

Stratified Sampling Monte Carlo Algorithm for Efficient BER Estimation in Long-Haul Optical Transmission Systems

Paolo Serena, Nicola Rossi, Marco Bertolini, *Student Member, IEEE*, and Alberto Bononi

Abstract—We propose an efficient stratified sampling (SS) algorithm for estimating the bit error rate (BER) of a digital communication system. Our algorithm efficiently exploits the observations of an approximate, but usually fast, model of the system under investigation to drive a clever Monte Carlo (MC) estimation based on SS. The proposed method is faster than standard MC even at BER in the range 10^{-3} to 10^{-5} . Moreover, it is possible to evaluate the estimated standard deviation of the measured BER, such as in an MC simulation, so that it is possible to associate a confidence to the results. We test the algorithm both in a simple optical system distorted by group velocity dispersion (GVD) and in more complex differential quadrature phase shift keying (DQPSK) systems. In the last case, we measured computational savings up to 70% compared with standard MC.

Index Terms—Differential quadrature phase shift keying (DQPSK), Monte Carlo (MC) methods, optical communication.

I. INTRODUCTION

THE bit error rate (BER) is the main performance parameter of a digital communication system. The growing complexity of optical links has increased the impact of nonlinear effects calling for efficient numerical algorithms for BER estimation.

An effective and widely adopted BER estimation method is the Karhunen–Loève (KL) algorithm [1] based on signal propagation through the split step Fourier method (SSFM), followed by semi-analytical receiver analysis using noise expansion in a suitable base. The KL algorithm gives the exact BER for common quadratic receivers in presence of additive Gaussian noise. Unfortunately, there are situations where the noise is not Gaussian, for instance in presence of nonlinear phase noise induced by the Kerr effect of the optical fiber [2]. An approximate extension of the KL algorithm to cope with nonlinear phase noise was proposed in [3] for moderate nonlinear phase rotations.

The most widespread BER estimation algorithm, working for any set-up with the desired error confidence, is so far the direct Monte Carlo (MC) error counting [4]. Unfortunately, MC simulations are very slow because they require transmission of long

bit patterns for accurate estimation. Such a problem can be alleviated at small BER values ($< 10^{-6}$) by multicanonical Monte Carlo (MMC) algorithms [5]. To our experience, MMC is however less efficient than MC at predecoder BER comparable or larger than 10^{-5} , as commonly found in systems employing forward error correction (FEC). One of the reasons is that the first cycle in MMC is indeed a pure MC. Besides, the MC algorithm is widely used in the community for three main reasons. First, it is very simple to implement; second, where MC is based on independent observations, its accuracy can be associated with a reliable and simple parameter, i.e., the estimated variance of the measured BER; third, it produces parallel measurements over the entire propagated bit-sequence, instead of a single-target bit as in MMC [6].

Motivated by such features of MC, in this study we present an improved MC method based on stratified sampling (SS) [4], [7], which can be more efficient than MC even at BERs in the range 10^{-3} to 10^{-5} . The idea of SS is to run independent MC estimations over disjoint subsets forming a partition of the sample space, motivated by the intuition that one needs a few samples to estimate the system observable in a given subset if the observable varies very little within such a subset. Clearly, the identification of a suitable partition is a key point for SS and in this paper we propose an original solution.

We will show that our practical implementation of the SS idea is a novel form of adaptive importance sampling (IS) [8].

Optimal IS-biasing requires a detailed knowledge of the system input/output relation, hence suboptimal, non-adaptive biasing is usually implemented in traditional IS, either by biasing the variance [9] or the mean [10] of the known input sample statistics. However, such simple approaches have several drawbacks. Biasing the variance leads to a dimensionality problem [11], so that the intrinsic IS gain disappears in problems with large-dimensional input spaces. Unfortunately, this is the case with standard optical simulations, where the input space dimension (i.e., the number of random variables that determine the decision statistic) easily exceeds 10^3 . Biasing the mean, instead, requires a signal-dependent optimization that makes its implementation and control more difficult [11].

The proposed SS algorithm is simple, stable, and has the same MC stopping criterion based on the estimated variance. It also has a significant speed gain compared with standard MC. As a case study, we apply it to a simple system distorted by only group velocity dispersion (GVD) and a more complex one with both nonreturn to zero (NRZ) and return to zero (RZ) differential quadrature phase shift keying (DQPSK) propagated along a dispersion managed ultralong haul system with wavelength di-

Manuscript received June 26, 2008; revised November 07, 2008. First published April 17, 2009; current version published June 26, 2009. This work was supported by a grant from Alcatel-Lucent Bell Labs, France.

The authors are with the Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Parma, V.le G. Usberti 181/A, 43100 Parma, Italy (e-mail: serena@tlc.unipr.it)

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JLT.2008.2010560

vision multiplexing (WDM). In both the cases, we compare the performance with standard MC.

Besides our specific applications in the optical field, our SS idea can, in principle, be applied to estimate the average value of any function of random variables.

The paper is organized as follows. In Section II, we introduce the basics of MC and SS. In Section III, we describe our original implementation of the SS algorithm. In Section IV, we investigate the performance of the SS algorithm. In Section V, we draw our conclusions. The Appendix contains a brief introduction of the SS concept, and to how it can be actually implemented as an IS algorithm.

II. BACKGROUND

Suppose I is a random variable with unknown average value $E[I]$ and variance σ^2 . $I = g(\mathbf{X})$ is a computationally expensive function of the random vector $\mathbf{X} = (x_1, \dots, x_M)$ taking values in the input space Ω . We wish to estimate $E[I]$. The maximum likelihood estimator of $E[I]$ is the sample mean,

$$\bar{I} = \frac{1}{n} \sum_{k=1}^n I^{(k)} \quad (1)$$

with $I^{(k)} = g(\mathbf{X}_k)$, $k = 1, \dots, n$, where \mathbf{X}_k are independent realizations of the random vector \mathbf{X} . Such an approach is also called MC estimation [4], [7], [12]. \bar{I} is an unbiased estimator of $E[I]$, with standard deviation [7]

$$\bar{\sigma} = \sqrt{\frac{\sigma^2}{n}}. \quad (2)$$

Since σ^2 is unknown, in order to estimate $\bar{\sigma}$ from simulation results, it is customary to replace the variance σ^2 in (2) with its maximum likelihood estimator, i.e., one computes

$$\sigma^2 \cong \frac{1}{n-1} \sum_{k=1}^n (I^{(k)} - \bar{I})^2.$$

Thanks to the \sqrt{n} factor in the denominator of (2), the standard deviation of \bar{I} decreases for increasing n , making \bar{I} a good estimator of the true average value for $n \gg 1$. In this case, it is possible to find n such that the estimated average value and the true one satisfy the following probability relation [12]:

$$Pr \{ \bar{I}(1 - \Delta_n) < E[I] < \bar{I}(1 + \Delta_n) \} = C \quad (3)$$

where C is the confidence level of the estimator, while Δ_n is the relative error at confidence C ,

$$\Delta_n = \frac{\varepsilon \bar{\sigma}}{\bar{I}} \quad (4)$$

being ε the $(1 - C)$ -percentile. By the central limit theorem [12], \bar{I} will have approximately Gaussian statistics, and thus $\varepsilon = \sqrt{2} \operatorname{erfc}^{-1}(1 - C)$ [12]. For instance, $\varepsilon = 1$ when $C \simeq 68\%$.

The procedure for an MC estimation is therefore the following. First, set a target relative error Δ with confidence level C . Second, compute ε . Third, run the MC estimation by increasing n until the relative error $\Delta_n \leq \Delta$.

Δ_n is a simple measure of the estimation accuracy at a given confidence level. Such observation is one of the reasons for the

success of MC among estimation algorithms. Another motivation for using MC relies on the fact that the accuracy is independent of the size of \mathbf{X} , i.e., of the dimensionality of the input space Ω . This is probably the major motivation for using MC to estimate integrals in spaces with large dimensionality, where the complexity of quadrature formulas is too large [4], [7].

Before continuing, let us tailor the previous general discussion to a digital system. I becomes the error indicator, equal to 1 when an error occurs and 0 otherwise. The desired average value of I , $E[I]$, is the probability of error. \mathbf{X} is a random vector containing all the random variables of the discrete equivalent model of the transmission system $g(\mathbf{X})$, including bits and discrete noise samples generated along the line. n is the number of samples, which does not necessarily coincide with the number of MC runs. Usually, in block simulations such as the ones using the SSFM and thus based on fast Fourier transform (FFT), at each run a block of N_b bits is propagated, and all of them can be used to update the sample mean.

The number of runs n required in MC estimation scales as $1/\text{BER}$. Although at BER values in the range 10^{-3} to 10^{-5} the number of required runs n may not look too large, for very "expensive" computation of the system output $g(\mathbf{X})$ we do have a strong motivation for looking for more efficient estimation algorithms.

A known improvement on MC is SS [4], [7], which runs r different MC samplings on r disjoint subsets (or strata) making up a partition of the input space Ω . The basics of SS are reviewed in the Appendix. The key idea is that if I is well known within a stratum, then samples $I^{(k)}$ from that stratum have a small variance, and running a reliable MC estimation in that stratum requires only a few samples, as per (2) and (4). We call p_s the probability of selecting stratum s , with $s = 1, \dots, r$, \bar{I}_s the sample mean of the MC estimation within stratum s , $\bar{\sigma}_s$ its corresponding standard deviation as in (2), while σ_s is the standard deviation of I conditioned on stratum s . $\bar{\sigma}_s$ and σ_s are related through (2). From the total probability law [12], we have that the global sample mean can be estimated as

$$\bar{I} = \sum_{s=1}^r p_s \bar{I}_s \quad (5)$$

while the variance of such estimator is

$$\bar{\sigma}^2 = \sum_{s=1}^r p_s^2 \bar{\sigma}_s^2. \quad (6)$$

Let n_s be the number of visits in stratum s so that $n = n_1 + \dots + n_r$. The basic optimization problem in SS is the following: for a given number of samples n , how should they optimally be subdivided into the r strata? That is, what are the best n_s , $s = 1, \dots, r$, so that $\bar{\sigma}$ is minimum? As shown in the Appendix, the answer is [7]

$$n_s = n \frac{\sigma_s p_s}{\sum_{k=1}^r \sigma_k p_k}. \quad (7)$$

The physical interpretation of the result is simple: concentrate most samples in strata of low knowledge, i.e., large variance. Using (7), it can be shown that $\bar{\sigma}$ for SS is smaller than the one for MC [7].

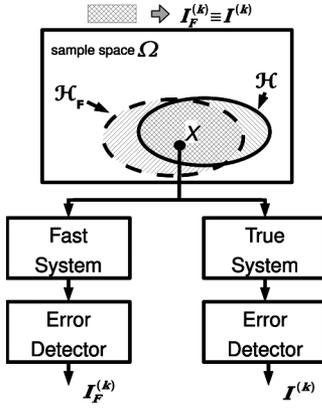


Fig. 1. Sketch of our implementation of an SS with two strata, \mathcal{H}_F , and \mathcal{H}_F^c , i.e., the sets in input space that provide/do not provide an error in the fast system, respectively. $I_F^{(k)}$ and $I^{(k)}$ are error indicators at run k of the fast/true system, respectively. \mathcal{H} is the error set induced by the true system. In this example, the input \mathbf{X} induces an error both in the fast and the true system.

III. THE COARSE/FINE APPROACH

The key issue in implementing SS is identifying strata of good knowledge. Our novel approach in the choice of the strata is described next. We partition the input space in just two strata. The strata are determined by a coarse, approximate model of the true system, which we call the “fast system”, since it is much faster to compute. The idea is sketched in Fig. 1. The true system $g(\cdot)$ maps \mathbf{X}_k to $I^{(k)}$ at each run k , and uniquely determines the true error set \mathcal{H} , i.e., the set of all input \mathbf{X} such that $\{I = 1\}$. The fast system, being an approximate model of the true system, maps \mathbf{X}_k to a different $I_F^{(k)}$, hence determining an approximate error set \mathcal{H}_F (stratum 1), which hopefully well overlaps with \mathcal{H} . Stratum 2 is the complementary set, i.e., $\mathcal{H}_F^c = \Omega \setminus \mathcal{H}_F$. The probability p_s , $s = 1, 2$, of selecting each stratum is estimated by running the fast system with independent realizations of \mathbf{X}_k at each run. Meanwhile, using a subset of the same \mathbf{X}_k samples, we run separate MC simulations using the true system within each stratum, and estimate the conditional BER \bar{I}_s and variance $\bar{\sigma}_s^2$ in each stratum. From such estimates we can estimate the global sample mean as in (5) and the variance of such an estimator as per (6).

In order to make our SS scheme efficient we have to visit each stratum with the optimal strategy (7). If we tested the true system using all \mathbf{X}_k samples used by the fast system in place of a subset, SS would visit stratum 1 with the probability of error p_1 of the fast system, thus like a standard MC simulation but different from the request of (7). In order to distribute the visits between the strata along the lines suggested by the optimal strategy, we have to skip some \mathbf{X}_k tested by the fast system. To this aim, the number of MC samples within each stratum is chosen by propagating in the true system sample \mathbf{X}_k in stratum 1 with probability $\alpha_1 = \min(1, \sigma_1/\sigma_2)$, and in stratum 2 with probability $\alpha_2 = \min(1, \sigma_2/\sigma_1)$. The motivation behind this choice is the following. Suppose that $\sigma_1 > \sigma_2$. Since $\alpha_1 = 1$ all \mathbf{X}_k yielding $I_F^{(k)} = 1$ (i.e., stratum 1) are accepted and therefore propagated also in the true system. n_F runs in the fast system yield on average $n_1 = n_F p_1$ runs in the true system in stratum 1. On the other hand, $\alpha_2 = \sigma_2/\sigma_1 < 1$; thus we have $n_F p_2 \alpha_2$

runs in stratum 2 of the true system. Hence, the ratio of visits, n_1 in stratum 1, to total number of runs in the true system n is

$$\frac{n_1}{n} = \frac{p_1}{p_1 + p_2 \frac{\sigma_2^2}{\sigma_1^2}}$$

exactly as prescribed by optimal SS, i.e., as derived from (7) with $s = 1$ and $r = 2$. Since the true standard deviations σ_1 and σ_2 are unknown, we use their sample estimates derived from propagation through the true system, which we *adaptively* update at each run through the true system.

Note that $\sigma_2 = \sqrt{\text{Var}[I | \mathcal{H}_F^c]}$ is the variance of the true error indicator when samples are in stratum 2 (\mathcal{H}_F^c), i.e., no error was signaled in the fast system. If the fast system is well chosen, i.e., if \mathcal{H}_F is reasonably close to \mathcal{H} , a gain in efficiency is achieved since for most samples in stratum 2 even the true system will have no errors, yielding $I = 0$ and therefore the variance σ_2 will be close to zero. Hence, most samples in stratum 2 will not be propagated in the true system. Such a gain in efficiency is similar to the one achieved by the improved IS method proposed in [13] and known as *IS with excision* [8].

However, our SS strategy has an overhead due to the runs into the fast system. Efficiency comes from a trade-off between overhead and skipped runs in the true system. The simulation ends when the relative accuracy in (4) is less than a given tolerance *TOL*.

The SS idea has been described so far for a single error observation, $I^{(k)}$. However, for a block simulation each \mathbf{X}_k contains N_b bits, hence for each run we have N_b different observations of $I^{(k)}$. We can take into account all of these observations by extending the definition of stratum 1 to the set “at least one error occurs in the fast system over the N_b bits.” Then the SS method can be applied along the same lines as before. Clearly, for large values of N_b and high error probabilities it is $p_1 \rightarrow 1$, thus the method becomes similar to standard MC.

The basic algorithm is described in Fig. 2. The rejection test is implemented by comparing a random number uniformly distributed in the unit interval (*rand*) with the ratio of the standard deviations evaluated over the available samples in the true system. In order to use the same \mathbf{X}_k both in the fast system and in the true one, we simply set the random state of the noise generator to the same value for both systems. Note that since I is a Bernoulli random variable, it is $\sigma^2 = \text{E}[I](1 - \text{E}[I])$, therefore the sample variance can be inferred from the sample mean. The sample mean can be updated recursively at each run as suggested in ([7], Section 2.7.3).

We start applying the rejection test only after the occurrence of one error in stratum 1. Such a condition ensures $\sigma_1 \neq 0$ so that $\alpha_1 \neq 0$. We cannot wait for an error also in stratum 2, since its probability can be very small if $\mathcal{H}_F \approx \mathcal{H}$. It may however happen that the estimate of σ_2 be zero, which implies that the rejection test will always reject visits in stratum 2 after the first error event in stratum 1. Prudently, in the accept/reject tests we prefer to use ratio $m = \min(\tilde{m}, \sigma_1/\sigma_2)$ instead of σ_1/σ_2 , being \tilde{m} a bound we impose so that stratum 2 gets visited with a finite probability even in the case of zero errors within it.

In the next section, we apply the algorithm to practical cases of optical transmission systems by discussing our proposal for the actual implementation of the fast system.

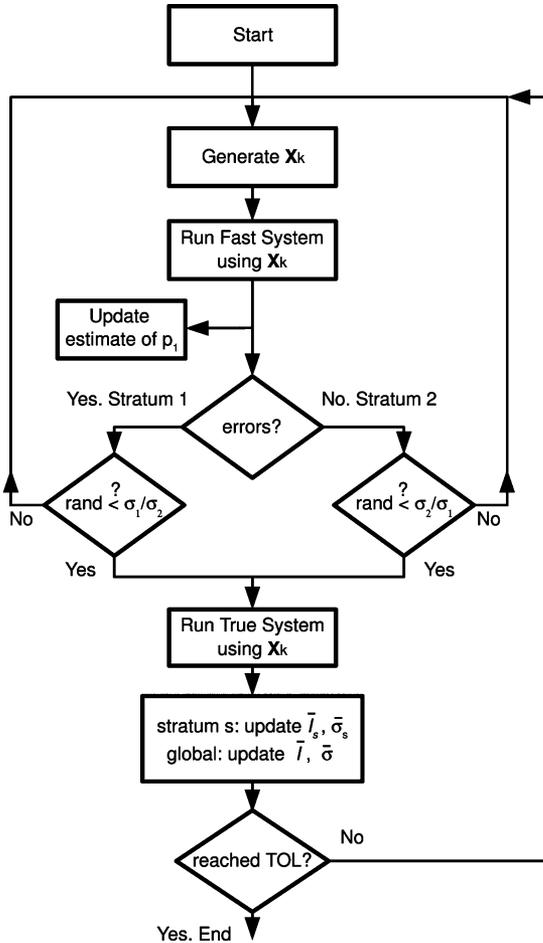


Fig. 2. Proposed stratified sampling algorithm.

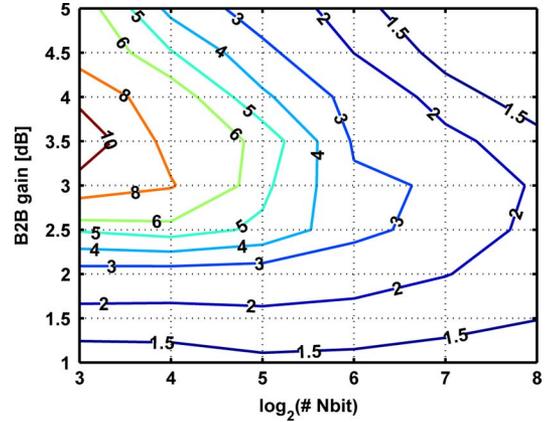
IV. APPLICATIONS

The proposed algorithm efficiency crucially depends on the choice of the fast system. In this section, we first discuss an implementation proposal in a simple case of a system affected by GVD, then we move to practical cases working in the nonlinear regime.

A. Only GVD

With only GVD the probability of error can be evaluated exactly by means of KL expansion of signal and noise [1]. However, applying the SS algorithm to such a case allows us to rapidly explore the behavior of several parameters in a wide range of values, and thus to understand the efficiency of the method. Hence, we use GVD as a simple source of intersymbol interference (ISI).

The true system under investigation is a purely linear fiber with cumulated dispersion D_t , followed by a noisy amplifier whose noise figure is chosen in order to have the desired signal-to-noise ratio at the receiver. The signal is a single-channel NRZ differential phase shift keying (DPSK), with bitrate $R_b = 10$ Gb/s. We now need an approximate and fast implementation of such a system. Since the BER is a function of the optical signal-to-noise ratio (OSNR), the GVD induced ISI degrades the BER in a similar way as a back-to-back (B2B) transmission with noise inflated. Hence, our fast system is a B2B transmission with a noise loading in


 Fig. 3. Ratio η between the number of runs into the true system required by MC and the one by SS, versus number of bits and B2B gain B . $D_t = 1000$ ps/nm, NRZ-DPSK at 10 Gb/s. Only GVD.

front of the receiver having noise figure B dB larger than the true system. Such an approach allows to save 2 FFT (direct and inverse) operations at each run. The optimal value of B is not known *a priori*, hence the SS-algorithm works efficiently if a computational gain appears in a wide range of B .

As a first test we fixed $D_t = 1000$ ps/nm, which at $\lambda = 1550$ nm corresponds to a normalized dispersion $\lambda^2 R_b^2 D_t / (2\pi c)$, where c is the speed of light. The OSNR was fixed at 8 dB over a bandwidth of 0.1 nm, which we verified using the KL method to yield $\text{BER} = 1.06 \times 10^{-4}$. The BER with zero GVD (standard B2B) was 7.5×10^{-6} , yielding an OSNR penalty w.r.t. B2B of 1 dB. The signal pattern was a random sequence of variable length in the range 2^3 to 2^8 bits. Each bit was discretized using 128 time samples. We measured the BER using standard MC and SS, both stopped with the same relative accuracy $\Delta = 0.1$ with confidence 95% ($\varepsilon \simeq 2$).

Fig. 3 shows contour levels of the ratio η between the number of bits measured in the true system using MC and the ones using SS, versus sequence length and B2B gain B . Note that η does not include the overhead of the fast system, which has negligible cost here compared to the true system. The contour levels are the average η among five simulations with different random seeds in order to exclude possible best/worst cases. From the figure we learn that the largest “computational gain” η is reached with short bit sequences, as discussed in Section III. Note that at best the average gain η is a factor 10 when $B = 3.5$ dB. However, a large gain exists in a wide range of B , showing that the SS improvement also appears away from the optimal setup. We observed a negligible difference between the SS estimated BER and the MC one at every point, consistently with the same relative accuracy of the simulation.

Depending on the random seed, in some cases stratum 2 was never visited, yielding a very fast answer about the BER. Nonetheless, the error on the BER was negligible, showing that the SS algorithm works efficiently. On the other hand, with some seeds we observed some errors in stratum 2 yielding a small ratio σ_1/σ_2 which slowed down the simulation time. However, we always observed a gain compared to standard MC.

In Fig. 4, we repeated the measurement by fixing the B2B gain to $B = 3$ dB and then varying the OSNR for $N_b = 16, 32$,

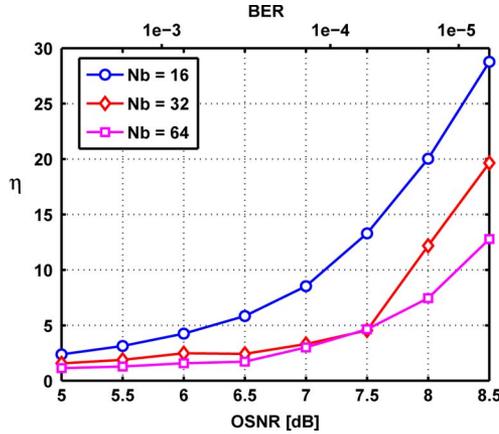


Fig. 4. SS gain η versus OSNR (dB @ 0.1 nm) using 16, 32, and 64 bits. $D_t = 750$ ps/nm, $B = 3$ dB. Only GVD.

and 64. The cumulated dispersion was $D_t = 750$ ps/nm, which gives an OSNR penalty of 0.5 dB @ BER = 10^{-5} . Note that the SS gain η is larger at higher OSNRs, i.e., smaller BERs. At very large BERs it is very likely to observe at least one error in the fast system, hence SS almost coincides with MC.

B. DQPSK in Nonlinear Regime

Having verified the SS algorithm in a simple case, we now move to a more practical one. We measured the BER of a five channel DQPSK system, shaped with both NRZ and RZ pulses [14]. Each channel had symbol rate $R_s = 20$ Gbaud ($R_b = 40$ Gb/s), and was spaced from the neighboring ones by 50 GHz. All channels were synchronous in time.

We simulated the true system using a fine SSFM with variable step size having a maximum nonlinear phase rotation per step of $\Delta\Phi = 3 \times 10^{-3}$ rad [15]. For the fast system we used a coarse SSFM with maximum phase rotation five times greater per step, which we separately verified to yield more than 1 dB difference in the Q -factor over a wide range of powers. The system under investigation was the periodic repetition of 100 km of $D = 8$ ps/nm/km fiber @ $\lambda = 1550$ nm. The fiber had attenuation 0.2 dB/km, effective area $63 \mu\text{m}^2$ and nonlinear coefficient $n_2 = 2.7 \times 10^{-20}$ m²/W. At the end of each period a linear dispersion compensating fiber gave the desired in-line residual dispersion while a noisy amplifier recovered the attenuation losses. Since the amplified spontaneous emission (ASE) noise has been introduced along the line, such a scheme allows to correctly reproduce the impact of nonlinear phase noise. Before transmission we inserted an ideal precompensating fiber of residual dispersion -290 ps/nm, while at the end of the link a postcompensating fiber brought the overall cumulated dispersion to zero. We used a standard DQPSK transmitter [16] with the Mach-Zehnders suitably modulated to have a de Bruijn quaternary signal¹ of 4^5 samples for each DQPSK channel. We used different de Bruijn seeds for different channels. The receiver was composed of an optical second-order super-Gaussian filter of bandwidth $1.5R_s$, a Mach-Zehnder delay interferometer, two ideal photodiodes and an electrical lowpass fifth-order Bessel filter, of bandwidth $0.65R_s$.

¹A De Bruijn quaternary sequence of 4^k symbols contains all subsequences of k symbols in alphabet 4 exactly once.

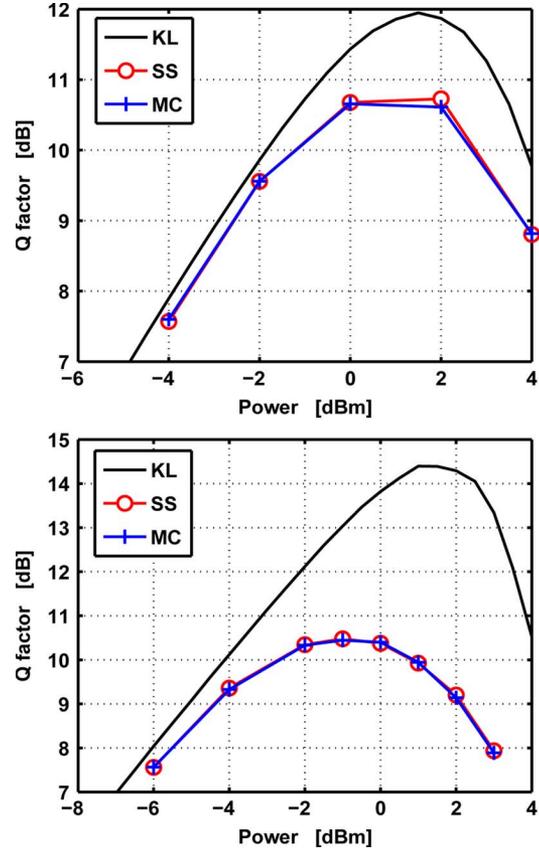


Fig. 5. Q -factor (dB) versus power (dBm) for a 10×100 km NRZ-DQPSK (top) and 20×100 km RZ-DQPSK (bottom). KL: Karhunen-Loève. MC: Monte Carlo. SS: Stratified sampling.

In order to test different configurations we analyzed a 20×100 km system with RZ pulses and amplifiers with noise figure $F = 4$ dB, and a 10×100 km system with NRZ pulses and $F = 9$ dB. In the first case we overlooked four-wave mixing (FWM), while in the second we accounted for all nonlinear effects in the SSFM. We used 25 ps/nm/span residual dispersion along the line for the NRZ case and 12.5 ps/nm/span for the RZ one. All BER measurements were taken on the central channel at 1550 nm, at a relative accuracy 0.1 with confidence level 68%, both with the MC and the SS algorithm. Fig. 5 (top) shows the Q -factor versus the channel power measured using MC (crosses), SS (circles) and the KL algorithm (no symbols) for the NRZ case. The KL algorithm used white noise statistics for the ASE noise. The Q factor is defined as

$$Q = 20 \log_{10} \left[\sqrt{2} \operatorname{erfc}^{-1} (2 \cdot \text{BER}) \right].$$

First, we note the inaccuracy of KL at large powers, which is due to the large nonlinear phase noise induced by the line. Second, MC and SS show comparable values, consistently with the same tolerance value. We repeated the same test for the RZ system obtaining again the same behavior, with a larger impairment due to phase noise, see Fig. 5 (bottom).

Having verified the consistency of the BER estimation, we moved to measure the algorithm efficiency. Since more than 98% of the simulation time was spent in the SSFM fiber propagation, we used as a fair cost-criterion the ratio between the number of FFTs used by standard MC and the ones by

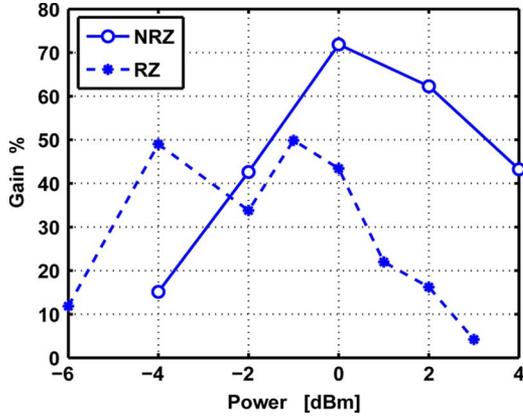


Fig. 6. Percentage reduction in number of FFT calls (gain) of the SS algorithm compared with MC at the same confidence level for the DQPSK system.

SS. Unlike the pure GVD case analyzed in Section IV.A, here we cannot declare costless for the runs within the fast system. Hence, with SS we have to sum the FFTs of the fast system to those of the true system. The FFT-reduction factor, i.e., gain, of SS w.r.t. MC, expressed in %, is shown in Fig. 6 versus power. Note that we averaged the gain over four different simulations with different seeds in order to exclude possible best/worst cases. We note that at best SS requires 70% less FFT than MC for NRZ, and 50% for RZ in absence of FWM. As a reference, with $P = 0$ dBm in the NRZ case the MC algorithm used 1.6×10^7 FFTs while SS, at the same accuracy, used 3.3×10^6 FFT in the fast system and 1.3×10^6 in the true system. It turns out that the tradeoff overhead/variance reduction of SS does give a significant computational time advantage.

Such an advantage depends on the reliability of the fast system. In Table I, we compare the performance of the SS algorithm by varying the nonlinear phase per step in the fast system. We analyzed the NRZ–DQPSK case with $P = 0$ dBm. The table collects the results in terms of the enhancement factor $G_\Phi \triangleq \Delta\Phi_F/\Delta\Phi_T$, where $\Delta\Phi_{F,T}$ is the nonlinear phase per step in the fast/true system, respectively. From the table we note that for decreasing fast system reliability, i.e., increasing G_Φ , the fast system calls for an increased number of runs n_F , according to the observation that the fast and the true system do not match very well and hence a larger number of trials is required for a correct estimation. However, a larger n_F does not necessarily mean a larger computational effort, since the number of FFTs in the fast system (FFT_F) decreases faster than the growth of n_F . At $G_\Phi = 20$ there is an inversion in the computational effort, i.e., the number of FFTs in the true system (FFT_T) exceeds FFT_F . From the table we also note that while n_1 , i.e., the number of runs in the stratum “errors in the fast system”, follows the behavior of n_F , n_2 is almost constant. This is an indicator of a high rejection ratio so that most of the samples in stratum 2 (no errors in the fast system) are collected during the initialization runs where SS waits for the first error in stratum 1.

We separately ran a very accurate MC simulation with the same tolerance but with confidence level of 99%. The BER was 3.23×10^{-4} . The BER estimated by the SS algorithm and shown in Table I is consistent with this value and with the confidence

TABLE I
PERFORMANCE OF SS ALGORITHM VERSUS. FAST SYSTEM ACCURACY G_Φ . NRZ–DQPSK CASE WITH $P = 0$ dBm. FFT_F , FFT_T : NUMBER OF FFTS IN THE FAST, TRUE SYSTEM, RESPECTIVELY. n_F , n_1 , n_2 NUMBER OF MC RUNS IN THE FAST SYSTEM, STRATUM 1 AND STRATUM 2, RESPECTIVELY

G_Φ	SS BER	FFT_F $\times 10^6$	FFT_T $\times 10^6$	n_F	n_1	n_2
2	$3.21 \cdot 10^{-4}$	8.2	1.22	308480	12288	10880
5	$3.14 \cdot 10^{-4}$	3.3	1.27	315136	13184	10880
7	$3.14 \cdot 10^{-4}$	2.4	1.29	315136	13568	10880
10	$2.88 \cdot 10^{-4}$	1.9	1.37	343552	14592	11392
20	$2.71 \cdot 10^{-4}$	1.0	1.67	366080	20864	10624

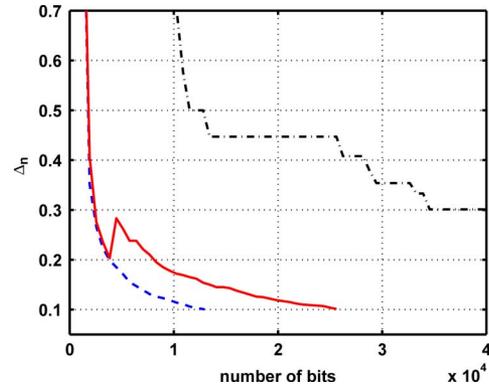


Fig. 7. Relative error Δ_n (4) at confidence 68% versus number of observed bits. NRZ–DQPSK case with $P = -2$ dBm, and $G_\Phi = 5$. Solid line: SS with $\tilde{m} = 50$. Dashed line: SS with $\tilde{m} = \infty$. Dashed-dotted line: MC.

used in the SS, i.e., we expect $\sim 68\%$ of measurements within $\Delta = 0.1$.

A final investigation regarded the impact of the bound \tilde{m} in the rejection ratio $m = \min(\sigma_1/\sigma_2, \tilde{m})$. As observed at the end of Section III, this forces SS to a nonzero visit in both strata in every case. Fig. 7 shows the behavior of Δ_n with $\varepsilon = 1$ (i.e., at confidence 68%) for a simulation with increasing number of bits and either $\tilde{m} = 50$ or $\tilde{m} = \infty$. The system was NRZ–DQPSK with $P = -2$ dBm. The solid line refers to SS with $\tilde{m} = 50$, the dashed line refers to SS with $\tilde{m} = \infty$ while the dashed-dotted one is for MC. Note that the MC curve decreases almost as $1/\sqrt{n}$ according to (2). In both SS cases, the first error in the true system occurred in stratum 1, hence after that occurrence m took the value \tilde{m} , according to our strategy detailed in Section III. Thus, after the first error in stratum 1, the SS algorithm never visited stratum 2 with $\tilde{m} = \infty$, while it seldom explored stratum 2 with $\tilde{m} = 50$. From the figure we note that the SS curve with $\tilde{m} = 50$ experiences an abrupt change around 4000 bits, which is not present in the $\tilde{m} = \infty$ case. Such a sharp raise corresponds to a visit in stratum 2 (i.e., at least one error over the block of bits). Such an error yields $\sigma_2 \neq 0$ which does decrease m to a value lower than 50. Consequently, the estimated uncertainty suddenly grows up and then starts to fall again toward small values as the number of runs increases. The $\tilde{m} = \infty$ case avoids such a behavior. The absence of visits in stratum 2 in the $\tilde{m} = \infty$ case does not imply a completely wrong BER estimation, since with a good choice of the fast system stratum 2 has a negligible impact. In this experiment we measured $\text{BER} = 1.41 \times 10^{-3}$ with $\tilde{m} = 50$ and $\text{BER} = 1.31 \times 10^{-3}$ with $\tilde{m} = \infty$. The BER with MC at confidence 99% was 1.45×10^{-3} . In conclusion, we suggest to use a finite value for \tilde{m} because (on average)

it avoids to get BER estimates without even visiting a stratum. The algorithm computational cost slightly increases, but still remains largely smaller than MC as clearly shown in Fig. 7.

V. CONCLUSIONS

We showed a fast and simple MC SS strategy for BER measurement. The idea is based on running an MC estimation after observing an approximate, but usually faster to simulate, description of the system under test, and then to use such information to drive the estimation on the true system. It was shown that our proposed implementation of the method is a new form of adaptive IS, which is similar in spirit to the “excision IS” method [8]. The algorithm showed a significant computational gain compared to standard MC at practical values of BER $\sim 10^{-3}$ to 10^{-5} . Moreover, the relative error of the proposed algorithm can be simply inferred from the estimated variance of the observable, like in an MC estimation, thus providing a statistical confidence on the results.

We verified the accuracy of the algorithm by comparing its predictions with the ones of standard MC at the same confidence level. We studied two possible implementations of the fast system, one in a simple case distorted by only GVD and a second in a practical WDM–DQPSK in the nonlinear regime.

The method can be applied to any system with random inputs. In such a case, SS needs to identify a fast system yielding an approximate description of the true system under investigation.

APPENDIX

This Appendix provides the basic connections among MC, SS, and IS.

Monte Carlo: Given the random variable $I(\mathbf{X})$, i.e., a scalar function of the random vector \mathbf{X} taking values in the multidimensional input space Ω , the MC method estimates the expected value $\mathbb{E}[I]$ of I by drawing n independent samples $I^{(1)} \equiv I(\mathbf{X}_1), \dots, I^{(n)} \equiv I(\mathbf{X}_n)$ from the known distribution $f_{\mathbf{X}}(\mathbf{x})$ and then computing the sample mean as in (1). The implicit assumption is that we are able to draw samples from $f_{\mathbf{X}}(\mathbf{x})$.

Stratified Sampling: The SS method instead assumes the input space Ω is partitioned into r disjoint subsets (strata) $\{\Gamma_1, \dots, \Gamma_r\}$ whose probabilities $p_s = \Pr\{\mathbf{X} \in \Gamma_s\}$, $s = 1, \dots, r$, are known. By the total probability law, the exact expectation can be written as $\mathbb{E}[I] = \sum_{s=1}^r p_s \mathbb{E}[I_s]$, where $\mathbb{E}[I_s]$ is shorthand notation for the conditional expectation $\mathbb{E}[I(\mathbf{X})|\mathbf{X} \in \Gamma_s]$. By similarity with the total probability law, the SS estimator of $\mathbb{E}[I]$ is

$$\bar{I}_{SS} \triangleq \sum_{s=1}^r p_s \underbrace{\left(\frac{1}{n_s} \sum_{k=1}^{n_s} I_s^{(k)} \right)}_{\bar{I}_s} \quad (8)$$

where \bar{I}_s is the MC estimator of $\mathbb{E}[I_s]$ in stratum s , based on drawing n_s independent samples $I_s^{(k)}$, $k = 1, \dots, n_s$ from the conditional distribution $f_{\mathbf{X}}(\mathbf{x}|\mathbf{x} \in \Gamma_s)$, with $\sum_{s=1}^r n_s = n$. Clearly SS implicitly assumes that drawing from such conditional distributions is feasible. Note that in SS the samples n_s in

each stratum are deterministic, known quantities. From (8) we get the variance of the SS estimator as

$$\bar{\sigma}_{SS}^2 \equiv \text{Var}[\bar{I}_{SS}] = \sum_{s=1}^r p_s^2 \frac{\sigma_s^2}{n_s} \quad (9)$$

where $\sigma_s^2 \equiv \text{Var}[I_s^{(k)}] = \text{Var}[I(\mathbf{X})|\mathbf{X} \in \Gamma_s]$. Equations (8), (9) are the explicit versions of (5), (6) in the text.

Now let $\alpha_s \triangleq n_s/n$ be the fraction of samples in stratum s , with the constraint $\sum_{s=1}^r \alpha_s = 1$. Thus $\alpha = \{\alpha_s\}_{s=1}^r$ is a probability mass function (PMF). Let us look for the optimal α by extremizing the Lagrangian $F = \bar{\sigma}_{SS}^2(\alpha) + \lambda(\sum_{s=1}^r \alpha_s)$, being λ the Lagrangian multiplier. Setting $\partial F/\partial \alpha_s = 0$ for all s yields the optimal $\alpha_s = p_s \sigma_s / \sqrt{n \lambda}$. Substitution into the constraint gives $\lambda = (\sum_{s=1}^r p_s \sigma_s)^2 / n$, hence finally the optimal fractions are

$$\alpha_s = \frac{p_s \sigma_s}{\sum_{j=1}^r p_j \sigma_j} \quad (10)$$

which yield (7) in the text.

Importance Sampling: Let us now keep the same strata partition $\{\Gamma_s\}_{s=1}^r$ as before, with known strata probability PMF $\mathbf{P} = \{p_s\}_{s=1}^r$. Let us warp the original input distribution $f_{\mathbf{X}}(\mathbf{x})$ to a new $f_{\mathbf{X}}^*(\mathbf{x})$, such that the warped strata probabilities are, for all s : $p_s^* \triangleq \int_{\Gamma_s} f_{\mathbf{X}}^*(\mathbf{x}) d\mathbf{x} = \alpha_s$. To obtain this warping, the warped input distribution must equal $f_{\mathbf{X}}^*(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) \alpha_s / p_s$ for all \mathbf{x} in stratum s . Hence the IS weight on stratum s is uniform and equal to

$$w_s \triangleq \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}}^*(\mathbf{x})} = \frac{p_s}{\alpha_s}.$$

Hence the IS estimator [8] of the expectation of RV I is $\bar{I}_{IS} \triangleq (\sum_{k=1}^n I^{(k)} w(\mathbf{X}(\mathbf{k}))) / n$. By collecting samples within each stratum we can rewrite it as

$$\bar{I}_{IS} = \frac{1}{n} \sum_{s=1}^r \left(\sum_{k=1}^{n_s} I_s^{(k)} w_s \right) = \sum_{s=1}^r p_s \left(\frac{1}{n \alpha_s} \sum_{k=1}^{n_s} I_s^{(k)} \right). \quad (11)$$

Now, in IS the number of warped samples n_s that fall in stratum s is a random variable, whose expectation is $\mathbb{E}[n_s] = n \alpha_s$, and by the law of large numbers $n_s \rightarrow \mathbb{E}[n_s]$ almost surely as the number of runs n goes to infinity. Hence, for large n the two estimators \bar{I}_{SS} in (8) and \bar{I}_{IS} in (11) do coincide.

The procedure that we adopted in the text, described in Fig. 2, does in fact implement the above mentioned IS version of SS. The warped $f_{\mathbf{X}}^*(\mathbf{x})$ is obtained from the original $f_{\mathbf{X}}(\mathbf{x})$ (whose samples feed the fast system) by the reject/accept mechanism that feeds samples \mathbf{X}_k to the true system.

REFERENCES

- [1] E. Forestieri, “Evaluating the error probability in lightwave systems with chromatic dispersion, arbitrary pulse shape and postdetection filtering,” *J. Lightw. Technol.*, vol. 18, no. 11, pp. 1493–1503, Nov. 2000.
- [2] K.-P. Ho, “Probability density of nonlinear phase noise,” *J. Opt. Soc. Am. B, Opt. Phys.*, vol. 20, no. 9, pp. 1875–1879, Sep. 2003.
- [3] P. Serena, A. Orlandini, and A. Bononi, “A parametric-gain approach to the analysis of single-channel DPSK/DQPSK systems with nonlinear phase noise,” *J. Lightw. Technol.*, vol. 24, no. 5, pp. 2026–2037, May 2006.

- [4] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Fortran 77: The Art of Scientific Computing*. New York, NY: Cambridge Univ. Press, 1992.
- [5] R. Holzlöhner and C. R. Menyuk, "Use of multicanonical Monte Carlo simulations to obtain accurate bit error rates in optical communications systems," *Opt. Lett.*, vol. 28, no. 20, pp. 1894–1896, Oct. 2003.
- [6] L. Gerardi, M. Secondini, and E. Forestieri, "Pattern perturbation method for multicanonical Monte Carlo simulations in optical communications," *IEEE Photon. Technol. Lett.*, vol. 19, no. 23, pp. 1934–1936, Dec. 2007.
- [7] G. S. Fishman, *Monte Carlo: Concepts, Algorithms and Applications*. New York: Springer-Verlag, 1999.
- [8] P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: A review of importance sampling techniques in communications systems," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 4, pp. 597–613, May 1997.
- [9] K. S. Shanmugam and P. Balaban, "A modified Monte-Carlo simulation technique for the evaluation of error rate in digital communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1916–1924, Nov. 1980.
- [10] D. Lu and K. Yao, "Improved importance sampling technique for efficient simulation of digital communication systems," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 1, pp. 67–75, Jan. 1988.
- [11] J.-C. Chen, D. Lu, J. S. Sadowsky, and K. Yao, "On importance sampling in digital communications-Part I: Fundamentals," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 3, pp. 289–299, Apr. 1993.
- [12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.
- [13] S. L. Smith and G. C. Orsak, "A modified importance sampling scheme for the estimation of detection system performance," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 1341–1346, Feb./Mar./Apr. 1995.
- [14] D. van den Borne, S. L. Jansen, G. D. Khoe, and H. de Waardt, "Line optimization in long-haul transmission systems with 42.8-Gbit/s RZ-DQPSK modulation," in *Proc. OFC, Paper OFD2*, pp. 5--10, Anaheim, CA, March.
- [15] O. V. Sinkin, R. Holzlöhner, J. Zweck, and C. R. Menyuk, "Optimization of the split-step Fourier method in modeling optical-fiber communications systems," *J. Lightw. Technol.*, vol. 21, no. 1, pp. 61–68, Jan. 2003.
- [16] A. H. Gnauck and P. J. Winzer, "Optical phase-shift-keyed transmission," *J. Lightw. Technol.*, vol. 23, no. 1, pp. 115–130, Jan. 2005.



Paolo Serena was born in Piacenza, Italy, in 1973. He received the Laurea degree in Electronic Engineering and the Ph.D. degree in information technology from the Università degli Studi di Parma, Parma, Italy, in 1999 and 2003, respectively.

Currently he is an Assistant Professor of Telecommunications at the Department of Information Engineering, University of Parma, where he teaches a course on advanced optical communications and a course on fundamentals of communication systems.

He is currently collaborating in research projects with Alcatel-Lucent, Bell Labs, France, on long haul WDM systems. He is the creator of Optilux, an open source platform for optical system simulations. His main research interests include nonlinear effects in optical fibers and advanced modulation formats.



Nicola Rossi was born in Cremona, Italy, in March 1980. He received the Laurea degree in telecommunication engineering from the University of Parma, Parma, Italy, in 2006. He is currently working toward a scholarship at the Department of Information Engineering, University of Parma. In 2008, he was a Visiting Student at Université Laval, Quebec, Canada. His current research interests include fast simulation techniques, multicanonical Monte Carlo algorithms and long haul transmission systems.



Marco Bertolini (S'08) was born in Reggio Emilia, Italy, in 1982. He received the Laurea degree (cum laude) and the Laurea magistrale degree (cum laude) from the University of Parma, Parma, Italy, in 2004 and 2006, respectively. He is currently working toward the Ph.D. degree in information technology at the Department of Information Engineering, University of Parma. In 2007, he was a Visiting Student at Université Laval, Quebec, Canada. His current research interests include optical coherent detection, advanced modulation formats, and long

haul transmission systems.



Alberto Bononi received the Laurea degree in electronics engineering (cum laude) from the University of Pisa, Italy, in 1988, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1992, and 1994, respectively.

Currently, he is an Associate Professor of Telecommunications at the School of Engineering, Università di Parma, Parma, Italy where he teaches courses in probability theory and stochastic processes, telecommunications networks, and optical communications.

From 1994 to 1996, he was an Assistant Professor in the Electrical and Computer Engineering Department at the State University of New York (SUNY), Buffalo, NY, where he teaches courses in electric circuits and optical networks. In 1997 and 1999, he was a Visiting Faculty at the Département de Génie Électrique, Université Laval, QC, Canada, doing research on fiber amplifiers. His current research interests include system design and performance analysis of high-speed all-optical networks, nonlinear fiber transmission for WDM systems, linear and nonlinear polarization mode dispersion, and transient gain dynamics in optical amplifiers.

Dr. Bononi is an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS.