Transactions Papers

Adaptive Importance Sampling for Performance Evaluation and Parameter Optimization of Communication Systems

David Remondo, Rajan Srinivasan, Victor F. Nicola, Wim C. van Etten, *Senior Member, IEEE*, and Henk E. P. Tattje

***Abstract*—In this paper, we present new adaptive importance sampling techniques based on stochastic Newton recursions. Their applicability to the performance evaluation of communication systems is studied. Besides bit-error rate (BER) estimation, the techniques are used for system parameter optimization. Two system models that are analytically tractable are employed to demonstrate the validity of the techniques. As an application to situations that are analytically intractable and numerically intensive, the influence of crosstalk in a wavelength-division multiplexing (WDM) crossconnect is assessed. In order to consider a realistic system model, optimal setting of thresholds in the detector is carried out while estimating error rate performances. Resulting BER estimates indicate that the tolerable crosstalk levels are significantly higher than predicted in the literature. This finding has a strong impact on the design of WDM networks. Power penalties induced by the addition of channels can also be accurately predicted in short run-times.**

***Index Terms*—Communication system performance, Monte Carlo methods, optical crosstalk, optical fiber communication, wavelength-division multiplexing.**

I. INTRODUCTION

PERFORMANCE evaluation and parameter optimization are major issues in the design of communication links and networks. The application of analytical techniques to the performance evaluation of complex communication systems 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 is an alternative that has received much attention in recent years.

The bit-error rate (BER) is a fundamental performance parameter in many digital communication systems. 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 [1], [2], reliability [3], queuing [4], detection [5]–[7], fading channels [8], [9], and other issues in digital communications [10]–[14].

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 occurs during simulation increases. An unbiased BER estimate is then obtained by weighting the results with the likelihood ratios of the actual to the IS densities.

The principle of IS is simple, but its effective application to particular systems is a research issue. The researcher must decide which type of pdf to use in IS and then has to find the pdf parameters 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 [5], [8], [10], [15] and techniques based on the large deviations theory [16]. 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.

This paper describes new adaptive techniques for IS parameter optimization. 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. A recently developed IS method with an adaptive implementation, referred to as the g-method, is described in [5]. It exploits knowledge of the distribution of the underlying random variables more fully and has an estimation performance superior to that of conventional IS methods. A related technique, and one of fundamental value in parameter optimization in communication systems, is the solution of the inverse IS problem [5]. It involves the minimization, through simulation, of a suitable stochastic objective function with respect to parameters of interest. These IS methods are described in Section II. Applications of

Paper approved by K. Townsend, the Editor for Computer-Aided Design of Communication Systems of the IEEE Communications Society. Manuscript received July 20, 1998; revised April 9, 1999. This work was supported in part by the Netherlands Organization for Scientific Research (NWO). This paper was presented in part at the Global Communications Conference (GLOBECOM), Sydney, Australia, November 1998, and in part at the IEEE 6th Symposium of the Benelux Joint Chapter on Communications and Vehicular Technology, Brussels, Belgium, October 1998.

D. Remondo, V. F. Nicola, W. C. van Etten, and Henke E. P. Tattje are with the University of Twente, 7500AE Enschede, The Netherlands (e-mail: remondo@cs.utwente.nl).

R. Srinivasan is with the Centre for Air Borne Systems, Defense Research and Development, Bangalore 560037, India.

Publisher Item Identifier S 0090-6778(00)03628-X.

the techniques to performance analysis and threshold optimization are contained in Sections III and IV. In Section III, a validation of adaptive IS is carried out using noncoherent on–off keying (OOK) and differential phase-shift keying (DPSK) receivers in additive white Gaussian noise (AWGN) as examples. Threshold optimization is demonstrated in the OOK receiver. The DPSK receiver involves a two-dimensional IS parameter optimization problem.

A wavelength-division multiplexing (WDM) network is studied in Section IV. New results are obtained that have an impact on the design of these networks. Using our IS methods, we demonstrate that the influence of crosstalk in a multiwavelength crossconnect is significantly lower than predicted by the methods available to date. Besides, we develop an algorithm that permits accurate evaluation of the system performance when the detection threshold is optimized. In this way, the effect of different system parameters, like the number of WDM channels, can be studied in a realistic manner.

II. Adaptive IS

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

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

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 weighting 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

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

where

$$I(\theta) = E\{g^2(X)W(X,\theta)\} = E_*\{g^2(X)W^2(X,\theta)\} \tag{3}$$

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. There is an absolute optimum biasing density, which is proportional to the original density in the region of interest and is zero elsewhere. This density cannot be used in practice because it requires previous knowledge of the quantity to be estimated [11]. However, this gives us a general criterion for the selection of an appropriate family of biasing densities: the tail probability should be enhanced while, at the same time, the biasing density should resemble the original density in the region of interest as much as possible (except for a proportionality constant). Together with this general criterion, also practical considerations are relevant, like the simplicity of the resulting likelihood ratio $W(x,\theta)$.

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). In an application, θ could represent a set of parameters (as in Section III-B).

B. Adaptive IS

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

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

where prime indicates derivative with respect to θ. Similarly

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

Estimators of these derivatives can be set up as

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

and

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

We can now use a root finding algorithm in the form of stochastic Newton formula recursions to estimate an optimum θ. Such an algorithm [6] is given by

$$\theta_{m+1} = \theta_m - \delta_\theta \frac{\hat{I}'(\theta_m)}{\hat{I}''(\theta_m)}, \qquad m = 1, 2, \cdots \tag{8}$$

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 value. For a large class of IS problems, the function $I(\theta)$ has a single minimum and the algorithm can locate it.

Stochastic optimization of IS parameters has been suggested before [8], [10], [15]. The algorithm therein uses the gradient descent technique by estimating the first derivative of $I(\theta)$. There is an essential difference between the procedures. The gradient descent algorithm takes a step in the direction of negative gradient, whereas in Newton recursion the steps are in the direction of the minimum of the function (see e.g., [17]). Further, it is well known that the (deterministic) Newton algorithm enjoys quadratic convergence provided the starting point is chosen carefully. Obviously, for quadratic functions there is single-step convergence. In the stochastic case, of course, this fast convergence is slowed down by the nondeterministic nature of the estimates of the derivatives. A statistical analysis of the convergence properties of the Newton algorithm is beyond the scope of the present work.

An application of the Newton algorithm to a multidimensional IS parameter optimization problem is described in Section III-B.

In many cases, the initial value of θ can be chosen so that the biased density is similar to the original one. However, there are problems where this may not work. To overcome this difficulty, the initial values of the IS parameters can be chosen according to the optimal values found when higher noise levels are present, as discussed in [10] and [8].

Function minimization algorithms that do not require derivatives, like Brent's method and the Golden Section Search method, do not yield satisfactory results in the minimization of the estimator variance.

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

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

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

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

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

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

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 [5] 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

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

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

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

This approach was proposed in [5] as the inverse IS problem. It was used therein along with adaptive biasing procedures to determine threshold multipliers for a practically important class of constant false alarm rate detectors. On the other hand, if p_τ represents an error probability in a communication system that is to be minimized, then the algorithm

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

can be used. The derivatives of the error probability estimates are obtained by analytically differentiating the right-hand term of (10) with respect to τ.

In some cases, it may not be possible to express $\hat{p}_\tau$ as a differentiable function of τ, for example when $g_\tau(x) = 1_\tau(x)$. In these situations, the indicator function can be approximated, for the purpose of obtaining derivative estimates, by

$$1_\tau(x) \approx [1 + \exp(c(\tau - x))]^{-1}, \qquad c > 0 \tag{15}$$

which is the well-known sigmoid function used in training artificial neural networks. Using an approximation here does not affect the unbiasedness of the estimator, but possibly its variance. The approximation becomes better for large values of the parameter c. Using (15) in (14), derivatives can be calculated. An application of this is demonstrated in Section III-A.

III. Validation of the Adaptive IS Method

In this section, we demonstrate the effectiveness and accuracy of adaptive IS by applying the techniques to two communication systems whose performances can be evaluated analytically. In the first system, we also optimize the detection threshold by using the approximation in (15).

A. Noncoherent OOK Receiver

Consider the well-known noncoherent OOK receiver with ideal filtering (see e.g., [18]). At the ideal sampling instants, the signal at the threshold comparator input is

$$y = \sqrt{[m + n_i]^2 + n_q^2} \tag{16}$$

where m represents the received signal, with amplitude $a \cdot A$ (the variable a represents the transmitted bit, taking the values 0 and 1 with equal probability), and the noise terms n_i and n_q are independent zero-mean Gaussian random variables with variance σ^2. The random variable y has a Rayleigh density when $a = 0$ and a Rice density when $a = 1$. The BER is then

$$P_e = \frac{1}{2}\int_\tau^{+\infty} \frac{y}{\sigma^2} \exp\left[-\frac{(y^2)}{2\sigma^2}\right] dy + \frac{1}{2}\int_0^\tau \frac{y}{\sigma^2} \exp\left[-\frac{(y^2 + A^2)}{2\sigma^2}\right] \cdot I_0\left[\frac{A \cdot y}{\sigma^2}\right] dy \tag{17}$$

where I_0 is the modified Bessel function of the first kind and τ is the detection threshold. This expression can be easily calculated numerically.

The IS simulation is carried out by using Gaussian biasing densities with modified parameters for the noise terms. This is a classical choice, which has often given relatively good results and has the advantage of the simplicity of the likelihood ratio (see e.g., [13]). When $a = 0$, the mean is kept at zero and the variance is increased by the same amount for each of the noise components, so that all directions undergo the same modification. When $a = 1$, the component n_q is not modified, but the

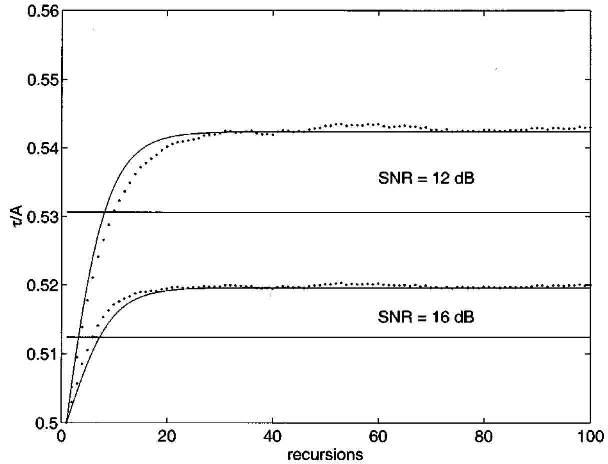


Fig. 1. OOK optimum threshold search.

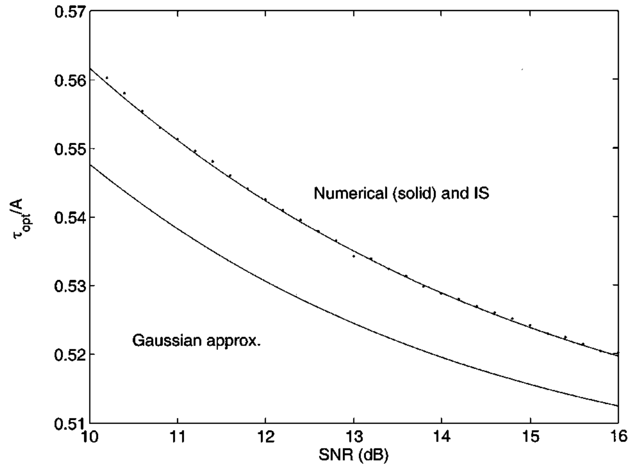


Fig. 2. OOK optimum threshold values.

mean of n_i takes a negative value. In this way, probability mass is moved in the direction of the region of interest. In each of the two cases, the optimum amount of parameter modification is found with the one-dimensional adaptive IS technique described in Section II-B. The first experiments were carried out with the usual value of $A/2$ for the detection threshold. The BER estimates, not shown here for reasons of space, were obtained in very short run-times [a satisfactory accuracy was achieved with a sample length of $K = 5000$ and only 15 iterations for the IS parameter optimization in (8)]. The obtained estimates agree with analysis [18] for a wide range of BER values.

The threshold that yields the minimum BER was obtained with the stochastic Newton recursion formula [see (14)] as explained in Section II-D. The derivatives in (14) were not obtained from the available closed-form formula for the BER in (17), but through derivation of the right-hand term of (10) and using (15) in the place of $g_\tau(\cdot)$. Although the parameter c in (15) has to be adjusted experimentally, we shall see that one single value is adequate for a wide range of signal-to-noise ratio (SNR) levels. Fig. 1 illustrates the convergence of the optimum threshold search for two SNR levels. The dashed lines represent the relative threshold values obtained from the usual Gaussian approximation (GA). The solid curves contain the outcomes of a Newton search algorithm that uses numerical integration. The dotted curves indicate the values obtained with the presented method, where the same value of c was used for both SNR levels. The figure shows that, although slower than numerical integration, our method yields correct values of the optimum threshold.

In Fig. 2 are shown the optimum threshold values obtained after 100 recursions of (14) for a wide range of SNR levels. The dashed curve gives the values obtained from the GA, the solid curve shows the outcomes of numerical integration, and the dots are the values generated with the presented adaptive IS technique. A single value of c was used for the whole range of SNR levels. The close agreement between the values obtained using adaptive IS and using numerical integration demonstrates the validity of the approximation in (15) as well as the effectiveness of the adaptive IS method for optimum threshold search.

B. DPSK Receiver

Due to the symmetry of a DPSK receiver, we can assume that the error probability when a ONE is transmitted is the same as for a transmitted ZERO. The BER [18] is

$$P_e = P(\alpha - \beta \leq 0) \tag{18}$$

where

$$\alpha = \sqrt{(A+\alpha_i)^2 + \alpha_q^2} \tag{19}$$

and

$$\beta = \sqrt{\beta_i^2 + \beta_q^2}. \tag{20}$$

The input signal amplitude is A and the variables α_i, α_q, β_i, and β_q are independent zero-mean Gaussian random variables with variance $\sigma^2/2$. The value of σ^2 corresponds to the noise power at the band-pass filter output.

Like in the previous section, the IS experiment uses Gaussian biasing densities with modified parameters. To enhance the detection error probability, a quantity b_1 is added to α_i (mean translation), the standard deviations of β_i and β_q are multiplied by a common factor b_2, and the remaining parameters are not modified. Hence, the IS technique involves a two-dimensional optimization (to determine the optimum values of b_1 and b_2).

Define the column vector

$$\underline{b} = (b_1, b_2)^T. \tag{21}$$

The stochastic Newton recursion formula is then

$$\underline{b}_{m+1} = \underline{b}_m - \delta \cdot J_m^{-1} \cdot \nabla I(\underline{b}_m). \tag{22}$$

In (22), J is the Hessian matrix of I, that is

$$J = \begin{pmatrix} I_{b_1 b_1} & I_{b_1 b_2} \\ I_{b_1 b_2} & I_{b_2 b_2} \end{pmatrix} \tag{23}$$

where the notation $I_{\mu\nu}$ indicates

$$I_{\mu\nu} = E\left\{1_\tau(X) \cdot \frac{\partial^2 W(X;\mu,\nu)}{\partial\mu\partial\nu}\right\} \tag{24}$$

with $\tau = 0$. The last factor in (22) is defined as

$$\nabla I(b_1, b_2) = \left(\frac{\partial I(b_1, b_2)}{\partial b_1} \quad \frac{\partial I(b_1, b_2)}{\partial b_2} \right)^T. \tag{25}$$

The derivatives are obtained analytically and the estimators of their expected values are implemented in parallel with the BER estimation. The results are shown in Fig. 3 (dots) together with the exact BER values (solid line) [18]. The IS estimates of the BER show a very good correspondence with the exact BER's, and each required only 3-s run-time on a Pentium 150-MHz processor, using 10^5 simulation samples.[1]

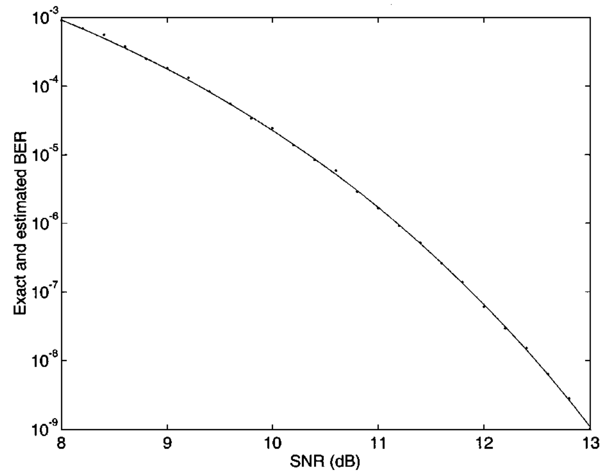


Fig. 3. BER of a DPSK receiver.

IV. Application to a WDM Network

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

A fundamental element in WDM systems is the wavelength router in an all-optical interconnection, also called optical cross-connect. For this purpose, an arrayed-waveguide grating (AWG) seems to be a good candidate [19]. In this device, crosstalk between different channels can arise. In principle, there will be in-band and out-band crosstalk components originating from channels having the same wavelength or different wavelengths, respectively. In practical situations, however, out-band crosstalk can be neglected with respect to in-band crosstalk due to the demultiplexing process before the receiver.

We shall employ the described IS techniques to determine the BER degradation due to in-band crosstalk. Results will be compared with the commonly used GA [20] and a recently developed Chernoff bound (CB) [21]. Furthermore, we shall present novel results on optimization of detector threshold setting, which turns out to have a relevant impact on system performance. Besides, an accurate assessment of the power penalty due to the introduction of additional WDM channels is given.

A. System Model

Consider the AWG schematic in Fig. 4. There are four nodes connected to the crossconnect. Each node includes a multiwavelength transmitter (four light sources and a multiplexer) and a multiwavelength receiver (a demultiplexer and four photo-detectors). The router can send any wavelength from any input port to any output port [20].

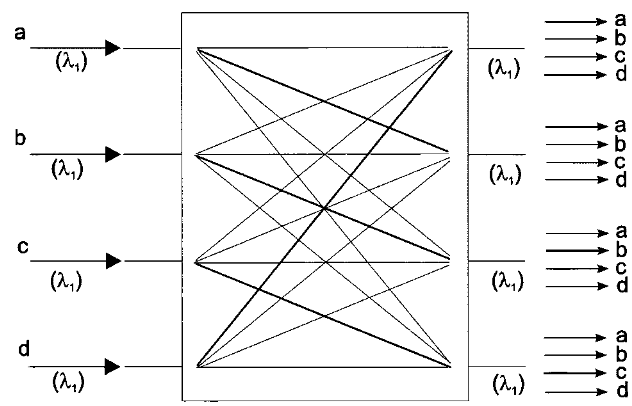


Fig. 4. Schematic of an AWG. Thick and thin lines indicate signal and crosstalk components, respectively. Only one wavelength is shown.

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 optical field of the desired input channel is

$$s_1(t) = a_1 E \times \cos(2\pi f_0 t), \qquad 0 \le t \le T \tag{26}$$

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

$$s_m(t) = \sqrt{\varepsilon} a_m E \times \cos(2\pi f_0 t + \phi_m(t)). \tag{27}$$

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 $a_m = 1$, $m = 2, \cdots, M$. The photocurrent generated by a photodiode with unity quantum efficiency will be then [20]

$$i_d = \frac{a_1 E^2}{2} + \sum_{m=2}^{M} \sqrt{\varepsilon} a_1 E^2 \times \cos(\phi_m) + \sum_{\substack{m,n=2 \\ m>n}}^{M} \varepsilon E^2 \times \cos(\phi_m - \phi_n) + (M-1)\varepsilon \frac{E^2}{2} + n_G \tag{28}$$

where n_G is the AWGN of the receiver, which is independent of the signal and the crosstalk components.

The usual GA [20] assumes that the third and fourth terms in the above sum can be neglected (small ε). When the decision threshold τ is set at half the ON-signal output current (symmetric setting, i.e., $\tau = E^2/4$) and the crosstalk components are relatively small, it can be assumed that the system BER is

[1]The algorithm was implemented in C-language.

equal to half the probability that the ON-symbol is detected erroneously. The BER is then given by

$$\mathrm{BER} \approx \frac{1}{2}\,\mathrm{erfc}\left(\frac{\tau}{\sqrt{2\sigma_G^2+(M-1)\varepsilon^2E^4}}\right). \tag{29}$$

Using the same assumptions, the CB [21] is

$$\mathrm{BER} < \frac{1}{2}\min_s\left[I_0^{M-1}(s\sqrt{\varepsilon}E^2)\exp\left(-s\tau+\frac{s^2\sigma_G^2}{2}\right)\right] \tag{30}$$

where σ_G^2 is the variance of the receiver noise.

B. BER Estimation Using IS

In contrast with the two approximate methods in (29) and (30), the IS experiments include all the terms in (28). Moreover, error probabilities were 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 floor when $\tau = E^2/4$.

Due to the relative complexity of the densities of the cosine functions in (28), the system was simulated with modified biasing densities for the $M-1$ phases ϕ_m, $m=2,\cdots,M$, 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 (28) yields the largest negative values and the third term yields the largest positive values. This strategy enhances the detection error probability: 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 (28) is zero, therefore the third term will tend to increase and thereby i_d. When the system error probability is very small, the optimal variance of the Gaussian densities can be expected to be much smaller than 2π, and therefore, the Gaussian 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

$$\begin{aligned} g_\tau(\phi_2,\cdots,\phi_M) &= \\ \frac{1}{2}\,\mathrm{erfc}&\left\{\frac{\alpha}{\sqrt{2}\sigma_G}\left[\tau-\frac{a_1E^2}{2}-\sum_{m=2}^{M}\sqrt{\varepsilon}a_1E^2\times\cos\phi_m\right.\right. \\ &\left.\left.-\sum_{\substack{m,n=2\\m>n}}^{M}\varepsilon E^2\times\cos(\phi_m-\phi_n)-(M-1)\,\varepsilon\frac{E^2}{2}\right]\right\} \end{aligned} \tag{31}$$

where $\alpha=1$ for the ONE hypothesis and $\alpha=-1$ for the ZERO hypothesis. The weighting function and its two first derivatives can be easily found analytically.

In Section IV-D we shall show the difference between the BER estimates provided by the approximate methods in (29) and (30) and the accurate results obtained with our IS techniques.

C. Realistic BER Estimation Through Threshold Optimization

The symmetric setting of the detection threshold τ in this WDM network is clearly far from the optimum, due to the fact that the probability distributions are very different for both input hypotheses. The system performance can be expected to improve significantly if the threshold setting is optimized for the expected crosstalk level. Actually, a realistic experimental setting will require adjusting the threshold. Therefore, a good performance analysis tool has to consider the WDM system with an optimized threshold setting.

Invoking the parameter optimization method described in Section II-D, we estimate the first two derivatives of the BER with

$$\begin{aligned} \frac{\partial\widehat{\mathrm{BER}}}{\partial\tau} &= \frac{1}{2}\frac{1}{K}\sum_{k=1}^{K} g'_\tau\left(\phi_2^{(k)},\cdots,\phi_M^{(k)}\right)\Big|_{\alpha=1} \\ &\quad\times W\left(\phi_2^{(k)},\cdots,\phi_M^{(k)}\right) \\ &\quad+\frac{1}{2}\frac{1}{K}\sum_{k=K}^{2K} g'_\tau\left(\phi_2^{(k)},\cdots,\phi_M^{(k)}\right)\Big|_{\alpha=-1} \\ &\quad\times W\left(\phi_2^{(k)},\cdots,\phi_M^{(k)}\right), \\ &\qquad \phi_m^{(k)}\sim f_{*m}, \qquad m=2,\cdots,M \end{aligned} \tag{32}$$

and

$$\begin{aligned} \frac{\partial^2\widehat{\mathrm{BER}}}{\partial\tau^2} &= \frac{1}{2}\frac{1}{K}\sum_{k=1}^{K} g''_\tau\left(\phi_2^{(k)},\cdots,\phi_M^{(k)}\right)\Big|_{\alpha=1} \\ &\quad\times W\left(\phi_2^{(k)},\cdots,\phi_M^{(k)}\right) \\ &\quad+\frac{1}{2}\frac{1}{K}\sum_{k=K}^{2K} g''_\tau\left(\phi_2^{(k)},\cdots,\phi_M^{(k)}\right)\Big|_{\alpha=-1} \\ &\quad\times W\left(\phi_2^{(k)},\cdots,\phi_M^{(k)}\right), \\ &\qquad \phi_m^{(k)}\sim f_{*m}, \qquad m=2,\cdots,M \end{aligned} \tag{33}$$

where f_{*m} is the modified density of the phase of ϕ_m, and the derivatives are with respect to τ.

The optimum threshold setting obtained from the above algorithm can be used to determine the corresponding BER curve. This is given in Section IV-D.

D. Results for the WDM Network

Let us first consider a four-channel WDM router. The receiver noise variance is $\sigma_G^2=(E^2/32)^2$ (i.e., $\mathrm{SNR}=21.1$ dB), and the threshold setting is symmetric. In Fig. 5, we show the BER as a function of the crosstalk-to-signal ratio $\mathrm{XSR}=10\times\log\varepsilon$; the "+" signs represent the values obtained with the GA (29), the dots are the outcomes of the CB (30), and the solid line is the result of applying our IS techniques. All the IS estimates shown were generated using the same rate factor δ_θ in (8) and yielded an accuracy better than $\pm3\%$ for 95% confidence level.

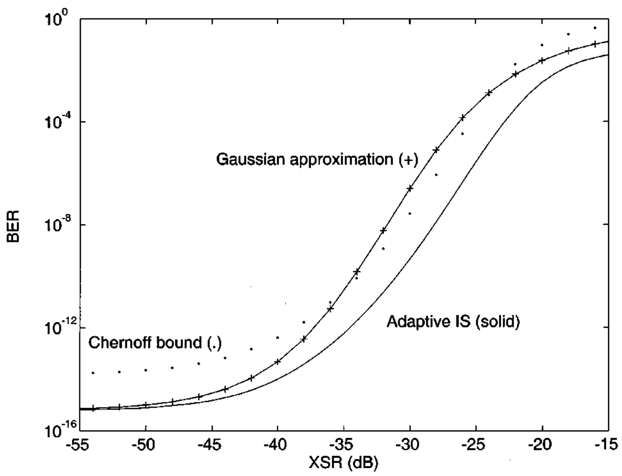


Fig. 5. BER calculated with the GA, the CB, and adaptive IS. The threshold setting is symmetric and $\mathrm{SNR} = 21.1$ dB.

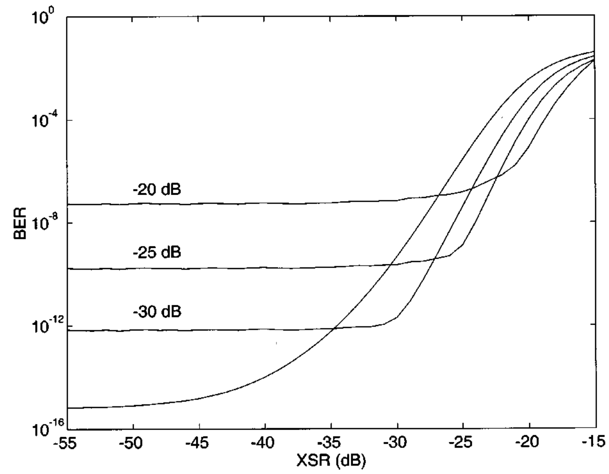


Fig. 6. Effect of threshold optimization. The curve without label corresponds to the symmetric threshold. The other curves were obtained with thresholds optimized at the indicated XSR values. $\mathrm{SNR} = 21.1$ dB.

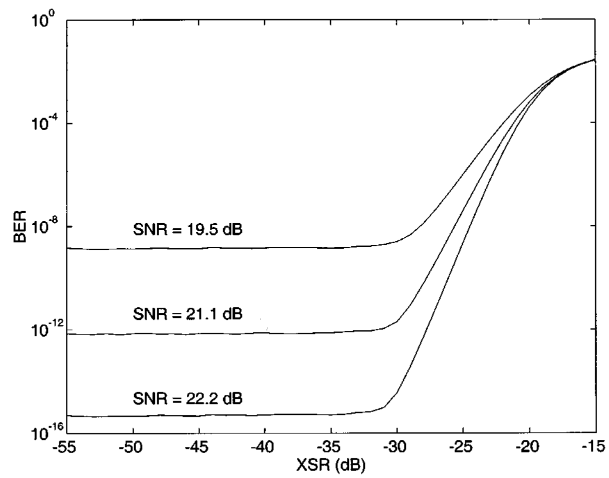


Fig. 7. Dependence of optimum threshold on SNR.

A run-time of 10 s per value was sufficient to achieve this accuracy.[2] This high accuracy is maintained through low BER values, indicating that the IS strategy is close to optimum.

As expected, the GA results in Fig. 5 coincide with the IS results for very low crosstalk levels. The lack of tightness of the CB can be observed: this upper bound is more than one order of magnitude above the true BER at practical crosstalk levels (around $\mathrm{XSR} = -25$ dB). The practical implication of the results in Fig. 5 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.

As stated above, a more realistic system model implies considering an optimized threshold setting. The influence of the threshold setting on the system performance is illustrated in Fig. 6, where the optimum threshold value was obtained by means of the IS-based technique described in Section IV-C. Shown in this figure are BER curves for symmetric threshold as well as for optimum thresholds obtained at three values [of the crosstalk-to-signal ratio (XSR)]: -20, -25, and -30 dB. The number of channels and the AWGN level are the same as in Fig. 5.

The influence of the threshold setting is quite significant. We observe that with optimum threshold the tolerable crosstalk level increases further by 3 dB, for a wide range of XSR values. This is about a 5-dB improvement with respect to the value predicted by the CB. In conclusion, the symmetric threshold assumption used in analyses available to date is quite pessimistic, and our techniques demonstrate that the threshold optimization has a major impact on system performance.

The optimum threshold search requires a relatively large run-time since it implies estimating a BER several times. Fortunately, the threshold that has been found to be the optimum for a particular amount of crosstalk and receiver noise is close to the optimum for a wide range of receiver noise levels. This is shown in Fig. 7, which contains BER curves corresponding to different SNR values, all obtained with the same threshold setting (the optimum setting for $\mathrm{XSR} = -30$ dB and the receiver noise variance used in Fig. 5).

An important parameter in WDM networks is the number of channels. Our techniques can be used to predict the impact of this parameter on the system performance. In Fig. 8, we observe the power penalty due to the introduction of an additional channel. The curves were obtained with $\mathrm{XSR} = -25$ dB and the threshold being optimized at $\mathrm{SNR} = 22$ dB for each of the curves. In this example, the introduction of a fifth channel requires an additional SNR of about 1 dB to maintain the BER at 10^{-9}.

The curves in Fig. 8 were obtained in about 9 min,[3] which is quite a reasonable run-time considering that all the beat-noise components in (28) were included and a realistic threshold setting was used.

V. Discussion and Conclusions

We have demonstrated the validity and practical applicability of new adaptive IS techniques for performance evaluation of

[2]This run-time was also obtained on a Pentium 150-MHz processor with C-language implementation.

[3]On a Pentium 150-MHz processor with C-language implementation.

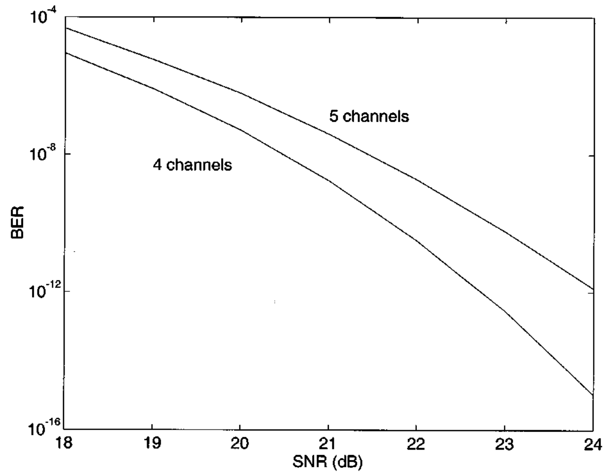


Fig. 8. Impact of the number of WDM channels on the system BER. $\mathrm{XSR} = -25$ dB.

communication systems. The algorithm for optimization of IS parameters, which is based on stochastic Newton recursions, was applied to the performance analysis of the noncoherent OOK and DPSK receivers in AWGN. The second system involved a two-dimensional IS parameter optimization. In both cases, the obtained BER estimates agreed with those obtained by numerical integration (OOK) or exact analysis (DPSK). Furthermore, accurate estimates were obtained in a few seconds.

A stochastic Newton search was applied on top of the IS experiment to determine the optimum detection threshold of the OOK receiver. In this case, the indicator function was approximated by a sigmoid function with parameter c. The algorithm converges for a wide range of values of c and, as expected, yields an unbiased estimator.

An important problem considered was the performance degradation in a WDM network due to crosstalk in optical crossconnects. Worst-case analysis (i.e., all interfering channels are ON) was carried out, and in contrast with the analyses in the literature, we included all the terms in (28). An appropriate IS biasing strategy was designed for both input signal hypotheses. Moreover, stochastic Newton recursions were combined with the g-method, reducing the IS parameter optimization problem to a single dimension and improving the estimator accuracy in the neighborhood of the BER floors. The short simulation run-times (a few seconds) required to yield accurate BER estimates demonstrated the effectiveness of the adaptive IS techniques employed.

We considered a WDM crossconnect with four channels in order to compare our estimates with the commonly used GA (e.g., [20]) and the recently developed CB [21]. In Fig. 5, we observe that at practical XSR levels, the GA is rather pessimistic and the CB is still one order of magnitude above the true BER.

In a WDM experiment, the setting of the detection threshold will not be symmetric. A more realistic performance evaluation tool was achieved here that optimizes the threshold by means of stochastic Newton search. The obtained results (Fig. 6) imply that for a four-channel WDM crossconnect, the tolerable XSR levels are about 5 dB higher than predicted in the literature. This has significant implications on the design of WDM networks.

Finally, our adaptive IS techniques were used to accurately predict the power penalty induced by the introduction of additional WDM channels in relatively short run-times, as shown in Fig. 8.

The impact of additional disturbances on the performance of the WDM crossconnect may be investigated in the future using the adaptive IS techniques presented in this paper.

REFERENCES

[1] 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. Select. Areas Commun.*, vol. 15, pp. 731–741, May 1997.

[2] K. B. Letaief, “The performance of optical fiber direct-sequence spread-spectrum multiple-access communications systems,” *IEEE Trans. Commun.*, vol. 43, pp. 2662–2666, Nov. 1995.

[3] 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.

[4] 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.

[5] R. Srinivasan, “Some results in importance sampling and an application to detection,” *Signal Processing*, pt. 1, vol. 65, pp. 73–88, Feb. 1998.

[6] ——, “Estimation and approximation of densities of I.I.D. sums via importance sampling,” *Signal Processing*, pt. 3, vol. 71, pp. 235–246, Dec. 1998.

[7] 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.

[8] 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. Select. Areas Commun.*, vol. 15, pp. 614–625, May 1997.

[9] K. B. Letaief, K. Muhammad, and J. S. Sadowsky, “Fast simulation of DS/CDMA with and without coding in multipath fading channels,” *IEEE J. Select. Areas Commun.*, vol. 15, pp. 626–639, May 1997.

[10] 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.

[11] P. J. Smith, M. Shafi, and H. Gao, “Quick simulation: A review of importance sampling techniques in communications systems,” *IEEE J. Select. Areas Commun.*, vol. 15, pp. 597–613, May 1997.

[12] K. B. Letaief, “Efficient evaluation of the error probabilities of spread-spectrum multiple-access communications,” *IEEE Trans. Commun.*, vol. 45, pp. 239–246, Feb. 1997.

[13] J.-C. Chen, D. Lu, J. S. Sadowsky, and K. Yao, “On importance sampling in digital communications—Part I: Fundamentals,” *IEEE J. Select. Areas Commun.*, vol. 11, pp. 289–299, Apr. 1993.

[14] 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.

[15] J. S. Stadler and S. Roy, “Importance sampling for detection of scale problems,” *IEEE J. Select. Areas Commun.*, vol. 11, pp. 309–316, Dec. 1993.

[16] 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.

[17] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[18] A. B. Carlson, *Communication Systems*, 3rd ed. New York: McGraw-Hill, 1986.

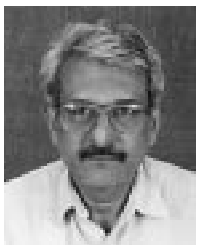
[19] C. G. P. Herben, C. G. M. Vreeburg, D. H. P. Maat, X. J. M. Leijtens, Y. S. Oei, F. H. Groen, J. J. G. M. van der Tol, P. Demeester, and M. K. Smit, “A compact integrated single phasor optical crossconnect on InP,” in *Proc. IEEE/LEOS Symp. Benelux'97*, Eindhoven, The Netherlands, Nov. 1997, pp. 41–44.

[20] H. Takahashi, K. Oda, and H. Toba, “Impact of crosstalk in an arrayed-waveguide multiplexer on $N \times N$ optical interconnection,” *J. Lightwave Technol.*, vol. 14, pp. 1097–1105, June 1996.

[21] 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, The Netherlands, Nov. 1997, pp. 169–172.

David Remondo was born in Madrid, Spain, in 1967. He received the M.S. degree in physics from the University of Cantabria, Santander, Spain, in 1990. He received a second M.S. degree in electrical engineering at the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1992. Since 1995, he has been working toward the Ph.D. degree at the University of Twente, Enschede, The Netherlands.

From 1992 to 1995, he worked at C.E.I.T. in Spain, where he was involved in research activities of the Electrical Engineering Department. His present research interests include performance analysis of optical fiber communications and mobile radio.



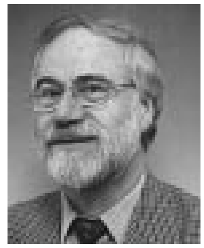
Rajan Srinivasan received the B.S. and Master of Technology (with distinction) degrees from the Indian Institute of Technology, Delhi, India, and the Ph.D. degree from the University of Aston, Birmingham, U.K., in 1976.

He is an electrical engineering educator and researcher, having lectured and carried out research at various academic institutions in the world. Currently, he is a Senior Research Scientist working for the Ministry of Defense of the Government of India. He has authored, singly and jointly, several papers in international journals and conferences. He is listed in the Millenium 2000 Edition of *Marquis Who's Who in the World*. His principal interests are in communications and statistical signal processing with present attention focused on applications of rare event theory and fast stochastic simulation.

In 1984, he received a Young Scientist Award from the URSI (International Union of Radio Science), Florence, Italy, for his work on distributed detection.

Victor F. Nicola received the B.S. and M.S. degrees in electrical engineering from Cairo University, Cairo, Egypt, and Eindhoven University of Technology, Eindhoven, The Netherlands, respectively. He received the Ph.D. degree in computer science from Duke University, Durham, NC.

From 1979, he held faculty and research staff positions at Eindhoven University and at Duke University. In 1987, he joined IBM Thomas J. Watson Research Center, Yorktown Heights, NY, as a Research Staff Member. Since 1993, he has been an Associate Professor at the Department of Electrical Engineering, University of Twente, Enschede, The Netherlands. His research interests include performance and reliability modeling, fault-tolerance, queueing theory, (rare event) simulation methodology, with applications to computer systems and telecommunication networks.



Wim C. van Etten (M'80–SM'91) was born in Zevenbergen, The Netherlands, in 1942. He received the M.Sc. and Ph.D. degrees in electrical engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, in 1969 and 1976, respectively.

From 1969 to 1970, he was employed by Philips Electronics, developing circuits for oscilloscopes. In 1970, he was appointed as Assistant Professor at Eindhoven University of Technology, Faculty of Electrical Engineering. From 1970 to 1976, he was engaged in research on the transmission of digital signals via coaxial and multiwire cables. Since 1976, he has been involved with research and education on optical fiber communications. In 1985, he was appointed Associate Professor at the Eindhoven University of Technology. In 1994, he was appointed Full Professor of Telecommunications at the University of Twente, Enschede, The Netherlands. He is author or co-author of over 60 papers in international journals, conferences, and symposia. He served as a member of organizing committees and program committees, as well as proceedings editor of various conferences and symposia. Moreover, he is first author of the textbook Fundamentals of Optical Fiber Communications(Englewood Cliffs, NJ: Prentice Hall, 1991). His current interests comprise optical communications, mobile communications, detection, and simulation of communication systems.

Prof. van Etten is President of the NERG (Nederlands Elektronica en Radiogenootschap; Dutch Institute of Electronic and Radio Engineers). He is a member of the Executive Committee of the IEEE Benelux Joint Chapter on Communications and Vehicular Technology.

Henk E. P. Tattje studied electrical engineering at the Polytechnic School of Utrecht, Utrecht, The Netherlands.

In 1966, he joined the University of Twente, Enschede, The Netherlands, as a Member of the Scientific Staff of the Department of Electrical Engineering in the Information, Communication and Systems Theory Group. Recently, he joined the Telecommunication Engineering Group. His main research interests include communication systems.

Ir. Tattje is a member of the Dutch Electronic and Radio Engineering Society (NERG).