Amplitude sensitivity limits of optical sampling for optical performance monitoring

Mihaela Dinu and Francesco Quochi

Bell Laboratories, Lucent Technologies
101 Crawfords Corner Road, Holmdel, New Jersey 07733
mdinu@lucent.com

Received 4 April 2002; revised manuscript received 2 July 2002

We derive limits on the sensitivity of optical sampling methods for optical performance monitoring, accounting both for the fundamental detection limit and for realistic detector and thermal noise sources during detection. Nonlinear and linear (homodyne) sampling are compared with optoelectronic monitoring techniques. © 2002 Optical Society of America

OCIS codes: 060.2330, 190.1900.

1. Introduction

The increasing complexity in wavelength-division-multiplexed (WDM) networks and the expected migration toward transparency makes the availability of reliable and flexible monitoring techniques a significant issue. The ideal monitoring technique would provide information on the signal quality of optical channels present in the fiber while simultaneously determining the emergence of impairments, including intermittent or nonrecurring errors. Because of the necessity of fast data recovery, monitoring techniques should ideally diagnose all WDM channels in a short time interval (of the order of 100 ms) while tapping off only a minimal amount of power, 1–3% or less, from each channel.

Depending on whether operation is performed on the data stream before or after clock and data recovery (CDR), monitoring techniques can be classified as analog or digital, respectively. Current diagnostic methods that operate on transmitted digital data monitor parity bytes included in the SONET/SDH headers (synchronous optical network and synchronous digital hierarchy, respectively) and infer the bit-error rate (BER) from parity byte errors. This method is direct and (relatively) accurate, since it operates on transmitted data, but it is protocol dependent. Extensions to other protocols (Gigabit Ethernet or Fibre Channel) employ similar conventions. In the optical or physical layer domain, analog monitoring is currently performed at each amplification site by measurement of the average power in each WDM channel by use of a scanning or array spectrometer. The noise floor is estimated to be equal to the optical power at wavelengths between the optical channels, thus allowing the optical signal-to-noise ratio (OSNR) to be determined.

The ultimate benchmark for performance in a network is the BER. Whereas digital monitoring methods allow the BER to be estimated, analog methods permit early identification of failure mechanisms through the measurement of amplified-spontaneous-emission (ASE) noise, channel power equalization, OSNR, and so on. Spectral techniques give only a time-averaged picture of the status of the channels and are not sensitive to intermittent impairments, or to impairments to the data stream only, such as jitter, loss of modulation in return-to-zero (RZ) data, or dispersion. On the other hand, the SONET header technique is evidently bit-rate and data-format dependent, and can be implemented only at O–E–O (optical–electronic–optical) regeneration sites. If migration to all-optical networks is to occur, alternative real-time BER estimation techniques that do not rely on the presence of a receiver (and therefore on data termination) have to be identified. To provide sensitivity to intermittent impairments, such a diagnostic technique needs to operate in real time, without
averaging of multiple bits. If the network is completely transparent, or the data format is not known, the necessary monitoring and diagnostic techniques would have to be format independent and bit-rate independent. In recent publications, optical sampling using a nonlinear conversion process and an ultrafast pulse as the temporal gate has been suggested as being such a technique, usable up to data rates as high as 300 Gbit/s. On the other hand, sampling using a reference receiver has been shown to exhibit significant sensitivity limitations. Therefore it is of interest to analyze the sensitivity limits of optical sampling in the context of optical performance monitoring, i.e., under the constraints of signal powers and data rates encountered in high-capacity WDM networks.

2. Optical Sampling

In essence, nonlinear optical sampling consists of the generation of an optical pulse by means of a nonlinear process whenever the signal of interest \( P_D(t) \) and a short gating pulse \( P_G(t) \) are both present on a nonlinear medium, as shown in Fig. 1(a). In the context of ultrafast pulse characterization, time-averaged optical sampling gives the cross correlation of the two pulses involved, and in the field of ultrafast dynamics in materials characterization, nonlinear optical sampling is known as (luminescence) upconversion.

Consider a RZ data stream with average power \( P_{D,ave} \) at the repetition rate \( f_D \) consisting of a random string of bits (ones and zeros) with the “1” pulses having a duration (FWHM) \( t_p \). In the case of 10-Gbit/s transmission, typical values are \( f_D = 10 \text{ GHz} \) and \( t_p = 33 \text{ ps} \). The gating pulse train consists of short (FWHM = \( t_G \)) pulses at the much smaller repetition rate \( f_G \). Under these conditions, the resulting samples \( P_s(t) \) are simply proportional to the power of the data pulse in the temporal window provided by the gate \( P_G(t) \):

\[
P_s(t) = \eta^{-1} \int P_D(t') P_G(t'-t) \, dt' \approx \eta \, P_{G,0} P_D(t' = t),
\]

where \( \eta \) is the nonlinear conversion efficiency (in inverse watts) of the nonlinearity, assumed to be a second-order process; \( P_{G,0} \) is the peak power of the gating pulse, and \( t_G \) is its duration. The samples are subsequently detected using relatively slow detectors and electronics, limited only by the repetition rate of the gating pulses \( f_G \). The conversion efficiency from the data pulses to the gated samples is \( \eta' = \eta P_{G,0} \); conversion efficiencies \( \eta' \) as high as \( 10^{-1} \) or larger have been demonstrated in periodically poled lithium niobate (PPLN) and highly nonlinear fibers by use of subpicosecond gating pulses with low average power levels.

In linear (homodyne) sampling, the signal and the gate interfere directly on a photodetector, as shown in Fig. 1(b). This technique was first proposed and demonstrated experimentally for optical performance monitoring of telecom signals by Dorrer et al. The gating pulses, at the same optical frequency as the data, can be considered a local oscillator. The photocurrent then consists of a photocurrent background related to the average signal and gate powers, plus an interference term,

\[
i(t) = \Re \left[ \int \{ P_D(t') + P_G(t'-t) \pm 2 \Re [ E_D(t') E_G^*(t'-t)] \} \, dt' \right],
\]

depending on the optical phase between the data and the gate optical fields; here \( \Re \) (in A/W) is the detector responsivity and \( E_D, E_G \) are the electric fields. The homodyne term can be measured by balanced homodyne detection, and the resulting photocurrent is

\[
i_{\text{bal}}(t) = \Re \int 2 \Re [ E_D(t') E_G^*(t'-t)] \, dt'.
\]

Thus linear (homodyne) sampling offers the same temporal resolution as nonlinear optical sampling with the added advantage of potentially high amplification (given by the magnitude of the local oscillator electric field \( E_G \)).
One benefit of using optical sampling to characterize a data stream stems from the virtually unlimited temporal resolution, determined only by the duration of the gate pulse. Optical sampling is thus independent of the data rate, of the specific transmission protocol, or of the data format [RZ or non-RZ (NRZ)] employed. The samples taken at the repetition rate of the gate can be used to build up the eye diagram, and the error probability can be estimated from the histogram of the ones and zeros by use of the usual $Q$-factor

$$Q = \frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0},$$  \hspace{1cm} (4)$$

where $\mu_0$ and $\mu_1$ are the mean levels for the 0 and 1 bits, respectively, and $\sigma_0$ and $\sigma_1$ are the standard deviations in the 0 and 1 bits. If the optical sampling setup replicates the bandwidth and sensitivity characteristics of the end receiver, then the relationship between the $Q$-factor and the BER is unambiguous. In this case, under the assumption of Gaussian noise and optimum decision level,$^{10}$ the BER is related to the $Q$-factor by the equation

$$BER = \frac{1}{2} \text{erfc} \left( \frac{Q}{\sqrt{2}} \right).$$ \hspace{1cm} (5)$$

For a more general sampling case, the sampling $Q$-factor will be related to the equivalent receiver $Q$-factor $Q_{\text{rec}}$ through a calibration curve.

An additional advantage of optical sampling is the possibility of evaluating the $Q$-factor even in the absence of clock recovery, by use of asynchronous histograms.$^{2,11}$ Because of its time resolution, its capability to detect intermittent impairments, and its scalability to high data rates, all-optical sampling has been suggested as an attractive solution for optical performance monitoring. However, no detailed study of the sensitivity limits of optical sampling has yet been attempted. The following sections evaluate the factors that influence the sensitivity of optical sampling under realistic network conditions.
3. Sensitivity of Sampling

The sensitivity of a detection scheme is usually defined as the minimum power necessary to achieve a given signal-to-noise ratio (SNR) or a given BER (typically $10^{-9}$). In the particular case of sampling, the required SNR will be a function of the type of impairment to be detected, but we can safely assume that a minimum requirement will be in the range of 20–30 dB. The following discussion will focus exclusively on 10 GHz RZ data as an example, for realistic comparisons with available optoelectronic solutions; the conclusions can be generalized to the NRZ format and scaled to higher data rates quite easily.

As usual, the SNR is defined simply as the ratio of the peak current of the photodetector to the total variance of the noise,

$$\text{SNR} = \frac{I_{pk}^2}{\sigma_I^2}, \quad (6)$$

where here we have defined the electrical SNR (ESNR).

We are interested in determining the upper limit to the SNR resulting only from the intrinsic noise sources related to the statistics of light and to the detection system. Measuring noise contributions such as ASE, signal-spontaneous beating, and spontaneous–spontaneous interference is the object of the optical monitoring system, and the sensitivity to such impairments will be limited by the intrinsic detectable SNR. Therefore, in the following, only the upper limit of the SNR, as determined by the detection conditions, will be considered.

In the case of nonlinear sampling, the peak value of the detected photocurrent is inversely proportional to the response time of the detection circuit $t_{\text{det}} = \left((2B_{\text{det}})^{-1}\right)$ and proportional to the gating pulsewidth $t_G$:

$$I_{pk} = \Re\eta'P_{tG}2B_{\text{det}} = \Re\eta'P_{tG}2B_{\text{det}}. \quad (7)$$

This equation is valid if $f_G \leq B_{\text{det}} \leq t_G^{-1}$, which is true in usual sampling situations.

The intrinsic noise sources consist of shot noise on the detected photocurrent, shot noise on the detector dark current $I_d$, and thermal (Johnson) noise on the resistor load $R_L$ (or the equivalent input noise on a transimpedance amplifier):

$$\sigma_I^2 = 2q(I_{pk} + I_d)B_{\text{det}} + \frac{4kT}{R_L}B_{\text{det}}. \quad (8)$$

If an optical filter is used for spectrally discriminating the samples from the data and the gate, which have a different wavelength, there will be no background photocurrent dependent on the average data and gate power. This constitutes an advantage of nonlinear sampling as compared with linear (homodyne) sampling. The SNR limit becomes

$$\text{SNR}_{\text{NLS}} = \frac{(\Re\eta'P_{tG}2B_{\text{det}})^2}{2q(\Re\eta'P_{tG}2B_{\text{det}} + I_d)B_{\text{det}} + \frac{4kT}{R_L}B_{\text{det}}}. \quad (9)$$

For an ideal detection system with negligible dark current and Johnson noise, we obtain the shot-noise limit for the sensitivity,

$$\text{SNR}_{\text{NLS}}|_{S-N} = \frac{\Re\eta'P_{tG}}{q} = \frac{\Re h\nu N_{\text{ph}}}{q}, \quad (10)$$

where $h\nu$ is the photon energy and $N_{\text{ph}}$ is the average number of photons generated by sampling within the time window $t_G$. Equation (10) shows that the signal-to-noise limit for sampling is ultimately determined by the shot-noise fluctuations of the photons gated within time $t_G$. Note also that, when thermal noise predominates, the SNR can be (counter-intuitively) improved by means of increasing the bandwidth of the detector. The shot-noise limit, however, does not depend on the detection bandwidth.

JON 1059 July 2002 / Vol. 1, No. 7 / JOURNAL OF OPTICAL NETWORKING 240
In the case of linear sampling with balanced detection, the signal power is given by the interference term in Eq. (3) and is equal to

$$I_{pk}^2 = 4R^2 P_{G0} P_D (t_G B_{det})^2.$$  \hfill (11)

Apart from Johnson noise and shot noise on the detector dark current, a noise source specific to balanced detection is shot noise on the total average photocurrent (before balancing):

$$\sigma_I^2 = 2R q (P_{D,ave} + P_{G0} t_G 2B_{det} + 2\sqrt{P_{G0} P_D t_G 2B_{det}} B_{det} + 4qI_d B_{det} + \frac{4kT}{R} B_{det}.$$  \hfill (12)

Again, in the ideal limit, when the Johnson noise and dark current can be neglected and $P_{G0} \gg P_D$, the SNR limit is

$$\text{SNR}_{LS} \mid_{S-N} = \frac{4R P_D t_G}{q} = \frac{4R h\nu}{q} N_{ph}.$$  \hfill (13)

This again is the shot-noise limit determined by the number of photons in the gating window only. Note the factor of 4 improvement that usually comes from homodyne detection: The shot-noise limit for linear optical sampling with background cancellation is 6 dB higher than the shot-noise limit for nonlinear sampling.

Let us compare the performance of the two optical sampling methods against sampling, using a reference receiver. In contrast to the sampling situations considered above, the bandwidth of the reference receiver is fixed to a value only slightly smaller than the data rate (typically $\sim 70\%$), in order to duplicate the behavior of the end (terminating) receiver. Thus the SNR limit of a receiver with bandwidth $B_{rec}$ of the order of magnitude of the data rate (RZ data with pulsewidth $t_D$ FWHM) is

$$\text{SNR}_{rec} = \frac{I_{pk}^2}{2q(I_{pk} + I_d) B_{rec} + \frac{4kT}{R} B_{rec}},$$  \hfill (14)

and $I_{pk} \approx R P_D t_D 2B_{rec}$, since the receiver integrates the entire energy of the data pulses. Then the shot-noise-limited performance of sampling with a reference receiver is

$$\text{SNR}_{rec} \mid_{S-N} = \frac{R P_D t_D}{q} = \frac{R h\nu}{q} N_{ph},$$  \hfill (15)

where $N_{ph}$ is the average number of photons in a bit.

The SNR in the shot-noise limit is depicted as a function of average data power in Fig. 2 for both linear and nonlinear (with $\eta' = 1.0$) sampling. For the graphs in Fig. 2(a), the bit rate of the 33% RZ data is $f_D = 10$ GHz, the detector responsivity is $R = 0.9$, and the bandwidth of the receiver is $B_{rec} = f_D$. The sampling pulses are $t_G = 3$ ps long. It is important to realize that the plots in Fig. 2(a) show the fundamental quantum limit to the sampling sensitivity for a given input signal power, as determined by the shot-noise fluctuations in the number of photons sampled (for coherent light; this limit can in principle be improved by use of amplitude-squeezed light).\footnote{12} Improved temporal resolution (smaller $t_G$) will inevitably come at the expense of reduced sensitivity. As opposed to traditional optical sampling, in an optical performance-monitoring context it is desirable\footnote{8} to restrict the duration of the gating pulses to a value comparable with the duration of the data pulses, to simulate an equivalent receiver. Enhanced temporal resolution may however be required in advanced single-channel diagnostics, or for higher data rates (monitoring applications at 160 Gbit/s require $t_G$ smaller than 2 ps).
Figure 2(b) shows the shot-noise-limited sensitivity for the same parameters with the exception of the gate pulsewidth, which was set to \( t_G = 33 \) ps in order to replicate the bandwidth of the receiver. The improvement by 10 dB in the sampling ESNR reflects the increase in number of photons generated in the sampling window; this improvement is rigorous only in the limit of gating pulses shorter than the data pulses. It should be pointed out that in a realistic monitoring situation the available average power is of the order of 1–3% of the line power (which can range from \(-25\) to \(-5\) dBm); i.e., an upper limit for the tap power available for monitoring is in the range of \(-20\) dBm.

An alternative way to interpret Fig. 2 would be to consider some physical limits for the nonlinear sampling of a data-stream tap with average power \( P_{D, ave} \sim -30 \) dBm, which is expected to be a typical value in monitoring applications. Since the fundamental limit on the SNR depends only on the product \( \eta' P_{D, ave} t_G \), nonlinear conversion efficiencies \( \eta' > 10^{-2} \) are necessary in order to attain useful (larger than 10) SNR values. Similar conclusions and limits also apply for asynchronous sampling, with the added caveat that the asynchronous (averaged) \( Q \)-factor depends sublinearly on the reference (synchronous) \( Q \)-factor,\(^4\) and therefore impairment detection would lose accuracy faster at low power levels.

To visualize the effect of shot noise on the eye diagram and histogram \( Q \), Fig. 3 shows eye diagrams calculated under the assumptions of shot noise only, for two different average data power levels \((-25\) and \(-35\) dBm), which correspond to a 1% tap at the beginning and at the middle of a typical (100 km) transmission fiber span with initial injected power of \(-5\) dBm. The sampling pulses are assumed to be \( t_G = 3 \) ps long, and the sampling efficiency \( \eta' = 1 \). From the eye diagrams, the OSNR of the “1” bits is estimated in the two cases to be 13.3 and 8.0 dB, respectively, which is in good agreement with the ESNR values shown in Fig. 2(a). Figure 3 confirms that a lower limit to the input power on an optical monitor based on optical sampling with \( t_G = 3 \) ps resolution and 25-dB SNR is set by fundamental quantum fluctuations at \(-25\) dBm. From the eye diagram in Fig. 3(b) it is also obvious that a loss in amplitude sensitivity is accompanied by a loss in timing (jitter) sensitivity.

To estimate the effect of dark current and Johnson noise introduced by a realistic de-
Fig. 4 presents calculated ESNR values for $t_G = 3$ ps and $t_G = 30$ ps gate pulses and nonlinear conversion efficiency $\eta'$ equal to unity. The sampling rate is $f_G = 100$ MHz, and the sampling detector bandwidth is $B_{\text{det}} = 100$ MHz. The peak power in the nonlinear sampling gating beam is $P_G^0 = 33$ W, which corresponds to an average power of $P_G = 10$ mW for 3-ps pulses and $P_G = 100$ mW for 30-ps pulses. The average gate power $P_G = 10$ mW in the linear sampling case. We assumed realistic values for the detector parameters nevertheless, taking care to optimize sensitivity. The values of the effective detector noise-equivalent power (NEP) for the linear and nonlinear sampling cases were $\text{NEP}_{\text{LS}} = 2 \text{ pA}/\sqrt{\text{Hz}}$, $\text{NEP}_{\text{NLS}} = 50 \text{ fA}/\sqrt{\text{Hz}}$, respectively, and the detection bandwidths were $B_{\text{det}} = 100$ MHz. For nonlinear sampling we considered a silicon avalanche photodiode (APD) with a multiplication factor of $M = 100$ and an excess noise factor $F_A = 3$. The SNR for the APD is estimated with Eq. (6) with the substitutions $I_{pk} \rightarrow I_{pk} M$, and the shot-noise term $\sigma^2 \rightarrow \sigma^2 M^2 F_A$. The effective receiver NEP was assumed to be $\text{NEP}_{\text{rec}} = 35 \text{ pA}/\sqrt{\text{Hz}}$, and its bandwidth is $B_{\text{rec}} = 10$ GHz. The dashed curves in Fig. 4 indicate the shot-noise-limited performance in the four cases considered. Figure 4 shows that in a realistic detection chain the sampling sensitivity can be degraded from its shot-noise-limited value. Attainment of the fundamental limit depends on the careful design of the sampling setup, with particular emphasis on minimizing detection noise. In the considered data power range, linear sampling attains shot-noise-limited sensitivity, whereas nonlinear sampling performance is degraded with respect to the shot-noise limit by the excess noise factor of the APD $F_A$. In fact, with conversion efficiencies $\eta'$, close to unity, and using low-noise detectors such as silicon APDs, optical sampling using sum-frequency generation in PPLN has attained (by our estimates) a SNR within a factor of approximately 8 dB (electrical) of the shot-noise limit.\(^7\) Note that both all-optical sampling schemes exhibit better performance at low powers than optoelectronic (reference receiver) sampling.

It is important to note that, for a given time resolution, the sensitivity of optical sampling is independent of the bit rate, as opposed to the reference receiver technique, in which an increase in bandwidth always results in a higher NEP; in contrast, slow detectors with small NEP can be used for sampling. The independence of the sampling sensitivity of the data bit rate is an important feature of this technique, which may become relevant as data rates increase to 40 Gbit/s and beyond. Another important comment is that, whereas the reference receiver $Q$-factor is directly correlated with the BER by means of Eq. (5), the...
Fig. 4. SNR limit (solid curves) for a detection system with the inclusion of thermal and dark noise, for linear and nonlinear sampling ($\eta' = 1.0$) compared against sampling with a reference receiver. Two different gating pulse widths are considered: (a) $t_G = 3$ ps and (b) $t_G = 30$ ps, respectively. The shot-noise limit is plotted with dashed curves for each case. For the linear sampling case, the solid and the dashed curves overlap.

$Q$-factor derived with optical sampling will in general be related to the BER by use of a calibration factor dependent on the sensitivity and bandwidth of the end receiver. When for the relevant power levels the sampling detector experiences a transition from the thermal-noise-limited to the shot-noise-limited regime, such as in Fig. 4(b), the calibration factor relating the sampling $Q$-factor to the end receiver $Q$-factor and thus to the BER may also be signal-power dependent.

4. Sensitivity of Sampling with Preamplification

Optical preamplification can be employed to reduce the effect of detection noise on sampling sensitivity. Let us assume that the tapped data stream of average power $P_{\text{ave}}$ undergoes optical amplification with an amplifier with gain $G$ and noise figure NF. The amplifier will introduce additional noise resulting from ASE, dependent on the population inversion $\chi = n_u/(n_u - n_g)$ between the ground and the upper states, which is related to the noise figure NF by $\text{NF} = \text{SNR}_{\text{in}} / \text{SNR}_{\text{out}}$ by

$$\text{NF} = 1 + \chi^* (G - 1)/G.$$  (16)

When we take as an example the case of nonlinear optical sampling, the signal-spontaneous beating introduces a current noise term:

$$\sigma^2_{I,D-s} = 2(\Re G\eta' P_{D} t_G^2 B_{\text{det}})[\Re 2\chi(G-1)\hbar v B_{\text{det}}].$$  (17)

The ESNR will be

$$\text{SNR}_{\text{NLS}} = \frac{(\Re G\eta' P_{D} t_G^2 B_{\text{det}})^2}{2q(I_{pk} + I_d)B_{\text{det}} + \frac{4kT}{R} B_{\text{det}} + \sigma^2_{I,D-s}}.$$  (18)
If the amplifier gain \( G \) is sufficiently large so that the signal-spontaneous noise term predominates, we recover the shot-noise limit for detection, degraded, however, by the noise figure of the amplifier:

\[
\text{SNR}_{\text{NLS}} = \frac{\left(\Re G\eta' P_D t G 2B_{\text{det}}\right)^2}{2\left|\Re G\eta' P_D t G 2B_{\text{det}}/\Re 2\chi(G-1)h\nu B_{\text{det}}\right|} = \frac{\eta'(P_D t G)}{(\text{NF})h\nu} = \frac{N_{\text{ph}}}{\text{NF}}. \tag{19}
\]

Similar expressions can be derived for linear sampling and for the reference receiver monitor. Estimated SNR values are plotted in Fig. 5 with the same parameters as in Fig. 4. The amplifier was assumed to have a gain \( G = 20 \text{ dB} \) and a noise figure \( \text{NF} = 5 \text{ dB} \). Under these assumptions, Fig. 5 shows that nonlinear sampling with a baseline electrical SNR better than 20 dB is possible only for input powers larger than \( P_{D,\text{ave}} = -32 \text{ dBm} \). On the other hand, preamplification degrades the sensitivity of linear sampling by the noise figure of the amplifier, because the shot-noise limit was already achieved for linear sampling even before amplification. Figure 5 shows that, although mitigating the effects of thermal noise, optical preamplification can only, even under the best conditions, degrade shot-noise-limited performance. In this sense, the shot-noise-limited performance in Fig. 2 still constitutes the fundamental guideline for the assessment of optical sampling.

Sensitivity limits and SNR floors can always be improved by reducing the detection bandwidth or, equivalently, by increasing the integration time. This approach is used in time-averaged sampling experiments (equivalent to data-gate cross correlation), in which a statistical average of the data pulse shape is detected. However, because a data stream consists of a (random) sequence of ones and zeros, temporal averaging has to be done over a time interval sufficient to overcome fluctuations in the data stream itself. Thus, to achieve 30-dB SNR without being affected by the random nature of the data, more than 1000 pulses have to be averaged, leading to integration times longer than 10 \( \mu \text{s} \) at each sampling point for a sampling rate \( f_G = 100 \text{ MHz} \). Therefore sensitivity improvements by averaging are intricately dependent on the data and sampling rates and on the original SNR; averaging will not automatically improve the SNR, as it does for purely repetitive signals.

5. Discussion

In practical situations it is necessary to consider additional sources of noise, related, for instance, to the gating pulses: Gain-switched lasers often used as gates experience amplitude noise and time jitter (up to several picoseconds) unless a seeding stabilization scheme is used. Another noise source could be the RIN (relative intensity noise) of the gating laser in the detection bandwidth; the effect of gating laser noise depends on the sampling process (linear or nonlinear) and also on the order of the nonlinearity. For instance, simulations indicate that the amplitude jitter of the gating laser has a relatively small effect on the nonlinear sampling \( Q \)-factor compared with the timing jitter, whereas the effect of amplitude jitter on the linear sampling \( Q \)-factor can be significant. To illustrate this point, Fig. 6 shows the sampling \( Q \)-factor degradation with gating laser time jitter and amplitude noise (laser RIN), for nonlinear and linear sampling, respectively. Unless otherwise mentioned, the detector and gating laser parameters are the same as for the simulations in Fig. 4. Balanced detection has been assumed for linear sampling, as discussed in Section 3. It is evident from Fig. 6 that, whereas nonlinear sampling is fairly robust to gating laser jitter, linear sampling sensitivity degrades rapidly with gating laser amplitude noise.

In this study, we have determined the fundamental limits to the sensitivity of optical sampling. Optical sampling sensitivity is ultimately limited by the shot noise on each bit of the data stream, and as the temporal resolution requirements increase and the available optical power in the data stream decreases, this limit becomes extremely relevant. Furthermore,
Fig. 5. SNR limits for linear and nonlinear sampling following preamplification with gain $G = 20$ dB and noise figure $NF = 5$ dB. The sampling efficiency is $\eta' = 1$. The gating pulse durations are (a) $t_G = 3$ ps and (b) $t_G = 30$ ps, respectively. The SNR for a reference receiver with preamplification is also shown. In the high-power region, the SNR is degraded from the shot-noise limit by the noise figure of the amplifier.

since the signal-to-noise requirements are much more stringent for the purposes of diagnostic techniques than at the receiver end, our estimates set constraints on the parameter space where optical sampling can be used for optical monitoring.

The sensitivity limitations delineated in the previous paragraphs have to be put in the context of the optical monitoring function, i.e., evaluating the quality of signal and detecting impairments in a network. For instance, shot noise introduces fluctuations only on the “1” bits of a RZ signal. These fluctuations will be directly reflected in the values of the quality factor $Q$ extracted from eye diagram histograms or in estimates for the data jitter. Detection of impairments monitored with the $Q$-factor or the eye diagram will be thus affected by shot noise, even if only trend estimation is important. Thermal noise on the detector, on the other hand, will limit the detectable ASE noise. Therefore, for practical purposes, optical sampling sensitivity requirements need to be defined in the context of the impairment to be monitored.

To conclude, we have presented a theoretical discussion of the fundamental and practical limits of optical sampling for optical performance monitoring under the realistic constraints of WDM optical networks. These estimates place fundamental limits on format-independent and data-rate-independent optical monitoring using all-optical sampling and may have to be taken into account in the design of future all-optical networks.

Acknowledgments

We thankfully acknowledge helpful discussions with C. Dorrer, H. Garcia, D. C. Kilper, and H. R. Stuart.
Fig. 6. (a) Nonlinear sampling $Q$-factors for sampling pulses with $t_G = 30$ ps for two values of the average data power, as a function of the sampling laser FWHM Gaussian time jitter. The solid curves show the shot-noise limit, whereas the dashed curves also take into account detector noise. (b) Linear sampling $Q$-factor for sampling pulses with $t_G = 3$ ps as a function of sampling laser amplitude noise.

References and Links
1998).