The use of multicanonical Monte Carlo simulations to obtain accurate bit error rates in optical communications systems

Ronald Holzlöhner and Curtis R. Menyuk
Department of Computer Science and Electrical Engineering,
University of Maryland Baltimore County, TRC 201-B,
1000 Hilltop Circle, Baltimore, Maryland 21250

Submitted to Opt. Lett. on May 12, 2003

We apply the multicanonical Monte Carlo (MMC) method to compute the probability distribution of the received voltage in a chirped return-to-zero (CRZ) system. When computing the probabilities of very rare events, the MMC technique greatly enhances the efficiency of Monte Carlo simulations by biasing the noise realizations. Our results agree with the covariance matrix method over 20 orders of magnitude. MMC can be regarded as iterative importance sampling that automatically converges toward the optimal bias so that it requires less a priori knowledge of the simulated system than does importance sampling. A second advantage is that the merging of different regions of pdfs in order to obtain the entire pdf is not necessary in many cases.

© 2003 Optical Society of America
OCIS codes: 060.2330, 060.5530, 270.2500, 000.5490

The accurate computation of bit error rates (BERs) and the probability distribution function (pdf) of the received voltage in optical communications systems depends on modeling very rare events with probabilities on the order of 10^{-16}–10^{-6}. When forward error correction (FEC) is used, it is still important to accurately model the voltage pdfs prior to correction.[1] One traditional way of computing BERs and eye diagrams is to run Monte Carlo simulations and extrapolate the results under the assumption that the electrical voltage at the receiver after narrow-band filtering is Gaussian distributed in the marks (ones) and spaces (zeros). If the noise realizations in the Monte Carlo simulation are picked in an unbiased way, as is commonly done, this method cannot yield information about the tails of the pdf within a reasonable computational time. As a consequence, system designers often use a simplified approach in which they assume that the optical noise spectrum at the receiver is white, effectively neglecting the nonlinear signal-noise interaction in the fiber. This simplification is inappropriate for many long-haul optical communications systems. By contrast, the covariance matrix method does take into account the nonlinear signal-noise interaction and works for nonlinear systems.[2–4] However, it can fail in principle when the nonlinear noise-noise interaction in the fiber becomes large.

In this contribution, we apply the multicanonical Monte Carlo (MMC) simulation technique that was proposed by Berg and Neuhaus[5] in 1992. MMC is closely related to importance sampling[6] in the sense that both methods increase the number of events in the tail region of the pdf by biasing them. In standard importance sampling, one needs to guess what regions of the state space are the major source of the bit errors and devise a set of biases that will preferentially sample these regions. The MMC method automatically determines the bias by using an iterative procedure. The iterative procedure uses a control quantity to update the set of biases for the next iteration, so that as the iteration number increases, there tends to be an approximately equal number of hits in each bin of the histogram of the control quantity.[5] Moreover, the merging of different regions of a distribution[7] is not necessary in many cases. Recently, Yevick[8] successfully applied the MMC method to compute polarization mode dispersion (PMD) emulator statistics. In this paper, we outline the MMC algorithm, and we then employ it to compute the received voltage pdfs in an optical communications system and compare these pdfs with the pdfs that we previously computed using the covariance matrix method.[3]

We introduce a state space Γ with a probability density ρ, *i.e.*, $\rho(\mathbf{z}) \geq 0$ and $\int_\Gamma \rho(\mathbf{z})\, d\mathbf{z} = 1$. When computing bit errors due to noise, Γ is the space of all possible amplified spontaneous emission (ASE) noise inputs at all amplifiers and all frequencies. We want to compute the pdf $p(V)$ for each value of the received voltage V. We partition Γ into M subspaces $\Gamma_k = \{\mathbf{z} \in \Gamma \,|\, (k-1)\Delta V \leq V(\mathbf{z}) < k\Delta V\}$, where ΔV is a small voltage difference, and $1 \leq k \leq M$. If $P_k = \int_{(k-1)\Delta V}^{k\Delta V} p(V)\, dV$ is the probability that $(k-1)\Delta V \leq V(\mathbf{z}) < k\Delta V$, then

$$P_k = \int_\Gamma \chi_k(\mathbf{z})\rho(\mathbf{z})\, d\mathbf{z}, \tag{1}$$

where $\chi_k(\mathbf{z}) = 1$ if $\mathbf{z} \in \Gamma_k$ and $\chi_k(\mathbf{z}) = 0$ otherwise. One can use a standard Monte Carlo simulation to approximate P_k by

$$P_k \approx \frac{1}{N} \sum_{i=1}^{N} \chi_k(\mathbf{z}^i), \tag{2}$$

where the $\mathbf{z}^i$ are N random sample points in Γ, selected according to the probability density ρ.

The goal of any biasing scheme, including MMC, is to reduce the variance of the sum in Eq. (2) by introducing a positive *biasing pdf* $\rho^*(\mathbf{z})$. We rewrite Eqs. (1) and (2) as[6]

$$\begin{aligned} P_k &= \int_\Gamma \chi_k(\mathbf{z}) \frac{\rho(\mathbf{z})}{\rho^*(\mathbf{z})}\rho^*(\mathbf{z})\,d\mathbf{z} \\ &\approx \frac{1}{N}\sum_{i=1}^{N} \chi_k(\mathbf{z}^{*,i})\frac{\rho(\mathbf{z}^{*,i})}{\rho^*(\mathbf{z}^{*,i})}, \end{aligned} \tag{3}$$

where the $\mathbf{z}^{*,i}$ are sampled from $\rho^*(\mathbf{z})$ instead of $\rho(\mathbf{z})$. The ratio $L = \rho(\mathbf{z}^{*,i})/\rho^*(\mathbf{z}^{*,i})$ is called the likelihood ratio. The variance of the sum in Eq. (3) is zero if the optimal biasing pdf $\rho^*(\mathbf{z}) = \rho^*_{\mathrm{opt}}(\mathbf{z}) = \chi_k(\mathbf{z})\rho(\mathbf{z})/P_k$ is used. However, $\rho^*_{\mathrm{opt}}(\mathbf{z})$ depends on P_k and hence is initially unknown. In standard importance sampling, one uses physical intuition to guess a biasing pdf that is close to ρ^*_{opt}. The MMC algorithm instead iterates over biasing pdfs $\rho^{*,j}$ that approach ρ^*_{opt}. We define $\rho^{*,j}$ for the jth iteration by

$$\rho^{*,j}(\mathbf{z}) = \frac{\rho(\mathbf{z})}{c^j P^j_k}, \qquad \mathbf{z} \in \Gamma_k. \tag{4}$$

The quantities P^j_k satisfy $P^j_k > 0$ and $\sum_{k=1}^{M} P^j_k = 1$, where we recall that M is the number of partitions of the state space, and c^j is an unknown constant that ensures $\int_\Gamma \rho^{*,j}(\mathbf{z})\,d\mathbf{z} = 1$. The vector of the P^j_k is the key quantity in the MMC algorithm and completely determines the bias. The P^j_k are updated after each iteration such that after a number of iterations, the expected number of samples in each bin of the histogram is $\langle \sum_{i=1}^{N} \chi_k(\mathbf{z}^{*,i})\rangle = N/M$ and is hence independent of k.[9] Substituting this value in Eqs. (3) and (4) yields $c^j \to M$ and $P^j_k \to P_k$.[10] We will discuss the assumptions under which this convergence can be achieved in a later publication.

Within each MMC *iteration* j, we employ the Metropolis algorithm[11] to produce a random walk of *samples* $\mathbf{z}^{*,i}$ whose distribution equals $\rho^{*,j}(\mathbf{z})$. We consider a Markov chain of transitions consisting of small steps in the noise space. Each transition goes from $\mathbf{z}^{*,i} = \mathbf{z}^*_a \in \Gamma_k$ to $\mathbf{z}^*_b = \mathbf{z}^*_a + \epsilon^j \Delta\mathbf{z}$, where $\Delta\mathbf{z}$ is random and symmetric, *i.e.,* it does not favor any direction in Γ, and the transition is accepted with probability π_{ab}. If a transition from $\mathbf{z}^{*,i} = \mathbf{z}^*_a$ to $\mathbf{z}^*_b$ is accepted, we set $\mathbf{z}^{*,i+1} = \mathbf{z}^*_b$, and otherwise we set $\mathbf{z}^{*,i+1} = \mathbf{z}^{*,i} = \mathbf{z}^*_a$.

In our simulations, ρ is the product of the pdfs of all ASE noise inputs at each amplifier and each frequency, which we assume to be independent identical Gaussian pdfs ρ_l with $\rho = \prod_{l=1}^{d} \rho_l$, where d is the dimension of Γ. We consider the perturbation of the noise component in each bit $z^*_{a,l}$ of $\mathbf{z}^*_a$ separately, and accept or reject it independently with the probability $\min\left[\rho_l(z^*_{b,l})/\rho_l(z^*_{a,l}),\, 1\right]$. We pick each perturbation Δz_l from a zero mean symmetric pdf. We obtain a trial state $\mathbf{z}^*_b$ in which only some of the components are different from their previous values in $\mathbf{z}^*_a$. Next, we compute k_b and finally accept the step from $\mathbf{z}^*_a$ to $\mathbf{z}^*_b$ with the probability $\min\left[P^j_{k_a}/P^j_{k_b},\, 1\right]$. The compound transition probability is hence

$$\pi_{ab} = \left\{ \prod_{l=1}^{d} \min\left[\frac{\rho_l(\mathbf{z}^*_{b,l})}{\rho_l(\mathbf{z}^*_{a,l})},\, 1\right] \right\} \min\left[\frac{P^j_{k_a}}{P^j_{k_b}},\, 1\right]. \tag{5}$$

The probability ratio π_{ab}/π_{ba} equals $\rho^{*,j}(\mathbf{z}^*_b)/\rho^{*,j}(\mathbf{z}^*_a)$, which is the detailed balance condition that ensures that the limiting (stationary) distribution for infinitely many steps of this random walk is $\rho^{*,j}$.[11]

In each iteration, the perturbation coefficient ϵ^j is constant for all samples. After each iteration, we adjust ϵ^j so that the *acceptance ratio* α, which is the ratio of the number of accepted steps to the total number of steps N, is close to 0.3. The minimum required number of samples N of this random walk depends on the average step size $\alpha\epsilon^j \langle |\Delta\mathbf{z}|\rangle$ and is hence system-dependent. The noise realizations are recorded in the histogram $H^{*,j}$, where $H^{*,j}_k = \sum_{i=1}^{N} \chi_k(\mathbf{z}^{*,i})$ is the number of the $\mathbf{z}^{*,i}$ in iteration j that fall into Γ_k. The P^k_j are updated after each MMC iteration using the recursion relations given in[9] based on the histogram $H^{*,j}$. As j increases, the expected number of samples $\langle H^{*,j}_k\rangle$ becomes independent of the bin number k, which implies that $P^j_k \to P_k$.

We note that we could in principle also pick all perturbations at the same time and accept the entire step with the probability $\min\left[\rho(\mathbf{z}^*_b)P^j_{k_a}/\rho(\mathbf{z}^*_a)P^j_{k_b},\, 1\right]$. The advantage of our individual perturbation method is that we can use significantly larger perturbations ϵ^j. However, this method only works in problems where the pdf ρ factors into independent pdfs ρ_l.

In the first iteration, one can set $P^1_k = 1/M$ or use an initial guess for the P^1_k. We have used both approaches. To update the P^j_k at the end of iteration j, we initially set P^{j+1}_1 to an arbitrary positive value and use the recursion relations[9]

$$P^{j+1}_{k+1} = \frac{P^{j+1}_k P^j_{k+1}}{P^j_k}\left(\frac{H^{*,j}_{k+1}}{H^{*,j}_k}\right)^{\widehat{g}^j_k}, \tag{6a}$$

$$\widehat{g}^j_k = \frac{g^j_k}{\sum_{l=1}^{j} g^l_k}, \qquad g^l_k = \frac{H^{*,l}_k H^{*,l}_{k+1}}{H^{*,l}_k + H^{*,l}_{k+1}}, \tag{6b}$$

where in addition we define $\widehat{g}^j_k = 0$ if $g^j_k = 0$ and $g^l_k = 0$ if $H^{*,l}_k + H^{*,l}_{k+1} = 0$. The exponent $0 \le \widehat{g}^j_k \le 1$ hence depends on all previous iterations. Finally, we normalize the P^{j+1}_k so that $\sum_{k=1}^{M} P^{j+1}_k = 1$.

We applied the MMC algorithm to a 10 Gb/s, single-channel chirped return-to-zero (CRZ) system of length 6,100 km that resembles a submarine system. Previously, we have computed the pdf of the received voltage in this system using the covariance matrix method and compared its results to standard Monte Carlo simulations.[3]

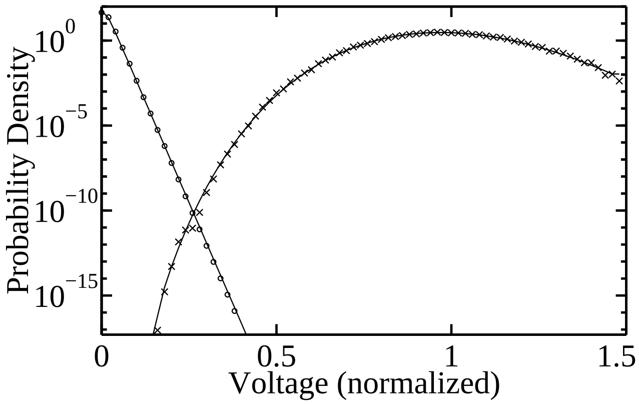

Figure 1. Pdfs of the low-pass filtered received voltage for the marks and the spaces in the CRZ system. Circles and crosses: Results from the last iteration of MMC simulations with 55,000 and 100,000 samples, respectively; solid lines: covariance matrix method.

However, we were only able to show agreement over about six orders of magnitude. By contrast, the MMC method enables us to show agreement over about 20 orders of magnitude.

The state space Γ is of dimension $d =$ (34 dispersion map periods) × (4 amplifiers per period) × (140 relevant frequencies) $\times\, 2 = 38{,}080$, where the last factor of 2 accounts for the real and imaginary parts of the optical noise. We launch a 32-bit PRBS signal and calculate the average pdf of the received voltage in the marks by averaging over all 16 marks, and we use the same procedure in the spaces. We consider $v(t_{1,n})$ and $v(t_{0,n})$, where $t_{1,n}$ and $t_{0,n}$ with $n = 1\ldots16$ are the points in time at the centers of the bit slots that contain a mark or a space, respectively, and $v(t)$ is the low-pass filtered received voltage as a function of time. We record the average pdfs of $v(t_{1,n})$ and $v(t_{0,n})$ in weighted histograms by quantizing $v(t_{1,n})$ and $v(t_{0,n})$. For each noise realization $\mathbf{z}^{*,i}$, we increment the value of the histogram in the appropriate bin by the likelihood ratio $L = \rho(\mathbf{z}^{*,i})/\rho^*(\mathbf{z}^{*,i}) = c^j P_k^j$, $\mathbf{z}^{*,i} \in \Gamma_k$, according to Eq. (3). The pdfs of $v(t_{1,n})$ and $v(t_{0,n})$ can be regarded as vertical slices through an eye diagram.

Figure 1 shows the average pdf of $v(t_{1,n})$ and $v(t_{0,n})$ in the last MMC iteration. The voltage is normalized by $\max_n[v(t_{1,n})]$ in the absence of noise. The solid line shows the result of the covariance matrix method, and the crosses and circles show the pdf from the MMC simulations for the marks and the spaces. The agreement is excellent.

In the following, we summarize additional details of the algorithm. To bias the marks, we seek to decrease the maximum voltage in the marks to close the eye. We therefore choose the control quantity to be $V = \min_n[v(t_{1,n})]$ for each noise realization. Conversely, we seek to increase the voltage in the spaces, and in that simulation we set $V = \max_n[v(t_{0,n})]$. In our simulations, the voltage V thus merely plays the role of a control quantity and has no other use.

We set $M = 50$. In the first iteration, we chose $N = N^{j=1} = 5000$ samples. The simulation covers a larger voltage range with each new iteration, and moreover the Metropolis random walk tends to accept more steps at voltages where $p(V)$ is large than in the tails of $p(V)$. We therefore increase the number of samples after each iteration so that $N^{j+1} = 1.15\,N^j$. In the simulation of the marks, we needed 55,000 total samples in 8 iterations and to simulate the spaces, we needed 101,000 total samples in 10 iterations. We considered the simulation to have sufficiently converged when $\max_k |(P_k^j - P_k^{j+1})/P_k^{j+1}| < 0.1$, which is hardly visible on a log scale.

In conclusion, we applied the MMC simulation technique to a CRZ system with a transmission distance of 6,100 km. We were able to compute the pdfs of the low-pass filtered received voltage over a range of 20 orders of magnitude. We compare this result with the covariance matrix method[2–4] and obtain excellent agreement, fully validating the covariance matrix method for this system. This result also demonstrates the usefulness of the MMC method for calculating the complete voltage pdf in a receiver, while at the same time accounting for the full nonlinear noise-noise interaction in the fiber.

The authors would like to thank J. Zweck, W. L. Kath, V. S. Grigoryan, and G. Biondini for helpful discussions. R. Holzlöhner's email is holzloehner@umbc.edu.

References

1. Y. Cai, J. M. Morris, T. Adalı, and C. R. Menyuk, J. Lightwave Technol. **21**, 727 (2003).
2. R. Holzlöhner, V. S. Grigoryan, C. R. Menyuk, and W. L. Kath, J. Lightwave Technol. **20**, 389 (2002).
3. R. Holzlöhner, C. R. Menyuk, W. L. Kath, and V. S. Grigoryan, IEEE Photon. Technol. Lett. **14**, 1079 (2002).
4. R. Holzlöhner, C. R. Menyuk, W. L. Kath, and V. S. Grigoryan, IEEE Photon. Technol. Lett. **15**, 688 (2003).
5. B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
6. P. M. Hahn and M. C. Jeruchim, IEEE Trans. Commun. **35**, 706 (1987).
7. G. Biondini, W. L. Kath, and C. R. Menyuk, IEEE Photon. Technol. Lett. **14**, 310 (2002).
8. D. Yevick, IEEE Photon. Technol. Lett. **15**, 224 (2003).
9. B. A. Berg, Nucl. Phys. Proc. Suppl. **63**, 982 (1998).
10. $P_k^j \propto 1/w^j(k)$, where $w^j(k)$ is Berg's[9] "weight". We choose to deal with the P_k^j rather than the $w^j(k)$, since we consider them to be more intuitive quantities.
11. N. Metropolis and S. Ulam, J. Am. Stat. Assoc. **44**, 335 (1949).