PERFORMANCE EVALUATION AND PARAMETER OPTIMIZATION OF WAVELENGTH DIVISION MULTIPLEXING NETWORKS WITH ADAPTIVE IMPORTANCE SAMPLING TECHNIQUES

David Remondo, Rajan Srinivasan, Victor F. Nicola, Wim C. van Etten, and Henk E.P. Tattje
University of Twente
Department of Electrical Engineering
Postbus 217
7500 AE Enschede
The Netherlands

ABSTRACT

In this paper new adaptive importance sampling techniques are applied to the performance evaluation and parameter optimization of a wavelength division multiplexing (WDM) network impaired by crosstalk in an optical cross-connect. Worst-case analysis is carried out including all the beat noise terms originated by in-band crosstalk. Both input signal hypotheses are considered. The accurate bit-error-rate estimates, which are obtained in short run-times, indicate that the influence of crosstalk is much lower than that predicted by previous analyses. This finding has a strong impact on the design of WDM networks. Besides, a method is used to optimize the detection threshold, which turns out to improve the system performance significantly. The presented techniques also allow us to determine the power penalty due to the introduction of additional WDM channels.

I. INTRODUCTION

In wavelength division multiplexing (WDM) systems several information channels can be transmitted along the same optical fiber by using different wavelengths. The main advantage of WDM is that communication networks can be easily reconfigured to adapt to varying traffic demands, without changing the physical layout.

A fundamental element in WDM networks is the all-optical wavelength router, also called optical cross-connect. For this purpose, an arrayed-waveguide grating (AWG) seems to be a good candidate (see, e.g. [1]). In this device, however, there will be crosstalk components originating from different information streams. The performance of WDM networks can be significantly degraded by this disturbance [2].

The application of analytical techniques to the performance evaluation of WDM systems impaired by crosstalk is usually very difficult and often requires excessive simplification of the system model. On the other hand, building a hardware prototype is expensive, time-consuming and relatively inflexible. Owing to these difficulties, computer simulation represents an attractive alternative.

The bit-error-rate (BER) is a fundamental performance parameter. The values of interest in optical communications are very small. Unfortunately, Monte Carlo simulation requires large run-times to yield accurate BER estimates. Therefore, it is desirable to find efficient variance-reduction techniques, such as those derived from importance sampling (IS), that lead to simulation speed-up.

IS has found application in a variety of fields, such as optical fiber communications [3]-[4], reliability [5], queuing [6], detection [7]-[8], fading channels [9]-[10], and other issues in digital communications [11]-[15].

IS involves running a Monte Carlo simulation where probability density functions (pdf's) are employed that are different from the actual ones, so that the probability that an error arises during simulation increases. An unbiased BER estimate is then obtained by weighing the results with the likelihood ratios of the actual to the IS densities.

The principle of IS is simple, but its efficient application to particular systems is a research issue. The researcher must decide which type of IS pdf to use, and then has to find the parameters of the pdf that yield a minimum estimator variance.

In general, the performance of the IS estimator closely depends on the choice of the IS pdf's and their parameters. Two main methodologies have been developed for the optimization of IS parameters: adaptive techniques [9], [11], [16], and techniques based on Large Deviations Theory [17]. The advantages of the former are its generality and applicability to a wide range of systems. The latter often requires difficult analysis that is possible only for relatively simple systems.

In this paper the search for optimal IS parameter values is made with new adaptive techniques based on stochastic Newton recursions. The techniques require some additional analytical work, but robust and easy-to-implement algorithms result. Therefore, simulation run-time is traded for algorithm design effort. Furthermore, a conditioning technique, referred to as the g-method [7], is combined with the adaptive IS algorithm, so that knowledge of the distribution of the underlying random variables is more fully exploited. A related technique is used for the optimization of system parameters. It involves the minimization, through simulation, of a suitable stochastic objective function with respect to parameters of interest [7]. All these IS techniques are briefly explained in Section II.

We use the presented IS techniques to determine the BER degradation due to crosstalk in the AWG. The system model is described in Section III. In practical situations, out-band crosstalk can be neglected with respect to in-band crosstalk due to the demultiplexing process made before the receiver [2]. Worst-case analysis is carried out including all the beat noise terms. Moreover, both input signal hypotheses are considered. Results are compared with the commonly used Gaussian approximation [2] and a recently developed Chernoff bound [18]. Furthermore, we present novel results on optimization of detector threshold setting, which turns out to have a relevant impact on system performance. Besides, an accurate assessment is given of the power penalty due to the introduction of additional WDM channels.

II. IMPORTANCE SAMPLING

A. Basics of IS

Consider estimating the quantity $G \equiv E\{g(X)\} < +\infty$, where $g(X)$ is a real-valued function. For notational convenience, we assume that X is a random variable with density f. The extension to random vectors is straightforward. An unbiased IS estimator $\hat{G}$ of G is given by

$$\hat{G} = \frac{1}{K}\sum_{k=1}^{K} g(X_k)W(X_k,\theta),\ X_k \sim f_*(x,\theta), \qquad (1)$$

where f_* denotes a biasing family of densities parameterized by θ, the function W is the likelihood ratio $W(x,\theta) = f(x)/f_*(x,\theta)$ used as a weighing function, and K is the IS simulation length. The notation $X \sim f$ denotes that X is drawn from a distribution with density f. The estimator variance is given by

$$\operatorname{var}\hat{G} = \frac{1}{K-1}\left[I(\theta) - G^2\right], \qquad (2)$$

where

$$\begin{aligned} I(\theta) &= E\{g^2(X)W(X,\theta)\} \\ &= E_*\{g^2(X)W^2(X,\theta)\} \end{aligned} \qquad (3)$$

and E_* denotes expectation with respect to f_*. If $g(\cdot)$ represents the indicator of some event, say $\{X \geq \tau\}$, then $G = P(X \geq \tau)$ and $\hat{G}$ is an estimator of a tail probability.

The first step in the application of IS is to select a family of densities $f_*(x,\theta)$ that enhances the tail probability in an adequate manner. In an application, θ could represent a set of parameters. Once $f_*(x,\theta)$ is chosen, the IS problem centers around determining the value of θ that minimizes the variance in (2) or equivalently $I(\theta)$ in (3).

B. Adaptive IS

The algorithmic minimization of $I(\theta)$ can be done in the following way. From (3) we have

$$\begin{aligned} I'(\theta) &= E\{g^2(X)W'(X,\theta)\} \\ &= E_*\{g^2(X)W'(X,\theta)W(X,\theta)\}, \end{aligned} \qquad (4)$$

where prime indicates derivative with respect to θ. Similarly,

$$I''(\theta) = E_*\{g^2(X)W''(X,\theta)W(X,\theta)\}. \qquad (5)$$

Estimators of these derivatives can be set up as

$$\hat{I}'(\theta) = \frac{1}{K}\sum_{k=1}^{K} g^2(X_k)W(X_k,\theta)W'(X_k,\theta),\ X_k \sim f_*, \qquad (6)$$

and

$$\hat{I}''(\theta) = \frac{1}{K}\sum_{k=1}^{K} g^2(X_k)W(X_k,\theta)W''(X_k,\theta),\ X_k \sim f_*. \qquad (7)$$

We can now use a root finding algorithm in the form of stochastic Newton formula recursions to estimate an optimum θ. Such an algorithm is given by

$$\theta_{m+1} = \theta_m - \delta_\theta \frac{\hat{I}'(\theta_m)}{\left|\hat{I}''(\theta_m)\right|},\quad m = 1,2,\ldots, \qquad (8)$$

where the rate factor δ_θ controls convergence speed and noisiness. As is typical of stochastic approximation procedures, convergence of this algorithm is characterized by a small random oscillation around the optimum. For a large class of IS problems the function $I(\theta)$ has a single minimum and the algorithm can locate it. The function $I(\theta)$ does not need to be convex. Other numerical methods are available (e.g. [9], [11]) that can be combined with IS simulation procedures to minimize the estimator variance. On the other hand, the Brent's method and the Golden Section Search method, which are meant for deterministic function minimization, do not yield satisfactory results.

C. The g-Method

In some applications, the system performance can be characterized as a probability in the form $p_\tau = P(Z + X \geq \tau)$, where Z is a random variable with known density, and X represents a random variable or function of random variables. The variable τ represents some system parameter, for example, a threshold level in a digital receiver. It is assumed here that Z and X are independent. Then we can write

$$p_\tau = E\{P(Z \geq \tau - X \mid X)\} = E\{g_\tau(X)\}, \qquad (9)$$

where $g_\tau(x) \equiv P(Z \geq \tau - x)$ is a continuous function of τ. In analogy with (1), we have the IS estimator $\hat{p}_\tau$ of p_τ as

$$\hat{p}_\tau = \frac{1}{K}\sum_{k=1}^{K} g_\tau(X_k) W(X_k, \theta), \; X_k \sim f_*(x, \theta). \qquad (10)$$

The estimator exploits knowledge of the density of Z, with IS being performed on X. In contrast to this is the normal IS estimator given by

$$\frac{1}{K}\sum_{k=1}^{K} 1_\tau(Z_k + X_k) W(Z_k, X_k, \theta),$$
$$Z_k \sim f_{Z*}(z, \theta), \; X_k \sim f_{X*}(x, \theta), \qquad (11)$$

where $1_\tau(y) = 1$ for $y \geq \tau$ and $1_\tau(y) = 0$ otherwise (this is usually called the indicator function). Here IS is performed on Z and X. It has been shown [7] that for any biasing scheme the estimator in (10) yields a smaller variance than that in (11).

D. Optimization of System Parameters

The differentiability, with respect to τ, of the estimator in (10) permits optimization of the system parameters to achieve a desired performance. Suppose that α_o is the desired value of performance probability p_τ, which is obtained at $\tau = \tau_o < +\infty$. To estimate τ_o we form the stochastic objective function

$$J(\tau) \equiv [\hat{p}_\tau - \alpha_o]^2 \qquad (12)$$

and minimize $J(\cdot)$ with respect to τ using the algorithm

$$\tau_{i+1} = \tau_i + \delta_\tau \frac{\alpha_o - \hat{p}_\tau(\tau_i)}{\hat{p}'(\tau_i)}, \quad i = 1,2,\ldots . \qquad (13)$$

This approach was proposed in [7] as the inverse IS problem. On the other hand, if p_τ represents an error probability in a communication system that is to be minimized, then the algorithm

$$\tau_{i+1} = \tau_i - \delta_\tau \frac{\hat{p}'(\tau_i)}{|\hat{p}''(\tau_i)|}, \quad i = 1,2,\ldots \qquad (14)$$

can be used. The estimates of the derivatives in these algorithms can be obtained from (10).

III. SYSTEM MODEL

Consider the schematic of a 4×4 AWG in Fig. 1. There are 4 nodes connected to the cross-connect. Each node includes a multi-wavelength transmitter (4 light sources and a multiplexer) and a multi-wavelength receiver (a demultiplexer and 4 photo-detectors). The router can send any wavelength from any input port to any output port [2].

Worst-case analysis implies considering that the interfering channels are in the ON-state. The out-band crosstalk is neglected. The phase of the desired optical signal is assumed to be zero without any loss of generality, and the phases of the interfering signals are independent and uniformly distributed in $[0, 2\pi)$. The extinction ratio is assumed infinite. The optical field of the desired input channel is

$$s_1(t) = a_1 E \times \cos(2\pi f_0 t), \quad 0 \leq t \leq T, \qquad (15)$$

where $a_1 \in \{0,1\}$ is the information bit, E is the pulse amplitude, and T is the symbol period. Each of the $M-1$ interfering channels has an optical field

$$s_m(t) = \sqrt{\varepsilon}\, a_m E \times \cos(2\pi f_0 t + \phi_m(t)). \qquad (16)$$

The factor ε accounts for the amount of crosstalk. Within the symbol period, the phase $\phi_m(t)$ is assumed to be constant, i.e. $\phi_m(t) = \phi_m$. Worst-case analysis implies, under both signal hypotheses, that $a_m = 1, \quad m = 2,\ldots,M$.

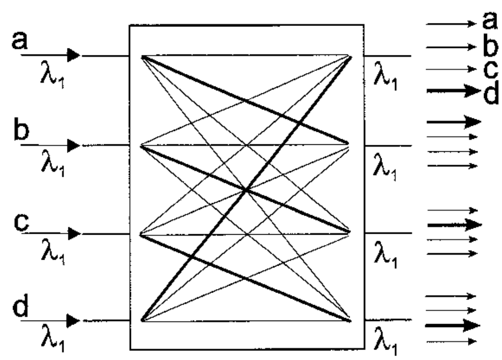

Fig. 1. Schematic of a 4×4 AWG. Thick and thin lines indicate signal and crosstalk components, respectively. Only one wavelength is shown.

The photocurrent generated by a photodiode with unity quantum efficiency at one of the outputs of an $M \times M$ AWG will be [2]

$$i_d = \frac{a_1 E^2}{2} + \sum_{m=2}^{M} \sqrt{\varepsilon}\, a_1 E^2 \times \cos(\phi_m)$$
$$+ \sum_{\substack{m,n=2 \\ m>n}}^{M} \varepsilon\, E^2 \times \cos(\phi_m - \phi_n) + (M-1)\varepsilon \frac{E^2}{2} + n_G, \quad (17)$$

where n_G is the additive white Gaussian noise (AWGN) of the receiver, which is independent of the signal and the crosstalk components. The second term in (17) is the signal-crosstalk beat noise and the third term is the crosstalk-crosstalk beat noise.

IV. ANALYSIS AND RESULTS

The usual Gaussian approximation [2] assumes that the third and fourth terms in (17) can be neglected (small ε). When the decision threshold τ is set at half the ON-signal output current (symmetric setting, i.e. $\tau = E^2/4$), the system BER can be assumed to be equal to half the probability that the ON-symbol is detected erroneously. The Gaussian approximation is then given by

$$\mathrm{BER} \approx \frac{1}{2}\mathrm{erfc}\left(\frac{\tau}{\sqrt{2\sigma_G^2 + (M-1)\varepsilon^2 E^4}}\right), \qquad (18)$$

where σ_G^2 is the variance of the receiver noise. The Chernoff bound [18] is

$$\mathrm{BER} < \frac{1}{2}\min_s\left[\mathrm{I}_0^{M-1}\left(s\sqrt{\varepsilon}E^2\right)\exp\left(-s\tau + \frac{s^2\sigma_G^2}{2}\right)\right], \qquad (19)$$

where $\mathrm{I}_0(\cdot)$ is the zero-order modified Bessel function of the first kind.

In contrast with the two approximate methods in (18) and (19), the IS experiments include all the terms in (17). Moreover, error probabilities are obtained for both the ON $(a_1 = 1)$ and the OFF $(a_1 = 0)$ signal hypotheses. Estimating the error probability for the OFF case represents a challenge because this probability possesses a very low BER floor when $\tau = E^2/4$.

The system was simulated with modified biasing densities for the $M-1$ phases of the interfering components. All modified phase densities were identical Gaussian pdf's, with means at π, and with common variance to be determined with the stochastic Newton formula (8). In this way, the probability densities of the phases are concentrated in the region where the second term in (17) yields the largest negative values and the third term yields the largest positive values. Under hypothesis $a_1 = 1$, the second term is much more significant than the third term, so that smaller values of i_d will become more probable. When $a_1 = 0$, the second term in (17) is zero, therefore the described modification of the phase densities will tend to increase the third term and thereby i_d. The tails outside the interval $[0, 2\pi)$ will only affect the estimator accuracy when the error probability is very high.

The g-method is applied to the AWGN component, n_G, hence reducing the IS parameter optimization problem to one dimension: the variance of the modified phase densities. The function defined in Section II-C becomes

$$g_\tau(\phi_2,\ldots,\phi_M) = \frac{1}{2}\mathrm{erfc}\left\{\frac{\alpha}{\sqrt{2}\sigma_G}\left[\tau - \frac{a_1E^2}{2} - \sum_{m=2}^{M}\sqrt{\varepsilon}a_1E^2\times\cos\phi_m - \sum_{\substack{m,n=2\\ m>n}}^{M}\varepsilon E^2\times\cos(\phi_m - \phi_n) - (M-1)\varepsilon\frac{E^2}{2}\right]\right\}, \qquad (20)$$

where $\alpha = 1$ for the ON hypothesis and $\alpha = -1$ for the OFF hypothesis. The weighing function and its two first derivatives can be easily found analytically.

Let us first consider a 4-channel WDM router with a symmetric threshold setting. In Fig. 2 we show the BER, for a particular value of the receiver noise variance σ_G^2, as a function of the crosstalk-to-signal ratio $\mathrm{XSR} = 10\times\log\varepsilon$. The different curves were obtained with the Gaussian approximation (18), the Chernoff bound (19), and our IS techniques. All the IS values shown where obtained with the same rate factor δ_θ, and yielded an accuracy better than $\pm 3\%$ for 95% confidence level. A run-time of 10 seconds per value on a Pentium PC was sufficient to achieve this accuracy. This high accuracy is maintained through low BER values, indicating that the IS strategy is close to the optimum.

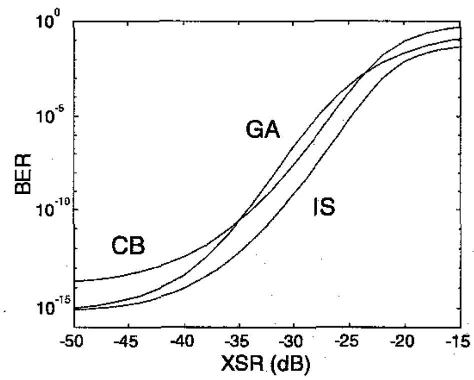

Fig. 2. BER curves obtained with the Gaussian approximation in [2] (GA), the Chernoff bound in [18] (CB), and our adaptive IS techniques.

As expected, the Gaussian approximation results shown in Fig. 2 coincide with the IS results for very low crosstalk levels. The lack of tightness of the Chernoff bound can be observed: this upper bound is more than one order of magnitude above the true BER at practical crosstalk levels (around $\mathrm{XSR} = -25\,\mathrm{dB}$). The practical implication of the results in Fig. 2 is that our techniques allow the network designer to employ optical crossconnects with almost twice as large crosstalk levels than those predicted by the approximation methods.

The performed IS experiments indicated that, with the symmetric threshold setting used, the probability of error for the OFF signal hypothesis is much smaller than that of the ON case. Therefore, threshold optimization can be expected to improve the BER significantly. We find the optimal threshold for a given XSR and AWGN level by applying the technique described in Section II-D.

The impact of the threshold setting can be appreciated in Fig. 3. Shown in this figure are BER curves for symmetric threshold as well as for optimal thresholds obtained at three XSR values: -20 dB, -25 dB, and -30 dB. The number of channels and the AWGN level are the same as in Fig. 2. The influence of the threshold setting is quite significant. We also observe that the tolerable crosstalk level increases further by 3 dB, for a wide range of XSR values. This is equivalent to a 5 dB improvement with respect to the value predicted by the Chernoff bound.

An important parameter in WDM networks is the number of channels. In Fig. 4 we can appreciate the power penalty due to the introduction of an additional channel. The curves were obtained with $XSR = -25\,dB$ and the threshold being optimized at $SNR = 22\,dB$ for each of the curves. In this example, the introduction of a fifth channel requires additional SNR of about 1 dB to maintain the BER at 10^{-9}.

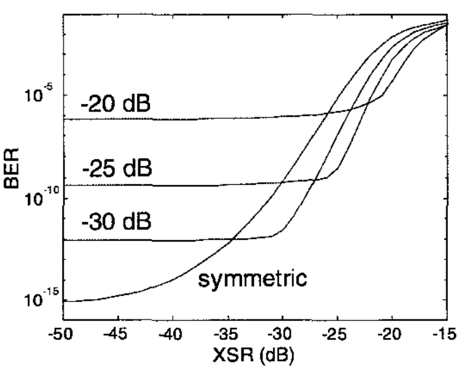

Fig. 3. Effect of threshold optimization. The threshold was optimized at the indicated XSR values.

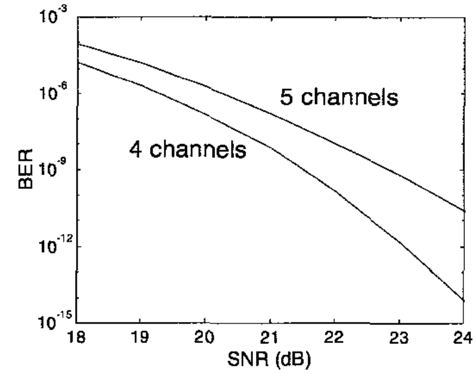

Fig. 4. Impact of the number of channels.

V. CONCLUSIONS

We investigated the performance degradation in a WDM network due to crosstalk in an AWG working as an optical cross-connect. Worst-case analysis was carried out and, in contrast with the approximation methods in the literature, all the beat-noise terms in (17) were included.

Appropriate IS strategies were developed that give accurate BER estimates in short simulation run-times. After conditioning with the g-method, the optimization of IS parameters was done using stochastic Newton recursions. BER estimates were obtained with high accuracy in 10 seconds run-time, for practical values of system parameters. The IS results indicate that, at practical XSR levels, more than twice as much crosstalk can be tolerated than predicted by the approximation methods.

Because both input signal hypotheses are included, different threshold settings could be considered, which turned out to have a strong impact on the system performance. Due to the effectiveness of the IS techniques, a minimum search algorithm can be used along with IS to perform threshold optimization. This resulted in a further relaxation of 3 dB for the crosstalk requirements.

Finally, the IS techniques also proved to be useful for the determination of the power penalty due to the introduction of additional WDM channels.

ACKNOWLEDGEMENTS

This work was supported in part by the Netherlands Organization for Scientific Research (NWO).

REFERENCES

[1] C.G.P. Herben, C.G.M. Vreeburg, D.H.P. Maat, X.J.M. Leijtens, Y.S. Oei, F.H. Groen, J.J.G.M. van der Tol, P. Demeester and M.K. Smit, "A Compact Integrated Single Phasar Optical Crossconnect on InP," in *Proc. IEEE/LEOS Symp. Benelux '97,* Eindhoven (NL), pp. 41-44, Nov. 1997.

[2] H. Takahashi, K. Oda and H. Toba, "Impact of Crosstalk in an Arrayed-Waveguide Multiplexer on NxN Optical Interconnection," *J. Lightwave Technol.,* vol. 14, pp. 1097-1105, Jun. 1996.

[3] N.B. Mandayam and B. Aazhang, "Gradient Estimation for Stochastic Optimization of Optical Code-Division Multiple-Access Systems: Part I-Generalized Sensitivity Analysis," *IEEE J. Selec. Areas Commun.,* vol. 15, pp. 731-741, May 1997.

[4] K. Ben Letaief, "The Performance of Optical Fiber Direct-Sequence Spread-Spectrum Multiple-Access Communications Systems," *IEEE Trans. Commun.,* vol. 43, pp. 2662-2666, Nov. 1995.

[5] V.F. Nicola, M.K. Nakayama, P. Heidelberger and A. Goyal, "Fast Simulation of Highly Dependable Systems with General Failure and Repair Processes," *IEEE Trans. Comput.,* vol. 42, pp. 1440-1452, Dec. 1993.

[6] P. Heidelberger, "Fast Simulation of Rare Events in Queueing and Reliability Models," *ACM Trans. Modeling and Comput. Simulation,* vol. 5, pp. 43-85, Jan. 1995.

[7] R. Srinivasan, "Some Results in Importance Sampling and an Application to Detection," *Signal Processing,* North Holland, vol. 65, no. 1, 1998.

[8] S.L. Smith and G.C. Orsak, "A Modified Importance Sampling Scheme for the Estimation of Detection System Performance," *IEEE Trans. Commun.,* vol. 43, pp. 1341-1346, Feb./Mar./Apr. 1995.

[9] W.A. Al-Qaq and J.K Townsend, "A Stochastic Importance Sampling Methodology for the Efficient Simulation of Adaptive Systems in Frequency Nonselective Rayleigh Fading Channels," *IEEE J. Selected Areas Commun.,* vol. 15, pp. 614-625, May 1997.

[10] K. Ben Letaief, K. Muhammad and J.S. Sadowsky, "Fast Simulation of DS/CDMA with and without Coding in Multipath Fading Channels," *IEEE J. Selected Areas Commun.,* vol. 15, pp. 626-639, May 1997.

[11] W.A. Al-Qaq, M. Devetsikiotis and J.K. Townsend, "Stochastic Gradient Optimization of Importance Sampling for the Efficient Simulation of Digital Communication Systems," *IEEE Trans. Commun.,* vol. 43, pp. 2975-2985, Dec. 1995.

[12] P.J. Smith, M. Shafi and H. Gao, "Quick Simulation: A Review of Importance Sampling Techniques in Communications Systems," *IEEE J. Selected Areas Commun.,* vol. 15, pp. 597-613, May 1997.

[13] K. Ben Letaief, "Efficient Evaluation of the Error Probabilities of Spread-Spectrum Multiple-Access Communications," *IEEE Trans. Commun.,* vol. 45, pp. 239-246, Feb. 1997.

[14] J.-C. Chen, D. Lu, J.S. Sadowsky and K. Yao, "On Importance Sampling in Digital Communications-Part I: Fundamentals," *IEEE J. Selected Areas Commun.,* vol. 11, pp. 289-299, Apr. 1993.

[15] R.J. Wolfe, M.C. Jeruchim and P.M. Hahn, "On Optimum and Suboptimum Biasing Procedures for Importance Sampling in Communication Simulation," *IEEE Trans. Commun.,* vol. 38, pp. 639-647, May 1990.

[16] J. Scott Stadler and S. Roy, "Importance Sampling for Detection of Scale Problems," *IEEE J. Selected Areas Commun.,* vol. 11, pp. 309-316, Dec. 1993.

[17] R.K. Bahr and J.A. Bucklew, "Quick Simulation of Detector Error Probabilities in the Presence of Memory and Nonlinearity," *IEEE Trans. Commun.,* vol. 41, pp. 1610-1617, Nov. 1993.

[18] A.M.J. Koonen and D.T. van Veen, "Estimating Error Probabilities caused by In-Band and Out-Band Crosstalk in Multi-Wavelength All-Optical Networks," in *Proc. IEEE/LEOS Symp. Benelux '97,* Eindhoven (NL), pp. 169-172, Nov. 1997.