, vol. 4, pp. 1280–1289, 1976.

[16] J. A. Salehi, "Code division multiple access techniques in optical fiber networks: Part I—Fundamental principles," *IEEE Trans. Commun.*, vol. 37, pp. 824–833, Aug. 1989.

[17] J. A. Salehi and C. A. Brackett, "Code division multiple access techniques in optical fiber networks: Part II—Systems performance analysis," *IEEE Trans. Commun.*, vol. 37, pp. 834–842, Aug. 1989.

[18] A. W. Lam and A. M. Hussain, "Performance analysis of code division multiple access optical communications with avalanche photodetectors," in *Proc. Conf. Inform. Sci. Syst.*, Mar. 1989.

[19] M. Brandt-Pearce and B. Aazhang, "Performance analysis of single-user and multi-user detectors for optical code division multiple access communication systems," *IEEE Trans. Commun.*, vol. 43, pp. 435–444, Feb./Mar./Apr. 1995.
[20] M. I. Reiman and A. Weiss, "Sensitivity analysis for simulations via likelihood ratios," *Oper. Res.*, vol. 37, pp. 830–844, Oct. 1989.
[21] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation.* New York: Wiley, 1990.
[22] R. Ash, *Real Analysis and Probability.* New York: Academic, 1972.



Narayan B. Mandayam (M'94) received the B.Tech. (Hons.) degree in 1989 from the Indian Institute of Technology, Kharagpur, and the M.S. and Ph.D. degrees in 1991 and 1994, respectively, from Rice University, Houston, TX, all in electrical engineering.

From 1994 to 1996, he was a Research Associate at the Wireless Information Network Laboratory (WINLAB), Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ. In September 1996, he joined the ECE Department at Rutgers as an Assistant Professor, and also serves as the Research Director for Radio Systems at WINLAB. His research interests are in communication theory, radio and optical multiple-access communications, discrete-event dynamic systems, and multimedia wireless communications for cellular and PCS.

Dr. Mandayam was a recipient of the Government of India National Talent Scholarship, and the Institute Silver Medal, awarded to the best graduating student by the EE Department, Indian Institute of Technology, Kharagpur.

Behnaam Aazhang (S'81–M'85–SM'91) was born in Bandar Anzali, Iran, on December 7, 1957. He received the B.S. (with highest honors), M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana–Champaign in 1981, 1983, and 1986, respectively.

From 1981 to 1985, he was a Research Assistant in the Coordinated Science Laboratory, University of Illinois. In August 1985, he joined the faculty of Rice University, Houston, TX, where he is now a Professor in the Department of Electrical and Computer Engineering. He has been a Visiting Professor at IBM Federal Systems Company, Houston, the Institut für Kommunikationstechnik at Swiss Federal Institute of Technology (ETH), Zürich, Switzerland, the Telecommunications Laboratory at the University of Oulu, Oulu, Finland, and at the U.S. Air Force Phillips Laboratory, Albuquerque, NM. His research interests are in the areas of communication theory, information theory, and their applications, with emphasis on multiple-access communications, cellular mobile radio communications, optical communication networks, and the efficient simulation of stochastic systems.

Dr. Aazhang is a recipient of the Alcoa Foundation Award 1993, the NSF Engineering Initiation Award 1987–1989, and an IBM Graduate Fellowship 1984–1985, and is a member of Tau Beta Pi and Eta Kappa Nu. He is currently serving as the Editor for Spread Spectrum Networks of IEEE TRANSACTIONS ON COMMUNICATIONS and as the Treasurer of the IEEE Information Theory Society. He served as the Secretary of the Information Theory Society from 1990 to 1993, and as the Publications Chairman of the 1993 IEEE International Symposium on Information Theory, San Antonio, TX.