JOURNAL OF
Optical Communications
and Networking

Proposal and performance analysis of an optical mode-division multiplexing random-access protocol for data centers

HOSSAM M. H. SHALABY[1,2]

[1]Faculty of Engineering, Electrical Engineering Department, Alexandria University, Alexandria 21544, Egypt
[2]Department of Electronics and Communications Engineering, Egypt-Japan University of Science and Technology (E-JUST), Alexandria 21934, Egypt (shalaby@ieee.org)

Received 3 June 2019; revised 24 July 2019; accepted 30 July 2019; published 26 August 2019 (Doc. ID 369085)

A random-access protocol for optical direct-detection mode-division multiplexing networks is proposed. The protocol is suitable for both data centers and on-chip communications systems. A mathematical description of the protocol is developed using both a Markov chain model analysis and an equilibrium point analysis. Several performance measures—specifically average network throughput, average packet delay, and average blocking probability—are derived. In our analysis, the effects of both multiplexing crosstalk and receiver noise are taken into account. The effect of several design parameters on the obtained performance measures are examined with the aid of a set of numerical examples.

https://doi.org/10.1364/JOCN.11.000501

1. INTRODUCTION

Mode division multiplexing (MDM) techniques are promising runners to meet the increasing traffic demand on optical networks, data centers, and on-chip communications [1–4]. Integrated photonic interconnect technology is more suitable for both data centers and on-chip communications. Indeed, this technology is free from the bandwidth-distance limitation that exists in electrical interconnects, making it a potential alternative for next-generation scalable data centers [5]. Accordingly, it is required to switch and route MDM channels through a reconfigurable network, which, when combined with wavelength division multiplexing (WDM), allows a more than 4 Tb/s data transmission rate on a multimode waveguide [4].

A large number of mode division (de)multiplexers have been proposed and implemented on silicon-on-insulator (SOI) platforms (e.g., microring resonators [6], Y-junctions [7], multimode interferometers [8], adiabatic couplers [9], asymmetrical directional couplers [10,11], bidirectional Bragg-grating-supported couplers [12–14], flying-bird mode-division demultiplexers [15], and photonic crystal based [16]).

In addition, a few reconfigurable approaches for mode division (de)multiplexing have been proposed for SOI platforms, which include microrings [4,17], multimode interference (MMI) couplers [18,19], Mach–Zehnder interferometers (MZIs) [20], and triple-silicon-waveguide directional couplers [21]. In [4], Stern *et al.* have proposed a 1×2 multimode switch for routing four data channels using reconfigurable microring mode (de)multiplexers. The device has a crosstalk of -16.8 dB and an aggregated bandwidth of 40 Gb/s. Melati *et al.* have reported a reconfigurable (de)multiplexer based on a balanced MZI with two 2×2 MMIs and a phase shifter (thermo-optic heater) [18]. They have measured a crosstalk of -20 dB and an operational bandwidth of at least 10 nm. In [17], Wang and Dai have proposed and demonstrated an on-chip reconfigurable optical add-drop multiplexer for both MDM and WDM simultaneously. It integrates a mode demultiplexer, four tunable microring resonator switches, and a mode multiplexer. In [19], Priti and Liboiron-Ladouceur have used a thermo-optically tuned MMI coupler, in which an aggregated bandwidth of 20 Gb/s is achieved at a BER of 10^{-12} with a crosstalk of -20 dB. Sun *et al.* have proposed and experimentally demonstrate an integrated tunable mode filter based on an SOI platform [20]. Their device consists of two switchable mode exchangers using an MZI and a fundamental-mode pass filter. The experimental results have shown that the extinction ratio and insertion loss are about 18 dB and 3 dB, respectively, in the wavelength range from 1530 nm to 1565 nm. Jiang *et al.* have proposed and optimized a reconfigurable mode (de)multiplexer based on a triple-silicon-waveguide directional coupler, consisting of a central waveguide with a $Ge_2Sb_2Se_4Te_1$ (GSST) upper layer and two outer silicon waveguides [21]. The device has a compact length of 17.65 μm, insertion losses of 0.16 and 0.96 dB, and mode crosstalk of -15.91 and -26.97 dB in ON and OFF states, respectively.

During the long period of research in the field of optical MDM systems, most focus has been oriented toward the physical layer of the communication network, and we know of no studies of the network or link layer, although this is an essential part in the design of next generation data centers. Accordingly, in this paper, we propose an optical mode-division multiplexing random-access link layer protocol suitable for both data centers and on-chip communications. Although no previous MDM link layer protocols exist, there have been many proposals of WDM and code division multiplexing random access protocols in the literature. Among them are slotted Aloha and reservation Aloha [22], round robin receiver/transmitter [23], stop-and-wait automatic repeat request [24], and selective retransmission [25]. In the selective retransmission technique, only corrupted packets are retransmitted at the end of the message, which makes it more efficient than other protocols. This motivates us to build our proposed MDM protocol based on this technique.

We develop a mathematical model and a theoretical study of the system performance of the proposed protocol. The effect of MDM crosstalk on the bit error rate (BER) is determined and taken into consideration in the system performance. Several performance measures are examined in this paper—specifically, the average system throughput, average packet delay, and average blocking probability. We adopt two theoretical methods in our derivations: the Markov chain model analysis and the equilibrium point analysis (EPA). The effects of both multiplexing crosstalk and receiver noise are taken into account in our analysis. Finally, a numerical study of the derived performance measures that takes into account the effect of changing several design parameters is presented.

This paper has six sections. In Section 2, we introduce a basic description of the system architecture, along with the proposed link-layer protocol. A mathematical model and theoretical study of the system is presented in Section 3. Specifically, derivations of both steady-state system throughput and average packet delay are given using a Markov chain model analysis. In our analysis, the effects of both multiplexing crosstalk and receiver noise are taken into account. To study the effect of message length on system performance, an EPA and a basic description of the state diagram of the proposed protocol are presented in Section 4. The average message delay is also derived in this section. Section 5 is maintained for a numerical study of the derived performance measures, taking into account the effect of changing several design parameters. Finally, we give our conclusion in Section 6.

2. SYSTEM ARCHITECTURE

The basic architecture of an optical MDM network is shown in Fig. 1. It is composed of a set of N nodes or servers, a multimode optical interconnect cloud, and a set of TE modes $\mathcal{M} \stackrel{\text{def}}{=} \{\text{TE}_0, \text{TE}_1, \ldots, \text{TE}_{N-1}\}$.

The type of transmission is a sort of broadcast-and-select, where a message transmitted by any node is received by all other nodes. Each node selects the appropriate message according to the protocol proposed below. We assume that nodes are located uniformly from each other so that the near-far effect can be neglected and each node can be assumed to receive an equal amount of transmitted power. In other words, the broadcast-and-select architecture is implemented here using a star topology [22,26], although ring and bus topologies can also be adopted. Modes are assigned to all nodes a priori (i.e., when a node subscribes to the network, it is given a mode from $\mathcal{M}$). Each node is equipped with a fixed mode multiplexer and a tunable mode demultiplexer (FT-TR). The transmitter generates an optical Manchester signal that represents its data. The tunable receiver is able to tune to any mode of all other transmitters. Following the protocol given below, the receiver of a node would filter out the undesired modes and pass the desired mode to a burst-mode reception to recover the transmitted packets. Burst-mode reception is suitable for many high-speed optical multiaccess networks [e.g., time-division multiple access (TDMA) passive optical networks and wavelength-division multiple access (WDMA) optical star networks] [27,28].

A. Proposed Optical MDM Protocol

Here we introduce the protocol description and its modeling assumptions.

1. Modeling Assumption

We have the following assumptions in our model for the optical MDM protocol:

- Time is slotted with slot size T_s.
- Messages arrive at a station (or node) with probability A, also called *node activity*.
- A message is composed of a number of packets.
  1. A packet is composed of K bits. One packet should fit in a time slot. Thus, the packet time slot $T_s = K T_b$, where T_b is the bit time duration.
  2. Message lengths are geometrically distributed with an average message length of L slots.
  3. Message lengths are used to model connection holding times of the system circuits.
- Each node is equipped with a memory to store the messages. The memory is freed once the stored messages are transmitted *successfully*.

2. MDM Protocol Description

The proposed optical MDM protocol is summarized as follows:

- When a node has a message to transmit, it first transmits a connection request to the destination station.
- The receiver of an idle station scans across all modes for connection requests in a round-robin (or polling) manner.
- At any time slot, if there is a connection request at a certain mode, the station proceeds to send an acknowledgment.
- If there are no requests and there is a message arrival, the station tunes its receiver to the destination mode and transmits a connection request on its signature mode.
- Once an acknowledgment is received, the station proceeds to transmit its message packet by packet (each packet occupies a time slot).

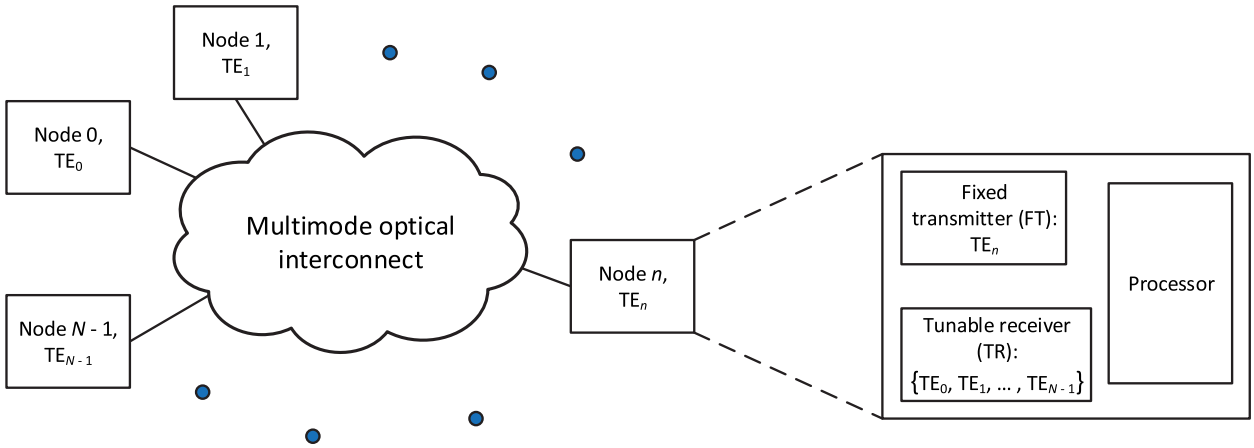

Fig. 1. Optical MDM network architecture.

• If a message is received successfully at the destination, the receiver transmits a positive acknowledgment.

In what follows, we incorporate the above protocol in a complete optical direct-detection system and analyze its performance. In our analysis, we are focusing on performance degradation due to both multiple-access interference and receiver shot and thermal noises.

Three main performance measures are evaluated: the average network throughput β, which determines the average number of successfully received packets within one packet slot; the average packet delay D, which can be evaluated using Little's formula [29]; and the blocking probability, which is determined using EPA.

B. Specific System Model

We consider a synchronous MDM network composed of N nodes (or information sources). The nth information source, $n \in \{0, 1, \ldots, N-1\}$, generates an independent and identically distributed (iid) binary data sequence $\{d_n\}$, where $d_n \in \{0, 1\}$ and $\Pr\{d_n = 0\} = \Pr\{d_n = 1\} = 1/2$. This sequence modulates the intensity of light pulses emitted from a light source. In addition, each node is assigned a unique TE mode selected from the set of modes $\{\mathrm{TE}_0, \mathrm{TE}_1, \ldots, \mathrm{TE}_{N-1}\}$. Specifically, node n, $n \in \{0, 1, \ldots, N-1\}$, transmits its data on the TE_n mode only. We denote the optical source frequency and bandwidth by f_c and $\Delta\nu$, respectively, and the average node power by P_{av}. In our analysis below, we assume that the zero's node is the desired node. At the receiver side of the desired node, a tunable detection scheme is used [20,30] that detects the received signal from node $n \in \{1, 2, \ldots, N-1\}$ (i.e., $n \neq 0$). The detected photocurrent is then time averaged to give the decision current. Finally, this decision current is examined to decide on the data.

3. THEORETICAL ANALYSIS USING THE MARKOV CHAIN MODEL

In this section, we adopt a Markov chain model to determine both average network throughput and average packet delay. We assume that all N nodes are having the same average activity A. Upon a successful connection request, a node transmits a packet (with probability A) using its signature mode at the beginning of a time slot to the destination. Here, the message is composed of one packet ($L = 1$ packet) and the length of a packet is K bits and corresponds to a slot duration. The intended receiver, once it receives a packet, transmits an acknowledgment to the sending node, indicating whether the packet is received successfully or not. If not, the transmitter enters a backlog state and retransmits the packet after a geometric distributed random delay time with average d slots. Because of the very short distances between the transceivers (of data centers and on-chip nodes), the transmission times for requests and acknowledgment control are neglected. Although we impose this assumption here to make the mathematics tractable, we relax it when analyzing the system using the EPA method in Section 4.

Assuming that at a given slot the number of backlogged nodes is $n \in \{0, 1, 2, \ldots, N\}$, the offered traffic and system throughput are

$$G(n) = (N-n)A + \frac{n}{d} = NA - n\left(A - \frac{1}{d}\right),$$

$$\beta(n) = \sum_{j=0}^{N-n}\sum_{i=0}^{n}(i+j)\,P_{c\mathrm{Pac}}(i+j)\,P_{\mathrm{bl}}(i|n)\,P_{\mathrm{th}}(j|n), \quad \textbf{(1)}$$

respectively, where the factor $1/d$ is the activity of a backlogged node. Here, $P_{c\mathrm{Pac}}(S)$ denotes the probability of a packet's success given S active nodes, $S \in \{1, 2, \ldots, N\}$. Also, $P_{\mathrm{th}}(j|n)$ and $P_{\mathrm{bl}}(i|n)$ denote the probabilities of j thinking (transmitting new packets) and i backlogged nodes, respectively, being active at a given time slot with n backlogged nodes. These are given by

$$P_{\mathrm{th}}(j|n) = \binom{N-n}{j} A^j (1-A)^{N-n-j},$$

$$P_{\mathrm{bl}}(i|n) = \binom{n}{i}\left(\frac{1}{d}\right)^i\left(1-\frac{1}{d}\right)^{n-i}, \quad \textbf{(2)}$$

where $j \in \{0, 1, \ldots, N-n\}$ and $i \in \{0, 1, \ldots, n\}$.

A. Statistics of Photocurrent

Assuming that there are S active nodes, $S \in \{1, 2, \ldots, N\}$, the resultant incident field on desired node photodiode PD_0 can be written as

$$E_R(t) = \sum_{n=0}^{S-1} \sqrt{P_n} g(t - d_n\tau) e^{j[\omega_c t + \phi_n(t)]} = e^{j\omega_c t}\sqrt{P_0} \sum_{n=0}^{S-1} g(t - d_n\tau)\sqrt{\xi_n} e^{j\phi_n(t)}, \quad (3)$$

where $P_0 = 2P_{\mathrm{av}}$; $\omega_c = 2\pi f_c$; $\tau = T_b/2$; $\xi_n = P_n/P_0$, $n \in \{0, 1, \ldots, S-1\}$, is the crosstalk to node 0; $g(t)$ is a normalized envelope of a single-node light field of spectral width $\Delta\nu$; and $\phi_n(t)$, $n \in \{0, 1, \ldots, S-1\}$, is the phase of the light source of the nth node, assumed to be a Wiener process [31,32]. We assume that $g(t)$ has the following properties:

$$g(t) = 0; \quad \text{if } t \notin [0, \tau], \qquad \frac{1}{\tau}\int_0^\tau g^2(t)\mathrm{d}t = 1. \quad (4)$$

If the photodiode responsivity is denoted by $\mathcal{R}$, then the average currents (over half-bit durations) of PD_0 are given by

$$I_j = \frac{\mathcal{R}}{\tau}\int_{j\tau}^{(j+1)\tau} |E_R(t)|^2 \mathrm{d}t, \quad (5)$$

for $j \in \{0, 1\}$. Using Eqs. (3) and (5), we get

$$I_j = \frac{\mathcal{R}P_0}{\tau}\int_{j\tau}^{(j+1)\tau} \left|\sum_{n=0}^{S-1} g(t - d_n\tau)\sqrt{\xi_n} e^{j\phi_n(t)}\right|^2 \mathrm{d}t. \quad (6)$$

Expanding the last equation, we get

$$I_j = \mathcal{R}P_0 \left[\sum_{n=0}^{S-1} |1 - j - d_n|\xi_n + 2\sum_{n=0}^{S-2}\sum_{m=n+1}^{S-1} \sqrt{\xi_n\xi_m}|1 - j - d_n| \cdot |1 - j - d_m| \times \frac{1}{\tau}\int_{j\tau}^{(j+1)\tau} g^2(t - j\tau)\cos(\phi_n(t) - \phi_m(t))\mathrm{d}t\right]. \quad (7)$$

We define the following random variable for any $j \in \{0, 1\}$ and any $n, m \in \{0, 1, \ldots, S-1\}$ with $n \neq m$ as

$$\chi_{nm}^j \overset{\mathrm{def}}{=} \frac{2}{\tau}\int_{j\tau}^{(j+1)\tau} g^2(t - j\tau)\cos(\phi_n(t) - \phi_m(t))\,\mathrm{d}t. \quad (8)$$

Substituting in Eq. (7), we get

$$I_j = \mathcal{R}P_0 \left[\sum_{n=0}^{S-1} |1 - j - d_n|\xi_n + \sum_{n=0}^{S-2}\sum_{m=n+1}^{S-1} \sqrt{\xi_n\xi_m}|1 - j - d_n||1 - j - d_m|\chi_{nm}^j\right]. \quad (9)$$

Using [32] and [31], the mean and variance of χ_{nm}^j are

$$\mu_{\chi_{nm}^j} \overset{\mathrm{def}}{=} E\{\chi_{nm}^j\} = 0,$$

$$\sigma^2_{\chi_{nm}^j} \overset{\mathrm{def}}{=} \mathrm{var}\{\chi_{nm}^j\} = \frac{1}{\tau^2}\int_{j\tau}^{(j+1)\tau}\int_{j\tau}^{(j+1)\tau} \times g^2(t - j\tau)g^2(u - j\tau)e^{-2\frac{|t-u|}{\tau_c}}\mathrm{d}t\mathrm{d}u, \quad (10)$$

respectively, where τ_c is the coherence time. For the special case where $g(t)$ is a rectangular envelope Eq. (10) reduces to

$$\sigma^2_{\chi_{nm}^j} = \frac{\tau_c}{\tau}\left[1 - \frac{\tau_c}{2\tau}\left(1 - e^{-\frac{2\tau}{\tau_c}}\right)\right] \overset{\mathrm{def}}{=} \sigma_\tau^2. \quad (11)$$

To get the expected value of I_j, we use Eqs. (10) and (11) in Eq. (9), taking into account the independence of the random variables χ_{nm}^j:

$$\mu_{I_j} \overset{\mathrm{def}}{=} E\{I_j\} = \mathcal{R}P_0\left[\sum_{n=0}^{S-1} |1 - j - d_n|\xi_n\right] = \mathcal{R}P_0\left[|1 - j - d_0| + \sum_{n=1}^{S-1} |1 - j - d_n|\xi_n\right]. \quad (12)$$

Assuming that the crosstalk $\xi_n = \xi$ for any $n \in \{1, 2, \ldots, S-1\}$, we get

$$\mu_{I_j} = \mathcal{R}P_0\left[|1 - j - d_0| + \xi\kappa_j\right], \quad (13)$$

where we define the interference random variable κ_j as

$$\kappa_j \overset{\mathrm{def}}{=} \sum_{n=1}^{S-1} |1 - j - d_n| \quad (14)$$

for any $j \in \{0, 1\}$. Noticing that

$$k_0 + k_1 = \sum_{n=1}^{S-1}(|1 - d_n| + |d_n|) = S - 1, \quad (15)$$

it follows that the interference random vector $\kappa \overset{\mathrm{def}}{=} (\kappa_0, \kappa_1)$ follows a binomial distribution:

$$P_\kappa(\ell) = \frac{1}{2^{S-1}}\frac{(S-1)!}{\ell_0!\ell_1!}, \quad (16)$$

where the vector $\ell = (\ell_0, \ell_1) \in \{0, 1, \ldots, S-1\}^2$ with $\ell_0 + \ell_1 = S - 1$. The variance of I_j can be easily evaluated as

$$\sigma^2_{I_j} \overset{\mathrm{def}}{=} \mathrm{var}\{I_j\} = (\mathcal{R}P_0)^2\sigma_\tau^2\left[\xi|1 - j - d_0|\kappa_j + \xi^2\frac{\kappa_j(\kappa_j - 1)}{2}\right]. \quad (17)$$

B. Statistics of the Decision Current

From the discussion in the last section, the decision current $I_D = I_1 - I_0$ has the following mean and variance, given desired node bit $d_0 = b \in \{0, 1\}$ and interference $\kappa = \ell = (\ell_0, \ell_1) \in \{0, 1, \ldots, S-1\}^2$ with $\ell_0 + \ell_1 = S - 1$:

$$\mu_{b,\ell} = \mathcal{R} P_0[(2b-1) + \xi(\ell_1 - \ell_0)],$$
$$\sigma^2_{b,\ell} = \sigma^2_{\text{PIIN}|b,\ell} + \sigma^2_{\text{sh}|b,\ell} + \sigma^2_T, \tag{18}$$

respectively, where $\sigma^2_{\text{PIIN}|b,\ell}$, $\sigma^2_{\text{sh}|b,\ell}$, and σ^2_T are the phase-induced intensity noise (PIIN), shot noise, and thermal noise variances, respectively. These are given by

$$\begin{aligned}\sigma^2_{\text{PIIN}|b,\ell} &= \sigma^2_{I_0} + \sigma^2_{I_1} \\ &= (\mathcal{R}P_0)^2\sigma^2_\tau \left[\xi(1-b)\ell_0 + \xi b \ell_1 \right. \\ &\quad \left. + \xi^2\left(\frac{\ell_0(\ell_0-1)}{2} + \frac{\ell_1(\ell_1-1)}{2}\right)\right], \\ \sigma^2_{\text{sh}|b,\ell} &= 2e B_e \left(\mu_{I_1} + \mu_{I_0}\right) \\ &= 2e B_e \mathcal{R} P_0 \left[1 + \xi(S-1)\right], \\ \sigma^2_T &= 4k_B T^\circ B_e / R_L \times 2 = 8 k_B T^\circ B_e / R_L,\end{aligned} \tag{19}$$

respectively. Here $e = 1.6 \times 10^{-19}$ C is the electron charge, $k_B = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, $B_e = 1/2\tau$ is the receiver electrical bandwidth, T° is the receiver noise temperature, and R_L is the receiver load resistor.

C. BER and Packet Success Probability

The bit-error probability $P_{e\text{Bit}}$ can be written as

$$P_{e\text{Bit}}(S) = \frac{1}{2} \sum_{\substack{\ell \in \{0,1,\ldots,S-1\}^2, \\ \ell_0 + \ell_1 = S-1}} \left[P_{e|\kappa,d_0}(\ell, 1) + P_{e|\kappa,d_0}(\ell, 0)\right] P_\kappa(\ell), \tag{20}$$

where $P_{e|\kappa,d_0}(\ell, b)$ is the conditional bit error rate (BER) given $\kappa = \ell \in \{0, 1, \ldots, S-1\}^2$ interferers and that the desired node has sent data bit $d_0 = b \in \{0, 1\}$. The last BER can be easily simplified to

$$P_{e\text{Bit}}(S) = \frac{1}{2} \sum_{\substack{\ell \in \{0,1,\ldots,S-1\}^2, \\ \ell_0 + \ell_1 = S-1}} P_\kappa(\ell)\, \text{erfc}\left(\frac{\mu_{1,\ell}}{\sigma_{1,\ell}\sqrt{2}}\right). \tag{21}$$

Accordingly, the average packet success rate given S active nodes is evaluated as

$$P_{c\text{Pac}}(S) = [1 - P_{e\text{Bit}}(S)]^K. \tag{22}$$

D. Steady-State Performance

To obtain the steady-state throughput and average packet delay, the above system can be described by a discrete-time Markov chain [33]. The chain consists of $N+1$ states depending on the number of backlogged nodes $n \in \{0, 1 \ldots, N\}$. The transition from one state to another occurs on a slot-by-slot basis. We determine the transition probability P_{nm} from state n to state m, where $n, m \in \{0, 1 \ldots, N\}$, of backlogged nodes as follows. Let k and l denote the number of thinking and backlogged nodes, respectively, being active at state n. It is obvious that there are $n - l$ backlogged nodes that are still idle and cannot succeed in transmission. For the system to jump to state m, $m - (n - l) = l + m - n$ nodes have to fail out of $k + l$ transmitting nodes. The remaining $k - m + n$ nodes have to succeed. Thus we obtain the transition probability for the system as

$$\begin{aligned}P_{nm} = \sum_{l=0\vee(n-m)}^{n} \sum_{k=0\vee(m-n)}^{N-n} & P_{\text{bl}}(l|n) P_{\text{th}}(k|n) \binom{k+l}{k-m+n} \\ & \times P_{c\text{Pac}}^{k-m+n}(k+l)[1 - P_{c\text{Pac}}(k+l)]^{l+m-n},\end{aligned} \tag{23}$$

where $x \vee y$ denotes the maximum of the two numbers x and y. A stationary probability distribution π_n, $n \in \{0, 1, \ldots, N\}$, always exists for the above irreducible Markov chain and can be obtained from the following set of equations:

$$\sum_{n=0}^{N} \pi_n = 1,$$
$$\sum_{n=0}^{N} \pi_n P_{nm} = \pi_m, \qquad \forall m \in \{0, 1, \ldots, N\}. \tag{24}$$

The steady-state system throughput β and average offered traffic G are given by

$$\beta = \sum_{n=0}^{N} \beta(n)\pi_n,$$
$$G = \sum_{n=0}^{N} G(n)\pi_n, \tag{25}$$

respectively. Finally, the average packet delay D can be determined from [29]

$$D = \frac{G}{\beta}. \tag{26}$$

4. EQUILIBRIUM POINT ANALYSIS

To study the effect of message length L on the system performance and determine the blocking probability, we adopt the EPA technique, which simplifies the problem and makes it more tractable. In this technique, the system is always assumed to be operating at an equilibrium point [34,35]. In other words, at any time slot, the expected number of nodes entering any state is always equal to the number of nodes departing from the state. In addition, EPA allows us to consider the effect of the control plane exchange (including the transmission times for both requests and acknowledgments), which has been neglected in Markov chain model.

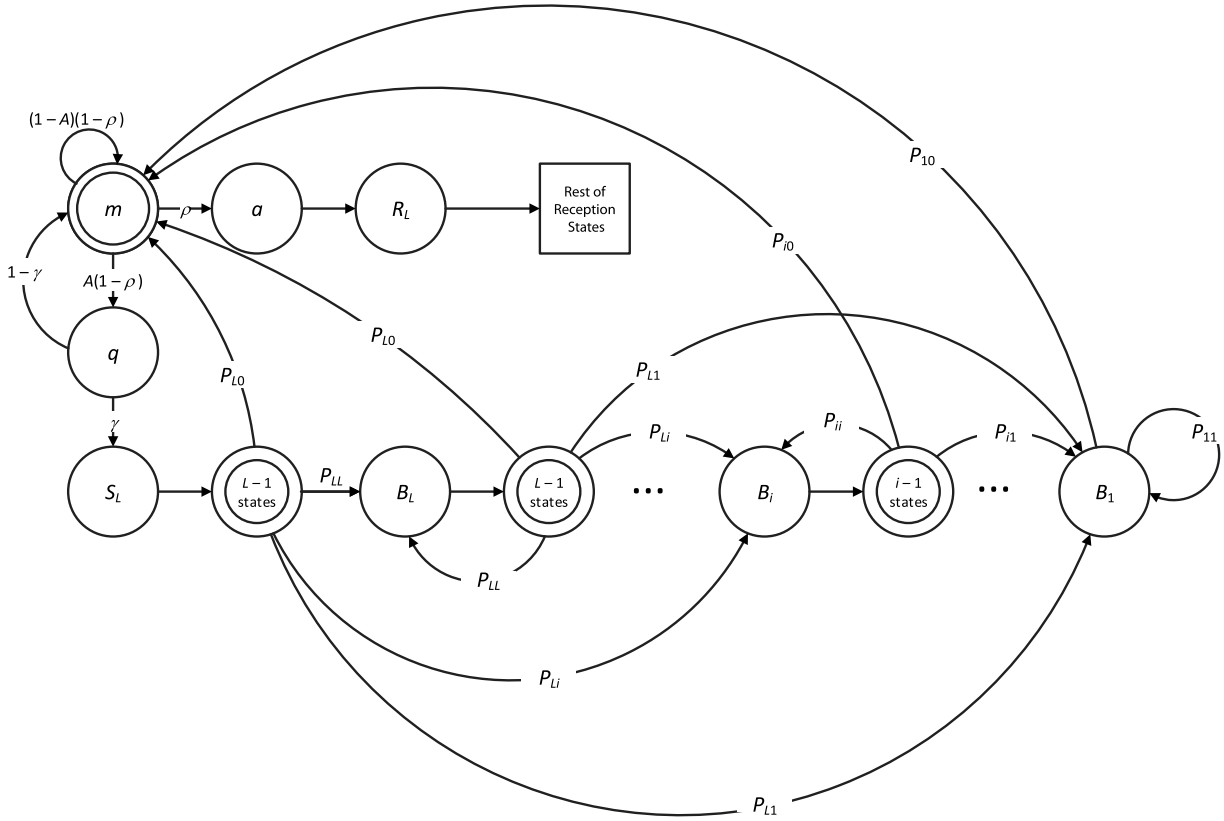

Fig. 2. State diagram of the proposed optical MDM protocol.

The state diagram of the proposed protocol is shown in Fig. 2. The states are labeled by the number of their nodes and are defined as follows:

- *Initial state*, m: A node is in state m if its receiver is scanning across the modes. Once tuned to a particular mode, if the station finds a request on that mode (an event that occurs with probability ρ), it proceeds to send an acknowledgment. If there is no request and there is an arrival (that occurs with probability A), the station enters the *requesting state*. If there is neither a request nor an arrival, the station remains as is.
- *Requesting state*, q: If there is a message arrival and no request, the station proceeds to send a connection request and tunes its receiver to the mode of the destination. If the station receives an acknowledgment (with probability γ), it enters the *transmission state*. If no acknowledgment is received, the station is *timed out* and goes back to the initial state m.
- *Acknowledgment state*, a: The station enters this state after the arrival of a connection request. The station sends an acknowledgment to the requesting station and enters the *reception state*.
- *Transmission state*, S_L: The station is in this state if it is transmitting its first packet of its new L-packets message. After sending its first packet, the station enters a sequence of $L-1$ states for transmitting the rest of the message (one state for each packet). All of these states have the same number of nodes as that of S_L. After transmitting the last packet, the station receives an acknowledgment about the status of the received packets. If the status indicates a successful transmission, the station returns to the initial state m. If the status indicates an unsuccessful transmission of i failed packets, $i \in \{1, 2, \ldots, L\}$, the station enters the backlogged state B_i and starts retransmitting the failed i packets. This process continues until the successful transmission of all packets.
- *Backlogged state*, B_n: A station is in state B_n, $n \in \{1, 2, \ldots, L\}$, if it is retransmitting the first packet of its n backlogged packets. After sending this packet, the station enters a sequence of $n-1$ states to retransmit the rest of the backlogged packets. All of these states have the same number of nodes as that of B_n. The station retransmits its first packet after waiting for a random delay time with an average of d slots.
- *Reception state*, R_L: The station enters this state if it is receiving the L-packets message for the first time. If all packets are received successfully, the station sends an acknowledgment to the transmitting station and returns to its initial state. If one or more packets are damaged, however, the station sends an ask-for-retransmission and enters a backlogged reception state that depends on the number of failed packets, like that of the backlogged transmission state.

A. Performance Analysis

From Fig. 2, we can write the following set of flow equations:

$$\begin{aligned} R_L &= a = \rho m, \\ q &= A(1-\rho)m, \\ S_L &= \gamma q. \end{aligned} \tag{27}$$

Noticing that the number of nodes receiving new messages should be equal to the number of nodes transmitting new messages, we get

$$R_L = S_L,$$
$$\rho m = \gamma q. \tag{28}$$

Substituting for q from Eq. (27), we get

$$\gamma A(1-\rho) = \rho. \tag{29}$$

The probability that a request is found by a scanning node ρ is equal to the probability that another node is in a requesting state q so that

$$\rho = \frac{q}{N} = \frac{A(1-\rho)}{N} m = \frac{A(1-\rho)}{N} \frac{R_L}{\rho} = \frac{A(1-\rho)}{N} \frac{S_L}{\rho}, \tag{30}$$

where we have used Eqs. (27) and (28) to justify the last two equalities. It can be noticed that Eq. (30) is a second-order equation in ρ, for which a solution is easily found by

$$\rho = \frac{1}{2}\left[\sqrt{u^2+4u} - u\right], \qquad \text{where } u = \frac{AS_L}{N}. \tag{31}$$

B. Number of Nodes in Backlogged States

The number of nodes in the backlogged state i, $i \in \{1, 2, \ldots, L\}$, is determined recursively from the flow equation at state B_i:

$$B_i = S_L P_{Li} + \sum_{n=i}^{L} B_n P_{ni}, \tag{32}$$

where P_{ni}, $n \in \{1, 2, \ldots, L\}$ and $i \in \{0, 1, \ldots, n\}$, is the probability of finding i damaged packets out of n received packets:

$$P_{ni} = \binom{n}{i} P_{c\mathrm{Pac}}^{n-i}(T)[1 - P_{c\mathrm{Pac}}(T)]^i, \tag{33}$$

where T is the total number of active (or transmitting) nodes:

$$T = LS_L + \sum_{n=1}^{L} (n+d) B_n, \tag{34}$$

where d is the average backlogged delay defined before. The average number of nodes in state S_L can be evaluated by the requirement that the total number of nodes in all states is equal to N:

$$N = m + a + q + 2T$$
$$= m + a + q + 2\left[LS_L + \sum_{n=1}^{L} (n+d) B_n\right]. \tag{35}$$

The factor of 2 in the last equation is because the number of receiving states is equal to that of transmitting states. The first three terms in Eq. (35) are given by

$$m + a + q = m + \rho m + A(1-\rho)m$$
$$= [1 + \rho + A(1-\rho)] \frac{S_L}{\rho}. \tag{36}$$

C. Steady-State Throughput

The steady state throughput $\beta(N, K, A, L)$ is defined as the number of successfully received packets per slot:

$$\beta(N, K, A, L) = TP_{c\mathrm{Pac}}(T), \tag{37}$$

where T is the total number of transmitting nodes, which is determined so the total number of nodes in all states is equal to N as described above. Also, the average packet delay is given by $D = NA/\beta$.

D. Blocking Probability

The blocking probability P_{blocking} is defined as the probability that an arrival is blocked. Accordingly, it is equal to the probability that the station is in the initial state m but there is a request for connection, and at the same time there is a message arrival A, or the station is not in the initial state m and there is a message arrival A:

$$P_{\text{blocking}} = \frac{m}{N}\rho A + \left(1 - \frac{m}{N}\right) A$$
$$= A - \frac{m}{N} A(1-\rho) = A - \rho, \tag{38}$$

where, we have used Eq. (30) to justify the last equality.

5. NUMERICAL RESULTS

In this section, we evaluate the performance of the proposed protocol using our developed measures given in previous sections. Here are the main parameters used in the evaluations: The data bit rate is $R_b = 0.5$ Gb/s, the linewidth of the light source is $\Delta\nu = 50$ kHz, its coherence time is $\tau_c = 1/\Delta\nu$, the receiver noise temperature is $T^\circ = 300$ K, the receiver load resistance is $R_L = 1$ kΩ, the photodiode responsivity is $\mathcal{R} = 1$, and the average received power is $P_{\text{av}} = -10$ dBm.

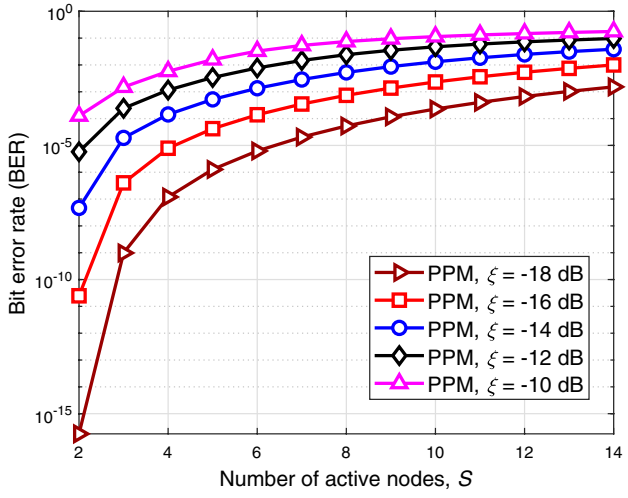

Fig. 3. Bit error rate for an MDM system adopting Manchester encoding.

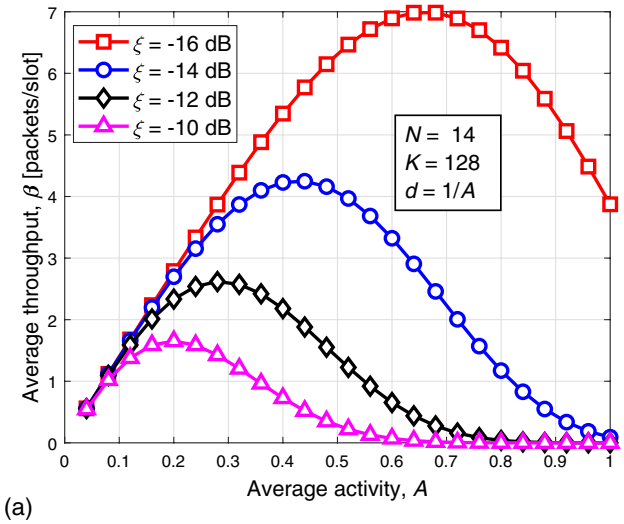

Fig. 4. Network performance for different crosstalk levels and average backlogged delay $d = 1/A$: (a) average network throughput versus node activity and (b) average packet delay versus throughput.

We start by evaluating the BER for different numbers of active nodes S and crosstalk ratios $\xi \in \{-18, -16, \ldots, -10\}$ dB. The results are plotted in Fig. 3. It is clear from the figure that the crosstalk seriously affects the BER performance. For example, at a BER of 10^{-5}, the supported numbers of active nodes are about 6, 4, 2, 2, and 1 for crosstalk of -18, -16, -14, -12, and -10 dB, respectively. Increasing the average power is not useful here as the system exhibits an error floor.

In Fig. 4(a), the throughput is plotted versus the average activity when assuming the same thinking and backlog activities $1/d = A$. The number of available nodes is $N = 14$ and the crosstalk ratios are $\xi \in \{-16, -14, -12, -10\}$ dB. Although $N = 14$ provides a poor BER as shown in Fig. 3, due to the burst nature of the network, not all subscribers are active at the same time. This depends on the activity A. A general trend of throughput is noticed, where it increases as A increases until it reaches a maximum value and then decreases when increasing A further. Indeed, as A increases above 0, more packets become available for transmission with a low interference (low number of active nodes) and the throughput increases until it reaches a peak. As the node activity increases further, the effect of multiple-access interference becomes significant and packet failures become more probable, leading to a throughput decay. The maximum achievable throughput is significantly dependent on the crosstalk level. For example, at $\xi = -16$ dB, this maximum is seven packets/slot, which is achieved at $A = 0.68$, while the maximum is only 4.25 packets/slot achieved at $A = 0.44$ when $\xi = -14$ dB. It is worth mentioning that the maximum achievable throughput cannot increase above seven packets/slot for this network of $N = 14$ nodes, where half of the nodes are transmitting and the other half are receiving. The corresponding average packet delay is plotted in Fig. 4(b) versus throughput when varying the average node activity. It is clear from the figure that the packet delay increases slowly when increasing the throughput until a point where throughput reaches its maximum. After this point, the packet delay grows fast. Indeed, after this point, packet failures become more probable and retransmissions become more frequent.

Figures 5(a) and 5(b) are similar to Figs. 4(a) and 4(b), but they assume an average backlogged delay $d = 2$ slots independent of the node activity A. There is a difference in the general trend of the throughputs in Fig. 5(a). If the crosstalk is not that high (e.g., $\xi \in \{-16, -14\}$ dB), after the throughput reaches its peak, it does not decay that fast and becomes almost constant because here the offered traffic $G(n)$ [cf. Eq. (1)] for large activities $A > 1/d$ is less than that for the previous case. This situation in turn introduces less interference and hence slow-decaying throughput. In contrast with the previous case, here the average packet delay in Fig. 5(b) does not grow fast after throughput reaches it maximum. Indeed, here the packet failures become less probable and retransmissions become less frequent because of reduced offered traffic.

The effect of changing the packet length K on the network throughput is studied in Fig. 6 for average backlogged delays $d = 1/A$ and $d = 5$, Figs. 6(a) and 6(b), respectively. The figures are plotted versus number of active nodes at a fixed activity $A = 0.5$ and a fixed crosstalk $\xi = -16$ dB for different packet lengths $K \in \{128, 256, 512, 1024\}$ bits. The figures show that at small numbers of active nodes, increasing the packet size has no effect on throughput, while, at large numbers of active nodes, increasing the packet size would reduce the throughput. This can be explained by examining Eq. (22), where, for large numbers of active nodes, the BER decreases and the packet success rate $P_{c\mathrm{Pac}}$ decreases by increasing the packet size K. This leads to more packet failures and reduced throughput.

Finally, the effect of changing the message length L on the network throughput, average delay, and blocking probability is studied in Fig. 7 for average activities $A = 1/d = 0.5$ and network size $N = 14$ nodes. It is clear from Fig. 7(a) that the throughput always reaches a saturation value as L increases without limit. Indeed, if nodes have messages of unlimited length to transmit, every node will be busy either transmitting or receiving messages and there is no change in the interference pattern. It is easy to verify that the asymptotic limit in this case is given by $(N/2)P_{c\mathrm{Pac}}(N/2)$ independent of the activity A. Indeed, for any $A > 0$, messages keep arriving to new nodes

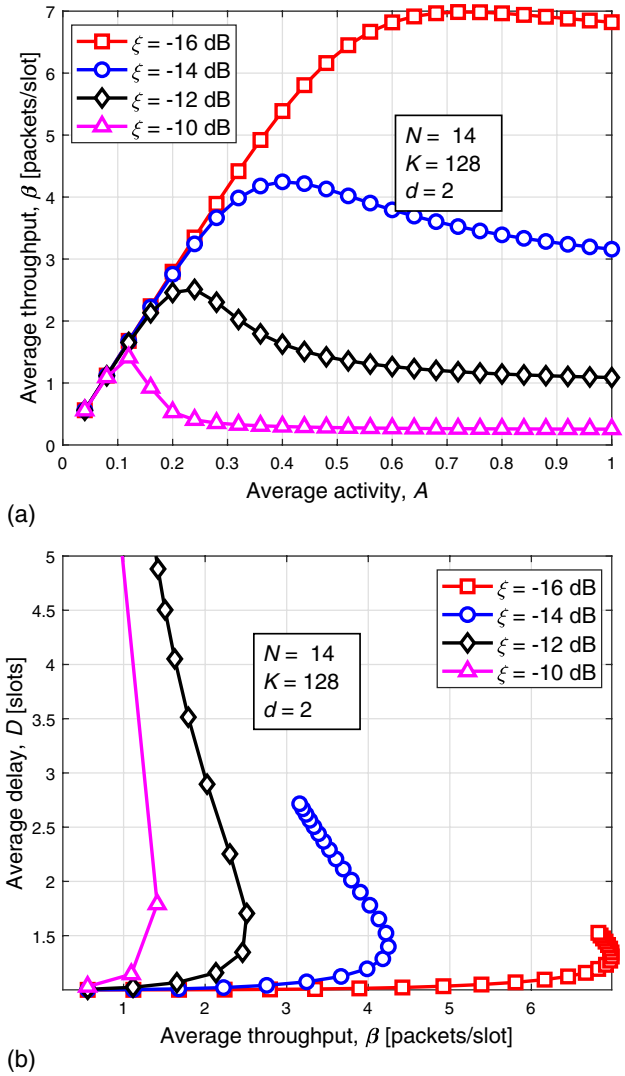

Fig. 5. Network performance for different crosstalk levels and average backlogged delay $d = 2$: (a) average network throughput versus node activity and (b) average packet delay versus throughput.

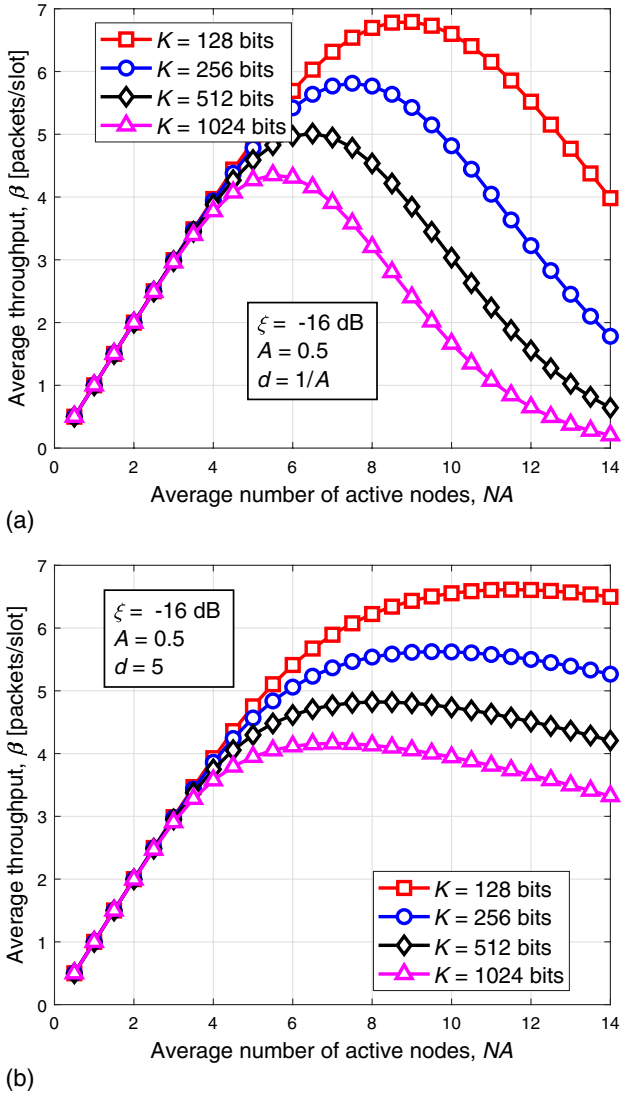

Fig. 6. Average network throughput versus number of active nodes for fixed activity $A = 1/2$ and different packet lengths: (a) average backlogged delay $d = 1/A$ and (b) average backlogged delay $d = 5$ slots.

that start transmitting and never stop. After a large enough number of slots, half of the nodes have long messages being transmitted to the other half. It can be concluded that small message lengths are suitable for systems with high crosstalk, while the inverse is true for systems with small crosstalk. Figure 7(b) confirms this conclusion as the delay increases

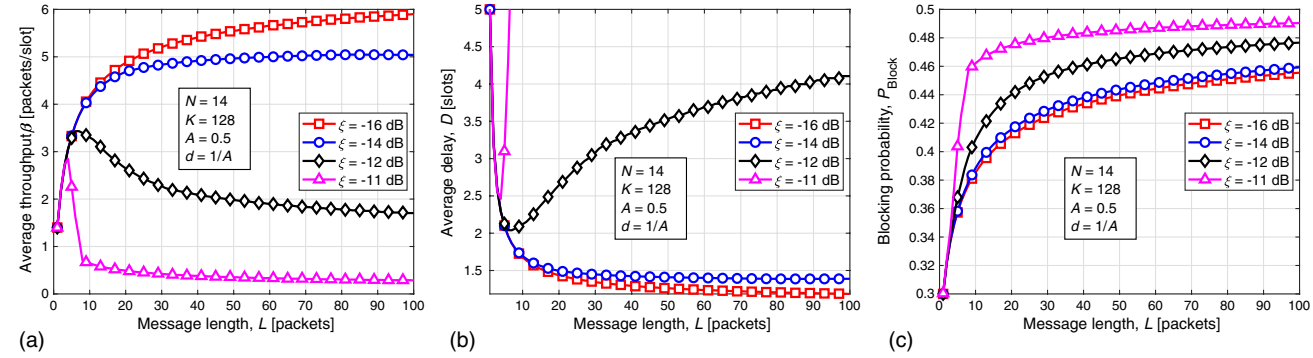

Fig. 7. Network performance versus message length for different crosstalk levels and average activities $A = 1/d = 1/2$: (a) average network throughput, (b) average packet delay and (c) blocking probability.

dramatically when increasing the message lengths for systems with high crosstalk, while the inverse occurs for systems with small crosstalk. The price paid in this case is the increase of blocking probability, as can be seen from Fig. 7(c). Indeed, nodes are now busy transmitting (or receiving) their long messages and cannot accept any more arrivals or send any more requests.

6. CONCLUSION

A random access protocol for optical direct-detection MDM networks has been proposed, and its performance has been analyzed and evaluated. Several performance measures—namely, average system throughput, average packet delay, and average blocking probability—have been derived based on both the Markov chain and equilibrium point analysis methods. In our analysis, we focused on the effect of both multiple-access interference and receiver noise, which is applicable for both data centers and on-chip communications systems. The effect of several design parameters on the system performance measures have been investigated and presented numerically for the proposed protocol. The following concluding remarks can be extracted from our results. Optimum values that maximize the throughput always exist, which are dependent on the number of active nodes and interference (crosstalk) patterns. These optimum values are also reachable by increasing the message lengths, for the case of small crosstalk, independent of the node activity. Small message lengths are suitable for systems with high crosstalk, while the inverse occurs for systems with small crosstalk.

REFERENCES

1. D. J. Richardson, J. M. Fini, and L. E. Nelson, "Space-division multiplexing in optical fibres," Nat. Photonics **7**, 354–362 (2013).
2. S. Berdagué and P. Facq, "Mode division multiplexing in optical fibers," Appl. Opt. **21**, 1950–1955 (1982).
3. A. Grieco, G. Porter, and Y. Fainman, "Integrated space-division multiplexer for application to data center networks," IEEE J. Sel. Top. Quantum Electron. **22**, 1–6 (2016).
4. B. Stern, X. Zhu, C. P. Chen, L. D. Tzuang, J. Cardenas, K. Bergman, and M. Lipson, "On-chip mode-division multiplexing switch," Optica **2**, 530–535 (2015).
5. Y. Chen, M. Kibune, A. Toda, A. Hayakawa, T. Akiyama, S. Sekiguchi, H. Ebe, N. Imaizumi, T. Akahoshi, S. Akiyama, S. Tanaka, T. Simoyama, K. Morito, T. Yamamoto, T. Mori, Y. Koyanagi, and H. Tamura, "A 25 Gb/s hybrid integrated silicon photonic transceiver in 28 nm CMOS and SOI," in *IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, California, 2015, pp. 402–404.
6. B. A. Dorin and W. N. Ye, "Two-mode division multiplexing in a silicon-on-insulator ring resonator," Opt. Express **22**, 4547–4558 (2014).
7. J. B. Driscoll, R. R. Grote, B. Souhan, J. I. Dadap, M. Lu, and R. M. Osgood, "Asymmetric Y junctions in silicon waveguides for on-chip mode-division multiplexing," Opt. Lett. **38**, 1854–1856 (2013).
8. L. Han, S. Liang, H. Zhu, L. Qiao, J. Xu, and W. Wang, "Two-mode de/multiplexer based on multimode interference couplers with a tilted joint as phase shifter," Opt. Lett. **40**, 518–521 (2015).
9. J. Xing, Z. Li, X. Xiao, J. Yu, and Y. Yu, "Two-mode multiplexer and demultiplexer based on adiabatic couplers," Opt. Lett. **38**, 3468–3470 (2013).
10. J. Liao, L. Zhang, M. Liu, L. Wang, W. Wang, G. Wang, C. Ruan, W. Zhao, and W. Zhang, "Mode splitter without changing the mode order in SOI waveguide," IEEE Photon. Technol. Lett. **28**, 2597–2600 (2016).
11. W. Jiang, "Ultra-compact and fabrication-tolerant mode multiplexer and demultiplexer based on angled silicon waveguides," Opt. Commun. **425**, 141–145 (2018).
12. H. M. H. Shalaby, "Bi-directional coupler as a mode-division multiplexer/demultiplexer," J. Lightwave Technol. **34**, 3633–3640 (2016).
13. O. M. Nawwar, H. M. H. Shalaby, and R. K. Pokharel, "Modeling, simulation, and fabrication of bi-directional mode-division multiplexing (BMDM) for silicon-on-insulator platform," Appl. Opt. **57**, 42–51 (2018).
14. H. M. H. Shalaby, "Bidirectional mode-division multiplexers with antireflection gratings," Appl. Opt. **57**, 476–484 (2018).
15. O. M. Nawwar, J. A. H. Odoeze, and H. M. H. Shalaby, "Rib waveguide as a three-mode demultiplexer for SOI," OSA Continuum **2**, 124–133 (2019).
16. O. M. Nawwar, H. M. H. Shalaby, and R. K. Pokharel, "Photonic crystal-based compact hybrid WDM/MDM (de)multiplexer for SOI platforms," Opt. Lett. **43**, 4176–4179 (2018).
17. S. Wang and D. Dai, "Silicon-based reconfigurable optical add-drop multiplexer for hybrid MDM-WDM systems," in *Optical Fiber Communication Conference (OFC)*, Los Angeles, California, 2017, paper Tu2C.1.
18. D. Melati, A. Alippi, and A. Melloni, "Reconfigurable photonic integrated mode (de)multiplexer for SDM fiber transmission," Opt. Express **24**, 12625–12634 (2016).
19. R. B. Priti and O. Liboiron-Ladouceur, "A reconfigurable multimode demultiplexer/switch for mode-multiplexed silicon photonics interconnects," IEEE J. Sel. Top. Quantum Electron. **24**, 1–10 (2018).
20. C. Sun, W. Wu, Y. Yu, X. Zhang, and G. T. Reed, "Integrated tunable mode filter for a mode-division multiplexing system," Opt. Lett. **43**, 3658–3661 (2018).
21. W. Jiang, "Reconfigurable mode (de)multiplexer using $Ge_2Sb_2Se_4Te_1$ based triple-silicon-waveguide," IEEE Photon. Technol. Lett. **30**, 2119–2122 (2018).
22. G. N. M. Sudhakar, N. D. Georganas, and M. Kavehrad, "Slotted Aloha and reservation Aloha protocols for very high-speed optical fiber local area networks using passive star topology," J. Lightwave Technol. **9**, 1411–1422 (1991).
23. H. M. H. Shalaby, "Performance analysis of an optical CDMA random access protocol," J. Lightwave Technol. **22**, 1233–1241 (2004).
24. Y. M. Abd-El-Malek, Z. A. El-Sahn, H. M. H. Shalaby, and E. A. El-Badawy, "New optical random access code-division multiple-access protocol with stop-and-wait automatic repeat request," Opt. Eng. **46**, 056007 (2007).
25. M. A. A. Mohamed, H. M. H. Shalaby, and E. A. El-Badawy, "Optical code-division multiple-access protocol with selective retransmission," Opt. Eng. **45**, 055007 (2006).
26. G. P. Agrawal, *Fiber-Optic Communication Systems*, 4th ed. (Wiley, 2010).
27. C. Su, L.-K. Chen, and K.-W. Cheung, "Theory of burst-mode receiver and its applications in optical multiaccess networks," J. Lightwave Technol. **15**, 590–606 (1997).
28. R. Koma, M. Fujiwara, J. Kani, S. Kim, T. Suzuki, K. Suzuki, and A. Otaka, "Demonstration of real-time burst-mode digital coherent reception with wide dynamic range in DSP-based PON upstream," J. Lightwave Technol. **35**, 1392–1398 (2017).
29. L. Kleinrock, *Queueing Systems, Volume 1: Theory* (Wiley, 1975).
30. W. Zhang and J. Yao, "A fully reconfigurable waveguide Bragg grating for programmable photonic signal processing," Nat. Commun. **9**, 1396 (2018).
31. M. M. Rad and J. A. Salehi, "Phase-induced intensity noise in digital incoherent all-optical tapped-delay line systems," J. Lightwave Technol. **24**, 3059–3072 (2006).
32. M. Noshad and K. Jamshidi, "Bounds for the BER of codes with fixed cross correlation in SAC-OCDMA systems," J. Lightwave Technol. **29**, 1944–1950 (2011).
33. A. Polydoros and J. Silvester, "Slotted random access spread-spectrum networks: an analytical framework," IEEE J. Sel. Areas Commun. **5**, 989–1002 (1987).

34. S. Tasaka and Y. Ishibashi, "A reservation protocol for satellite packet communication—a performance analysis and stability considerations," IEEE Trans. Commun. **32**, 920–927 (1984).
35. J. P. Jue, M. S. Borella, and B. Mukherjee, "Performance analysis of the rainbow WDM optical network prototype," IEEE J. Sel. Areas Commun. **14**, 945–951 (1996).

Hossam M. H. Shalaby (S'83–M'91–SM'99) was born in Giza, Egypt, in 1961. He received B.S. and M.S. degrees from Alexandria University, Alexandria, Egypt, in 1983 and 1986, respectively, and a Ph.D. degree from the University of Maryland at College Park in 1991, all in electrical engineering. In 1991, he joined the Electrical Engineering Department at Alexandria University, and was promoted to professor in 2001. Since November 2017, he has been on leave from Alexandria University, where he is a professor in the Department of Electronics and Communications Engineering (ECE), School of Electronics, Communications, and Computer Engineering, Egypt-Japan University of Science and Technology (E-JUST), New Borg EL-Arab City, Alexandria, Egypt. From September 2010 to August 2016, he was the chair of the ECE department, E-JUST. From December 2000 to 2004, he was an adjunct professor with the Faculty of Sciences and Engineering, Department of Electrical and Information Engineering, Laval University, Quebec, QC, Canada. From September 1996 to January 1998, he was with the Electrical and Computer Engineering Department, International Islamic University Malaysia, and from February 1998 to February 2001, he was with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He worked as a consultant at SysDSoft company, Alexandria, Egypt, from 2007 to 2010. His research interests include optical communications, silicon photonics, optical CDMA, and quantum information theory. He has served as a student branch counselor at Alexandria University, IEEE Alexandria, and North Delta Subsection, from 2002 to 2006, and served as chairman of the student activities committee of IEEE Alexandria Subsection from 1995 to 1996. He received an SRC fellowship from 1987 to 1991 from the Systems Research Center, Maryland; State Excellence Award in Engineering Sciences in 2007 from the Academy of Scientific Research and Technology, Egypt; Shoman Prize for Young Arab Researchers in 2002 from the Abdul Hameed Shoman Foundation, Amman, Jordan; State Incentive Award in Engineering Sciences in 1995 and 2001 from the Academy of Scientific Research and Technology, Egypt; University Excellence Award in 2009 from Alexandria University; and University Incentive Award in 1996 from Alexandria University. He is a senior member of the IEEE Photonics Society and The Optical Society (OSA).