

Guaranteeing the BER in Transparent Optical Networks Using OOK Signaling

Poompat Saengudomlert, *Student Member, IEEE* and Muriel Médard, *Senior Member, IEEE*

Abstract—A network consisting of transparent optical nodes (TONs) can provide high speed end-to-end communication paths with very low bit-error rates (BERs). However, owing to component crosstalk and other degradations at TONs, the BER of a particular communication path traversing several TONs can be degraded by a few orders of magnitude even in the absence of component failure. Monitoring the quality-of-service (QoS) of a communication path has typically relied on sporadic BER testing and operation monitoring by the nodes using probe signals. Intermittent BER testing cannot provide continuous monitoring of the network QoS. On the other hand, the use of probe signals is not sensitive enough to detect the BER degradation. This work investigates a novel approach of monitoring service degradation at individual nodes using a wrap-around device which taps and compares signals from the input and the output at each TON along the lightpath. We propose a modification using hard limiters at TON inputs and derive the BER value that this modified method can guarantee in the presence of signal degradation due to coherent crosstalk at TONs.

Index Terms—Alarm systems, bit-error rate, crosstalk, digital communications, network monitoring, transparent optical networks.

NOMENCLATURE

SNR	SNR of the lightpath.
σ^2	Variance of the real and imaginary parts of AWGN at any TON at any bit time.
M	Number of TONs in the lightpath.
T	Number of bits transmitted in an observation period.
$J_i^{(k)}$	Coherent crosstalk signal (real part) at node k for the i th bit transmission period.
α	Threshold value for a device alarm.
β	Threshold value for a BER alarm.
\hat{B}	Average BER in an observation period below which a BER alarm is not generated with very high probability (see also \tilde{B} and $\overline{FP}_{\text{BER}}$).
\tilde{B}	Average BER in an observation period above which a BER alarm is generated with very high probability (see also $\overline{FN}_{\text{BER}}$).
$\overline{FP}_{\text{BER}}$	Upper limit on FP, the probability of concluding that $\text{BER} > \hat{B}$ when in fact $\text{BER} \leq \hat{B}$.
$\overline{FN}_{\text{BER}}$	Upper limit on FN, the probability of concluding that $\text{BER} \leq \hat{B}$ when in fact $\text{BER} > \hat{B}$.
$A(\cdot)$	Probability function of a device-alarm generation at a TON.
$B(\cdot)$	End-to-end BER function of a transmitted bit.

Manuscript received March 8, 2001; revised December 20, 2001. This work was supported by the Defense Advanced Research Projects Agency for the study of optical network security.

The authors are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: tengo@mit.edu; medard@mit.edu).

Publisher Item Identifier S 0733-8716(02)03058-5.

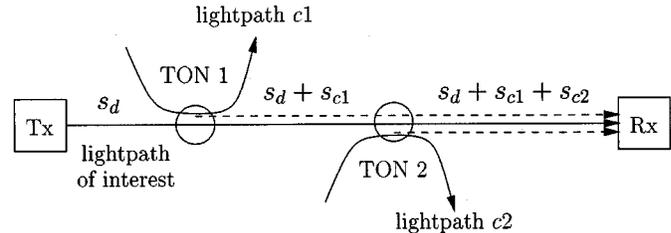


Fig. 1. Degradation of data signals due to coherent crosstalk.

I. INTRODUCTION

CONSIDER A communication path traversing several network nodes. In an electronic network, each intermediate node (between transmitting and receiving nodes) generally detects and regenerates (retransmits) traversing data signals. In an optical network, a high optical link bandwidth allows a transmission rate so high that electronic processing of data units (e.g., Internet protocol (IP) packets, ATM cells) cannot be done in real time. For example, an ATM cell of 53 bytes must be processed under 42.4 ns if an optical link transmits at 10 Gb/s. Consequently, it is desirable to have certain intermediate nodes be transparent optical nodes (TONs). A TON is a node which does not perform optical-to-electronic or electronic-to-optical conversion. Thus, signals traversing TONs remain in the optical domain. An end-to-end communication path through several TONs is referred to as a lightpath. A communication service through a lightpath can be made transparent to the formats (protocols) of transmitted data. However, since TONs do not detect and regenerate data signals, there is no error detection and correction at intermediate nodes. Instead, data signals experience cumulative degradation at TONs along the lightpath.

Common TONs consist of wavelength selective switches (WSSs), optical amplifiers, optical multiplexers, and demultiplexers. Certain switching TONs exhibit crosstalk effects [3]–[6], namely a light signal designated (switched) to one output fiber leaks onto another output fiber. Optical amplifiers are used to compensate for power loss within the network. The most common amplifiers are erbium doped fiber amplifiers (EDFAs), which are also broadband noise sources in the communication path. In addition, EDFAs exhibit gain competition, namely signals on different wavelengths must compete for limited power resources. An unusually high-powered signal on one wavelength can result in lower amplifier gains for all other wavelengths sharing the same EDFA.

A data signal traversing TONs can be degraded as a result of crosstalk and gain competition at TONs. Fig. 1 illustrates coherent crosstalk in a lightpath. Coherent crosstalk occurs when

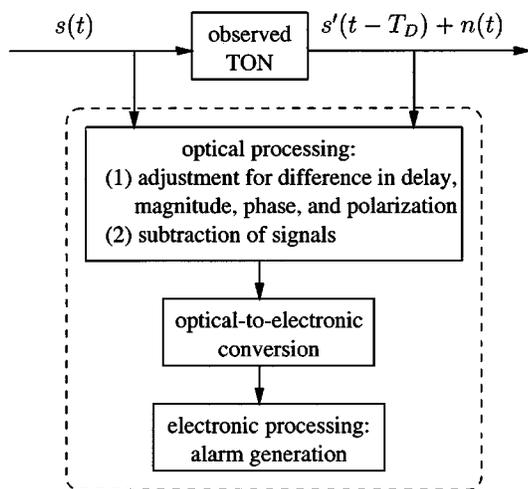


Fig. 2. A wrapping service monitoring device at a TON.

crosstalk signals have the same frequency as data signals and thus directly alter the data signal waveforms. Neglecting signal attenuation and all noise sources, the receiver receives the data signal s_d together with a crosstalk signal s_{c1} introduced at TON 1 by lightpath c_1 and a crosstalk signal s_{c2} introduced at TON 2 by lightpath c_2 . We concentrate our analysis on coherent crosstalk since noncoherent crosstalk due to gain competition is device specific and is outside the scope of this work.

Section II describes existing supervisory methods that may be used to monitor the quality-of-service (QoS) of a communication path. We argue that these methods are not sufficient to detect small bit-error rate (BER) degradations. Section III describes a novel BER monitoring system proposed in [1] and [2], and provides the setup for the analysis that follows. Section IV contains the analysis leading to the BER guarantees provided by the system discussed in Section III. To improve the sensitivity of the system, we propose in Section V a modification using hard limiters at TON inputs. Section VI provides a summary and future directions.

II. EXISTING SUPERVISORY METHODS

This section briefly discusses existing supervisory methods that may be used to detect BER degradation. References [7] and [8] provide detailed discussion on this topic. Two main approaches exist for monitoring the lightpath QoS. The first relies on the use of pilot signals, which are known signals inserted to travel along the same paths as data signals but are distinguishable from data signals. As an alternative, we can use an optical time domain reflectometer (OTDR) to detect the echo of pilot signals at the transmitter [9]. The second approach is based on observing the statistical parameters such as optical signal average power, optical signal frequency spectrum, and the BER of a lightpath [10]–[12].

Coherent crosstalk can degrade the BER without causing degradation to pilot signals. Pilot signal with frequencies close to those of data signals are referred to as subcarrier multiplexed (SCM) signals. In some cases, SCM signals are recovered by detecting them from the superposition with data signals. However, SCM signals are generally modulated at a much

lower rate than data signals [13], [14]. Therefore, a small percentage of data bits can be degraded without appreciable degradation of SCM signals.

Consider the second supervisory mechanism based on observing statistical parameters. A power detector may be used to detect BER degradation if crosstalk signals change the average power of data signals. If a crosstalk signal at a TON is present only 1% of the time, the BER may deteriorate by several orders of magnitude (up to 10^{-2}) while the average received signal power does not change significantly. By the same arguments, an optical spectrum analyzer (OSA) cannot detect BER degradation due to a sporadic crosstalk signal which is present only at a small percentage of time.

The use of a BER tester (BERT) is a standard procedure to detect BER degradation. However, BER testing is only performed sporadically on test patterns. Moreover, the use of a BERT requires a long detection time. For example, a BERT takes several hundreds of seconds to distinguish between a BER less than 10^{-10} and a BER greater than 10^{-8} for a transmission rate of 1 Gb/s.

Short detection time is desirable in several aspects. Since optical transmission rates are extremely high, a large amount of data can be affected before any action takes place. If retransmission is to be performed after the detection of excessive BER, long detection time implies the use of a large amount of buffer storage. As we shall see, our proposed method of detecting BER degradation is based on the detection of signal degradation at TONs rather than the detection of error bits themselves. As a result, our method can operate at roughly 6 orders of magnitude faster than a BERT.

III. SETUP FOR THE ANALYSIS

A. Novel Method for Detecting BER Degradation

This section describes the method for detecting BER degradation along a lightpath as proposed in [1] and [2]. The detection system is constructed by installing, at each TON along a lightpath, a service monitoring device which wraps around its TON as shown in Fig. 2. This wavelength-selective device compares the signals from input and output taps on a particular WDM channel. The relation between these two signals yields a diagnostic of TON operation. Our analysis will concentrate on a single wavelength division multiplexing (WDM) lightpath using on-off keying (OOK) signaling with coherent detection.

The overhead associated with this novel method includes monitoring devices at all TONs along the lightpath, additional control information about operational status of each TON, and additional transmit power to overcome tap losses. Although we concentrate on using a monitoring device for a single WDM channel, when there are lightpaths (on the same fiber) going through the same TON input–output pair, it is possible to modify the device using optical demultiplexers so that the device can simultaneously observe signals on several lightpaths taken from the same taps.

Lightpaths which do not have the same TON input–output pair require different monitoring devices at the TON. If only a fraction of lightpaths require service monitoring at any time, we can time-share the devices through the use of optical switches.

In addition, the monitoring device can be adjusted through its control software to handle possibly different bit rates without any change in its hardware.

In Fig. 2, we denote the input and output signals by $s(t)$ and $s'(t - T_D)$, where T_D is the TON delay and s' is used instead of s in consideration of the changes in magnitude (tap losses and amplification), phase, and polarization. Denote the sum of signal and noise at TON output by $s'(t - T_D) + n(t)$. In the optical domain, the device takes two tapped signals and compensates for the difference in delay, magnitude, phase and polarization under normal operation. It then performs signal subtraction and optical-to-electronic conversion using coherent detection. In general, the magnitude, phase, and polarization of data signals do not vary rapidly; and we can assume, in the absence of crosstalk, that the device yields an output proportional to $n(t)$. For notational simplicity, we drop the time notation below.

If there is a degradation J introduced at the TON, the device output will be proportional to $n + J$. When the output exceeds a preset threshold, the device generates an alarm notifying excessive signal degradation at the TON. With such a wrapping device installed at each node along the lightpath, we consider, in each bit transmission period, that a “device alarm” is generated if at least one TON in the lightpath generates an alarm. In each observation period, we count the total number of device alarms. If this total is higher than a preset threshold, a “BER alarm” (notifying an excessive BER) is generated.

Note that the BER we consider in this work is the probabilistic BER, as opposed to the actual observed BER. The use of this probabilistic measure allows us to detect BER degradation quickly without having to observe actual bit errors. In optical transmission, the BER guarantee is generally very low, e.g., 10^{-9} , therefore observing actual bit errors can be slow. For example, given the transmission rate of 1 Gbps and the BER of 10^{-9} , we expect only one bit error in each second of observation.

The performance criteria of the BER monitoring system are the false positive probability (FP), which is the probability of generating a BER alarm when the BER is below the guaranteed level, and the false negative probability (FN), which is the probability of not generating a BER alarm when the BER exceeds the guaranteed level. Preliminary results on the performance of this BER monitoring system are given in [2]. However, the results in [2] are based on direct detection of optical signals and the assumption that coherent crosstalk signals are constant throughout an observation period and are equal at all TONs. In this work, we drop these assumptions, i.e., we shall consider coherent detection of optical signals and allow crosstalk signals to vary with time and TON locations.

Our goal is to provide a BER guarantee for each observation period. In particular, let \tilde{B} denote the desired level of BER guarantee. When the BER exceeds \tilde{B} , we want the BER alarm generated with low FN. When the BER is below \tilde{B} , we want no BER alarm with low FP.

B. Notations

For a complex signal s , let s_R and s_I denote its real and imaginary parts. Let S^{OFF} and S^{ON} denote the OFF and ON signals for OOK signaling. Assume that the OFF and ON signals

are equally likely. By expressing a complex signal s in the form (s_R, s_I) , we can write S^{OFF} as $(0,0)$ and S^{ON} as $(S_R^{\text{ON}}, S_I^{\text{ON}})$.

Let M denote the number of TONs in the lightpath. Assume that there is additive white Gaussian noise (AWGN) at each TON with variance σ^2 for both real and imaginary parts. If we choose S^{ON} to be $(\sqrt{P}, 0)$, in the absence of crosstalk, the end-to-end BER is equal to $Q(\sqrt{P}/2\sqrt{M}\sigma^2)$, where Q is the complementary cumulative distribution function of a zero-mean, unit-variance Gaussian random variable. P is related to a particular signal-to-noise ratio (SNR) by the relation $\text{SNR} = (1/2)(P/2M\sigma^2)$.

Let $J_i^{(k)}$ be the coherent crosstalk signal at node k during the i th bit transmission in an observation period. Using the fact that crosstalk effects propagate toward downstream TONs, the end-to-end BER of the i th bit (i.e., the probability of error for the i th bit) is

$$\text{BER}_{\text{end-to-end}} = \frac{1}{2} Q \left(\frac{\frac{\sqrt{P}}{2} - \sum_{k=1}^M J_{i,R}^{(k)}}{\sqrt{M}\sigma^2} \right) + \frac{1}{2} Q \left(\frac{\frac{\sqrt{P}}{2} + \sum_{k=1}^M J_{i,R}^{(k)}}{\sqrt{M}\sigma^2} \right) \quad (1)$$

where $J_{i,R}^{(k)}$ add up constructively along the lightpath for the worst-case scenario.

Let d denote the device output at a particular TON. We want to generate an alarm when coherent crosstalk signal is large, i.e., $|d_R| > \alpha$, where α is a preset threshold. The probability of an alarm generation at node k at the i th bit time is

$$P\{\text{Device alarm}\} = Q \left(\frac{\alpha - J_{i,R}^{(k)}}{\sigma} \right) + Q \left(\frac{\alpha + J_{i,R}^{(k)}}{\sigma} \right). \quad (2)$$

Let T denote the length of an observation period in units of bit transmission periods, and \tilde{B} denote the average BER associated with this observation period. Among all crosstalk scenarios with $\text{BER} = \tilde{B}$, the one yielding the smallest total number of device alarms in an observation period has the smallest probability of a BER-alarm generation, and is thus the worst-case scenario for detection. To provide a BER guarantee of \tilde{B} , we need to check that our monitoring system performs satisfactorily under the worst-case scenario corresponding to the BER of \tilde{B} .

We shall describe each scenario in terms of crosstalk signals at all TONs at all bit transmission periods, i.e., $J_{i,R}^{(k)}$, $1 \leq k \leq M$, $1 \leq i \leq T$. For notational simplicity, we shall drop the subscript R and assume that $J_i^{(k)}$ refers to the real component for the rest of this paper.

For a fixed \tilde{B} , deriving the worst-case scenario is equivalent to finding the crosstalk signals which are the free variables in the following optimization problem

$$\begin{aligned} & \text{minimize} && \frac{1}{T} \times \text{expected number of device alarms} \\ & && \text{in } T \text{ bit periods} \\ & \text{subject to} && \text{average BER in } T \text{ bits} = \tilde{B}. \end{aligned} \quad (3)$$

For future reference, let $A(x)$ and $B(x)$ denote the device-alarm probability function and the end-to-end BER function whose expressions in (1) and (2) are presented again below

$$\begin{aligned} B(x) &= \frac{1}{2} Q\left(\frac{\frac{\sqrt{P}}{2} - x}{\sqrt{M\sigma^2}}\right) + \frac{1}{2} Q\left(\frac{\frac{\sqrt{P}}{2} + x}{\sqrt{M\sigma^2}}\right) \\ A(x) &= Q\left(\frac{\alpha - x}{\sigma}\right) + Q\left(\frac{\alpha + x}{\sigma}\right) \end{aligned} \quad (4)$$

IV. ANALYSIS ON THE BER GUARANTEES

This section provides mathematical analysis leading to BER guarantees by the monitoring system described in Section III-A. All the proofs of various lemmas are given in the Appendix.

We proceed by solving problem (3), which we now express more specifically. Since signal degradation propagates along the lightpath, the end-to-end BER for the i th transmitted bit is $B(\sum_{k=1}^M J_i^{(k)})$. Thus, the average BER in an observation period is $(1/T) \sum_{i=1}^T B(\sum_{k=1}^M J_i^{(k)})$. In each bit transmission period, we consider that a device alarm is generated if at least one TON in the lightpath generates a device alarm. The expected number of device alarms in an observation period is, therefore, $\sum_{i=1}^T [1 - \prod_{k=1}^M (1 - A(J_i^{(k)}))]$. Thus, problem (3) becomes

$$\begin{aligned} \text{minimize} \quad & \frac{1}{T} \sum_{i=1}^T \left[1 - \prod_{k=1}^M (1 - A(J_i^{(k)})) \right] \\ \text{subject to} \quad & \tilde{B} = \frac{1}{T} \sum_{i=1}^T B\left(\sum_{k=1}^M J_i^{(k)}\right) \end{aligned} \quad (5)$$

whose feasible solutions do not form a convex set. Therefore, common techniques based on convex optimization do not apply.

A. Worst-Case Scenario for $M = 1$ and $T > 1$

Assume a single TON in the lightpath. With only one TON, we can drop the superscript and define $J_i = J_i^{(1)}$, $1 \leq i \leq T$. We start with a simple case with $T = 2$. Problem (5) becomes

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} (A(J_1) + A(J_2)) \\ \text{subject to} \quad & \tilde{B} = \frac{1}{2} (B(J_1) + B(J_2)). \end{aligned} \quad (6)$$

The Lagrange multiplier method gives us the following constraints for an extremum:

$$A'(J_i) = \lambda B'(J_i), \quad i = 1, 2, \text{ eqno}(7)$$

where $A'(J_i)$ is a derivative of function $A(J_i)$ with respect to J_i .

Because $A'(J_i)$ and $B'(J_i)$ are even, we shall consider only nonnegative values of J_i . Since $A'(0) = B'(0) = 0$, zero is a solution to (7). Lemma 1 states that, under Assumption 1, there is at most one positive solution to (7).

Assumption 1: Let $\eta = \sqrt{P}/2$. Assume that $\sigma < \alpha < \eta - \sqrt{3}\sigma$.

Lemma 1: Given $\lambda > 0$ and that Assumption 1 holds, there is at most one positive solution to the equation $A'(J) = \lambda B'(J)$ in $(0, \eta - \sigma)$.

Assumption 1 generally holds when the SNR is sufficiently high since $\text{SNR} = (1/2)(2\eta^2/M\sigma^2)$ and, therefore, σ is small compared to η . Empirical data suggest that lemma 1 hold regardless of Assumption 1.

Given that only zero and one other positive value are solutions to (7), there are two forms of solutions for an extremum, namely $J_1 = J_2$, and J_1 (or J_2) = 0 with J_2 (or J_1) > 0 . Let \tilde{J} be such that $B(\tilde{J}) = \tilde{B}$ and J^s be such that $(1/2)B(0) + (1/2)B(J^s) = \tilde{B}$. It follows that the extrema are (\tilde{J}, \tilde{J}) , $(0, J^s)$, and $(J^s, 0)$.

Given the locations of the extrema, we can investigate their properties by evaluating the bordered Hessian determinant $|H|_{(J_1, J_2)}$ of the Lagrangian

$$|H|_{(J_1, J_2)} = \begin{vmatrix} 0 & -B'(J_1) & -B'(J_2) \\ -B'(J_1) & L''(J_1) & 0 \\ -B'(J_2) & 0 & L''(J_2) \end{vmatrix} \quad (8)$$

where $L(J_1, J_2, \lambda) = (1/2)(A(J_1) + A(J_2)) - \lambda((1/2)(B(J_1) + B(J_2)) - \tilde{B})$. The following lemma asserts that (\tilde{J}, \tilde{J}) corresponds to a local maximum, and $(J^s, 0)$ and $(0, J^s)$ correspond to local minima.

Lemma 2: Under Assumption 1, $|H|_{(J^s, 0)} = |H|_{(0, J^s)} < 0$ and $|H|_{(\tilde{J}, \tilde{J})} > 0$, for $\tilde{B} < (1/2)B(0) + (1/2)B(\eta - \sigma)$.

Since the constraint set defined in (6) is a contour curve of $(B(J_1) + B(J_2))/2$ in the first quadrant on the (J_1, J_2) plane, and this curve connects the two end points $(J^s, 0)$ and $(0, J^s)$, the absence of a local minimum elsewhere guarantees that $(J^s, 0)$ and $(0, J^s)$ are indeed global minima for the continuous cost function in (6). We conclude the result in Theorem 1.

Theorem 1: Under Assumption 1, given a single TON ($M = 1$) in the lightpath, the observation of 2 bit periods ($T = 2$), and the average BER $\tilde{B} < (1/2)B(\eta - \sigma) + (1/2)B(0)$, a scenario in which only 1 out of 2 bits is affected by coherent crosstalk has the smallest expected total number of device alarms and is thus the worst-case scenario.

We can extend the result from theorem 1 to the case with $T > 2$, problem (5) becomes

$$\begin{aligned} \text{minimize} \quad & \frac{1}{T} \sum_{i=1}^T A(J_i) \\ \text{subject to} \quad & \tilde{B} = \frac{1}{T} \sum_{i=1}^T B(J_i). \end{aligned} \quad (9)$$

The Lagrange multiplier method gives the same condition as in (7), namely $A'(J_i) = \lambda B'(J_i)$, $i = 1, \dots, T$. We have from Lemma 1 that under Assumption 1 there exists at most one positive solution to $A'(J_i) = \lambda B'(J_i)$ in $[0, \eta - \sigma)$. We shall assume that $\tilde{B} < (1/T)B(\eta - \sigma) + ((T-1)/T)B(0)$ to guarantee that $J_i \in [0, \eta - \sigma)$ for all i in all scenarios.

Out of scenarios with the average BER \tilde{B} , define a scenario in which t out of T bits are affected by crosstalk signals of constant magnitude J_i^s as scenario t . All possible scenarios are $1, 2, \dots, T$. Lemma 1 tells us that an extremum corresponds to one of the T scenarios just described. Let A_t denote the expected total number of device alarms in an observation period corresponding to scenario t . Lemma 3 states that $A_T > A_{T-1} > \dots > A_1$, yielding Theorem 2.

Lemma 3: Under assumption 1, given $M = 1$, $T > 1$, and $\tilde{B} < (1/T)B(\eta - \sigma) + ((T-1)/T)B(0)$, we have that $A_T > A_{T-1} > \dots > A_1$.

Theorem 2: Under Assumption 1, given $M = 1$, $T > 1$, and $\tilde{B} < (1/T)B(\eta - \sigma) + ((T-1)/T)B(0)$, a scenario in which only 1 out of T bits is affected by coherent crosstalk has the smallest expected total number of device alarms and is thus the worst-case scenario.

B. Worst Case Scenario for $M > 1$ and $T = 1$

With $T = 1$, we can drop the subscript and define $J^{(k)} = J_1^{(k)}$, $1 \leq k \leq M$. Also, define $C(J) = 1 - A(J)$. Considering only nonnegative values of $J^{(k)}$ for the extrema, we define $\tilde{J} > 0$ to be such that $B(\tilde{J}) = \tilde{B}$. Note that \tilde{J} is unique since $B(\cdot)$ is strictly increasing in $[0, \infty)$. Problem (5) becomes

$$\begin{aligned} & \text{minimize} && -\prod_{k=1}^M C(J^{(k)}) \\ & \text{subject to} && \tilde{J} = \sum_{k=1}^M J^{(k)}. \end{aligned} \quad (10)$$

The Lagrange multiplier method yields the constraint for an extremum

$$\begin{aligned} -\lambda &= C'(J^{(1)}) \prod_{k \neq 1} C(J^{(k)}) = C'(J^{(2)}) \prod_{k \neq 2} C(J^{(k)}) \\ &= \dots = C'(J^{(M)}) \prod_{k \neq M} C(J^{(k)}). \end{aligned}$$

Multiplying through by $[\prod_{k=1}^M C(J^{(k)})]^{-1}$, we have

$$\frac{C'(J^{(1)})}{C(J^{(1)})} = \frac{C'(J^{(2)})}{C(J^{(2)})} = \dots = \frac{C'(J^{(M)})}{C(J^{(M)})}.$$

Lemma 4 states that $C'(J)/C(J)$ is a strictly decreasing one-to-one function in $[0, \alpha)$.

Lemma 4: The function $C'(J)/C(J)$ strictly decreases in $[0, \alpha)$.

It follows that for any value of $\tilde{J} < M\alpha$, an extremum must have equal components, namely $(\tilde{J}/M, \dots, \tilde{J}/M)$, and is thus unique. To show that $(\tilde{J}/M, \dots, \tilde{J}/M)$ is indeed the global minimum, it is sufficient to show that the cost function at another point in the constraint set is larger than at the extremum. Consider the point $(3\tilde{J}/M, -\tilde{J}/M, \tilde{J}/M, \dots, \tilde{J}/M)$, we have that

$$\begin{aligned} & -C(3\tilde{J}/M)C(-\tilde{J}/M)C(\tilde{J}/M)^{M-2} \\ &= -C(3\tilde{J}/M)C(\tilde{J}/M)^{M-1} > -(C(\tilde{J}/M))^M \end{aligned}$$

where the equality follows from the fact that $C(J)$ is symmetric, and the inequality from the fact that $C(3\tilde{J}/2) < C(\tilde{J}/2)$ since $C(J)$ is strictly decreasing in $[0, \infty)$. We have thus established Theorem 3.

Theorem 3: Given $M > 1$, $T = 1$, and $\tilde{B} < B(M\alpha)$, a scenario in which the crosstalk signals at M TONs are equal

has the smallest expected total number of device alarms and is thus the worst-case scenario.

C. Worst Case Scenario for $M > 1$ and $T > 1$

As in the last section, define $C(J) = 1 - A(J)$. Problem (5) becomes

$$\begin{aligned} & \text{minimize} && -\frac{1}{T} \sum_{i=1}^T \prod_{k=1}^M C(J_i^{(k)}) \\ & \text{subject to} && \tilde{B} = \frac{1}{T} \sum_{i=1}^T B\left(\sum_{k=1}^M J_i^{(k)}\right). \end{aligned} \quad (11)$$

In any optimal solution, we must have, for each i , the minimum value of $-\prod_{k=1}^M C(J_i^{(k)})$ subject to the value of $\sum_{k=1}^M J_i^{(k)}$. Otherwise, we can obtain the better solution by reassigning the values of $J_i^{(k)}$ such that their sum remains the same but $-\prod_{k=1}^M C(J_i^{(k)})$ is strictly smaller.

Let J_i denote $\sum_{k=1}^M J_i^{(k)}$. Let $C_M(J_i)$ denote the maximal value of $\prod_{k=1}^M C(J_i^{(k)})$ for the i th bit interval subject to the constraint on $\sum_{k=1}^M J_i^{(k)}$. (In Section IV-B, empirical data suggest that $C_M(J_i) = (C(J_i/M))^M$, i.e., in the worst case scenario, coherent crosstalk signals at M TONs are equal. Theorem 3 verifies the relationship given the constraint $J_i < M\alpha$.) In terms of $C_M(J)$, problem (11) becomes

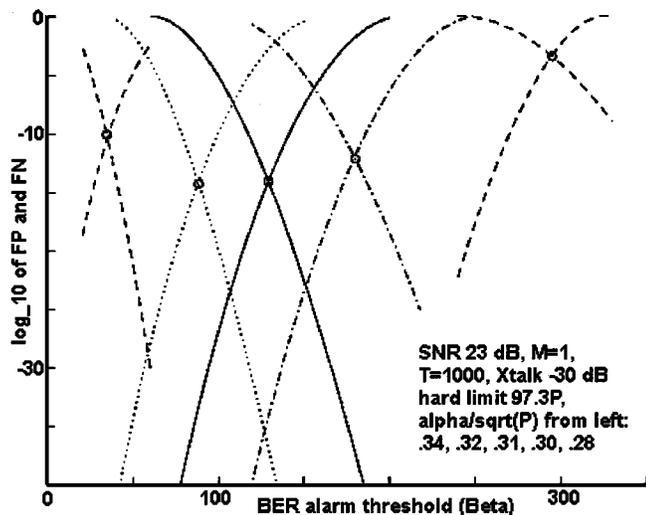
$$\begin{aligned} & \text{minimize} && -\frac{1}{T} \sum_{i=1}^T C_M(J_i) \\ & \text{subject to} && \tilde{B} = \frac{1}{T} \sum_{i=1}^T B(J_i). \end{aligned} \quad (12)$$

Comparing (12) with (9), we can consider problem (12) as if we were to find the worst case scenario for $M = 1$ and $T > 1$, but with the function $-C_M(J)$ instead of $A(J)$ as in Section IV-A.

Using the approach in Section IV-A, the Lagrange multiplier method yields the constraint for an extremum $C'_M(J_i) = \lambda B'(J_i)$, $i = 1, \dots, T$. Empirical data suggests the similar result as in Section IV-A, namely the curve $-C'_M(J)$ intersects $B(J)$ at zero and one other positive value in $[0, \infty)$, and the extrema correspond to scenarios in which some bits are not degraded while the others are degraded equally (denoted as scenarios 1 to T in Section IV-A). However, we have no theoretical proof of this claim. The difficulty lies in the complexity of an expression for $C'_M(J)$. Numerical verification in several cases leads us to the following conjecture: *given $M > 1$, $T > 1$, the worst-case scenario is the one in which only 1 out of T bits is affected by coherent crosstalk, and the crosstalk signals at M TONs are equal.*

V. BER MONITORING SYSTEM WITH HARD LIMITERS

For a low BER guarantee, under the worst-case scenario found in Section IV, a single severely degraded bit would only generate a single extra device alarm (compared to the case with no degradation at all) which leads to high values for FP and FN.

Fig. 3. FP and FN as functions of β for different values of α ($M = 1$).

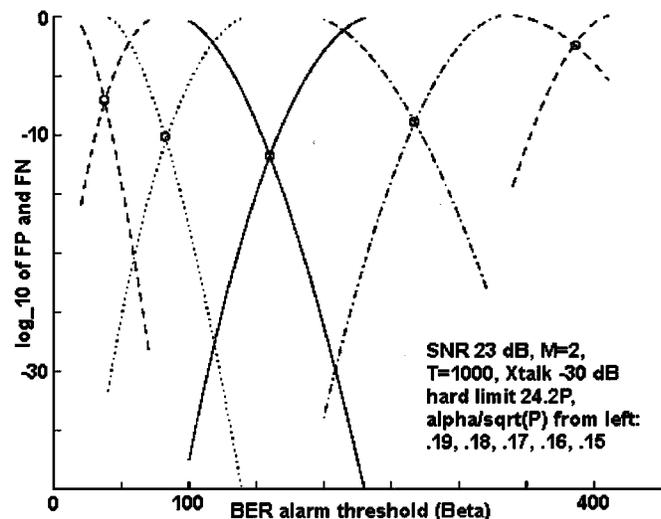
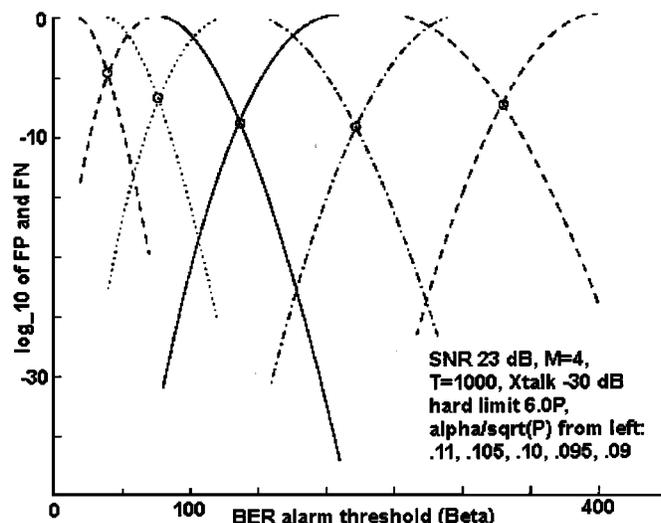
Therefore, we propose the use of hard limiters at TON inputs to improve the performance of the BER monitoring system.

A hard limiter is a device that limits the power of signals passing through it. For crosstalk signals to degrade few bits significantly, the instantaneous crosstalk signal power in those bit periods must be very high. But very high instantaneous crosstalk signal power results from very high instantaneous input power, which are prevented by hard limiters. Moreover, when the hard limit is sufficiently higher than the ON level signal power of OOK, the use of hard limiters does not interfere with normal operation.

In the presence of hard limiters, multiple bits must be degraded for the BER to exceed the guaranteed level. Given that at least t ($t > 1$) bits must be degraded, it follows that the worst-case scenario is the one in which t out of T bits are equally affected and crosstalk signals at M TONs are equal. When the hard limit is sufficiently low (the value of t is sufficiently high), our BER monitoring system will be able to detect the worst-case scenario for some BER guarantee.

We conclude with some numerical examples. Assign the following parameters: transmission rate = 1 Gbps, SNR = 23 dB, $\hat{B} = 10^{-12}$, $\tilde{B} = 10^{-8}$, $\overline{FP}_{\text{BER}} = \overline{FN}_{\text{BER}} = 10^{-8}$, and a TON crosstalk level of -30 dB. Figs. 3–5 show some FP and FN curves as functions of β for different values of α when $T = 1000$ and $M = 1, 2$ and 4 respectively. Due to difficulty in exact numerical computation, we approximate the values of FP and FN by their upper bounds using Chernoff bound approximation, which is reasonably accurate for small values of FP and FN.

FP is decreasing while FN is increasing with β . For a particular pair of FP and FN, if their intersection (denoted by a circle in figures) lies below 10^{-8} ($\overline{FP}_{\text{BER}}$ and $\overline{FN}_{\text{BER}}$), then the corresponding values of α and β can be used to construct a BER monitoring system. More specifically, when $M = 2$, we can choose from Fig. 4 the values $\alpha = 0.17\sqrt{P}$ and $\beta = 155$, or the values $\alpha = 0.16\sqrt{P}$ and $\beta = 85$. Both corresponding BER monitoring systems are able to detect a BER above 10^{-8} in $1 \mu\text{s}$ ($T = 1000$ at 1 Gbps), which is approximately 6 orders of magnitudes faster than a BERT described in Section II.

Fig. 4. FP and FN as functions of β for different values of α ($M = 2$).Fig. 5. FP and FN as functions of β for different values of α ($M = 4$).

Our examples also show that there is a desirable range for the values of α . When α is set too small, even under normal operation, we expect a large number of device alarms in each observation period, and therefore need to set β quite large to keep FP small. Consequently, such a large β yields too large FN (since FN is increasing with β). On the other hand, when α is set too large, even if β can be set small to keep FP small, we expect only few device alarms in the presence of BER degradation. Consequently, FN is still too large.

In general, the required hard limit decreases as M increases, since for the same BER degradation, the amount of signal degradation at individual nodes can be smaller for a larger M . Our examples suggest that there is a limit on the number of TONs in a lightpath above which the hard limit gets too close to the ON power level under normal operation.

VI. CONCLUSION AND FUTURE DIRECTIONS

We derived guaranteed BERs provided by the novel BER monitoring system in [1] and [2]. In doing so, we found that the

worst-case coherent crosstalk scenario for the BER monitoring system to detect is the one in which only 1 out of T bits in the observation period is affected by coherent crosstalk signals of the same magnitude at all M TONs in the lightpath. Based on this result, we suggested the use of hard limiters at TON inputs to improve the system performance.

The BER monitoring system considered in this work offers several advantages. The system can detect BER degradation much faster than a BERT (6 orders of magnitude faster in our examples). The monitoring devices can be installed at existing TONs so no modification of existing infrastructure is required. Although our analysis is based on OOK signaling, the same method can be applied to other transmission schemes as well. Finally, since the system does not differentiate between the sources of signal degradation, it can be used to detect malicious users who intentionally cause service disruption.

Several related issues require further consideration. First, a similar kind of analysis to this work can be done on BER degradation due to crosstalk effects as well as gain competition at EDFAs. Another issue has to do with actual implementation of the BER monitoring system. While the installation of service monitoring devices at TONs to monitor different WDM lightpaths seems straightforward in a WDM network with static wavelength routing, the implementation becomes more complex if we allow dynamic wavelength routing. Finally, there are also issues regarding how to provide appropriate responses after the detection of BER degradation.

APPENDIX PROOFS OF VARIOUS LEMMAS

Lemma A1: Given that $\lambda > 0$, $\eta > \alpha$, consider the function $f(x)$ of the form

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-((x-\alpha)^2/2\sigma^2)} - \lambda \frac{1}{\sqrt{2\pi\sigma^2}} e^{-((x-\eta)^2/2\sigma^2)}. \quad (13)$$

$f(x)$ has the following characteristics:

- i) There is a unique solution to $f(x) = 0$.
- ii) $f(x)$ has a unique local maximum and a unique local minimum.
- iii) Let x^{\max} and x^{\min} denote the locations of the maximum and the minimum respectively. We have $x^{\max} < \alpha < \eta < x^{\min}$.
- iv) As x increases from $-\infty$ to x^{\max} , $f(x)$ strictly increases from the limit value of zero to $f(x^{\max})$ and has a single saddle point where $f(x)$ changes from convexity (\smile) to concavity (\frown).
- v) As x increases from x^{\max} to x^{\min} , $f(x)$ strictly decreases.
- vi) As x increases from x^{\min} to ∞ , $f(x)$ strictly increases from $f(x^{\min})$ to the limit value of zero and has a single saddle point at which $f(x)$ changes from convexity to concavity.
- vii) $f(x)$ has three saddle points in $(-\infty, \infty)$. In addition to the two saddle points in iv) and vi), there is the third saddle point in (x^{\max}, x^{\min}) at which $f(x)$ changes from concavity to convexity.

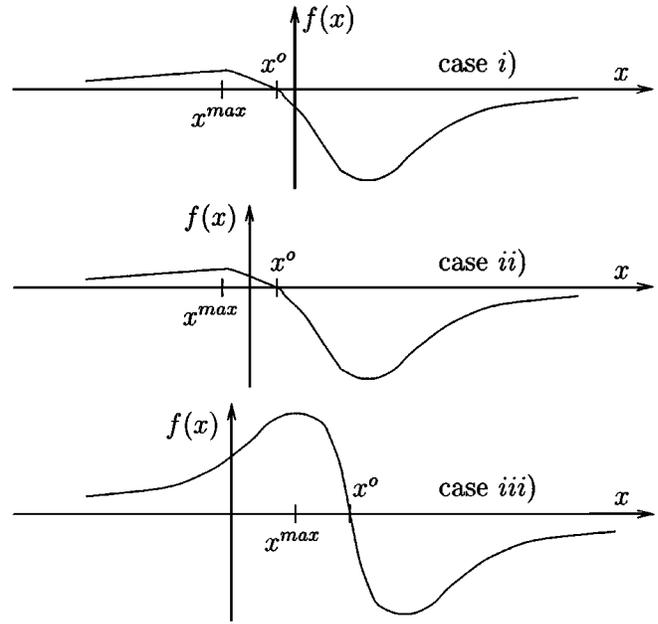


Fig. 6. Three different cases of $f(x)$ for the proof of Lemma 1.

Proof: Fig. 6 shows some example curves of $f(x)$.

- i) It is straightforward to solve $f(x) = 0$ to obtain a unique solution $x^o = (\eta^2 - \alpha^2 - 2\sigma^2 \ln \lambda) / 2(\eta - \alpha)$.
- ii), iii) Consider the first derivative of $f(x)$

$$f'(x) = -\frac{x-\alpha}{\sqrt{2\pi\sigma^3}} e^{-((x-\alpha)^2/2\sigma^2)} + \lambda \frac{x-\eta}{\sqrt{2\pi\sigma^3}} e^{-((x-\eta)^2/2\sigma^2)}. \quad (14)$$

For $x \in [\alpha, \eta]$, $f'(x) < 0$ and thus $f(x)$ strictly decreases in $[\alpha, \eta]$. As $x \rightarrow -\infty$, the first term dominates and we have $\lim_{x \rightarrow -\infty} f'(x) > 0$. As $x \rightarrow \infty$, the second term dominates and we have $\lim_{x \rightarrow \infty} f'(x) > 0$.

The fact that $\lim_{x \rightarrow -\infty} f(x) = 0$ and $\lim_{x \rightarrow -\infty} f'(x) > 0$ tell us that $f(x)$ increases from 0 as x increases from $-\infty$. Since $f(x)$ is continuous and has a zero crossing (solution) at x^o , $f(x)$ must have at least one local maximum. Since $f(x)$ strictly decreases in the range $[\alpha, \eta]$, there exists a local maximum x^{\max} and $x^{\max} < \alpha$. Similar arguments show that there exist a local minimum x^{\min} and $x^{\min} > \eta$.

Consider the expression of $f'(x)$ in (14). Denote the term

$$\frac{x-\alpha}{\sqrt{2\pi\sigma^3}} e^{-((x-\alpha)^2/2\sigma^2)}$$

by $f_1(x)$ and the term

$$\lambda \frac{x-\eta}{\sqrt{2\pi\sigma^3}} e^{-((x-\eta)^2/2\sigma^2)}$$

by $f_2(x)$. Consider their ratio

$$\frac{f_1(x)}{f_2(x)} = \frac{(x-\alpha)}{\lambda(x-\eta)} \times \frac{e^{-((x-\alpha)^2/2\sigma^2)}}{e^{-((x-\eta)^2/2\sigma^2)}}, \quad x \neq \eta.$$

We can think of $f_1(x)/f_2(x)$ as the product of two quantities whose derivatives are given by

$$\begin{aligned} \frac{d}{dx} \left(\frac{x - \alpha}{\lambda(x - \eta)} \right) &= -\frac{\eta - \alpha}{\lambda(x - \eta)^2} < 0, \quad x \neq \eta, \\ \frac{d}{dx} \left(\frac{e^{-((x-\alpha)^2/2\sigma^2)}}{e^{-((x-\eta)^2/2\sigma^2)}} \right) &= -\frac{\eta - \alpha}{\sigma^2} \\ &\cdot e^{-((2x(\eta-\alpha)+\alpha^2-\eta^2)/2\sigma^2)} < 0. \end{aligned}$$

Since $f_1(x)/f_2(x)$ is a product of two strictly decreasing functions, it is strictly decreasing in $(-\infty, \eta)$ and in (η, ∞) where it is well defined.

Since $f_1(x)/f_2(x)$ strictly decreases in $(-\infty, \eta)$, x^{\max} is the only value of x in $(-\infty, \eta)$ such that $f_1(x)/f_2(x) = 1$, i.e., $f'(x) = 0$. Since $f(x)$ strictly decreases in $[\alpha, \eta]$, there cannot be an extremum of $f(x)$ in $[\alpha, \eta]$. Therefore, x^{\max} is the only extremum (which happens to be a local maximum) in $(-\infty, \alpha)$.

By the same arguments, x^{\min} is the only value of x in (η, ∞) such that $f'(x) = 0$. Thus x^{\min} is the only extremum (which happens to be a local minimum) in (η, ∞) .

iv), vi), vii) Since there is no extremum in $(-\infty, x^{\max})$, $\lim_{x \rightarrow -\infty} f'(x) > 0$, and $f(x)$ is differentiable in $(-\infty, x^{\max})$, $f(x)$ strictly increases in $(-\infty, x^{\max})$. Similarly, since there is no extremum in (x^{\min}, ∞) , $\lim_{x \rightarrow \infty} f'(x) > 0$, and $f(x)$ is differentiable in (x^{\min}, ∞) , $f(x)$ strictly increases in (x^{\min}, ∞) .

Consider the second derivative of $f(x)$

$$\begin{aligned} f''(x) &= \frac{1}{\sqrt{2\pi\sigma^5}} [(x - \alpha)^2 - \sigma^2] e^{-((x-\alpha)^2/2\sigma^2)} \\ &\quad - \lambda \frac{1}{\sqrt{2\pi\sigma^5}} [(x - \eta)^2 - \sigma^2] e^{-((x-\eta)^2/2\sigma^2)}. \end{aligned} \quad (15)$$

Since $\lim_{x \rightarrow -\infty} f''(x) > 0$ and $\lim_{x \rightarrow \infty} f''(x) < 0$; therefore, there must be an odd number of saddle points [with $f''(x) = 0$] in $(-\infty, \infty)$. Let $g_1(x)$ and $g_2(x)$ denote the first and the second terms in (15)

$$\begin{aligned} g_1(x) &= \frac{1}{\sqrt{2\pi\sigma^5}} [(x - \alpha)^2 - \sigma^2] e^{-((x-\alpha)^2/2\sigma^2)}, \\ g_2(x) &= \lambda \frac{1}{\sqrt{2\pi\sigma^5}} [(x - \eta)^2 - \sigma^2] e^{-((x-\eta)^2/2\sigma^2)}. \end{aligned}$$

Similar to the proof of ii) and iii). Consider the ratio $g_1(x)/g_2(x)$ as the product of two quantities

$$\frac{g_1(x)}{g_2(x)} = \frac{1}{\lambda} \frac{(x - \alpha)^2 - \sigma^2}{(x - \eta)^2 - \sigma^2} \times \frac{e^{-((x-\alpha)^2/2\sigma^2)}}{e^{-((x-\eta)^2/2\sigma^2)}}, \quad x \neq \eta + \sqrt{\sigma}.$$

Before proceeding further, note that it is straightforward to show $f''(\eta + \sqrt{\sigma}) \neq 0$, i.e., $\eta + \sqrt{\sigma}$ is not a saddle point. The derivatives of the two factors above are

$$\begin{aligned} \frac{d}{dx} \left(\frac{1}{\lambda} \frac{(x - \alpha)^2 - \sigma^2}{(x - \eta)^2 - \sigma^2} \right) &= \frac{-2(\eta - \alpha) [(x - \eta)(x - \alpha) + \sigma^2]}{\lambda [(x - \eta)^2 - \sigma^2]^2}, \quad x \neq \eta + \sqrt{\sigma}, \end{aligned}$$

$$\begin{aligned} \frac{d}{dx} \left(\frac{e^{-((x-\alpha)^2/2\sigma^2)}}{e^{-((x-\eta)^2/2\sigma^2)}} \right) &= -\frac{\eta - \alpha}{\sigma^2} e^{-((2x(\eta-\alpha)+\alpha^2-\eta^2)/2\sigma^2)} < 0. \end{aligned} \quad (16)$$

The term $2(\eta - \alpha)/\lambda[(x - \eta)^2 - \sigma^2]^2$ in (16) is strictly positive since $\lambda > 0$ and $\eta > \alpha$. Consider the following equation which comes from the remaining factor in (16):

$$- [(x - \eta)(x - \alpha) + \sigma^2] = 0. \quad (17)$$

Equation (17) has either zero or two real solutions. If there were no real solution to (17), the left-hand side of (17) is always negative. Consequently, $g_1(x)/g_2(x)$ will be a product of two strictly decreasing functions, and is thus strictly decreasing and equal to 1 (i.e., $f''(x) = 0$) at most once. Since we have argued above that there must be an odd number of saddle points in $(-\infty, \infty)$, it follows that there is a single saddle point in this case. If there are two real solutions to (17) denoted by x_1 and x_2 , the ratio $g_1(x)/g_2(x)$ is decreasing in $(-\infty, x_1)$, increasing in (x_1, x_2) , and decreasing in (x_2, ∞) . Thus there can be either one saddle point or three saddle points in this case.

Notice that $f''(x^{\max}) < 0$ and $f''(x^{\min}) > 0$ since a twice differentiable function must be concave at a local maximum and convex at a local minimum. Since $\lim_{x \rightarrow -\infty} f''(x) > 0$, there is at least one saddle point where $f(x)$ changes from convexity to concavity in $(-\infty, x^{\max})$. Similarly, since $\lim_{x \rightarrow \infty} f''(x) < 0$, there is at least one saddle point where $f(x)$ changes from convexity to concavity in (x^{\min}, ∞) .

Therefore, the case of one saddle point is impossible. There must be three saddle points. One is in $(-\infty, x^{\max})$. Another is in (x^{\min}, ∞) . It follows that the third saddle point can only be in (x^{\max}, x^{\min}) , or else $f(x)$ would be convex at x^{\max} if there were two saddle points in $(-\infty, x^{\max})$ and vice versa.

v) Since there are a unique local maximum, a unique local minimum, and no other extreme point, it follows that the continuous function $f(x)$ is strictly decreasing in (x^{\max}, x^{\min}) . \square

Lemma A2: Under Assumption 1, $f(x)$ as defined in (13) has a strictly decreasing second derivative in $(x^s, \alpha]$, where x^s denotes the unique saddle point in $(-\infty, x^{\max})$.

Proof: Consider the third derivative of $f(x)$

$$\begin{aligned} f'''(x) &= \frac{1}{\sqrt{2\pi\sigma^7}} (\alpha - x) [(x - \alpha)^2 - 3\sigma^2] e^{-((x-\alpha)^2/2\sigma^2)} \\ &\quad - \lambda \frac{1}{\sqrt{2\pi\sigma^7}} (\eta - x) [(x - \eta)^2 - 3\sigma^2] e^{-((x-\eta)^2/2\sigma^2)}. \end{aligned} \quad (18)$$

Under Assumption 1, it is straightforward to show from (18) that $f'''(x) < 0$ in $(\alpha - \sqrt{3}\sigma, \alpha]$. If $x^s \geq \alpha - \sqrt{3}\sigma$, we are done. Otherwise, it remains to show that $f'''(x) < 0$ in $(x^s, \alpha - \sqrt{3}\sigma]$.

Let $h_1(x)$ and $h_2(x)$ denote the first and the second terms of $f'''(x)$ on the right-hand side of (18), i.e., $f'''(x) = h_1(x) - h_2(x)$. We can express $h_1(x)/h_2(x)$ as the product of two quantities

$$\frac{h_1(x)}{h_2(x)} = \frac{1}{\lambda} \frac{(x - \alpha) [(x - \alpha)^2 - 3\sigma^2]}{(x - \eta) [(x - \eta)^2 - 3\sigma^2]} \times \frac{e^{-((x-\alpha)^2/2\sigma^2)}}{e^{-((x-\eta)^2/2\sigma^2)}}$$

whose derivatives are shown in (19) at the bottom of the next page.

We now show that the right-hand side of (19) is negative in $(x^s, \alpha - \sqrt{3}\sigma]$. Define $h_3(x)$ as follows:

$$h_3(x) = (x - \eta) [(x - \eta)^2 - 3\sigma^2] [3(x - \alpha)^2 - 3\sigma^2] - (x - \alpha) [(x - \alpha)^2 - 3\sigma^2] [3(x - \eta)^2 - 3\sigma^2].$$

Note that $h_3(x)$ has the same sign as the right-hand side of (19) since the denominator is always positive. For $x \in (x^s, \alpha - \sqrt{3}\sigma)$, it is easy to verify that all the quantities in square brackets are positive. Using the fact that $x - \eta < x - \alpha$, we can bound $h_3(x)$ as follows:

$$\begin{aligned} h_3(x) &< (x - \alpha) [(x - \eta)^2 - 3\sigma^2] [3(x - \alpha)^2 - 3\sigma^2] \\ &\quad - (x - \alpha) [(x - \alpha)^2 - 3\sigma^2] [3(x - \eta)^2 - 3\sigma^2] \\ &= (x - \alpha) \times 6\sigma^2 [(x - \eta)^2 - (x - \alpha)^2] < 0, \\ &\quad x \in (x^s, \alpha - \sqrt{3}\sigma), \end{aligned}$$

where the last inequality follows from the fact that $0 < x < \alpha < \eta$ in $(x^s, \alpha - \sqrt{3}\sigma)$. At $x = \alpha - \sqrt{3}\sigma$, using the fact that $\alpha < \eta$,

$$\begin{aligned} h_3(\alpha - \sqrt{3}\sigma) &= (\alpha - \sqrt{3}\sigma - \eta) \\ &\quad \cdot \left[(\alpha - \sqrt{3}\sigma - \eta)^2 - 3\sigma^2 \right] \left[3(\sqrt{3}\sigma)^2 - 3\sigma^2 \right] < 0. \end{aligned}$$

Thus, $h_3(x) < 0$ in $(x^s, \alpha - \sqrt{3}\sigma]$. It follows that $h_1(x)/h_2(x)$, being a product of two strictly decreasing functions, is strictly decreasing in $(x^s, \alpha - \sqrt{3}\sigma]$. Consequently, there can be at most one point in $(x^s, \alpha - \sqrt{3}\sigma]$ such that $f'''(x) = 0$ ($h_1(x)/h_2(x) = 1$). We now show that $f'''(x^s) < 0$ ($h_1(x^s)/h_2(x^s) < 1$), which implies that $f'''(x) < 0$ ($h_1(x)/h_2(x) < 1$) in $(x^s, \alpha - \sqrt{3}\sigma]$ since $h_1(x)/h_2(x)$ is decreasing.

To show that $f'''(x^s) < 0$, we rewrite and bound (18) as follows

$$\begin{aligned} f'''(x) &= -\frac{x - \alpha}{\sigma^2} \left[\frac{(x - \alpha)^2 - \sigma^2}{\sqrt{2\pi}\sigma^5} \right] e^{-((x - \alpha)^2/2\sigma^2)} \\ &\quad + \lambda \frac{x - \eta}{\sigma^2} \left[\frac{(x - \eta)^2 - \sigma^2}{\sqrt{2\pi}\sigma^5} \right] e^{-((x - \eta)^2/2\sigma^2)} \\ &\quad + \frac{2}{\sigma^2} \frac{x - \alpha}{\sqrt{2\pi}\sigma^3} e^{-((x - \alpha)^2/2\sigma^2)} \\ &\quad - \lambda \frac{2}{\sigma^2} \frac{x - \eta}{\sqrt{2\pi}\sigma^3} e^{-((x - \eta)^2/2\sigma^2)} \\ &< -\frac{x - \eta}{\sigma^2} f''(x) - \frac{2}{\sigma^2} f'(x) \end{aligned}$$

where we have used (14) and (15), and the fact that $x - \alpha > x - \eta$ to construct the inequality. By definition, $f''(x^s) = 0$. Since

$x^s < x^{\max}$, we have from Lemma A1 part iv) that $f'(x^s) > 0$. It follows from the above expression that $f'''(x^s) < 0$. \square

Lemma 1: Given $\lambda > 0$, and that Assumption 1 holds, there is at most one positive solution to the equation $A'(J) = \lambda B'(J)$ in $(0, \eta - \sigma)$.

Proof: In what follows, we use x in place of J as a function argument. Using $f(x)$ defined in (13), we can express $A'(x) - \lambda B'(x)$ as

$$\begin{aligned} A'(x) - \lambda B'(x) &= \frac{1}{\sqrt{2\pi}\sigma^2} \left[e^{-((x - \alpha)^2/2\sigma^2)} - \lambda e^{-((x - \eta)^2/2\sigma^2)} \right] \\ &\quad - \frac{1}{\sqrt{2\pi}\sigma^2} \left[e^{-((x + \alpha)^2/2\sigma^2)} - \lambda e^{-((x + \eta)^2/2\sigma^2)} \right] \\ &= f(x) - f(-x). \end{aligned}$$

Based on the results of lemma A1, let x^o , x^{\max} and x^{\min} denote the solution (zero crossing), the local maximum and the local minimum of $f(x)$, respectively.

Let x^* denote the smallest positive intersection of $f(x^*)$ and $-f(-x^*)$, i.e., $f(x^*) - f(-x^*) = 0$. Note that x^* may or may not exist. For convenience, let $f_-(x)$ denote $f(-x)$. We shall show that there is at most one intersection of $f(x)$ and $f_-(x)$ in $(0, \eta - \sigma)$ in three separate cases which cover all the possibilities, as shown in Fig. 6.

i) $x^o \leq 0$: We shall show that no intersection of $f(x)$ and $f_-(x)$ exists in $(0, x^{\min})$. Since $x^{\min} > \eta$ [Lemma A1 part iii)], the result will imply that there is no intersection in $(0, \eta - \sigma)$.

Consider two separate cases: $-x^o \geq x^{\min}$ and $-x^o < x^{\min}$. When $-x^o \geq x^{\min}$, we have $f(0) = f_-(0)$ and $f(x)$ is decreasing while $f_-(x)$ is increasing in $(0, x^{\min})$. Therefore, there is no intersection in $(0, x^{\min})$. When $-x^o < x^{\min}$, we have that $f(0) = f_-(0)$ and $f(x)$ is decreasing while $f_-(x)$ is increasing in $(0, -x^o)$. Therefore, there cannot be an intersection in $(0, -x^o)$. In $[-x^o, x^{\min})$, $f_-(x)$ is nonnegative while $f(x)$ is strictly negative. Therefore, there cannot be an intersection in $(0, x^{\min})$. In conclusion, there is no intersection in $(0, x^{\min})$.

ii) $x^o > 0$, $x^{\max} \leq 0$: First note that there is no intersection of $f(x)$ and $f_-(x)$ in $(0, -x^{\max}]$ since $f(0) = f_-(0) > 0$ and $f(x)$ remains below $f(0)$ [$f(x)$ decreases from a positive value $f(0)$ to a negative value $f(x^{\min})$ and remains negative as $x \rightarrow \infty$] while $f_-(x)$ is increasing in $(0, -x^{\max}]$.

Consider three separate intervals: $(0, \alpha]$, $(\alpha, \alpha + \sigma]$, and $(\alpha + \sigma, \eta - \sigma)$. We shall show that there is no intersection in any of them.

• Consider the interval $(0, \alpha]$. If $-x^{\max} \geq \alpha$, we are done since there is no intersection in $(0, -x^{\max}]$. When

$$\frac{d}{dx} \left(\frac{e^{-((x - \alpha)^2/2\sigma^2)}}{e^{-((x - \eta)^2/2\sigma^2)}} \right) = -\frac{\eta - \alpha}{\sigma^2} e^{-(2x(\eta - \alpha) + \alpha^2 - \eta^2)/2\sigma^2} < 0,$$

$$\frac{d}{dx} \left(\frac{1}{\lambda} \frac{(x - \alpha)[(x - \alpha)^2 - 3\sigma^2]}{(x - \eta)[(x - \eta)^2 - 3\sigma^2]} \right) = \frac{(x - \eta)[(x - \eta)^2 - 3\sigma^2] [3(x - \alpha)^2 - 3\sigma^2] - (x - \alpha)[(x - \alpha)^2 - 3\sigma^2] [3(x - \eta)^2 - 3\sigma^2]}{\lambda(x - \eta)^2 [(x - \eta)^2 - 3\sigma^2]^2} \quad (19)$$

$-x^{\max} < \alpha$, the intersection, if exists, must lie in $(-x^{\max}, \alpha]$.

Let x^s denote the unique saddle point in $(-\infty, x^{\max})$. Since $x^{\max} < \alpha$ [Lemma A1 part iii)], there exists a point $x^{\max} + \delta$ with $\delta > 0$ such that $x^{\max} + \delta \leq \alpha$. We have that

$$\begin{aligned} |f'(x^{\max} + \delta)| &= - \int_{x^{\max}}^{x^{\max} + \delta} f''(x) dx \\ &> - \int_{x^{\max} - \delta}^{x^{\max}} f''(x) dx \\ &= |f'(x^{\max} - \delta)| \end{aligned} \quad (20)$$

where the inequality holds in the case $x^{\max} - \delta > x^s$ since $f''(x)$ is negative and is decreasing (Lemma A2) in $(x^s, \alpha]$; and in the case $x^{\max} - \delta < x^s$ since $f''(x)$ is non-negative in $(x^{\max} - \delta, x^s]$ and is negative and decreasing in $(x^s, \alpha]$.

Let $\xi \in (-x^{\max}, \alpha]$ and $\delta' = x^{\max} + \xi$. We have that

$$\begin{aligned} f(\xi) - f_-(\xi) &= \int_{x^{\max}}^{\xi} f'(x) dx + \int_{-\xi}^{x^{\max}} f'(x) dx \\ &= \int_{x^{\max} - \delta'}^{x^{\max}} f'(x) dx + \int_{x^{\max}}^{x^{\max} + \delta'} f'(x) dx \\ &\quad + \int_{x^{\max} + \delta'}^{\xi} f'(x) dx \\ &= \int_{x^{\max} - \delta'}^{x^{\max}} |f'(x)| dx - \int_{x^{\max}}^{x^{\max} + \delta'} |f'(x)| dx \\ &\quad + \int_{x^{\max} + \delta'}^{\xi} f'(x) dx \\ &< 0 + \int_{x^{\max} + \delta'}^{\xi} f'(x) dx < 0 \end{aligned} \quad (21)$$

where the first inequality is a consequence of (20), and the second inequality follows from Lemma A1 parts iii) and v) that $f'(x) < 0$ in $(x^{\max}, \alpha]$. Equation (21) shows there is no intersection in $(-x^{\max}, \alpha]$. In conclusion, there is no intersection in $(0, \alpha]$.

For later arguments, we now show that if $-x^{\max} < \alpha$, then $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < f'_-(\alpha)$.

Given $-x^{\max} < \alpha$, (21) tells us that $f(\alpha) < f_-(\alpha)$. Let $\delta > 0$ be such that $x^{\max} - \delta = -\alpha$, then (20) tells us that $|f'_-(\alpha)| = |f'(-\alpha)| < |f'(x^{\max} + \delta)|$. Since $f(x)$ is strictly decreasing in (x^{\max}, α) [Lemma A1 parts iii) and v)], we have $f'(x^{\max} + \delta) < f'_-(\alpha)$. Finally, since $f(x)$ is strictly concave in (x^{\max}, α) (Lemma A2) and $x^{\max} + \delta < \alpha$, we have $f'(\alpha) < f'(x^{\max} + \delta) < f'_-(\alpha)$.

For convenience, we define $f(x) = f_\alpha(x) + f_\eta(x)$, where

$$\begin{aligned} f_\alpha(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-((x-\alpha)^2/2\sigma^2)} \\ f_\eta(x) &= -\lambda \frac{1}{\sqrt{2\pi\sigma^2}} e^{-((x-\eta)^2/2\sigma^2)}. \end{aligned} \quad (22)$$

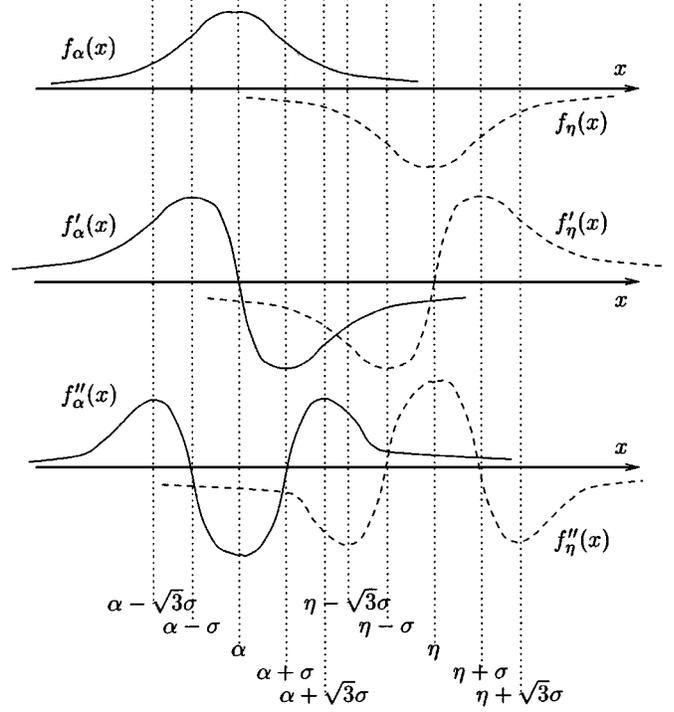


Fig. 7. $f_\alpha(x)$ and $f_\eta(x)$ as defined in (22).

Fig. 7 shows example curves of $f_\alpha(x)$ and $f_\eta(x)$ together with their derivatives.

- Consider the interval $(\alpha, \alpha + \sigma]$. If $-x^{\max} \geq \alpha + \sigma$, there is no intersection in $(0, \alpha + \sigma]$ since there is no intersection in $(0, -x^{\max}]$. So we consider the case when $-x^{\max} < \alpha + \sigma$.

Under Assumption 1, $f(x)$ is strictly decreasing and strictly concave [$f'(x) < 0$ and $f''(x) < 0$] in $(\alpha, \alpha + \sigma]$ since $f_\alpha(x)$ and $f_\eta(x)$ both have negative first and second derivatives (see Fig. 7). On the other hand, the contribution of $f_\alpha(x)$ to $f_-(x)$ is decreasing but convex, while the contribution of $f_\eta(x)$ is increasing. An additional condition that $f'(\alpha) < f'_-(\alpha)$ will lead to the conclusion that $f(x)$ is decreasing faster than $f_-(x)$ at every point in $(\alpha, \alpha + \sigma]$.

For $-x^{\max} < \alpha$, we have shown that $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < f'_-(\alpha)$. For $-x^{\max} > \alpha$, we have $f(\alpha) < 0 < f_-(\alpha)$, and $f'(\alpha) < 0 < f'_-(\alpha)$. In both cases, we have $f(\alpha) < f_-(\alpha)$ and $f(x)$ decreases faster than $f_-(x)$ at every point in $(\alpha, \alpha + \sigma]$. Therefore, there is no intersection in $(\alpha, \alpha + \sigma]$.

Consequently, given that $-x^{\max} < \alpha + \sigma$, we have $f(\alpha + \sigma) < f_-(\alpha + \sigma)$ and $f'(\alpha + \sigma) < f'_-(\alpha + \sigma)$.

- Consider the interval $(\alpha + \sigma, \eta - \sigma)$. If $-x^{\max} \geq \eta - \sigma$, there is no intersection in $(0, \eta - \sigma)$ since there is no intersection in $(0, -x^{\max}]$. So we consider the case when $-x^{\max} < \eta - \sigma$.

In $(\alpha + \sigma, \eta - \sigma)$, the contributions of $f_\alpha(x)$ to $f(x)$ and $f_-(x)$ are both decreasing and convex (see Fig. 7). However, at any particular point $x \in (\alpha + \sigma, \eta - \sigma)$, the contribution of $f_\alpha(x)$ to $f(x)$ decreases faster than its contribution to $f_-(x)$. On the other hand, the contribution of $f_\eta(x)$ to $f(x)$ is decreasing and concave, while its

contribution to $f_-(x)$ is increasing. An additional condition that $f'(\alpha+\sigma) < f'_-(\alpha+\sigma)$ will lead to the conclusion that $f(x)$ is decreasing faster than $f_-(x)$ at every point in $(\alpha + \sigma, \eta - \sigma)$.

For $-x^{\max} < \alpha + \sigma$, we have shown that $f(\alpha + \sigma) < f_-(\alpha + \sigma)$ and $f'(\alpha + \sigma) < f'_-(\alpha + \sigma)$. For $-x^{\max} > \alpha + \sigma$, we have $f(\alpha + \sigma) < 0 < f_-(\alpha + \sigma)$ and $f'(\alpha + \sigma) < 0 < f'_-(\alpha + \sigma)$. In both cases, we have that $f(\alpha + \sigma) < f_-(\alpha + \sigma)$ and $f(x)$ decreases faster than $f_-(x)$ at every point in $(\alpha + \sigma, \eta - \sigma)$. Therefore, there cannot be an intersection in $(\alpha + \sigma, \eta - \sigma)$.

iii) $x^o > 0, x^{\max} > 0$: First note that there is no intersection of $f(x)$ and $f_-(x)$ in $(0, x^{\max}]$ since $f(0) = f_-(0)$ and $f(x)$ is increasing while $f_-(x)$ is decreasing in $(0, x^{\max}]$. Consider two separate intervals: (x^{\max}, α) and $[\alpha, \eta - \sigma)$.

iii) $x^o > 0, x^{\max} > 0$: First note that there is no intersection of $f(x)$ and $f_-(x)$ in $(0, x^{\max}]$ since $f(0) = f_-(0)$ and $f(x)$ is increasing while $f_-(x)$ is decreasing in $(0, x^{\max}]$. Consider two separate intervals: (x^{\max}, α) and $[\alpha, \eta - \sigma)$.

- We now show that there can be at most one intersection in (x^{\max}, α) . Suppose there is an intersection denoted by x^* . We now argue that $f'(x^*) < f'_-(x^*)$. The intersection must occur in (x^{\max}, x^o) since $f(x) < 0$ and $f_-(x) \geq 0$ in $[x^o, \infty)$. In (x^{\max}, x^o) , both $f(x)$ and $f_-(x)$ are decreasing. Since $f(x^{\max}) > f_-(x^{\max})$, for $f(x)$ and $f_-(x)$ to intersect, $f(x)$ has to decrease faster than $f_-(x)$ at x^* .

Consider a point $x \in (x^*, \alpha)$ and let $\epsilon = x - x^*$. We want to show $f'(x) < f'_-(x)$ as follows:

$$\begin{aligned}
& f'(x) - f'_-(x) \\
&= f'(x^*) + \int_{x^*}^{x^*+\epsilon} f''(x) dx - f'_-(x^*) \\
&\quad - \int_{x^*}^{x^*+\epsilon} f''_-(x) dx \\
&= f'(x^*) + \int_{x^*}^{x^*+\epsilon} f''(x) dx - f'_-(x^*) \\
&\quad - \int_{-x^*}^{-x^*-\epsilon} f''(x) dx \\
&< 0 + \int_{x^*}^{x^*+\epsilon} f''(x) dx \\
&\quad - \int_{-x^*}^{-x^*-\epsilon} f''(x) dx \tag{23} \\
&< 0. \tag{24}
\end{aligned}$$

Equation (23) follows from $f'(x^*) < f'_-(x^*)$. It remains to prove (24).

Let x^s denote the unique saddle point in $(-\infty, x^{\max})$ [Lemma A1 part iv)]. If $x^s < -x^* - \epsilon$, then (24) holds as a consequence of Lemma A2. If $-x^* - \epsilon \leq x^s < -x^*$, we have that

$$\int_{-x^*-\epsilon}^{x^s} f''(x) dx > 0$$

$$\begin{aligned}
& \int_{x^s}^{-x^*} f''(x) dx > \int_{x^s}^{-x^*} f''(x) dx \\
& \int_{-x^s}^{x^*+\epsilon} f''(x) dx < 0
\end{aligned}$$

where the first inequality holds since $f(x)$ is convex in $(-\infty, x^s)$. The second inequality holds as a consequence of Lemma A2. The last inequality holds since $f(x)$ is concave in $(x^s, \alpha) \supset (-x^s, \alpha)$. Combining the above inequalities, (24) holds.

If $-x^* < x^s$, we have that

$$\int_{-x^*-\epsilon}^{-x^*} f''(x) dx > 0 \quad \int_{x^*}^{x^*+\epsilon} f''(x) dx < 0$$

where the first inequality holds since $f(x)$ is convex in $(-x^* - \epsilon, -x^*)$ [Lemma A1 part iv)]. The second inequality holds since $\alpha > x^* > x^{\max} > x^s$ and thus $f(x)$ is concave in $(x^*, x^* + \epsilon)$ (Lemma A2). Combining the above inequalities, (24) holds.

Therefore, after the intersection point x^* , $f(x)$ decreases faster than $f_-(x)$ at all points in (x^*, α) . It follows that there can be at most one intersection point in (x^{\max}, α) . In addition, if an intersection exists in (x^{\max}, α) , then $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < f'_-(\alpha)$.

- Consider the interval $[\alpha, \eta - \sigma)$. We shall show that, given an intersection of $f(x)$ and $f_-(x)$ in $(0, \alpha)$, there can be no additional intersection in $[\alpha, \eta - \sigma)$. In addition, given no intersection in $(0, \alpha)$, there is at most one intersection in $[\alpha, \eta - \sigma)$.

We first show that $-x^s < \alpha$, i.e., $f_-(x)$ is convex in $[\alpha, \infty)$. Since x^s is the only saddle point in $(-\infty, x^{\max})$ where $f(x)$ changes from convexity to concavity [Lemma A1 part iv)], it is sufficient to show that $f''(-\alpha) > 0$. From (15),

$$\begin{aligned}
& f''(-\alpha) \\
&= \frac{1}{\sqrt{2\pi}\sigma^5} \left[(4\alpha^2 - \sigma^2)e^{-(4\alpha^2/2\sigma^2)} \right. \\
&\quad \left. - \lambda(\alpha^2 + \eta^2 + 2\alpha\eta - \sigma^2) \right. \\
&\quad \left. \cdot \exp\left(-\frac{\alpha^2 + \eta^2 + 2\alpha\eta}{2\sigma^2}\right) \right] \\
&> \frac{1}{\sqrt{2\pi}\sigma^5} \left[(4\alpha^2 - \sigma^2)e^{-(4\alpha^2/2\sigma^2)} - \frac{\alpha}{\eta} \right. \\
&\quad \left. \cdot \exp\left(-\frac{\alpha^2}{2\sigma^2} + \frac{\eta^2}{2\sigma^2}\right) \right. \\
&\quad \left. \cdot (\alpha^2 + \eta^2 + 2\alpha\eta - \sigma^2) \right. \\
&\quad \left. \cdot \exp\left(-\frac{\alpha^2 + \eta^2 + 2\alpha\eta}{2\sigma^2}\right) \right] \\
&= \frac{e^{-(2\alpha^2/2\sigma^2)}}{\sqrt{2\pi}\sigma^5} \left(2\alpha^2 \left[e^{-(2\alpha^2/2\sigma^2)} - e^{-(2\alpha\eta/2\sigma^2)} \right] \right. \\
&\quad \left. + \left[\alpha^2 e^{-(2\alpha^2/2\sigma^2)} - \alpha\eta e^{-(2\alpha\eta/2\sigma^2)} \right] + (\alpha^2 - \sigma^2) \right. \\
&\quad \left. \left[e^{-(2\alpha^2/2\sigma^2)} - \frac{\alpha}{\eta} e^{-(2\alpha\eta/2\sigma^2)} \right] \right) > 0. \tag{25}
\end{aligned}$$

The first inequality holds since $x^{\max} > 0$, and thus $f'(0) > 0$ [Lemma A1 part iv)], from which the expression for $f'(0)$ in (14) yields $\lambda < (\alpha/\eta)e^{-(\alpha^2/2\sigma^2)+(\eta^2/2\sigma^2)}$. The second inequality results from the fact that all quantities in square brackets of (25) are positive and $\alpha > \sigma$ by assumption 1. Since $\alpha < \eta$, the quantities in the first and the third square brackets are positive. The quantity in the second square bracket is positive since $xe^{-(x/\sigma^2)}$ is decreasing in (σ^2, ∞) .

Lemma A1 parts iii) and vii) tell us that $[\alpha, \eta - \sigma) \subset (x^{\max}, x^{\min})$, and there is a single saddle point in (x^{\max}, x^{\min}) . Thus, in $[\alpha, \eta - \sigma)$, there is at most one saddle point, which we denote by x^t when exists. If x^t exists, we consider two separate intervals $[\alpha, x^t)$ and $[x^t, \eta - \sigma)$. Otherwise, we consider the interval $[\alpha, \eta - \sigma)$. We show that in each interval, the following two properties hold.

- 1) If an intersection occurs before the interval, there is no additional intersection in the interval.
- 2) If no intersection occurs before the interval, there is at most one intersection in the interval.

We prove the two properties below. Keep in mind that $f(x)$ and $f_-(x)$ are decreasing in $[\alpha, \eta - \sigma)$.

— Consider the first interval: $[\alpha, x^t)$ (or $[\alpha, \eta - \sigma)$ if x^t does not exist). We now show that, given an intersection before α , there is no additional intersection in $[\alpha, x^t)$.

Let $x \in [\alpha, x^t)$ and $\epsilon = x - \alpha$. We show $f'(x) < f'_-(x)$ below

$$\begin{aligned} f'(x) - f'_-(x) &= f'(\alpha) + \int_{\alpha}^{\alpha+\epsilon} f''(x) dx \\ &\quad - f'_-(\alpha) - \int_{\alpha}^{\alpha+\epsilon} f''_-(x) dx \\ &< \int_{\alpha}^{\alpha+\epsilon} f''(x) dx - \int_{\alpha}^{\alpha+\epsilon} f''_-(x) dx < 0. \end{aligned}$$

The first inequality holds since we have shown that if an intersection x^* exists and $x^* < \alpha$, then $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < f'_-(\alpha)$. The second inequality holds since $f''(\alpha) < 0$ [$f''(x^{\max}) < 0$ and Lemma A2] and thus $f(x)$ is concave in $[\alpha, x^t)$, while $f_-(x)$ is convex in $[\alpha, \infty)$. Since $f(\alpha) < f_-(\alpha)$ and $f(x)$ decreases faster than $f_-(x)$ at every point in $[\alpha, x^t)$, there is no additional intersection in $[\alpha, x^t)$. In addition, since $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < f'_-(\alpha)$, we have that $f(x^t) < f_-(x^t)$ and $f'(x^t) < f'_-(x^t)$.

If there is no saddle point in $[\alpha, \eta - \sigma)$, the facts that $f'(\alpha) < f'_-(\alpha)$ and $f(x)$ is concave while $f_-(x)$ is convex imply that $f(x)$ decreases faster than $f_-(x)$ at every point in $[\alpha, \eta - \sigma)$, and no additional intersection is possible in $[\alpha, \eta - \sigma)$.

Given that there is no intersection x^* such that $x^* < \alpha$, we now show there is at most one intersection in $[\alpha, x^t)$. As a reminder, we have in this

case that $f(\alpha) > f_-(\alpha)$ since $f(0) = f_-(0)$, $f'(0) > 0 > f'_-(0)$, and no intersection has occurred before α .

Let x^+ denote the smallest point of intersection, if exists, in $[\alpha, x^t)$. Since $f(\alpha) > f_-(\alpha)$ and both $f(x)$ and $f_-(x)$ are decreasing, for $f(x)$ and $f_-(x)$ to intersect, we must have $f'(x^+) < f'_-(x^+)$.

We argue that $f(x)$ decreases faster than $f_-(x)$ at every point in (x^+, x^t) since $f'(x^+) < f'_-(x^+)$ and $f(x)$ is concave while $f_-(x)$ is convex. Therefore, there can be no additional intersection in (x^+, x^t) . Moreover, given that x^+ exists in $[\alpha, x^t)$, we have that $f(x^t) < f_-(x^t)$ and $f'(x^t) < f'_-(x^t)$.

If there is no saddle point in $[\alpha, \eta - \sigma)$, we can let x^+ denote the smallest intersection point in $[\alpha, \eta - \sigma)$. The facts that $f'(x^+) < f'_-(x^+)$ and $f(x)$ is concave while $f_-(x)$ is convex imply that $f(x)$ decreases faster than $f_-(x)$ at every point in $(x^+, \eta - \sigma)$ and no additional intersection is possible in $(x^+, \eta - \sigma)$.

— Consider the interval $[x^t, \eta - \sigma)$ (this case does not exist if x^t does not exist). We now show that, given an intersection $x^* < x^t$, there is no additional intersection in $[x^t, \eta - \sigma)$.

In $[x^t, \eta - \sigma)$, $f(x)$ is decreasing and convex (by the definition of x^t). Similarly, $f_-(x)$ is decreasing and convex. However, in $[x^t, \eta - \sigma)$, the contribution of $f_{\alpha}(x)$ to $f(x)$ is decreasing faster than its contribution to $f_-(x)$ (see Fig. 7). On the other hand, the contribution of $f_{\eta}(x)$ to $f(x)$ is decreasing while its contribution to $f_-(x)$ is increasing. Therefore, $f(x)$ decreases faster than $f_-(x)$ at every point in $[x^t, \eta - \sigma)$. Since we have shown that $f(x^t) < f_-(x^t)$ when there is an intersection x^* such that $x^* < x^t$, it follows that there is no additional intersection in $[x^t, \eta - \sigma)$.

Given that there is no intersection x^* such that $x^* < x^t$, we now show that there is at most one intersection in $[x^t, \eta - \sigma)$. As a reminder, we have in this case that $f(x^t) > f_-(x^t)$ since $f(0) = f_-(0)$, $f'(0) > 0 > f'_-(0)$, and no intersection has occurred before x^t .

Let x^+ denote the smallest point of intersection, if exists, in $[x^t, \eta - \sigma)$. Since $f(x^t) > f_-(x^t)$ and both $f(x)$ and $f_-(x)$ are decreasing, for $f(x)$ and $f_-(x)$ to intersect, we must have $f'(x^+) < f'_-(x^+)$.

We argue that $f(x)$ decreases faster than $f_-(x)$ at every point in $(x^+, \eta - \sigma)$ since $f'(x^+) < f'_-(x^+)$ and we have argued that the overall contribution of $f_{\alpha}(x)$ and $f_{\eta}(x)$ to $f(x)$ is decreasing faster than the overall contribution of $f_{\alpha}(x)$ and $f_{\eta}(x)$ to $f_-(x)$ in $[x^t, \eta - \sigma) \supset (x^+, \eta - \sigma)$. Therefore, there can be no additional intersection in $(x^+, \eta - \sigma)$.

We conclude that, given an intersection before α , there is no additional intersection in $[\alpha, \eta - \sigma)$. On the other hand, given no intersection before α , there is at most one intersection in $[\alpha, \eta - \sigma)$.

Combining the results for the intervals $(0, x^{\max}]$, (x^{\max}, α) , and $[\alpha, \eta - \sigma)$, we conclude that there is at most one intersection in $(0, \eta - \sigma)$. \square

Lemma 2: Under Assumption 1, $|H|_{(J^s, 0)} = |H|_{(0, J^s)} < 0$ and $|H|_{(\tilde{J}, \tilde{J})} > 0$, for all $\tilde{B} < (1/2)B(0) + (1/2)B(\eta - \sigma)$.

Proof: We first show that $|H|_{(J^s, 0)} = |H|_{(0, J^s)} < 0$ for all $J^s > 0$. From (8)

$$|H|_{(J^s, 0)} = -\frac{1}{2} \left[A''(0) - \frac{A'(J^s)}{B'(J^s)} B''(0) \right] (B'(J^s))^2.$$

It suffices to show that $A'(x)/B'(x) < A''(0)/B''(0)$ for all $x > 0$ as follows:

$$\begin{aligned} \frac{A'(x)}{B'(x)} &= \frac{e^{-(x-\alpha)^2/2\sigma^2} - e^{-(x+\alpha)^2/2\sigma^2}}{e^{-(x-\eta)^2/2\sigma^2} - e^{-(x+\eta)^2/2\sigma^2}} \\ &= \frac{\alpha e^{-(\alpha^2/2\sigma^2)} \left[\frac{\sigma^2}{\alpha} e^{x\alpha/\sigma^2} - \frac{\sigma^2}{\alpha} e^{-(x\alpha/\sigma^2)} \right]}{\eta e^{-(\eta^2/2\sigma^2)} \left[\frac{\sigma^2}{\eta} e^{x\eta/\sigma^2} - \frac{\sigma^2}{\eta} e^{-(x\eta/\sigma^2)} \right]} \\ &< \frac{\alpha e^{-(\alpha^2/2\sigma^2)}}{\eta e^{-(\eta^2/2\sigma^2)}} = \frac{A''(0)}{B''(0)} \end{aligned}$$

where the inequality follows from the facts that $\alpha < \eta$, and $g(y) = (1/y)e^{xy} - (1/y)e^{-xy}$ is increasing in $(0, \infty)$ for all $x > 0$. It is straightforward to verify the last equality from (4).

We now show that $|H|_{(\tilde{J}, \tilde{J})} > 0$, for all $\tilde{J} < \eta - \sigma$. From (8)

$$|H|_{(\tilde{J}, \tilde{J})} = - \left[A''(\tilde{J}) - \frac{A'(\tilde{J})}{B'(\tilde{J})} B''(\tilde{J}) \right] (B'(\tilde{J}))^2. \quad (26)$$

Define $G(x) = (A'(x)/B'(x))B''(x) - A''(x)$. It suffices to show that $G(x) < 0$ in $(0, \eta - \sigma)$. Given an intersection between the curves $A'(x)$ and $\lambda B'(x)$ at $x^* \in (0, \eta - \sigma)$, we have $\lambda = A'(x^*)/B'(x^*)$. We can therefore think of $(A'(x^*)/B'(x^*))B''(x^*)$ as the slope of the $\lambda B'(x)$ curve and $A''(x^*)$ as the slope of the $A'(x)$ curve at x^* .

We have shown that $\lambda = A'(x^*)/B'(x^*) < A''(0)/B''(0)$, or equivalently $\lambda B''(0) < A''(0)$. Since $\lambda B'(0) = A'(0) = 0$, the condition $\lambda B''(0) < A''(0)$ yields $\lambda B'(0^+) < A'(0^+)$, where $h(0^+)$ represents $\lim_{x \rightarrow 0^+} h(x)$. The facts that x^* is the only intersection of $\lambda B'(x)$ and $A'(x)$ in $(0, \eta - \sigma)$ (Lemma 1) and $\lambda B'(0^+) < A'(0^+)$ imply that at x^* the slope of $\lambda B'(x)$ [equal to $(A'(x^*)/B'(x^*))B''(x^*)$] is larger than the slope of $A'(x)$ [equal to $A''(x^*)$] since $\lambda B'(x)$ intersects $A'(x)$ from below. Equivalently, $G(x^*) < 0$ for all $x^* \in (0, \eta - \sigma)$.

The condition $\tilde{B} < (1/2)B(0) + (1/2)B(\eta - \sigma)$ ensures that $J^s < \eta - \sigma$ and $\tilde{J} < \eta - \sigma$. Thus, given the condition, it follows that $|H|_{(J^s, 0)} = |H|_{(0, J^s)} < 0$ and $|H|_{(\tilde{J}, \tilde{J})} > 0$. \square

Lemma A3: Let $1/2 < \gamma < 1$, and $\tilde{B} < \gamma B(\eta - \sigma) + (1-\gamma)B(0)$. Under Assumption 1, in the modified optimization problem with $M = 1$ and $T = 2$ given below

$$\begin{aligned} &\text{minimize} \quad \gamma A(J_1) + (1-\gamma)A(J_2) \\ &\text{subject to} \quad \tilde{B} = \gamma B(J_1) + (1-\gamma)B(J_2) \end{aligned}$$

there are extrema of the forms (\tilde{J}, \tilde{J}) and $(J^s, 0)$, where $B(\tilde{J}) = \tilde{B}$ and $\gamma B(J^s) + (1-\gamma)B(0) = \tilde{B}$. (\tilde{J}, \tilde{J}) corresponds to the global maximum, while $(J^s, 0)$ corresponds to a local minimum and yields a lower cost than (\tilde{J}, \tilde{J}) .

Proof: Similar to the work in Section IV-A, we only consider nonnegative values of J_1 and J_2 . The condition $\tilde{B} < \gamma B(\eta - \sigma) + (1-\gamma)B(0)$ ensures that both J_1 and J_2 are in $[0, \eta - \sigma)$. By differentiating the Lagrangian with respect to J_i ,

we have the same conditions for an extremum as in problem (6), namely $A'(J_i) = \lambda B'(J_i)$, $i = 1, 2$. Lemma 1 tells us that an extremum is one of the following: (\tilde{J}, \tilde{J}) , $(J^s, 0)$ and $(0, J_2^s)$, where \tilde{J} and J^s are defined above and $\gamma B(0) + (1-\gamma)B(J_2^s) = \tilde{B}$. It remains to investigate the bordered Hessian determinant $|H|_{(J_1, J_2)}$.

For the extremum (\tilde{J}, \tilde{J}) , $|H|_{(J_1, J_2)}$ is exactly the same as (26) and is thus positive for all $\tilde{J} < \eta - \sigma$. Thus, (\tilde{J}, \tilde{J}) is a local maximum. For the extremum $(J^s, 0)$

$$|H|_{(J_1, J_2)} = -\gamma \left[A''(0) - \frac{A'(J^s)}{B'(J^s)} B''(0) \right] (B'(J^s))^2.$$

For the extremum $(0, J_2^s)$

$$|H|_{(J_1, J_2)} = -(1-\gamma) \left[A''(0) - \frac{A'(J_2^s)}{B'(J_2^s)} B''(0) \right] (B'(J_2^s))^2.$$

In either case, we have $|H|_{(J_1, J_2)} < 0$ as shown in the proof of Lemma 2. Thus $(J^s, 0)$ and $(0, J_2^s)$ corresponds to local minima.

Similarly to the arguments used to construct theorem 1, since the constraint set is a contour curve of $\gamma B(J_1) + (1-\gamma)B(J_2)$ in the first quadrant of (J_1, J_2) plane and connects the two end points $(J^s, 0)$ and $(0, J_2^s)$, the absence of an extremum elsewhere together with the continuity of cost function imply that (\tilde{J}, \tilde{J}) is a global maximum and thus has a higher cost than $(J^s, 0)$. \square

Lemma 3: Let J_t^s be the coherent crosstalk signal associated with scenario t in which t out of T bits are equally affected and BER is equal to \tilde{B} . Let A_t be the expected total number of device alarms in an observation period corresponding to scenario t . Under Assumption 1, given $M = 1$, $T > 1$, and $\tilde{B} < (1/T)B(\eta - \sigma) + ((T-1)/T)B(0)$, we have that $A_T > A_{T-1} > \dots > A_1$.

Proof: The condition $\tilde{B} < (1/T)B(\eta - \sigma) + ((T-1)/T)B(0)$ guarantees that $J_i \in [0, \eta - \sigma)$ for all i , $1 \leq i \leq T$. We shall prove the lemma by induction. First, compare the scenarios T and $T-1$ yielding the expected total number of device alarms A_T and A_{T-1} respectively. Notice that comparing these two scenarios is equivalent to comparing the two extrema in the modified optimization problem of lemma A3 with $\gamma = (T-1)/T$. The two extrema are (J_T^s, J_T^s) for scenario T and $(J_{T-1}^s, 0)$ for scenario $T-1$. Lemma A3 tells us $A_T > A_{T-1}$.

Now compare scenarios $T-j$ and $T-j-1$, $1 \leq j \leq T-2$. Consider the problem of minimizing the expected total number of device alarms over the $T-j$ bits that are degraded in scenario $T-j$. Note that the expected number of device alarms for the remaining bits is the same in both scenarios, namely $jA(0)$. Comparing these two scenarios in the problem is equivalent to comparing the two extrema in Lemma A3 with $\gamma = (T-j-1)/(T-j)$ and the BER constraint set to $\tilde{B} - (j/T)B(0)$. The extrema of interest are (J_{T-j}^s, J_{T-j}^s) and $(J_{T-j-1}^s, 0)$. Lemma A3 tells us that $A_{T-j} > A_{T-j-1}$, completing the induction. \square

Lemma A4: For the function $A(J)$ defined in (4): i) There is a unique positive solution J^* such that $A''(J^*) = 0$, ii) $J^* > \alpha$, iii) $A'(J) > 0$ in $[0, \infty)$.

Proof: The function

$$-A'(x) = \frac{1}{\sqrt{2\pi}\sigma} \left[e^{-(x+\alpha)^2/2\sigma^2} - e^{-(x-\alpha)^2/2\sigma^2} \right]$$

has the same form as $f(x)$ defined in (13) in Lemma 3A. In particular, here $-\alpha$ takes the place of α in (13), while α takes the place of η in (13). Therefore, various properties from Lemma 3A hold for the function $-A'(x)$. From Lemma 3A, we know that $-A'(x)$ has only one local maximum located before $-\alpha$ ($<-\alpha$) and one local minimum after α ($>\alpha$). We also know that $-A'(x)$ strictly increases (from the limit value of 0) in the interval from $-\infty$ to the local maximum, strictly decreases in the interval from the local maximum to the local minimum, and strictly increases (to the limit value of 0) in the interval from the local minimum to ∞ .

i), ii) It follows that in the interval from zero to the local minimum, $-A'(x)$ strictly decreases and, therefore, $-A''(x) < 0$. At the local minimum, we have $-A''(x) = 0$. In the interval from the local minimum to ∞ , $-A'(x)$ strictly increases and, thus, $-A''(x) > 0$. Therefore, we have $-A''(x) = 0$ only at the local minimum J^* of $-A'(x)$. In addition, the location of this local minimum is in the positive direction of α ($>\alpha$).

iii) It is straightforward to show from (4) that $A'(0) = 0$. $-A'(x)$ decreases and is thus negative in $[0, J^*)$. In addition, $-A'(x)$ increases monotonically to the limit value of 0 in (J^*, ∞) . Therefore, $-A'(x) < 0$ or equivalently $A'(x) > 0$ in $[0, \infty)$. \square

Lemma 4: The function $C'(J)/C(J)$ strictly decreases in $[0, \alpha)$.

Proof: Let J^* be the unique positive solution to $A''(x) = 0$ as described in lemma A4. Since $J^* > \alpha$, it suffices to show that $C'(J)/C(J)$ strictly decreases in $[0, J^*)$. Consider the derivative of $C'(J)/C(J)$

$$\begin{aligned} \frac{d}{dJ} \left(\frac{C'(J)}{C(J)} \right) &= \frac{C(J)C''(J) - (C'(J))^2}{(C(J))^2} \\ &= \frac{-(1 - A(J))A''(J) - (A'(J))^2}{(C(J))^2} \end{aligned} \quad (27)$$

which is continuous since it is the derivative of a continuous function $C'(J)/C(J)$. Note that $1 - A(J) > 0$ for all J since from (4) $A(J) < 1$ for all finite J . In addition, lemma A4 tells us that $A''(J) > 0$ in $[0, J^*)$. It follows from the above expression that the derivative of $C'(J)/C(J)$ is negative in $[0, J^*)$. \square

REFERENCES

- [1] M. Médard, S. R. Chinn, and P. Saengudomlert, "Attack detection in all-optical networks," in *Proc. OFC'98*, ThD4, pp. 272–273.
- [2] —, "Attack and failure detection in transparent optical nodes," Lincoln Laboratory–Internal Rep., Lexington, MA, 1998.
- [3] L. A. Buckman, L. P. Chen, and K. Y. Lau, "Crosstalk penalty in all-optical distributed switching networks," *IEEE Photonics Technol. Lett.*, vol. 9, Feb. 1997.

- [4] C. S. Li and F. Tong, "Crosstalk and interference penalty in all-optical networks using static wavelength routers," *J. Lightwave Technol.*, vol. 14, June 1996.
- [5] C. Saxtoft and P. Chidgey, "Error rate degradation due to switch crosstalk in large modular switched optical networks," *IEEE Photonics Technol. Lett.*, vol. 5, no. 7, July 1993.
- [6] C. X. Yu, W. K. Wang, and S. D. Brorson, "System degradation due to coherent crosstalk in WDM network nodes," in *Proc. OFC'98*, WM 30, pp. 212–213.
- [7] M. Médard, D. Marquis, and S. R. Chinn, "Attack detection methods for all-optical networks," in *Proc. Internet Society's Symp. Network and Distributed System Security (NDSS)*, Mar. 1998, session 3, paper 1.
- [8] P. Saengudomlert, "Analysis and detection of jamming attacks in an all-optical network," MIT Master thesis, Dept. Electrical Engineering and Computer Science, MIT, Cambridge, MA, 1998.
- [9] M. Sumida, "OTDR performance enhancement using a quaternary FSK modulated probe and coherent detection," *IEEE Photon. Technol. Lett.*, vol. 7, pp. 336–338, Mar. 1995.
- [10] I. Katzela, G. Ellinas, and T. E. Stern, "Fault diagnosis in the linear lightwave network," in *Proc. LEOS Summer Topical Meetings*, Aug. 7–11, 1995, pp. 41–42.
- [11] N. Schroff and M. Schwartz, "Fault detection/identification in the linear lightwave network," Columbia Univ., CU/CTR/TR 243-91-24, 1991.
- [12] A. V. Yakovlev, "An optical-fiber system for transmitting confidential information," *Telecommun. Radio Eng.*, vol. 10, 1995.
- [13] G. R. Hill *et al.*, "A transport network layer based on optical network elements," *J. Lightwave Technol.*, vol. 11, May/June 1993.
- [14] A. Kloch, B. Mikkelsen, and K. E. Stubkjaer, "Pilot tones in WDM networks with wavelength converters," in *Proc. OFC'97*, paper TuE6, pp. 24–25.

Poompat Saengudomlert (S'94) is a Ph.D. candidate at the Laboratory for Information and Decision Systems in the Electrical Engineering and Computer Science (EECS) Department, Massachusetts Institute of Technology (MIT), Cambridge, MA. He received the B.S.E. degree in electrical engineering from Princeton University, Princeton, NJ, in 1996, and the M.S. degree in EECS from MIT, in 1998.

His research interests lie in communication networks. Since 1998, he has been conducting researches on the design of high-speed optical networks under the MIT Lincoln Laboratory's Next Generation Internet project.

Muriel Médard (S'91–M'95–SM'02) she received the B.S. degree in EECS and in mathematics, in 1989, the B.S. degree in humanities, in 1990, the M.S. degree in EE, in 1991, and the Sc.D. degree in EE, in 1995, from the Massachusetts Institute of Technology (MIT), Cambridge, MA.

She is an Assistant Professor in the Electrical Engineering and Computer Science Department and a member of the Laboratory for Information and Decision Systems, MIT. She was previously an Assistant Professor in the Electrical and Computer Engineering Department and a member of the Coordinated Science Laboratory at the University of Illinois, Urbana-Champaign. From 1995 to 1998, she was a Staff Member at MIT Lincoln Laboratory in the Optical Communications and the Advanced Networking Groups. Her research interests are in the areas of reliable communications, particularly for optical and wireless networks.

She was the winner of the 2002 IEEE Leon Kirshmayor Best Paper Award and received a NSF Career Award, in 2000.