Stratified-Sampling Estimation of PDL-Induced Outage Probability in Nonlinear Coherent Systems

Nicola Rossi, Paolo Serena, *Member, IEEE*, and Alberto Bononi, *Senior Member, IEEE*

Abstract—We present a novel adaptive stratified-sampling (SS) algorithm to efficiently assess the polarization-dependent loss (PDL) induced outage probability (OP) in presence of fiber non-linearity and polarization mode dispersion. We show that, at standard OP values, the SS algorithm more than quadruples the efficiency in simulation time with respect to the standard Monte Carlo method at equal accuracy. We applied the SS algorithm to polarization-division multiplexing quadrature phase-shift keying transmissions in both dispersion-managed and dispersion-unmanaged links, showing a non-negligible impact of PDL using typical ITU-T recommendations. Finally, we show that estimating the OP by neglecting the interaction of nonlinearity and PDL can be largely optimistic.

Index Terms—Outage probability, polarization-dependent loss (PDL), stratified sampling.

I. INTRODUCTION

THE design of coherent optical transmission links must account for the polarization scattering induced by the optical devices along the line. In this scenario, while polarization mode dispersion (PMD) manifests as a unitary transformation that can be efficiently compensated by digital signal processing at the receiver, polarization dependent loss (PDL) induces a non-recoverable degradation of the signal to noise ratio (SNR).

Early studies on PDL investigated such an SNR degradation in the linear regime, where semi-analytical modeling is available [1]–[4]. However, the system designer is interested in the extra margin that must be allocated to the PDL penalty in realistic conditions that include the Kerr nonlinearity. Recent works [5]–[7] pointed out that the PDL/Kerr-effect interplay may have a significant impact on such a margin in terrestrial coherent links. New investigations are thus mandatory.

Moreover, PDL and PMD are slow random phenomena, thus analyzing the average Q-factor degradation does not provide information on the outage probability (OP). When the target OP is $\leq 10^{-5}$, brute force OP Monte Carlo (MC) simulations become infeasible and advanced methods are mandatory to speed up the computation [8]. The problem is more serious in the nonlinear regime, where simulating the propagation in an optical link, especially with wavelength division multiplexing (WDM), is extremely time consuming.

A solution for an efficient numerical simulation of PDL-induced OP in the linear regime was proposed by Tao *et al.* by a semi-analytical algorithm [9]. Biondini *et al.* [10] used importance sampling (IS) to estimate the probability of rare events of first/second order uncompensated PMD, which can be used to derive system outages. In [11] we proposed the stratified sampling (SS) algorithm for PDL-induced OP both in the linear and nonlinear regimes.

In this work we improve on the SS algorithm in [11] and show for the first time that it is possible to efficiently assess the OP down to 10^{-6} in practical cases including PDL, PMD and Kerr effect. We tested the algorithm for a WDM system with polarization division multiplexed quadrature phase shift keying (PDM-QPSK) channels modulated at 128 Gbit/s propagating in an optical link with different scenarios. In particular we investigated a dispersion managed (DM) and a dispersion unmanaged (DU) link in presence/absence of PMD and/or Kerr effect.

The paper is organized as follows. In Section II we review the main procedure to estimate the OP and discuss some pitfalls about its numerical simulation; in Section III we introduce the SS idea; in Section IV we present the numerical setup adopted in the simulations; in Section V we show the average Q-factor performance of the DM and DU links; in Section VI we show results on OP estimations by the novel SS algorithm. Finally in Sections VII and VIII we provide our main comments and conclusions.

II. ESTIMATION OF OUTAGE EVENTS

The bit error rate (BER) is the average rate of bit errors in the detected data stream. Averaging is used to summarize the impact of the stochastic parameters affecting the received signal into a unique quality of transmission (QoT) parameter. For this reason, the measurement time of the simulation or experiment should be kept sufficiently long to get stable BER estimates. However, most of the times the stochastic variations of the disturbing parameters operate over widely different time scales. For instance, amplified spontaneous emission (ASE) noise has very fast dynamics of the order of [THz], while PDL/PMD vary with frequencies of the order of [kHz] [12]. Averaging BER over millisecond-long times is thus not sufficient for the system designer, since it cannot give information about the rate of out-of-service events [13]. A more interesting QoT parameter is given by the OP, defined as the probability that the BER, averaged over a period longer than the dynamics of the rapidly varying parameters but still shorter than the dynamics of the slowly varying parameters, is larger than a reference value BER_{FEC} set by the forward error correction (FEC) threshold.

From the previous discussion we understand that it is mandatory to estimate the OP by running two nested cycles, i.e., an outer cycle over slowly varying parameters (called *slow* random variables (RVs)), and an inner cycle over rapidly varying parameters (*fast* RVs). In the outer cycle we average with respect to (w.r.t.) PDL, PMD and laser state of polarization; in the inner cycle we average w.r.t. ASE noise and information symbols. We have thus the following definitions [2]:

$$\mathrm{BER}(X) = E[\mathcal{I}(\hat{b}_k \neq b_k)|X] \qquad (1)$$

$$\mathrm{OP} = E[\underbrace{\mathcal{I}(\mathrm{BER}(X) > \mathrm{BER}_{\mathrm{FEC}})}_{\mathcal{I}_O(X)}] \qquad (2)$$

where: X indicates the set of slow RVs, b_k is the transmitted bit, $\hat{b}_k$ is the detected bit; $E[.|X]$ indicates statistical conditional expectation over the fast RVs (estimated by time-averaging over the short time-scale); $\mathcal{I}$ is the indicator function (which equals 1 when the argument is true and 0 otherwise); $\mathcal{I}_O(X)$ the indicator that BER exceeds the FEC threshold; and $E[.]$ in (2) indicates statistical expectation over the slow RVs X (estimated by time-averaging over the long time-scale).

When a reliable semi-analytical BER model for the expectation in (1) w.r.t. the fast RVs is not available, the BER in (1) can be estimated by running a MC simulation of a data stream of n bits (possibly split in parallel MC simulations of shorter length) with a fixed configuration of slow RVs. It is worth noting that, for independent errors in the n-bit pattern, the MC measurement variance is $\sigma^2 = \mathrm{BER}(1 - \mathrm{BER})/n$. The standard deviation σ must be much smaller than the estimated BER value while spanning the values of X, otherwise the OP estimated value will be corrupted by the BER measurement noise [7]. Hence, the number of MC runs must be chosen in order to cap on σ^2, which is unknown at the beginning of the simulation. For this reason, a conservative choice with large accuracy for the BER estimation is recommended, especially for a link with small PDL and PMD leading to a small OP.

In this work, we always estimated the conditional BER (1) by using the MC algorithm. The key innovation of this paper pertains to an efficient algorithm to estimate the unconditional expectation (2), which is discussed in the next section.

III. SS FOR OP ESTIMATION

To introduce the ideas, let us consider for simplicity the case where the only slow RV is the PDL, and concentrate on the unconditional expectation (2). We plan to apply SS [8] to the estimation of such an expectation, to be performed in the outer cycle of the overall simulation. We will use a generalization of the SS algorithm proposed in [14], which can be seen as a form of IS, where the input sample space (i.e., the set of all possible slow RV realizations) is partitioned into disjoint subsets, or strata, and independent (local) MC simulations are performed within each stratum.

A block diagram describing the proposed adaptive SS-IS OP estimation is shown in Fig. 1. The key point is the availability of a "fast system" yielding (at negligible computational cost) a control variable C highly correlated with the "true
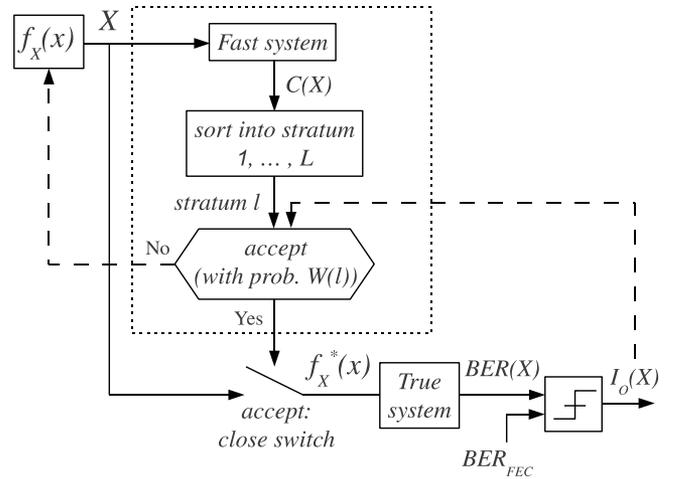

Fig. 1.   Block diagram of proposed adaptive SS-IS OP estimation. The dotted box is the SS control unit of the accept/reject input sieve that implements the warped PDL input distribution $f_X^*(X)$. If the stratum test is passed (with probability $W(l)$), the corresponding PDL realization X is accepted for propagation in the true system, yielding a BER sample and an outage indicator $\mathcal{I}_O$.

system" output (the BER in our case, at high computational cost). The variable $C(X)$ is next classified into strata $\mathcal{S}_C(l)$ (whose image in the X space is $\mathcal{S}_X(l)$) for $l = 1, \ldots, L$. The visiting probability of stratum l is the unwarped probability $P_l = P\{C \in \mathcal{S}_C(l)\} \equiv P\{X \in \mathcal{S}_X(l)\}$. The OP in stratum l is $\mathrm{OP}(l) = E[\mathcal{I}_O(X)|X \in \mathcal{S}_X(l)]$. If the standard deviation $\sigma_l = \sqrt{\mathrm{OP}(l)(1 - \mathrm{OP}(l))}$ of the Bernoulli indicator $\mathcal{I}_O$ in stratum l is known for the given stratification for all $l = 1, \ldots, L$, then the SS algorithm biases (warps) P_l according to the SS optimal rule [8], [14]:

$$P_l^* = \frac{\sigma_l P_l}{\sum_{i=1}^{L} \sigma_i P_i}, \quad l = 1, \ldots, L. \qquad (3)$$

This warping is achieved by the accept/reject block that operates by accepting an unwarped sample from stratum l with probability:

$$W(l) = \frac{\frac{P_l^*}{P_l}}{\sum_{i=1}^{L} \frac{P_i^*}{P_i}} = \frac{\sigma_l}{\sum_{i=1}^{L} \sigma_i}, \quad l = 1, \ldots, L. \qquad (4)$$

The second equality in (4) can be used to make the SS algorithm adaptive by updating the $\mathrm{OP}(l)$ estimates based on the feedback of the $\mathcal{I}_O$ observable, as shown in Fig. 1. To clarify the expression of $W(l)$, suppose to generate M samples. The number of samples in stratum l after the accept/reject block is $M_{\mathrm{w}}(l) \simeq M P_l W(l)$, hence, thanks to the rejections, the relative frequency of samples in stratum l at the output of the accept/reject block is:

$$\frac{M_{\mathrm{w}}(l)}{\sum_{i=1}^{L} M_{\mathrm{w}}(i)} \xrightarrow[M \to \infty]{} P_l^*, \quad l = 1, \ldots, L$$

as desired. $W(l)$ thus changes the input distribution of PDL expressed by $f_X(X)$ into a warped probability density function (PDF) $f_X^*(X)$ at the input of the true system, with $X = [(\vartheta_1, \varepsilon_1), \ldots, (\vartheta_N, \varepsilon_N)]^T$ where $(\vartheta_i, \varepsilon_i)$ are azimuth and
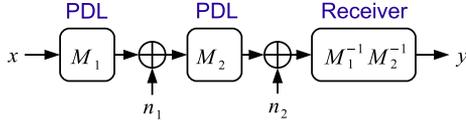
Fig. 2. Basic transmission corrupted by PDL and noise.

Fig. 3. Sketch of SS idea. We divide the (ρ^h, ρ^t) plane into L frames (strata) identified by multiples of the median value of the corresponding PDL. We then run an MC simulation in each stratum, by increasing the number of samples in frames showing poorer accuracy. Contour levels represent an example of the joint PDF of PDL (ρ^h, ρ^t), for $N = 30$ elements having PDL $\rho_i = 0.8$ dB, $i = 1, \ldots, 30$.

ellipticity of the ith PDL element. The efficiency of the warped PDF in speeding up the OP estimation depends on the ability of the fast system to capture the true system behavior.

The crucial point of SS is the choice of the partition of the input sample space. Optimal partitioning requires finding strata having observables with zero standard deviation within each stratum, so that ideally one sample is sufficient to estimate (2) on each stratum. To get close to this ideal target, we suggest to stratify the slow RVs sample space according to the way PDL builds up along the link, rooting our idea on the following observations. Assuming for instance the basic equalized system of Fig. 2, where $M_{1,2}$ represent the PDL matrices of the first and second half of the link while $n_{1,2}$ are additive independent noise sources, the output signal is:

$$y = x + M_1^{-1} n_1 + (M_2 M_1)^{-1} n_2 . \tag{5}$$

The total PDL ρ^t is a function of the eigenvalues of $M_1 M_2$ [15]. Such a PDL determines the contribution of n_2 to the variance of y, however it does not provide information about the contribution of n_1. Such information is given by the PDL of M_1 (referred to as ρ^h), i.e., by the eigenvalues of M_1, which give the PDL cumulated in the first half of the link. By considering $n_{1,2}$ as an additive distortion, either due to ASE or nonlinearity, we can summarize the impact of PDL on this basic two-section system by the pair (ρ^h, ρ^t). In a more general multi-span optical system, for the sake of simplicity we may think to virtually divide it in two halves, and proceed as before. Of course this idea is not fully comprehensive of the impact of PDL. For instance, the nonlinear effect contributed by the first half might be correlated with that in the second half. Nevertheless, stratifying the sample space in terms of the values taken by the pair (ρ^h, ρ^t) is expected to be a reasonable SS strategy. Hence, in our case the fast system was a two-trunk approximation of the true link, yielding a control variable $C = (\rho^h, \rho^t)$. In [11] we proposed a similar but less efficient SS algorithm, by stratifying the PDL on the basis of the eigenvalues of M_1^{-1} and M_2^{-1}, i.e., the first and the second half of the link. Such an approach is less efficient since large eigenvalues of M_1^{-1} and M_2^{-1} do not directly translate into large eigenvalues of $(M_2 M_1)^{-1}$.

We proceeded by subdividing both the ρ^h and ρ^t axes into regular grids, thus stratifying the sample space in frames as depicted by Fig. 3. In particular, we used the discretization $\eta_{h,t} = \rho^{h,t}/\rho_m^{h,t} = [1, 1.5, 2, 2.5, 2.75]$, where ρ_m^h and ρ_m^t are the median values of the PDL after propagation in the first half and in the total link, respectively. This choice leads to $L = 6$ different strata. As an example, in Fig. 3 we also reported the contour levels of the joint PDF $f_{HT}(\rho^h, \rho^t)$ of $N = 30$ elements, each having PDL [15] $\rho_i = 0.8$ dB. The probability P_l that a random PDL realization falls in stratum l is given by the integral
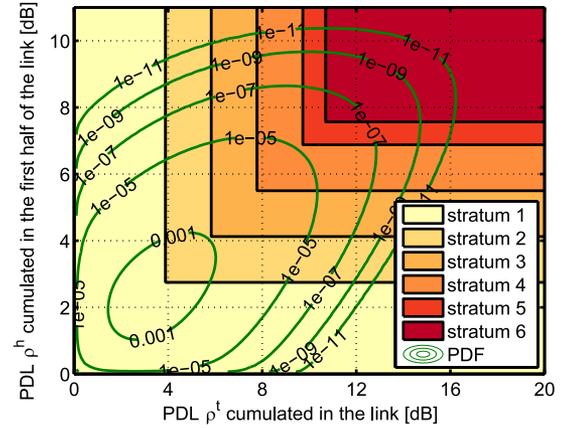
of $f_{HT}(\rho^h, \rho^t)$ over the corresponding frame (stratum). Since OP(l) is the outage probability of stratum l, the system outage probability is OP $= \sum_{l=1}^{L} P_l \cdot$ OP(l).

As each PDL realization is associated with a stratum of the sample space, a standard MC algorithm to estimate outage events would sample each stratum according to its probability of occurrence. This is strongly inefficient as we are interested in rare events of PDL. The proposed SS algorithm samples instead each stratum on the basis of the stratum uncertainty. The law establishing the optimal frequency of visits for the stratum l depends also on P_l [14], which is non trivial to compute due to the expression of $f_{HT}(\rho^h, \rho^t)$. Using the total probability law we have

$$P_l = \iint_{\mathcal{S}_C(l)} f_{HT}(h, t)\,dh\,dt = \iint_{\mathcal{S}_C(l)} f_T(t \mid H = h) f_H(h)\,dh\,dt \tag{6}$$

where $\mathcal{S}_C(l)$ is the frame identifying stratum l in the space $(h, t) = (\rho^h, \rho^t)$. It can be proved that for a finite number $N/2$ of PDL elements, each having PDL ρ_i, it is [16]:

$$f_H(h) = \frac{2h}{\pi} \int_0^\infty r \sin(hr) \operatorname{Sa}^{\frac{N}{2}}(\rho_i r)\,dr \tag{7}$$

with $\operatorname{Sa}(x) = \sin(x)/x$. For a chain of N elements where the first $N/2$ are fixed, we can apply the same reasoning leading to (7) and recall that the chain of $N/2$ fixed elements can be collected into one equivalent element, thus obtaining:

$$f_T(t \mid H = h) = \frac{2t}{\pi} \int_0^\infty r \sin(tr) \operatorname{Sa}(hr) \operatorname{Sa}^{\frac{N}{2}}(\rho_i r)\,dr . \tag{8}$$

A closed-form expression of (7) was first reported in [16]:

$$f_H(h) = \frac{h}{2\rho_i^2 \left(\frac{N}{2} - 2\right)!} \sum_{s=0}^{\lfloor \frac{N m}{2} \rfloor} (-1)^s \binom{N/2}{s} \left(\frac{N}{2} m - s\right)^{\frac{N}{2} - 2} \tag{9}$$

where $m = \frac{1}{2}(1 - \frac{2h}{N\rho_i})$. An expression of (8) can be found after expanding $\sin(tr)\sin(hr)$ by the prosthaphaeresis formula, obtaining:

$$f_{\mathrm{T}}(t \mid H = h) = \frac{t}{\pi h \rho_i}\left[G\left(\frac{t-h}{\rho_i}, \frac{N}{2}\right) - G\left(\frac{t+h}{\rho_i}, \frac{N}{2}\right)\right] \quad (10)$$

where

$$G(\delta, N) \triangleq \int_0^\infty \cos(\delta y)\left(\frac{\sin(y)}{y}\right)^N dy. \quad (11)$$

A closed form expression of (11) for $N > 1$ is obtained as [17]:

$$G(\delta, N) = \frac{\pi/2}{(N-1)!}\sum_{s=0}^{\lfloor\frac{N-1}{2}\rfloor}(-1)^s\binom{N}{s}\left\{\pm\left(\frac{N}{2} - s + \frac{\delta}{2}\right)^{N-1}\right.$$
$$\left.\pm\left(\frac{N}{2} - s - \frac{\delta}{2}\right)^{N-1} + \mu(-1)^{\frac{N}{2}}\binom{N}{\frac{N}{2}}\left|\frac{\delta}{2}\right|^{N-1}\right\} \quad (12)$$

where: the positive sign is to be taken in front of a bracket if the expression within the bracket is itself positive and the negative sign in the opposite case; $\mu = 1$ if N is even, 0 otherwise. Once the PDFs (9), (10) are available, P_l in (6) can be found by numerical two-dimensional quadrature integration.

Despite the analytical closed-form formulas, both expressions (9) and (12) are based on ratios and differences of very large numbers when $N \gg 1$ that are difficult to evaluate. In such cases, quadrature integration of (7) and (8), e.g., by the function quadgk of MATLAB, may be numerically more stable.[1] Alternatively, following the same steps of [16], one may resort to the asymptotic approximation for large N:

$$f_{\mathrm{T}}(t \mid H = h) \sim \frac{t}{h\sqrt{N}\rho_i}\sqrt{\frac{3}{\pi}}\left(e^{-\frac{3(h-t)^2}{N\rho_i^2}} - e^{-\frac{3(h+t)^2}{N\rho_i^2}}\right) \quad (13)$$

while (9) can be approximated by a Maxwellian distribution [16]. A check of the correctness of (9) and (10) for $N = 10$, 30 and $\rho_i = 0.5$ dB is reported in Fig. 4. We chose a random setup with $\rho^h = 0.54$ dB and $\rho^h = 0.76$ dB for the $N = 10$ and $N = 30$ cases, respectively.

A pseudo-code of the SS algorithm is reported in Fig. 5. Before starting the SS algorithm, we performed short (transient) MC true-system runs in each stratum to roughly estimate the standard deviation σ_l, $l = 1, \ldots, L$ (vector "sigma" in Fig. 5). A stratum is explored by generating PDL from the unwarped PDF and accepting only samples within that stratum. We started the transient from stratum 1 and stopped it at the first stratum in which we counted five outages (OUTthr = 5 in Fig. 5). If less than five outages were counted in stratum l after collecting $N_{\max}(l)$ samples, we moved to the next stratum $l + 1$. After the transient, we ran the true SS algorithm, and updated the

---

[1]In this work with N limited to 35 spans we always used (9) and (12). It is worth noting that already after $N = 42$ and $\rho_i = 0.5$ dB, (9) and (12) start to fail.
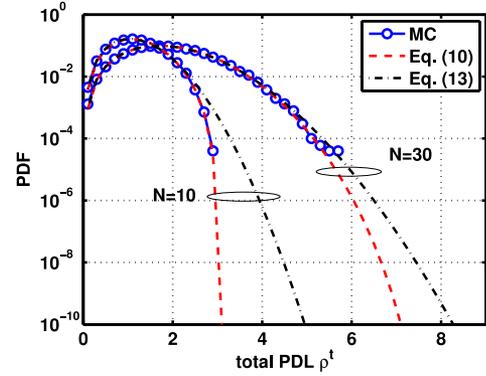

Fig. 4.   PDF $f_{\mathrm{T}}(t \mid H = h)$ of the total PDL conditioned to a random choice of the PDL cumulated in the first half of the link. Circles: MC simulations. Dashed lines: exact solution (10). Dash-dotted lines: asymptotic solution (13) for large N.

accept-reject probability $W(l)$, eq. (4), at the end of each BER estimation.

Since the estimated standard deviation of the OP is available, the SS simulation can be stopped with the same criterion as MC simulations, i.e., when the standard deviation of outage events is a specific fraction of the estimated OP.

IV. SIMULATIONS SETUP

We simulated with the open source software Optilux [18] both a DM and a DU link, each in presence or absence of PMD. In the DM case the optical link was composed of $N = 30$ spans, each of length 100 km, attenuation 0.2 dB/km, dispersion 4 ps/nm/km, nonlinear index $\gamma = 1.5$ 1/W/km and residual dispersion per span 30 ps/nm. Before transmission we added a pre-compensation of -495 ps/nm. The DU link was identical, except for a total of $N = 35$ spans with neither in-line compensation nor pre-compensation. The overall dispersion was set to zero by the receiver dispersion equalizer.

When present, PMD was emulated through 50 random waveplates per transmission fiber and a PMD coefficient of $0.13 \text{ ps}/\sqrt{\text{km}}$. PDL was emulated by lumped elements placed before the transmission fibers. The nonlinear signal-noise interaction along the link was accounted for by flat-gain noisy amplifiers with 6 dB noise figure. The WDM comb consisted of 15 channels with 32 Gbaud PDM-QPSK modulation, with 50 GHz spacing. In the DM case we used for each PDM-QPSK channel 1024 purely random symbols, while 2048 were used in the DU case [19], [20]. Fiber propagation used the split-step Fourier algorithm applied to the Manakov-PMD equation [21]. At the receiver, coherent detection of the central channel used a digital signal processor [22] including: analog to digital conversion with bandwidth 17 GHz, two samples-per-symbol data-aided least squares butterfly equalization with 7 taps, and Viterbi and Viterbi phase estimation with 15 taps.

The performance was measured in terms of the Q-factor obtained by inverting the BER. We counted at least 400 errors for each BER estimation. An outage event was declared when the Q-factor was below the FEC threshold of 6.25 dB, i.e., when

```
define strata;              % L strata (see, e.g., Fig. 3)
P = [P₁, P₂, ..., P_L];     % strata prob. from (6)
N = zeros(1,L);             % strata number of visits: vector
OUTcount = zeros(1,L);      % number of outage events
l = 1;                      % stratum index
while max(OUTcount) < OUTthr        % run transient.
   generate unwarped PDL in stratum l;  % (reject if not in l)
   Q-factor = Q_estimate(PDL); % inner cycle over fast RVs
   N(l) = N(l)+1;           % update visit counter of stratum l
   if Q-factor < FECthr,
      OUTcount(l) = OUTcount(l)+1;
   end
   if N(l) == Nmax(l),
      l = l+1; % move to next stratum
   end
end        % end transient.
OPstrata = OUTcount./N;       % vector of conditional OPs
OPest = sum(P.*OPstrata);               % System OP (scalar)
sigma = sqrt(OPstrata.*(1-OPstrata));   % std of strata OP
OPdev = sqrt(sum(sigma.*P)^2/sum(N)); % std(OPest) [8, p. 300]
ACC_OP = OPdev/OPest;         % estimation accuracy
while ACC_OP >= targetACC % outer cycle (slow RVs)
   P* = sigma.*P/sum(sigma.*P);       % update eq. (3)
   generate warped PDL % (accept-reject in Fig. 1)
   Q-factor = Q_estimate(PDL);
   update N(l),sigma(l),OPest,OPdev,ACC_OP; % like above
end

function Q-factor = Q_estimate(PDL) % true-system BER (1)
   ACC_Q = Inf;                 % sample-estimated Q accuracy
   while ACC_Q >= targetACCerg % inner cycle (fast RVs)
      generate fast RVs;        % ASE and symbols
      run SSFA;
      estimate Q-factor;            % pure MC
      update ACC_Q
   end
return
```

Fig. 5.   MATLAB-style pseudo-code of adaptive SS algorithm.

Fig. 6.   Average Q-factor of a 15-channel 32 Gbaud PDM-QPSK system, in presence/absence of PMD = 0.13 ps/$\sqrt{\text{km}}$ and rms PDL 2.74 dB. (Top) 30×100 km DM link. (Bottom) 35×100 km DU link.

the BER was above 2×10^{-2}. The OP was estimated after observing at least 25 outage events.

V. AVERAGE Q-FACTOR

Fig. 6 shows the average Q-factor of the central of 15 WDM channels versus signal power. The average Q-factor was obtained with the standard MC algorithm, by averaging over 10 seeds of slow RVs. In the DM scenario, Fig. 6 (top), a root-mean-square (rms) PDL of 2.74 dB reduces the best Q-factor by about 0.3 dB in both absence and presence of a typical PMD of 0.13 ps/$\sqrt{\text{km}}$. It is worth noting that, without PMD, the detrimental effect of PDL on the average Q-factor vanishes at large powers, as a consequence of the cross polarization modulation reduction induced by PDL [7]. When PMD is present, the best Q-factor improves by 0.4 dB because of the polarization decorrelation induced by PMD. However, now the beneficial effect of PDL at large powers disappears, thus making the PDL penalty on the average Q slightly increase with the input power.

Fig. 6 (bottom) shows the DU case. Now, in absence of PDL, PMD does not affect the average performance because of the large amount of dispersion cumulated along the link. By turning
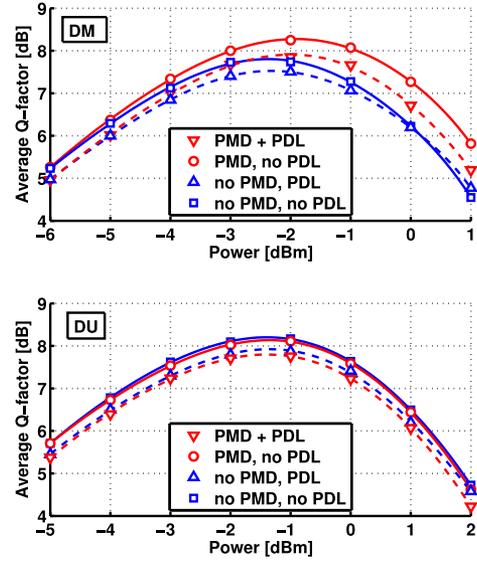
PDL on, we observe that the PDL penalty on the Q-factor is confined within 0.4 dB both in presence and absence of PMD.

Hence by just looking at the average Q-factor we would erroneously conclude that PDL/PMD have a marginal effect on system performance. We must instead look at the OP, which will be estimated in the next section at a power 1 dB beyond that at the best average Q-factor.

VI. OUTAGE PROBABILITY

In order to test the proposed SS algorithm and get a feeling of its computational gain with respect to standard MC, we first applied it to a linear system with $\rho_{\text{rms}}^t = 3.3$ dB, no PMD, at an optical signal to noise ratio (OSNR) of 12.8 dB over a bandwidth of 0.1 nm. We performed 20 parallel MC simulations of 8000 Q-factor samples (runs) each. The MC estimation of the OP, averaged over all the 160 000 Q-samples, was 2.3×10^{-3} and was declared the *true* value. Moreover, while the 20 simulations were in progress, we also measured the running OP estimate, and calculated the OP mean-square error (MSE) w.r.t. the true OP after each run. We then ran 20 SS simulations and again computed the running OP MSE. Please note that with SS we also accounted for the initial transient in the number of runs.

Fig. 7 shows the OP root-MSE (RMSE) over true OP versus the number of runs. We observe that the SS algorithm slashes the simulation time by more than a factor 4 (it was by a factor 2 in [11]) compared to the MC algorithm at an RMSE/OP value of 4×10^{-1}, i.e., at MSE 10^{-6}, and the gain in efficiency increases at lower RMSE.

Having verified the SS efficiency in a simple case where the *true* OP was computable with great accuracy, we next applied the SS algorithm to a realistic nonlinear DM transmission. Fig. 8 shows with solid lines the SS-estimated OP versus PDL, both without and with PMD (rms differential group delay (DGD)
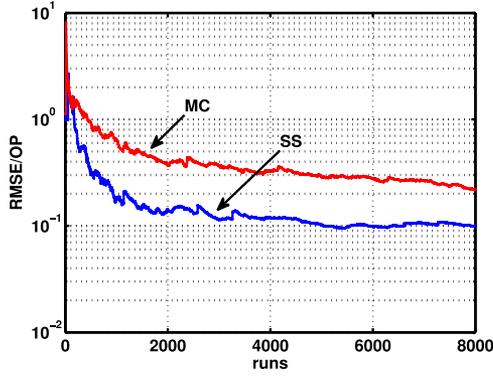
Fig. 7. OP rms error normalized to true OP versus number of runs for MC and SS algorithms, at OP = 2.3×10^{-3}, with $\rho_{\mathrm{rms}}^t = 3.3$ dB and OSNR = 12.8 dB/0.1 nm. Linear propagation without PMD.
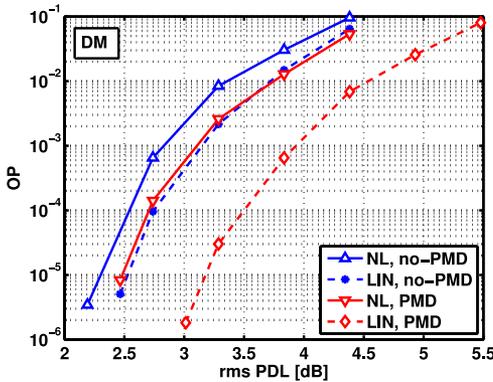

Fig. 8. OP versus PDL rms of a 15-channel 32 Gbaud PDM-QPSK system in a 30×100 km DM link with dispersion 4 ps/nm/km, in presence/absence of PMD = 0.13 ps/$\sqrt{\mathrm{km}}$. $P = -1.5$ dBm and OSNR = 15.7 dB/0.1 nm.

of 7.12 ps) and at transmitted power $P = -1.5$ dBm. For all SS simulations we used the values $\eta_{h,t} = [1, 1.5, 2, 2.5, 2.75]$, which define $L = 6$ strata, with the heuristic choice $N_{\max}(l) = l \times 100$ for $l = 1, \ldots, L$ as a compromise between computational speed and accuracy. It is worth noting that, in absence of PMD, the tolerable PDL to achieve an OP below 10^{-3} is $\rho_{\mathrm{rms}}^t = 2.8$ dB, i.e., the PDL generated by 30 lumped elements with typical 0.5 dB of PDL each [23]. When the target is instead an OP below 10^{-5}, the tolerable rms PDL becomes 2.3 dB. The presence of 0.13 ps/$\sqrt{\mathrm{km}}$ of PMD significantly lowers the OP in absolute terms, but improves by only 0.2 dB the PDL tolerance at fixed OP.

We also tried to see if the true OP values could be estimated by an equivalent linear system [2]. Hence we turned off fiber nonlinearity and set the amplifiers noise figure so as to have the same average Q-factor as the nonlinear system in absence of PDL, i.e., $Q_{\mathrm{avg}} = 7.6$ dB without PMD and $Q_{\mathrm{avg}} = 8.2$ dB with PMD, as shown in Fig. 6 (top). By this way, if PDL did not interact with Kerr nonlinearity, we would observe the same Q-factor distribution, hence resulting in the same OP as the nonlinear system. However, as shown by the dashed curves in Fig. 8, this does not happen, thus implying a significant coupling of PDL and fiber nonlinearity [7] that rules out the use of an
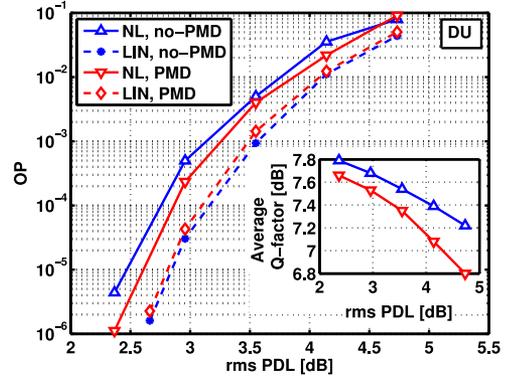

Fig. 9. OP versus rms PDL of a 15-channel 32Gbaud PDM-QPSK system in a 35×100 km DU link with dispersion 4 ps/nm/km, in presence/absence of PMD = 0.13 ps/$\sqrt{\mathrm{km}}$. $P = -0.5$ dBm and OSNR = 16.1 dB/0.1 nm. Inset: Average Q-factor versus rms PDL.

equivalent linear system to guess the true OP in presence of nonlinearity.[2]

We finally moved to estimate the OP in the DU scenario. Fig. 9 shows with solid lines the SS-estimated OP versus PDL, at an rms DGD both zero and 7.69 ps, and power $P = -0.5$ dBm. It can be noted that, in absence of PMD, the tolerable rms PDL to achieve an OP below 10^{-5} is $\rho_{\mathrm{rms}}^t = 2.45$ dB, and the presence of PMD along the link increases this value by 0.1 dB. Dashed lines in Fig. 9 depict the OP predicted by the linear equivalent system, which significantly underestimates the true OP.

We can see that the presence of PMD along the link lowers the OP notwithstanding a lower average Q-factor, as shown in the inset of Fig. 9. This fact suggests that the variance of the Q-factor is reduced in presence of PMD. Please note that all OPs below 10^{-3} were SS-estimated with a number of Q-samples varying between 2500 and 10 000, after observing a number of outage events varying between 25 and 200. It is worth noting that, when using standard MC, observing 25 outage events with 10 000 Q-samples corresponds, on average, to an estimated value OP = $25/10\,000 = 2.5 \times 10^{-3}$.

VII. DISCUSSION

The proposed SS strategy does not control directly the PDL of each span, but just their cumulative behavior by stratifying the sample space of ρ^t and ρ^h in a few strata, whatever the link length. This way, it is possible to run efficient simulations even when many PDL elements are present along the link, thus by-passing the typical dimensionality problem of IS [10], [24], [25]. A comparison of the proposed algorithm with other IS-based techniques is out of the scope of this paper. However, we remark the intrinsic simplicity of the code, the ability to work even in the nonlinear regime, and its efficiency at practical levels of OP.

It is worth noting that the efficiency of SS is maximum when the set of outage events looks like a rectangle in the (ρ^h, ρ^t)

---

[2] Please note that Fig. 5 in [11] differs from Fig. 8 because the solid red curve in [11] had PMD erroneously emulated with one waveplate per span. This led to an erroneous conjecture in [11].
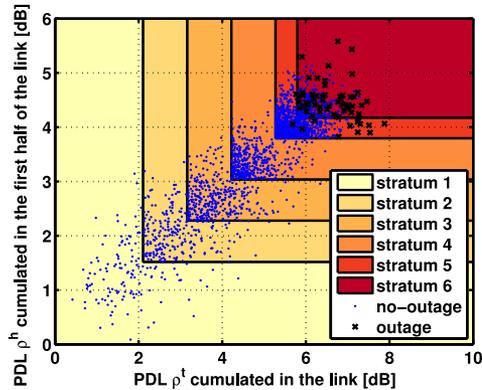
Fig. 10. Distribution in the (ρ^h, ρ^t) plane of SS-driven 2500 PDL realizations, from which we computed the OP in Fig. 9 for the nonlinear DU system without PMD at $\rho^t_{\mathrm{rms}} = 2.4$ dB. Cross: Outage. Dot: No-outage.

plane. In this case, the SS algorithm is essentially a stochastic search of the boundary of such a rectangle. If additional information about the shape of the set of outage events is available, the stratification can be changed from frames to other shapes better fitting the expected contour of outage events.

For the nonlinear DU system without PMD used to compute the OP in Fig. 9 at $\rho^t_{\mathrm{rms}} = 2.4$ dB, Fig. 10 shows the distribution in the (ρ^h, ρ^t) plane of SS-driven PDL realizations. Crosses/dots show PDL realizations inducing/not-inducing outage. We see that the set of PDL realizations leading to outage overlaps pretty well with only two strata. However, the OP per stratum is not always 1 or 0, as required by the optimal zero-variance IS estimator. Insight about this behavior can be obtained by looking at eq. (5). Intuitively, outage events occur for "large" PDL realizations, identified by the eigenvalues of M_1^{-1} and $(M_2 M_1)^{-1}$ and captured by the proposed (ρ^h, ρ^t) stratification. However, an uncertainty still remains because of the eigenvectors' random orientation, whence the gap between the proposed algorithm and the optimal IS.

VIII. CONCLUSION

We presented a SS algorithm to efficiently asses the PDL-induced OP in presence of fiber nonlinearity and PMD along the link. We showed the effectiveness of the proposed algorithm compared to standard MC, with a factor 4 savings in computation time. We applied the proposed algorithm to both DM and DU links, in presence and absence of PMD, by estimating the OP down to 10^{-5}. Our results show that speeding up simulations by considering a linear equivalent system with the same margin over FEC threshold as the true system leads to a significant underestimation of the actual OP.

REFERENCES

[1] B. Huttner, C. Geiser, and N. Gisin, "Polarization-induced distortions in optical fiber networks with polarization-mode dispersion and polarization-dependent losses," *IEEE J. Sel. Topics Quantum Electron.*, vol. 6, no. 2, pp. 317–329, Mar./Apr. 2000.

[2] C. Xie, "Polarization-Dependent loss induced penalties in PDM-QPSK coherent optical communication systems," presented at the Optical Fiber Communication Conf., San Diego, CA, 2010, Paper OWE6.

[3] C. Xie, "Impact of nonlinear and polarization effects in coherent systems," *Opt. Exp.*, vol. 19, no. 26, pp. B915–B930, Dec. 2011.

[4] M. Shtaif, "Performance degradation in coherent polarization multiplexed systems as a result of polarization dependent loss," *Opt. Exp.*, vol. 16, no. 18, pp. 13918–13932, Sep. 2008.

[5] O. Vassilieva, I. Kim, and M. Sekiya, "Statistical analysis of the interplay between nonlinear and PDL effects in coherent polarization multiplexed systems," presented at the European Conf. Optical Communication, Amsterdam, The Netherlands, 2012, Paper We.3.C.4.

[6] O. Vassilieva, S. Oda, T. Hoshida, J. Rasmussen, and M. Sekiya, "Experimental investigation of the statistics of the interplay between nonlinear and PDL effects in coherent polarization multiplexed systems," presented at the Optical Fiber Communication Conf., Anaheim, CA, USA, 2013, Paper OM3B.6.

[7] N. Rossi, P. Serena, and A. Bononi, "Polarization dependent loss impact on coherent optical systems in presence of fiber nonlinearity," *IEEE Photon. Technol. Lett.*, vol. 26, no. 4, pp. 334–337, Jan. 2014.

[8] G. S. Fishman, *Monte Carlo: Concepts, Algorithms and Applications*. New York, NY, USA: Springer-Verlag, 1999.

[9] Z. Tao, L. Dou, T. Hoshida, and J. Rasmussen, "A fast method to simulate the PDL impact on dual-polarization coherent systems," *IEEE Photon. Technol. Lett.*, vol. 21, no. 24, pp. 1882–1884, Dec. 2009.

[10] G. Biondini, W. L. Kath, and C. R. Menyuk, "Importance sampling for polarization-mode dispersion: Techniques and applications," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1201–1215, Apr. 2004.

[11] N. Rossi, P. Serena, and A. Bononi, "Stratified-sampling estimation of outage probability in nonlinear coherent systems with polarization dependent loss," presented at the European Conf. Optical Communication, London, U.K., 2013, Paper Th.1.D.6.

[12] M. Karlsson, J. Brentel, and P. A. Andrekson, "Long-Term measurement of PMD and polarization drift in installed fibers," *J. Lightw. Technol.*, vol. 18, no. 7, pp. 941–951, Jul. 2000.

[13] P. J. Winzer, M. Pfenningbauer, and R.-J. Essiambre, "Coherent crosstalk in ultradense WDM systems," *J. Lightw. Technol.*, vol. 23, no. 4, pp. 1734–1744, Apr. 2005.

[14] P. Serena, N. Rossi, M. Bertolini, and A. Bononi, "Stratified sampling monte carlo algorithm for efficient BER estimation in long-haul optical transmission systems," *J. Lightw. Technol.*, vol. 27, no. 13, pp. 2404–2411, Jul. 2009.

[15] J. N. Damask, *Polarization Optics in Telecommunications*. New York, NY, USA: Springer, 2004.

[16] M. Karlsson, "Probability density functions of the differential group delay in optical fiber communication systems," *J. Lightw. Technol.*, vol. 19, no. 3, pp. 324–331, Mar. 2001.

[17] J. O. Irwin, "On the frequency distribution of the mean of samples from a population having any law of frequency with finite moments, with special reference to Pearson's Type II," *Biometrika*, vol. 19, nos. 3/4, pp. 225–239, Dec. 1927.

[18] P. Serena, M. Bertolini, and A. Vannucci. (2009). Optilux Toolbox. [Online]. Available: http://www.optilux.sourceforge.net.

[19] J.-C. Antona, E. Grellier, A. Bononi, S. Petitreaud, and S. Bigo, "Revisiting binary sequence length requirements for the accurate emulation of highly dispersive transmission systems," presented at the European Conf. Optical Communication, Brussels, Belgium, Sep. 2008, Paper We.1.E.3.

[20] E. Grellier, J.-C. Antona, A. Bononi, and S. Bigo, "Revisiting binary sequence length requirements to accurately emulate optical transmission systems in highly dispersive regime," *Proc. SPIE Opt. Transmiss., Switching, Subsyst. VI*, Hangzhou, China, Oct. 2008, vol. 7136, pp. 713613-1–713613-10, 2008.

[21] D. Marcuse, C. R. Menyuk, and P. K. A. Wai, "Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence," *J. Lightw. Technol.*, vol. 15, no. 9, pp. 1735–1746, Sep. 1997.

[22] S. J. Savory, "Digital coherent optical receivers: Algorithms and subsystems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1164–1179, Sep./Oct. 2010.

[23] *ITU-T Rec. G.680: Physical Transfer Functions of Optical Networks Elements*, Int. Telecommun. Union, Geneva, Switzerland, Jul. 2007.

[24] R. Y. Rubinstein, *Simulation and the Monte Carlo Method*. New York, NY, USA: Wiley, 1981

[25] J. Chen, D. Lu, J. Sadovsky, and K. Yao, "On importance sampling in digital communications. I. Fundamentals," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 3, pp. 289–299, Apr. 1993.

Authors' biographies not available at the time of publication.