PAM Performance Analysis in Multicast-Enabled Wavelength-Routing Data Centers

Citation for the original published paper (version of record):
Rastegarfar, H., Yan, L., Szczerba, K. et al (2017)
PAM Performance Analysis in Multicast-Enabled Wavelength-Routing Data Centers
Journal of Lightwave Technology, 35(13): 2569-2569
http://dx.doi.org/10.1109/jlt.2017.2691278

PAM Performance Analysis in Multicast-Enabled Wavelength-Routing Data Centers

Houman Rastegarfar, Li Yan, Krzysztof Szczerba, and Erik Agrell

***Abstract*—Multilevel pulse amplitude modulation (M-PAM) is gaining momentum for high-capacity and power-efficient cloud computing. Compared to the classic on-off keying (OOK) modulation, high-order PAM yields better spectral efficiency but is also more susceptible to physical layer degradation effects. We develop a cross-layer analysis framework to examine PAM transmission performance in data center network environments supporting both optical multicasting and wavelength routing. Our analysis is conducted on a switch architecture based on an arrayed-waveguide grating (AWG) core and distributed broadcast domains, exhibiting different physical paths, and random, uncontrolled crosstalk noise. Reed-Solomon coding with rate adaptation is incorporated into PAM transceivers to compensate for impairments. Our Monte Carlo simulations point to the significant impact of AWG crosstalk on higher-order PAM in wavelength-reuse architectures and the importance of code rate adaptation for signals traversing multiple routing stages. According to our study, 8-PAM offers the highest effective bit rates for signals terminating in one broadcast domain and performs poorly when considering interdomain connectivity. On the other hand, the impairment-induced degradation of interdomain capacity for 4-PAM can be limited to 20.4%, making it better suited for connections spanning two broadcast domains and a crosstalk-rich stage. Our results call for software-defined PAM transceiver designs in support of both modulation order and code rate adaptation.**

***Index Terms*—Arrayed waveguide grating (AWG), bit error rate (BER), data center, forward error correction (FEC), goodput, optical multicast, physical layer, pulse amplitude modulation (PAM).**

I. Introduction

MULTICAST data transmission schemes have recently been receiving attention in developments towards cloud data center networks. The need for efficient and simultaneous transmission of the same information copy to a large number of data center nodes is being driven by many applications that benefit from execution parallelism and cooperation, including the MapReduce type of algorithms for processing data and applications that use distributed file systems for storage [1], [2]. Multicasting helps to minimize the network load, increase the throughput of bandwidth-hungry computations, accelerate the execution of delay-sensitive applications, and save on network communication resources and energy requirements [2]–[5]. Despite their advantages, the existing electronically-switched data centers are not efficient in supporting multicast traffic delivery and call for complex hardware configurations [1], [2], [5]. The lack of proper multicast mechanisms in existing data centers is further stressed out by the exponential growth of cloud traffic and the overwhelming challenges of electronic switching technologies in terms of scale and power consumption.

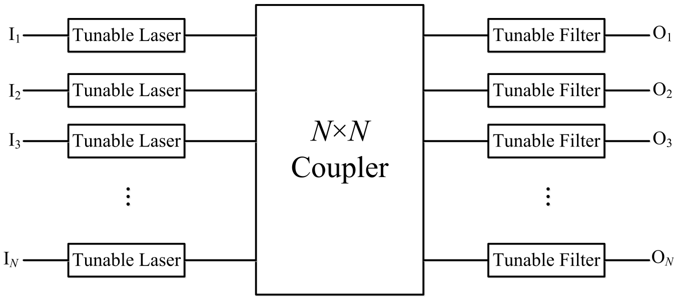


Fig. 1. Optical multicasting based on star coupler and tunable transceivers.

Optical switching has been proposed to support scalability, bit rate transparency, and low energy footprints in data centers [6]. Not only does it help to establish high-capacity point-to-point connections, but it can also be utilized for traffic multicasting [1], [7]–[14]. Fig. 1 depicts the key optical multicasting building blocks, including a star coupler as the broadcast medium and tunable transceivers. In order to support multicast traffic delivery, the designated receivers should all be tuned onto the wavelength of the transmitter node and the star coupler performs the message replication albeit with a splitting loss. To avoid collisions each active input port should carry a distinct wavelength. In general, it is not required that lasers be tunable; however, transmitter tunability is desired when the number of available wavelengths is less than the port count or the broadcast structure should contribute to a wavelength-routing design. The problem with the baseline multicast configuration is the lack of scalability due to the limited port count of the coupler, the limited tuning range of tunable lasers and filters, and losses. To achieve optical multicast scalability, innovations from both architectural design and physical-layer signal transmission perspectives are desired.

From a network architecture point of view, a modular multicast switch design can overcome the limited transceiver tunability and coupler port count bottlenecks. Designs that incorporate wavelength routing to interconnect distributed broadcast domains are a viable solution for a data center environment, primarily due to compatibility with tunable transceivers and footprint and speed constraints. Wavelength routing based on arrayed waveguide grating (AWG), a passive device that allows

H. Rastegarfar (e-mail: rastegarfar@gmail.com), L. Yan, and E. Agrell are with the Department of Signals and Systems, Chalmers University of Technology, 412 96 Gothenburg, Sweden.

K. Szczerba (e-mail: krzysztof.szczerba@finisar.com) is with Finisar Corporation, 1389 Moffet Park Dr., Sunnyvale, California 94089, USA.

Manuscript received January 16, 2017; revised January 16, 2017.

for the parallel switching of wavelength-division-multiplexed (WDM) signals, has already been proposed as a disruptive switching technology for data centers [15]–[18]. The use of AWG as the core of an optical multicast switch fabric results in spatial parallelism and reuse of spectral resources in multiple broadcast domains, a desirable consequence of the AWG cyclic routing property [13]. In realizing such architectures; however, the impact of AWG crosstalk should not be overlooked [19].

From a signal transmission perspective, higher-order modulation formats can be employed to enhance the multicast capacity on a per channel basis. On-off keying (OOK) is the prevalent binary modulation being used in data centers and high-performance computing systems today. Its low spectral efficiency, on one hand, and the need to keep up with the ever-increasing capacity demands in warehouse-scale computing, on the other, have sparked a flurry of activities toward short-reach optical interconnects operating at 100 Gbps and above [20]. M-level pulse amplitude modulation (M-PAM), based on intensity modulation and direct detection (IM/DD), is a promising candidate for spectral efficiency improvements in optical interconnects due to its simplicity and data center power and cost constraints. Multilevel PAM is being developed for both 850 nm and 1550 nm optical interconnects [21]–[24]. To compensate for the sensitivity of this modulation scheme to physical-layer impairments, short block-length error-correcting codes with rate adaptation [25], [26] can be introduced to help minimize the redundancy overheads and processing latencies.

In this paper, we assess the performance of pulse amplitude modulation under data center network impairments for a distributed switching scenario combining wavelength routing and optical multicasting. Previous work on optical-layer switch performance analysis has examined OOK modulation only [19], [27], [28]. Data center studies have also neglected code rate adaptation. We combine these two aspects into our analysis framework. More specifically, the contribution of this work is threefold. We 1) propose a modular switch architecture with an AWG core and distributed broadcast domains that supports data center traffic locality and allows for different levels of connectivity; 2) develop a cross-layer analysis framework based on mathematical formulations and Monte Carlo simulations to quantify PAM performance in a data center switching environment; and 3) introduce rate adaptation to adjust the coding overheads in accordance with the signal path and random crosstalk impairments. To the best of our knowledge this is the first work to investigate the joint performance of PAM and code rate adaptation in optical data centers.

The remainder of the paper is organized as follows. Section II illustrates the proposed multicast-enabled, wavelength-routing switch architecture and a distributed scheduling algorithm. Section III details the cross-layer performance analysis methodology. Section IV analyzes the Monte Carlo simulation results for two sets of physical paths, comparing the impact of various modulation orders, redundancy overhead thresholds, crosstalk levels, and symbol rates. Finally, Section V concludes the paper and highlights future directions.

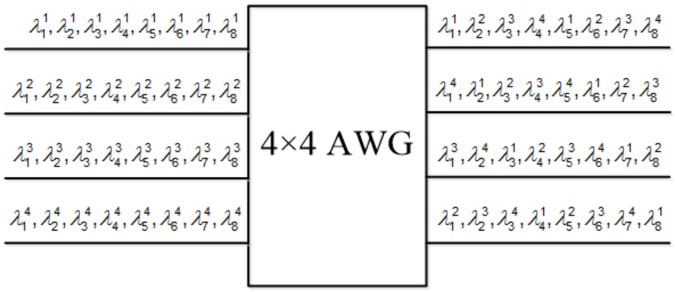

$\lambda_1^1,\lambda_2^1,\lambda_3^1,\lambda_4^1,\lambda_5^1,\lambda_6^1,\lambda_7^1,\lambda_8^1$
$\lambda_1^1,\lambda_2^2,\lambda_3^3,\lambda_4^4,\lambda_5^1,\lambda_6^2,\lambda_7^3,\lambda_8^4$
$\lambda_1^2,\lambda_2^2,\lambda_3^2,\lambda_4^2,\lambda_5^2,\lambda_6^2,\lambda_7^2,\lambda_8^2$
$\lambda_1^4,\lambda_2^1,\lambda_3^2,\lambda_4^3,\lambda_5^4,\lambda_6^1,\lambda_7^2,\lambda_8^3$
4×4 AWG
$\lambda_1^3,\lambda_2^3,\lambda_3^3,\lambda_4^3,\lambda_5^3,\lambda_6^3,\lambda_7^3,\lambda_8^3$
$\lambda_1^3,\lambda_2^4,\lambda_3^1,\lambda_4^2,\lambda_5^3,\lambda_6^4,\lambda_7^1,\lambda_8^2$
$\lambda_1^4,\lambda_2^4,\lambda_3^4,\lambda_4^4,\lambda_5^4,\lambda_6^4,\lambda_7^4,\lambda_8^4$
$\lambda_1^2,\lambda_2^3,\lambda_3^4,\lambda_4^1,\lambda_5^2,\lambda_6^3,\lambda_7^4,\lambda_8^1$

Fig. 2. The cyclic routing pattern of a 4×4 AWG with two FSRs.

II. Switch Architecture and Scheduling

The purpose of this work is to characterize the joint performance of pulse amplitude modulation and adaptive forward error correction in optical data center switches that exhibit random crosstalk. We propose a switch architecture and scheduling algorithm with several interesting properties that allow us to conduct a comprehensive physical-layer study under various impairments. The proposed design has the following key properties.

1) It supports both unicast and multicast traffic patterns.
2) It allows for wavelength routing, enabling spatial wavelength reuse and a scalable and modular multicast approach.
3) Despite wavelength routing, it does not depend on expensive wavelength conversion blocks and operates based on wavelength tunability at the edge.
4) It can support optical circuit switching and/or optical packet switching.
5) It supports data center traffic locality, avoids long physical paths as much as possible, and can work well with programmable transceivers.

Note that our proposed design is a forward-looking switching solution that relies on the commercialization of adaptive transceivers and large-scale photonic integration, making it a candidate for next-generation optical data center networks. The core of the architecture is an AWG that enables the passive routing of optical signals in a power- and space-efficient fashion. Fig. 2 depicts the cyclic routing of a 4×4 AWG, covering two free spectral ranges (FSRs). The cyclic routing pattern of the AWG implies that a signal with wavelength index $i = 1, 2, ...$ on input fiber port $j = 1, 2, ..., N$ of an $N \times N$ AWG is routed to its output port $1 + \mathrm{mod}(i - j, N)$. The routing pattern of the AWG allows for multiple signals with the same wavelength to coexist within the AWG, resulting in in-band (intraband) crosstalk. In this paper, we only consider one FSR; i.e., each input port can connect to each output port via a single wavelength.

A. Physical Layer Architecture

Fig. 3 depicts the proposed data center switch architecture, comprising tunable transceivers (Tx,Rx) and filters (TFs), optical amplifiers, wavelength selective switches (WSSs), $K \times K$ star couplers, and an $N \times N$ AWG as the core of the switch fabric. The design is modular and can support up to N broadcast domains as in Fig. 3(a). It can interconnect $N \times (K-1)$ computing nodes. Note that this design does not support recirculation buffering and all buffering should be performed electronically at the edge of the switch.

The hardware structure of the broadcast domains in Fig. 3(a) is detailed in Fig. 3(b). The tunable transmitters and filters should be able to tune over $K-1$ wavelengths for nonblocking intradomain connectivity. Sub-microsecond tuning speed of such devices allows for optical packet switching within each domain; however, packet switching is not a requirement for the proposed design. Tunable devices should also support wavelength routing across the AWG. For maximum spectral reuse, AWG wavelengths are shared for intradomain communications. For instance, with $N = K$, a total of N wavelengths are used to interconnect $N \times (N-1)$ nodes.

Each domain is equipped with a multiwavelength input and output port to interface with other domains through the AWG. Due to the broadcast property of the star coupler, the multiwavelength output port also carries the signals that belong to intradomain connections (i.e., connections whose destinations belong to the same broadcast domain D_i). Hence, a WSS (1×2 with one output port unused) follows each domain to block multiple undesired signals. In other words, the WSS only allows the interdomain connections to pass through the AWG. As the WSS is a relatively slowly reconfigurable device (reconfiguration time on the order of 10's of milliseconds), interdomain packet switching is not possible arbitrarily. However, with the WSS reconfigured, it is possible to perform rapid packet/burst switching between the interconnected domains across the AWG.

The splitting loss of star couplers along with the limited launch power of optical transmitters call for optical amplification in multicast switch architectures [1]. We propose the use of semiconductor optical amplifier (SOA) as an integrated solution within each broadcast domain for amplifying single-wavelength signals and erbium doped fiber amplifier (EDFA) on interdomain fibers for amplifying WDM signals. The choice of gain for these amplification stages has a significant impact on the physical layer performance. Optimization could be performed for instance to minimize the average bit error rate (BER) subject to amplification constraints. When tweaking the gains, the limitations on amplifier output power should be taken into account. In this study, we consider an SOA to compensate for the loss of a star coupler and an EDFA to compensate for the losses due to a star coupler and a tunable filter. The TFs and the AWG act as band-limiting filters for amplifiers.

B. Distributed Scheduling Algorithm

To characterize the physical layer performance of the switch architecture in Fig. 3, we propose a distributed, greedy scheduling algorithm that governs the generation of random crosstalk terms. A different algorithm could lead to different throughput and crosstalk levels; however, in this paper we lock the architecture and scheduling to focus on physical later scenarios within a unique framework.

We adopt an offline scheduling approach; i.e, the scheduling decisions regarding connection requests take place all at once. Each input port can either have no demand or can request to transmit to one output port following a certain distribution (unicast traffic). The scheduling algorithm is distributed (hence, scalable) in the sense that it runs in parallel within each domain. In servicing demands, priority is given to interdomain connection requests due to more stringent interdomain connectivity constraints imposed by the AWG-based switch fabric.

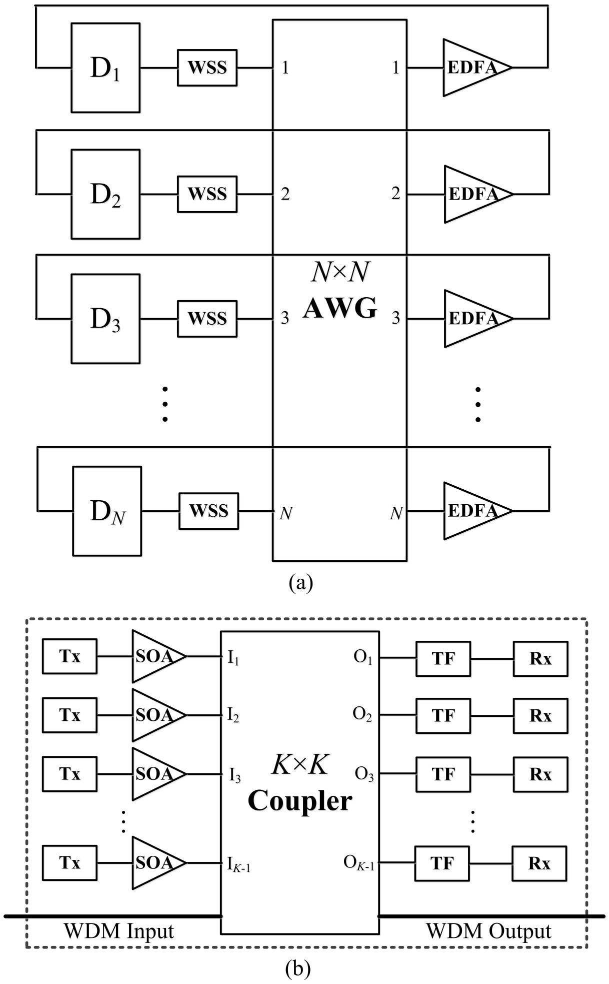


Fig. 3. (a) AWG-based optical multicast switch architecture, and (b) the architecture of broadcast domain D_i within the switch.

In scheduling interdomain connections, each output port $O_i, 1 \le i \le K-1$, within domain D_j holds the index of the domains (other than D_j) that request it. In matching output ports, during each step priority is given to the output port with the least number of domains trying to subscribe to it. This is because matching such an output port with one of its demanding (source) domains potentially blocks the minimum number of connections requesting domain D_j. This mechanism accounts for the greedy nature of the proposed scheduling algorithm. The scheduling steps that are run concurrently per domain are as follows. Ties are broken randomly.

AWG-Based Multicast Switch Scheduling Steps:

1) Serving interdomain connection requests in domain D_d:

```
for $it = 1$ to $T_{\max}$
    Generate traffic
    Allocate resources to interdomain requests
    Allocate resources to intradomain requests
    Quantify AWG crosstalk per established connection
    Quantify out-of-band crosstalk per connection
    Calculate BER per connection
    Consider the impact of adaptive coding overheads
    Collect BER and goodput statistics
end
Compile simulation results
```

Fig. 4. Pseudocode of the Monte Carlo switch simulator for a given input load.

repeat until all destination ports are examined

   a) Select the least demanded output port O in D_d.
   b) Randomly pick one of the domains D_s requesting O.
   c) See whether a wavelength with index $l = 1 + \mathrm{mod}$ $(s + d - 2, N)$ is unoccupied in both D_s and D_d. If so, pick one of the input ports I in D_s that requests O, assign λ_l to the transmitter at I and the receiver at O to complete the matching, and block any further subscription attempts from D_s to D_d. Otherwise, block all requests from D_s to D_d.
   d) Based on the outcome of step (c), update the connection request index list for the nodes in D_d.

2) Serving intradomain connection requests in domain D_d: for any unmatched output port O do
   a) Find the input ports in D_d that request O. Pick one input port I randomly and block the remaining inputs.
   b) Based on first-fit wavelength assignment, search for a free wavelength that can be used by both I and O. If such a wavelength exists, complete the matching. Otherwise, block the request from I.

III. Performance Analysis Methodology

The PAM performance analysis we conduct is based on a mix of mathematical calculations and Monte Carlo simulations. The mathematical framework of our analysis is presented in the Appendix. The contribution from several noise terms is deterministic. However, due to the random nature of connection requests, the loading of the AWG and star couplers within the architecture in Fig. 3 is stochastic. Hence, we conduct Monte Carlo simulations to evaluate the noise variance due to crosstalk. Our simulations are cross-layer in the sense that higher-layer scheduling is employed to model the performance in the physical layer and in turn the impact of the physical layer on a higher-layer measure (i.e., switch capacity) is quantified.

Fig. 4 illustrates the various steps in our Monte Carlo simulator for a given load (i.e., the probability of a an active connection request on each input port of the optical switch). The simulation consists of a number of iterations (denoted by $T_{\max}$). Each iteration starts by performing the scheduling tasks as elaborated in Section II–B. This in turn would determine the loading of the AWG and the star couplers. To calculate the in-band crosstalk noise variance for each interdomain connection (which passes through the AWG), the wavelength of the connection is examined and the number of adjacent and nonadjacent connections with similar wavelength is noted. In an $N \times N$ AWG, the ports adjacent to port i are indexed $1 + \mathrm{mod}(i - 2, N)$ and $1 + \mathrm{mod}(i, N)$. As for out-of-band (interband) crosstalk due to the nonideal receiver filter shape, we consider all interfering wavelengths coexisting within the connection's destination domain and use (17) to calculate the noise variance. In doing so, the wavelength and the received power of interfering connections are taken into account. Note that as connections can traverse different paths, their received power could be different and this should be included in evaluating the out-of-band crosstalk contribution.

Due to physical layer impairments, it is common for connections to not meet the target BER values for error-free operation (i.e., BER smaller than 10^{-12}). Hence, forward error correction (FEC) becomes necessary. The introduction of FEC results in overhead and latency, although with short block-length codes it is feasible to overcome the latency penalties. We adopt *code rate adaptation*. Assuming a target post-FEC BER and a maximum tolerable pre-FEC BER ($\mathrm{BER}_{\mathrm{pre,th}}$), we try to allocate redundancy just as much as needed. In our simulations, if the calculated pre-FEC BER per connection is less than or equal to $\mathrm{BER}_{\mathrm{pre,th}}$, a proper code rate will be picked to achieve the target post-FEC BER. If the pre-FEC BER is greater than $\mathrm{BER}_{\mathrm{pre,th}}$, the connection is deemed irretrievable and will not contribute to the goodput (i.e., bit rate excluding coding redundancy). The overall goodput is calculated by summing up the net rate of all connections for a given $\mathrm{BER}_{\mathrm{pre,th}}$.

IV. PAM Performance Evaluation Results

In this section, we analyze the performance of PAM in the presence of random crosstalk for the design in Fig. 3. Due to the short link lengths in a data center network [19], we ignore the dispersion and nonlinear effects. We consider the impact of thermal noise, shot noise, laser intensity noise, amplified spontaneous emission (ASE) noise, in-band AWG crosstalk, and out-of-band crosstalk arising due to nonindeal filtering at the receiver. We model these effects and their interactions as nine Gaussian noise terms (see the Appendix).

The fixed simulation parameters, included in Table I, are based on datasheets for commercially available products and recent demonstrations [19], [29]–[34]. A wide variety of choices exist regarding pulse shape [35]–[37]. In this work, we assume a non-return-to-zero (NRZ) Gaussian pulse shape and a matched-filter response for the receiver. We follow the rule of thumb for NRZ receivers and consider the receiver electrical filter 3-dB bandwidth, B_{e}, to be equal to 2/3 the symbol rate, R_{s} [37, ch. 4]. R_{s} is set to 28 GBaud unless stated otherwise. As well, unless stated otherwise, the default values for AWG adjacent (R_{AX}) and nonadjacent (R_{NX}) port crosstalk ratios are -30 dB and -35 dB, respectively.

In a simulation run, each node can generate a request with a fixed non-zero probability (i.e., load). Due to traffic locality in

TABLE I
TABLE OF SIMULATION PARAMETERS

Parameter	Definition	Value
k_B	Boltzmann's constant	1.38×10^{-23} J/K
h	Planck's constant	6.6261×10^{34} Js
q	Elementary charge	1.6×10^{-19} C
R_{L}	Receiver resistance	50Ω
T	Receiver temperature	300 K
F_{e}	Receiver amplifier noise figure	5 dB
R	Photodetector responsivity (PIN)	1 A/W
ER	Modulator extinction ratio	10 dB
RIN	Average laser RIN spectral density	-145 dB/Hz
B_{o}	Optical filter bandwidth	50 GHz
F_{EDFA}	EDFA noise figure	5 dB
F_{SOA}	SOA noise figure	6 dB
N	AWG port count	64
K	Star coupler port count	64
L_{A}	AWG insertion loss	6 dB
L_{C}	$K \times K$ coupler loss	$3\log_2 K + 1$ (dB)
L_{W}	WSS insertion loss	6 dB
L_{F}	Tunable filter insertion loss	3 dB
P_{S}	Average transmitter launch power	3 dBm

data centers [38], we assume that the request made by a node is uniformly destined to one of the destinations within its domain (excluding itself) with probability $3/4$. With probability $1/4$ the request is made to a node within a domain uniformly randomly picked from the other broadcast domains. The value for $T_{\max}$ in Fig. 4 is set to 100. We consider three modulation orders: 2-PAM, 4-PAM, and 8-PAM. 2-PAM represents the conventional OOK modulation, which we use as a baseline to compare the performance of higher-order modulation schemes with.

We begin our discussion by investigating how AWG crosstalk can deteriorate the BER performance for PAM. Fig. 5 depicts the average interdomain BER vs. load for various AWG crosstalk levels, considering the architecture of Fig. 3, the parameters in Table I, and 4-PAM modulation. The average BER for interdomain connections is reported as only these connections get affected by the AWG and undergo more severe signal degradation.

In Fig. 5, three crosstalk scenarios are compared: AWG crosstalk ratios of (-25 dB,-30 dB), (-30 dB,-35 dB), and a hypothetical case where the AWG exhibits zero crosstalk. In a realistic scenario, due to the passage of connections through two broadcast domains, the average BER is well beyond a desirable error-free target (e.g., 10^{-12}), implying the importance of FEC for PAM operation in multicast data centers. With non-zero crosstalk ratios, the BER becomes large and load dependent. In fact crosstalk-induced noise terms (either in-band or out-of-band) are the only noise terms that depend on the loading of the system, since the number of interferers increases with load. However, AWG crosstalk showed a much stronger impact compared to out-of-band receiver crosstalk in our simulations. For the launch power of 3 dBm, laser intensity noise, signal-in-band crosstalk beat noise, and thermal noise were observed as the three dominant noise terms (for both $R_{\mathrm{s}} = 28$ GBaud and $R_{\mathrm{s}} = 10$ GBaud).

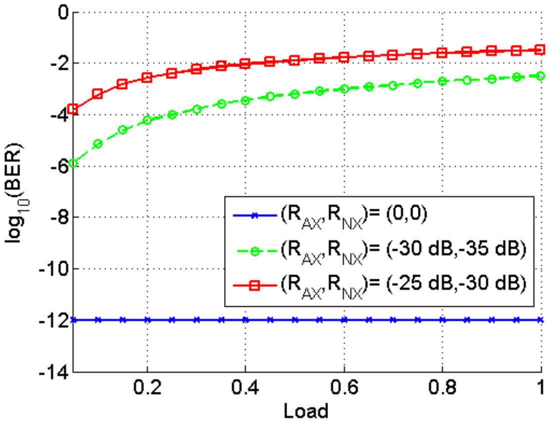


Fig. 5. interdomain bit error rate versus load for various AWG crosstalk levels (4-PAM modulation).

Comparing the BER levels in Fig. 5, one can easily conclude that large AWG crosstalk could be a significant barrier to wavelength-routed optical multicasting. With finite crosstalk ratios in place, the BER grows several orders of magnitude larger. Even the 5 dB difference between the crosstalk ratios, i.e., (-25 dB,-30 dB) vs. (-30 dB,-35 dB), could pose a significant impact. In fact with 4-PAM modulation, $R_{\mathrm{s}} = 28$ GBaud, and a hypothetically ideal physical layer (PHY) that introduces no impairments, the optical switch in our simulations could yield an interdomain throughput (i.e., the sum of the bit rates of all interdomain connections) of 41.6 Tbps. Considering zero AWG crosstalk and proper FEC (with a pre-FEC threshold of 3×10^{-2}), the same goodput could be achieved, corresponding to zero physical layer penalty. With (-30 dB,-35 dB) crosstalk ratios, the achievable goodput was 33.1 Tbps (corresponding to a 20.4% degradation). With (-25 dB,-30 dB), a remarkably small goodput of 6.6 Tbps was achieved, which translates to an 84.1% physical layer penalty. This analysis illustrates how sensitive the performance of PAM could be to crosstalk noise levels.

Now we draw our attention to the achievable goodput of the switch architecture in Fig. 3, studying the trade-off between modulation order and FEC overhead. We perform code rate adaptation based on Reed-Solomon code with a block length of 255 bytes, i.e., RS(255,k) [39], [40]. Fig. 6 depicts RS(255,k) code rate versus pre-FEC BER for a target post-FEC BER=10^{-12}. The short codeword and the high-speed optical links ensure that the decoding latency is negligible for data center applications (the transmission of 255 bytes over a 56 Gbps link using 4-PAM only takes 36.4 ns). We consider both ideal and imperfect physical layers (based on the parameters defined earlier). For imperfect PHY, three possible values of maximum pre-FEC BER threshold ($\mathrm{BER}_{\mathrm{pre,th}}$) are considered: 10^{-3}, 10^{-2}, and 3×10^{-2}. These correspond to minimum permissible code rates of 0.87, 0.59, and, 0.20, respectively (implying the feasibility of a wide range of overhead ratios). Allowing for larger FEC overheads with code rate adaptation helps to provide more room for accommodating connections. With FEC code rate adaptation we provide just enough overhead (by referring to Fig. 6) to move from a connection's pre-FEC BER to the target post-

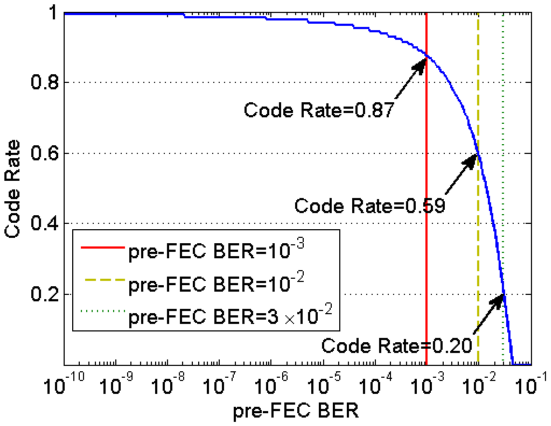


Fig. 6. RS(255,k) code rate versus pre-FEC BER for a target post-FEC BER=10^{-12}.

FEC BER.

Fig. 7 depicts the overall switch goodput (i.e., due to both intradomain and interdomain connections) vs. load, for different modulation orders and $\mathrm{BER}_{\mathrm{pre,th}}$ values. Considering the ideal PHY alone, we can easily note the relationship between capacity and modulation format (e.g., with 8-PAM the goodput is three times the goodput achieved with 2-PAM). However, when an imperfect PHY is in place, modulation schemes perform quite differently. Higher-order modulation results in higher spectral efficiency but is also more susceptible to impairments (given the same signal power). In this regard, 2-PAM remains quite resilient to impairments, 4-PAM behaves differently for different $\mathrm{BER}_{\mathrm{pre,th}}$ values, and 8-PAM performance is almost insensitive to variations in $\mathrm{BER}_{\mathrm{pre,th}}$. At full load, the degradation in goodput due to physical layer impairments is equal to zero for 2-PAM. For 4-PAM, this varies between 6% and 26.4% depending on the used FEC threshold and for 8-PAM it remains almost fixed around 30%. This implies that with 8-PAM, a large number of connections suffer from a BER greater than 3×10^{-2}, which cannot be retrieved.

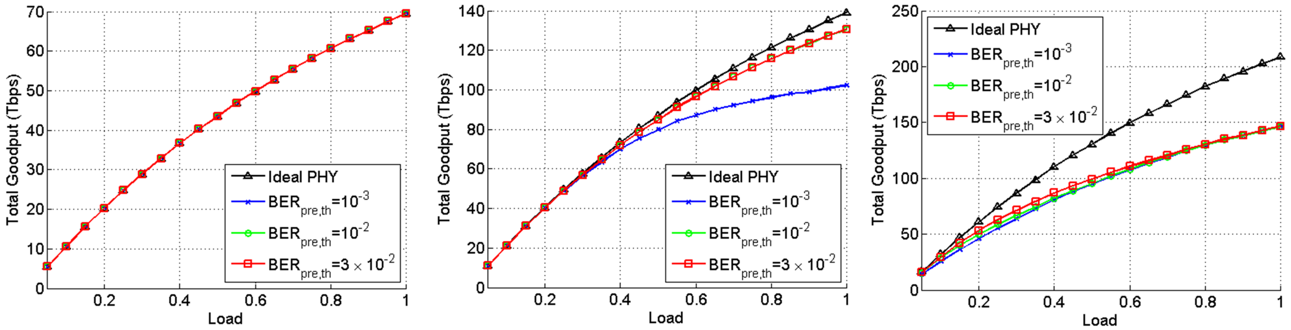
According to Fig. 7, irrespective of the physical layer degradation effects, the overall switch goodput increases with an increase in the modulation order when code rate adaptation is used. However, the values reported in Fig. 7 correspond to contributions from both intradomain and interdomain connections. These connections are affected by the physical layer quite differently as they have to traverse very different paths. As interdomain connections are prone to more impairments and their performance is key to the scalability of optical multicast, we specifically look at the evolution of interdomain goodput versus load in Fig. 8. Fig. 8(b) and Fig. 8(c) indicate that higher-order PAM schemes suffer significantly from AWG crosstalk compared with OOK. Unlike 2-PAM, 8-PAM is extremely vulnerable when it comes to end-to-end, interdomain routing and FEC cannot fully compensate for the physical layer penalties. With 8-PAM, the imperfect PHY turns the architecture of Fig. 3 into a segmented broadcast switch with virtually no AWG in place. PAM-4 provides the best performance among the three choices provided that proper $\mathrm{BER}_{\mathrm{pre,th}}$ is considered. A $\mathrm{BER}_{\mathrm{pre,th}}$ of 10^{-3} makes the physical layer penalty on interdomain connectivity as large as 88.5% whereas $\mathrm{BER}_{\mathrm{pre,th}} = 3 \times 10^{-2}$ limits this value to 20.4%. Indeed, 4-PAM accompanied by code rate adaptation with $\mathrm{BER}_{\mathrm{pre,th}} = 3 \times 10^{-2}$ results in the best interdomain performance among the studied scenarios.

It is interesting to note that the performance of 8-PAM could be significantly improved in the absence of AWG crosstalk. Our simulations indicate that if the AWG crosstalk ratios are set to zero, an interdomain goodput of 55.9 Tbps can be achieved with code rate adaptation (at $\mathrm{BER}_{\mathrm{pre,th}} = 3 \times 10^{-2}$). Compared to 62.3 Tbps with ideal PHY, this translates to 10.3% degradation. This result once again highlights the importance of developments toward low-crosstalk AWGs for wavelength-routing switch fabrics. Nonetheless, in line with the emergence of programmable data centers [41], an effective solution could be the design of PAM transceivers that can adapt the modulation order and code rate depending on physical and application layer constraints. For instance, for the proposed architecture in this paper, a node could switch between 4-PAM and 8-PAM modulation schemes depending on whether it wants to communicate with a node within its own domain or elsewhere. As well, the node can decide to switch between the two schemes for interdomain communications depending on the AWG loading. With adequately low crosstalk, it can make use of 8-PAM for a faster interdomain connectivity.

To conclude our analysis, we also study the behavior of interdomain goodput as a function of symbol rate. A decrease in symbol rate decreases the maximum achievable rate (with ideal PHY), but also can reduce the power of several noise terms (see the Appendix), leading to an improved BER. Here, we compare four bandwidth settings (i.e., $R_{\mathrm{s}} = 10$, 16, 22, and 28 GBaud). Fig. 9 depicts the interdomain goodput versus load when $\mathrm{BER}_{\mathrm{pre,th}}$ is set to 3×10^{-2}. Similar trends as for $R_s = 28$ GBaud in Fig. 8 can be observed for other symbol rates. Although BER values drop with a decrease in R_{s}, we should not neglect that the in-band crosstalk noise power, a dominant noise contributor, is bandwidth independent. The improvements in pre-FEC BER are simply not strong enough to unravel the physical layer impact. For 2-PAM and 8-PAM, the interdomain goodput degradation is almost symbol-rate independent. The strongest impact applies to 4-PAM where degradation varies between 17.7% (at $R_{\mathrm{s}} = 10$ GBaud) and 20.4% (at $R_{\mathrm{s}} = 28$ GBaud) at full load.

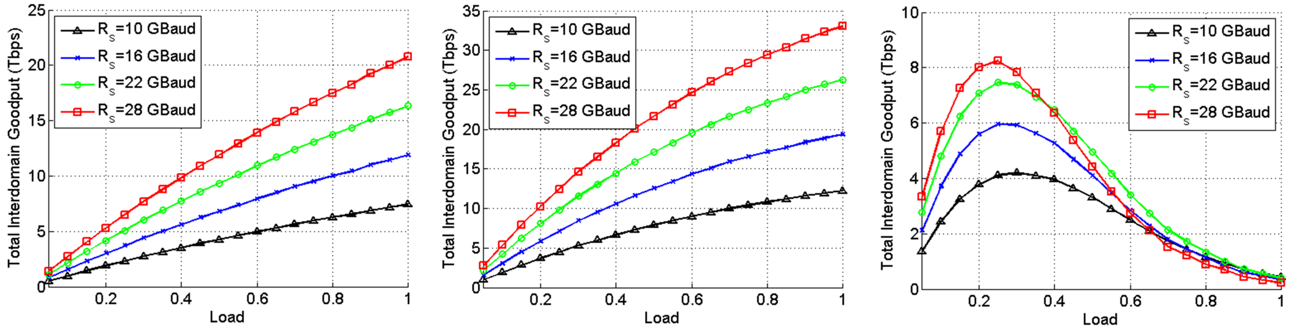
V. Conclusion

We developed a cross-layer framework for assessing the joint performance of pulse amplitude modulation and code rate adaptation in data center environments that support both wavelength routing and optical multicasting. Our analysis pointed to the significant impact of AWG crosstalk on higher-order PAM in distributed, wavelength-reuse architectures. We found that code rate adaptation based on short block-length codes is vital for high-order PAM signals that should traverse multiple routing stages.



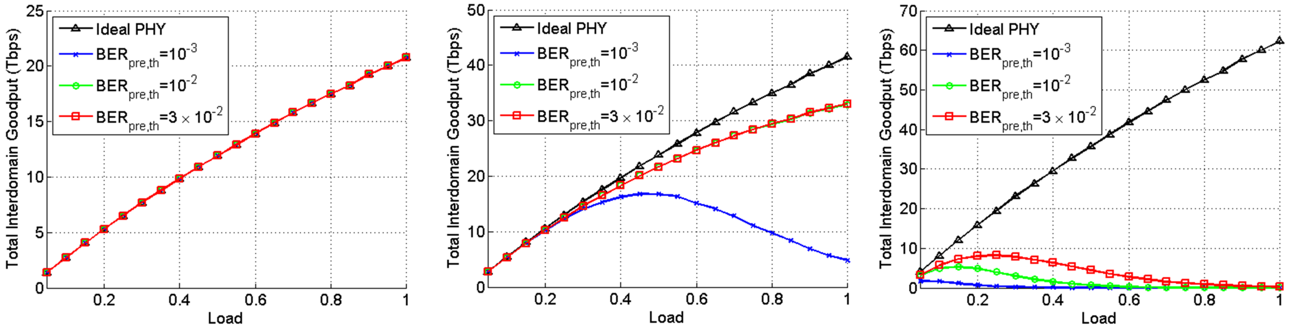
(a) (b) (c)

Fig. 7. Overall goodput versus load for (a) 2-PAM, (b) 4-PAM, and (c) 8-PAM signaling with FEC code rate adaptation. The symbol rate is fixed at 28 GBaud.



(a) (b) (c)

Fig. 8. Total interdomain goodput versus load for (a) 2-PAM, (b) 4-PAM, and (c) 8-PAM signaling.



(a) (b) (c)

Fig. 9. Impact of symbol rate on interdomain goodput for (a) 2-PAM, (b) 4-PAM, and (c) 8-PAM signaling. $\mathrm{BER}_{\mathrm{pre,th}}$ is set to 3×10^{-2}.

With regard to the settings in this paper, 8-PAM was found to provide the best performance for connections terminating in one broadcast domain, whereas 4-PAM resulted in higher effective bit rates for connections spanning two broadcast domains and a crosstalk-rich stage. PAM transceivers capable of adapting both modulation order and code rate in line with physical layer constraints and higher layer requirements can serve as a promising, disruptive technology for next-generation programmable data centers. Future work should investigate the joint optimization of code rate and modulation order for such transceivers, considering various code types (e.g., Bose, Ray-Chaudhuri and Hocquenghem (BCH) and convolutional codes) and block lengths. Besides, there is a need for a suite of optical data center architectures and scheduling algorithms that support multicast traffic, optimizing the cross-layer performance and costs.

Appendix

We present the mathematical framework for calculating the pre-FEC bit error rate of PAM signals. Let P_{S} be the average launch power of the transceivers in Fig. 3. For an M-PAM modulation scheme with a finite extinction ratio ER, we consider equidistant symbols and calculate $P_{\mathrm{S},i}$, i.e., the average launch power per symbol i, $i = 0, 1, ..., M-1$, as

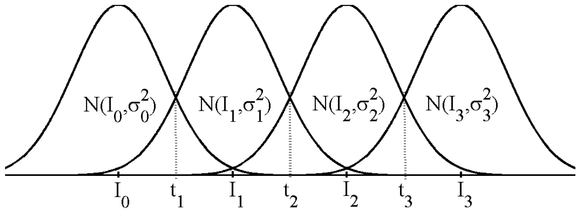


Fig. 10. Schematic of 4-PAM symbols with decision thresholds t_i.

$$P_{\mathrm{S},i} = \frac{2P_{\mathrm{S}}}{ER+1} \times \left[1 + \frac{i(ER-1)}{M-1}\right]. \tag{1}$$

The average received power per symbol i, $P_{\mathrm{R},i}$, is $P_{\mathrm{S},i}/L_{\mathrm{path}}$ where L_{path} is the intradomain/interdomain path loss and can be calculated as

$$L_{\mathrm{path}} = \begin{cases} \frac{L_{\mathrm{C}}L_{\mathrm{F}}}{G_{\mathrm{SOA}}} & \text{(intra)} \\ \frac{L_{\mathrm{C}}^2 L_{\mathrm{W}} L_{\mathrm{A}} L_{\mathrm{F}}}{G_{\mathrm{SOA}} G_{\mathrm{EDFA}}} & \text{(inter)} \end{cases}. \tag{2}$$

In (2), G_{SOA} and G_{EDFA} denote SOA and EDFA gains, respectively. For the definition of the other parameters, please refer to Table I. The average photocurrent per transmitted symbol can be calculated as $I_i = RP_{\mathrm{R},i}$. Fig. 10 denotes a set of 4-PAM symbols at the photodetection stage, assuming Gaussian noise distribution per symbol i with mean I_i and variance σ_i^2. The decision threshold t_i is determined through intersecting the distributions for symbols $i-1$ and i. For M-PAM modulation, let us define $t_0 = -\infty$ and $t_M = +\infty$. The probability of detecting symbol j given that symbol i has been transmitted can be calculated as [21]

$$P_{ij} = \frac{1}{2}\mathrm{erfc}\left(\frac{t_j - I_i}{\sigma_i\sqrt{2}}\right) - \frac{1}{2}\mathrm{erfc}\left(\frac{t_{j+1} - I_i}{\sigma_i\sqrt{2}}\right). \tag{3}$$

Hence, the BER for M-PAM modulation can be expressed as

$$\mathrm{BER} = \frac{1}{M}\sum_{i=0}^{M-1}\sum_{j=0,j\neq i}^{M-1}\frac{d_{ij}}{\log_2 M}P_{ij} \tag{4}$$

where d_{ij} is the Hamming distance between the labels for symbols i and j [21]. We consider Gray code labeling in this work.

The noise variance for each symbol i is

$$\sigma_i^2 = \sigma_{\mathrm{T},i}^2 + \sigma_{\mathrm{S},i}^2 + \sigma_{\mathrm{I},i}^2 + \sigma_{\mathrm{SIG-SP},i}^2 + \sigma_{\mathrm{SP-SP},i}^2 + \sigma_{\mathrm{SIG-IB},i}^2 + \sigma_{\mathrm{IB-IB},i}^2 + \sigma_{\mathrm{IB-SP},i}^2 + \sigma_{\mathrm{OB-OB},i}^2 \tag{5}$$

where the nine constituents denote thermal noise, shot noise, laser intensity noise, signal-spontaneous beat noise, spontaneous-spontaneous beat noise, signal-in-band crosstalk beat noise, in-band crosstalk-crosstalk beat noise, in-band crosstalk-spontaneous beat noise, and out-of-band crosstalk-crosstalk beat noise variance, respectively. Equations (6)–(10) are used to calculate the first five noise variance terms [34, ch. 5,6].

$$\sigma_{\mathrm{T},i}^2 = \frac{4k_B T F_{\mathrm{e}} B_{\mathrm{e}}}{R_{\mathrm{L}}}, \tag{6}$$

$$\sigma_{\mathrm{S},i}^2 = 2qI_iB_{\mathrm{e}}, \tag{7}$$

$$\sigma_{\mathrm{I},i}^2 = I_i^2 \times RIN \times B_{\mathrm{e}} \tag{8}$$

are the thermal, shot, and laser intensity noises, respectively. The signal-spontaneous and spontaneous-spontaneous beat noise variances are

$$\sigma_{\mathrm{SIG-SP},i}^2 = \frac{2RI_iP_{\mathrm{ASE}}B_{\mathrm{e}}}{B_{\mathrm{o}}}, \tag{9}$$

$$\sigma_{\mathrm{SP-SP},i}^2 = \frac{R^2P_{\mathrm{ASE}}^2\left(2B_{\mathrm{o}} - B_{\mathrm{e}}\right)B_{\mathrm{e}}}{2B_{\mathrm{o}}^2}. \tag{10}$$

In (9) and (10), P_{ASE} denotes the total ASE noise power that is applied to the receiver. For an interdomain connection, it is equal to

$$P_{\mathrm{ASE}} = \frac{\big(N_{\mathrm{SOA}}G_{\mathrm{EDFA}}/(L_{\mathrm{C}}L_{\mathrm{W}}L_{\mathrm{A}})\big) + N_{\mathrm{EDFA}}}{L_{\mathrm{C}}L_{\mathrm{F}}} \tag{11}$$

where N_{SOA} and N_{EDFA} are the ASE noise power generated by an SOA or an EDFA within the optical filter bandwidth, respectively. For an intradomain connection, $N_{\mathrm{EDFA}} = 0$. We calculate the noise generated by an optical amplifier with gain G as $N = Fh\nu(G-1)B_{\mathrm{o}}$ where F is the amplifier noise figure and ν is the optical carrier frequency [19].

We extend the analysis in [19] to derive the crosstalk beat noise expressions in (5). To do so, we need to consider the averaging over M possible symbols as well as the evolution of the out-of-band crosstalk noise in the absence of a receiver demultiplexer. Note that in this analysis we assume the power of signal and crosstalk terms to be concentrated solely on the optical carrier frequency (i.e., monochromatic field assumption). The total photocurrent for PAM symbol i in the presence of crosstalk terms can be expressed as [19]

$$I_{\mathrm{T},i} = R\left|\sqrt{P_{\mathrm{R},i}}e^{j(\omega_{\mathrm{S}}t+\theta_{\mathrm{R},i})} + \sum_{r=1}^{N_{\mathrm{IB}}}\sqrt{P_{\mathrm{IB},r}}e^{j(\omega_{\mathrm{S}}t+\theta_{\mathrm{IB},r})} + \sum_{s=1}^{N_{\mathrm{OB}}}\sqrt{P_{\mathrm{OB},s}}e^{j(\omega_{\mathrm{OB},s}t+\theta_{\mathrm{OB},s})}\right|^2. \tag{12}$$

In (12), the three terms correspond to the optical field of the desired signal, in-band crosstalk terms (developed due to the cyclic routing pattern of the AWG and coexisting on the same carrier frequency as the signal itself) and out-of-band crosstalk terms, respectively. ω_{S} is the angular carrier frequency of the signal as well as the in-band crosstalk terms. $P_{\mathrm{R},i}$ and $\theta_{\mathrm{R},i}$ are the received power and phase of the i^{th} PAM symbol. $P_{\mathrm{IB},r}$ and $\theta_{\mathrm{IB},r}$ denote the received power and phase of the r^{th} in-band crosstalk term. Similarly, $P_{\mathrm{OB},s}$, $\omega_{\mathrm{OB},s}$, and $\theta_{\mathrm{OB},s}$ denote the received power, frequency, and phase of the s^{th} out-of-band crosstalk term that leaks to the electrical receiver filter. N_{IB} ($\leq N-1$ for an $N \times N$ AWG) is the total number of in-band crosstalk terms. N_{OB} is the total number of out-of-band crosstalk terms that the desired symbol sees at its output port.

As the AWG is intrinsically a bandpass filter that is followed by the receiver filter, we only consider those interdomain terms that get routed due to the AWG routing pattern and neglect the impact of nonideal AWG filtering when considering out-of-band crosstalk terms.

The expansion of (12) results in several DC components followed by a large number of high-pass terms. The carrier frequencies of time-varying terms are integer multiples of the optical channel spacing. As the output of the photodetection stage is applied to a low-pass electrical filter (with bandwidth B_{e}), only the DC terms are assumed to contribute to the final photocurrent and high-pass beating terms are filtered away. Hence, we calculate the low-pass photocurrent for symbol i as

$$\begin{aligned} I_{\mathrm{LP},i} = RP_{\mathrm{R},i} + 2R\sum_{r=1}^{N_{\mathrm{IB}}}\sqrt{P_{\mathrm{R},i}P_{\mathrm{IB},r}}\cos\left(\theta_{\mathrm{R},i}-\theta_{\mathrm{IB},r}\right) \\ + R\left|\sum_{r=1}^{N_{\mathrm{IB}}}\sqrt{P_{\mathrm{IB},r}}e^{j\theta_{\mathrm{IB},r}}\right|^2 + R\sum_{s=1}^{N_{\mathrm{OB}}}P_{\mathrm{OB},s}. \end{aligned} \tag{13}$$

The variance of second, third, and fourth terms in (13) corresponds to signal-in-band crosstalk beat noise variance, in-band crosstalk-crosstalk beat noise variance, and out-of-band crosstalk-crosstalk beat noise variance, respectively. To calculate these variances, we consider equiprobable symbols, uniformly distributed phases, and worst case polarization alignment for crosstalk terms and the signal. We define the variable P_{IB} (i.e., the average optical in-band crosstalk power) as

$$P_{\mathrm{IB}} = \sum_{r=1}^{N_{\mathrm{IB}}}P_{\mathrm{IB},r} = \left(N_{\mathrm{AX}}R_{\mathrm{AX}} + N_{\mathrm{NX}}R_{\mathrm{NX}}\right)P_{\mathrm{R,inter}}. \tag{14}$$

where $P_{\mathrm{R,inter}}$ is the average received power per interdomain connection, N_{AX} is the number of adjacent crosstalk terms in the AWG, N_{NX} is the number of nonadjacent crosstalk terms, and R_{AX} and R_{NX} denote adjacent and nonadjacent AWG crosstalk power ratios, respectively. Now, the in-band signal-crosstalk and crosstalk-crosstalk beat noise variances can be expressed as

$$\sigma^2_{\mathrm{SIG-IB},i} = \frac{2I_i^2P_{\mathrm{IB}}}{R}, \tag{15}$$

$$\sigma^2_{\mathrm{IB-IB},i} = R^2P_{\mathrm{IB}}^2. \tag{16}$$

We assume the beating of the in-band crosstalk terms and the ASE noise to follow the same pattern as the beating of the signal and the ASE noise and use (9) and (16) to derive the crosstalk-spontaneous beat noise variance

$$\sigma^2_{\mathrm{IB-SP},i} = \frac{2R^2P_{\mathrm{IB}}P_{\mathrm{ASE}}B_{\mathrm{e}}}{B_{\mathrm{o}}}. \tag{17}$$

As discussed in the beginning of Section IV, we assume a Gaussian amplitude response for the electrical receiver filter (matched filter) and calculate the out-of-band crosstalk-crosstalk beat noise variance as

$$\begin{aligned} \sigma^2_{\mathrm{OB-OB},i} &= R^2\sum_{s=1}^{N_{\mathrm{OB}}}P^2_{\mathrm{OB},s} \\ &= \sum_{s=1}^{N_{\mathrm{OB}}}\left(\frac{1}{M}\sum_{j=0}^{M-1}I_j^2\right)H^2(f_S-f_s) \end{aligned} \tag{18}$$

where S denotes the wavelength index of the desired signal. $H(f_S-f_s)$ is a power transfer function for estimating the ratio of power leaked from an interfering channel s (with optical carrier frequency f_s) onto the desired channel S (with optical carrier frequency f_S). Defining $\Delta f = f_S - f_s$, we express this function as

$$H(\Delta f) = e^{-4\ln 2\left(\frac{\Delta f}{B_e}\right)^2}. \tag{19}$$

The full width at half maximum (FWHM) of $H(\Delta f)$ can be calculated as B_{e}, which corresponds to the electrical receiver filter bandwidth.

References

[1] P. Samadi, V. Gupta, J. Xu, H. Wang, G. Zussman, and K. Bergman, "Optical multicast system for data center networks," *Optics Express*, vol. 23, no. 17, pp. 22 162–22 180, Aug. 2015.

[2] D. Li, M. Xu, Y. Liu, X. Xie, Y. Cui, J. Wang, and G. Chen, "Reliable multicast in data center networks," *IEEE Transactions on Computers*, vol. 63, no. 8, pp. 2011–2024, Aug. 2014.

[3] Z. Guo, J. Duan, and Y. Yang, "On-line multicast scheduling with bounded congestion in fat-tree data center networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 1, pp. 102–115, Jan. 2014.

[4] W.-K. Jia, "A scalable multicast source routing architecture for data center networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 1, pp. 116–123, Jan. 2014.

[5] D. Li, Y. Li, J. Wu, S. Su, and J. Yu, "ESM: efficient and scalable data center multicast routing," *IEEE/ACM Transactions on Networking (TON)*, vol. 20, no. 3, pp. 944–955, Jun. 2012.

[6] C. Kachris, K. Kanonakis, and I. Tomkos, "Optical interconnection networks in data centers: recent trends and future challenges," *IEEE Communications Magazine*, vol. 51, no. 9, pp. 39–45, Sep. 2013.

[7] F. Yan, W. Hu, W. Sun, W. Guo, Y. Jin, H. He, Y. Dong, and S. Xiao, "Nonblocking four-stage multicast network for multicast-capable optical cross connects," *Journal of Lightwave Technology*, vol. 27, no. 17, pp. 3923–3932, Sep. 2009.

[8] Y. Yang, J. Wang, and C. Qiao, "Nonblocking WDM multicast switching networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 11, no. 12, pp. 1274–1287, Dec. 2000.

[9] W. Ni, C. Huang, Y. L. Liu, W. Li, K.-W. Leong, and J. Wu, "POXN: a new passive optical cross-connection network for low-cost power-efficient datacenters," *Journal of Lightwave Technology*, vol. 32, no. 8, pp. 1482–1500, Apr. 2014.

[10] H. Wang, Y. Xia, K. Bergman, T. S. E. Ng, S. Sahu, and K. Sripanidkulchai, "Rethinking the physical layer of data center networks of the next decade: using optics to enable efficient *-cast connectivity," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 3, pp. 52–58, Jul. 2013.

[11] J. Chen, Y. Gong, M. Fiorani, and S. Aleksic, "Optical interconnects at the top of the rack for energy-efficient data centers," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 140–148, Aug. 2015.

[12] K. Keykhosravi, H. Rastegarfar, and E. Agrell, "Multicast scheduling for optical data center switches with tunablity constraints," in *IEEE International Conference on Computing, Networking and Communications (accepted)*, Jan. 2017.

[13] M. Maier, M. Scheutzow, and M. Reisslein, "The arrayed-waveguide grating-based single-hop WDM network: an architecture for efficient multicasting," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 9, pp. 1414–1432, Nov. 2003.

[14] Q. Huang and W.-D. Zhong, "Wavelength-routed optical multicast packet switch with improved performance," *Journal of Lightwave Technology*, vol. 27, no. 24, pp. 5657–5664, Dec. 2009.

[15] H. Rastegarfar, L. A. Rusch, and A. Leon-Garcia, "WDM recirculation buffer-based optical fabric for scalable cloud computing," *Journal of Lightwave Technology*, vol. 32, no. 21, pp. 3451–3465, Nov. 2014.

[16] Y. Yin, R. Proietti, X. Ye, C. J. Nitta, V. Akella, and S. J. B. Yoo, "LIONS: an AWGR-based low-latency optical switch for high performance computing and data centers," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 19, no. 2, Mar./Apr. 2013.

[17] K.-I. Sato, H. Hasegawa, T. Niwa, and T. Watanabe, "A large-scale wavelength routing optical switch for data center networks," *IEEE Communications Magazine*, vol. 51, no. 9, pp. 46–52, Sep. 2013.

[18] Z. Cao, R. Proietti, M. Clements, and S. J. B. Yoo, "Experimental demonstration of flexible bandwidth optical data center core network with all-to-all interconnectivity," *Journal of Lightwave Technology*, vol. 33, no. 8, pp. 1578–1585, Apr. 2015.

[19] H. Rastegarfar, A. Leon-Garcia, S. LaRochelle, and L. A. Rusch, "Cross-layer performance analysis of recirculation buffers for optical data centers," *IEEE/OSA Journal of Lightwave Technology*, vol. 31, no. 3, pp. 432–445, Feb. 2013.

[20] C. Cole, "Beyond 100G client optics," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. s58–s66, Feb. 2012.

[21] K. Szczerba, P. Westbergh, J. Karout, J. S. Gustavsson, Å. Haglund, M. Karlsson, P. A. Andrekson, E. Agrell, and A. Larsson, "4-PAM for high-speed short-range optical communications," *Journal of Optical Communications and Networking*, vol. 4, no. 11, pp. 885–894, Nov. 2012.

[22] H. Mardoyan, M. A. Mestre, R. Rios-Müller, A. Konczykowska, J. Renaudier, F. Jorge, B. Duval, J.-Y. Dupuy, A. Ghazisaeidi, P. Jennevé, M. Achouche, and S. Bigo, "Single carrier 168-Gb/s line-rate PAM direct detection transmission using high-speed selector power DAC for optical interconnects," *Journal of Lightwave Technology*, vol. 34, no. 7, pp. 1593–1598, Apr. 2016.

[23] S. Kanazawa, T. Fujisawa, K. Takahata, Y. Nakanishi, H. Yamazaki, Y. Ueda, W. Kobayashi, Y. Muramoto, H. Ishii, and H. Sanjoh, "56-Gbaud 4-PAM (112-Gbit/s) operation of flip-chip interconnection lumped-electrode EADFB laser module for equalizer-free transmission," in *Optical Fiber Communication Conference*, Mar. 2016, paper W4J–1.

[24] J. Lee, S. Shahramian, N. Kaneda, Y. Baeyens, J. Sinsky, L. Buhl, J. Weiner, U.-V. Koc, A. Konczykowska, J.-Y. Dupuy, F. Jorge, R. Aroca, T. Pfau, and Y.-K. Chen, "Demonstration of 112-Gbit/s optical transmission using 56 GBaud PAM-4 driver and clock-and-data recovery ICs," in *Proc. European Conference on Optical Communications (ECOC)*, Sep. 2015, paper 0604.

[25] D. A. A. Mello, A. N. Barreto, T. C. de Lima, T. F. Portela, L. Beygi, and J. M. Kahn, "Optical networking with variable-code-rate transceivers," *Journal of Lightwave Technology*, vol. 32, no. 2, pp. 257–266, 2014.

[26] L. Yan, E. Agrell, and H. Wymeersch, "Sensitivity comparison of time domain hybrid modulation and rate adaptive coding," in *Optical Fiber Communication Conference*, Mar. 2016, paper W1I.3.

[27] R. Gaudino, G. A. G. Castillo, F. Neri, and J. M. Finochietto, "Can simple optical switching fabrics scale to terabit per second switch capacities?" *IEEE/OSA Journal of Optical Communications and Networking*, vol. 1, no. 3, pp. B56–B69, Aug. 2009.

[28] Q. Xu, H. Rastegarfar, Y. B. M'Sallem, A. Leon-Garcia, S. LaRochelle, and L. A. Rusch, "Analysis of large-scale multi-stage all-optical packet switching routers," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 4, no. 5, pp. 412–425, May 2012.

[29] Long-Reach DWDM SFP Transceiver, https://www.finisar.com/sites/default/files/downloads/fwlf1631r_long-reach_dwdm_sfp_transceiver_spec_revc1.pdf, Oct. 2016.

[30] +/-800 ps/nm (40km) Tunable XFP Optical Transceiver, https://www.finisar.com/sites/default/files/downloads/finisar_ftlx6611mcc_xx_and_ftlx6614mcc_xx_-800_ps_nm_40km_tunable_xfp_optical_transceiver_product_specification_rev_b1.pdf, Oct. 2016.

[31] Wavelength Selective Switch 1x2/2x1, https://www.components76.com/pdf_dir/63_pr_file_1.pdf, Oct. 2016.

[32] WaveShaper 16000S, https://www.finisar.com/sites/default/files/downloads/waveshaper_16000s_product_brief_11_14_0.pdf, Oct. 2016.

[33] 64x64-Channel Uniform-Loss and Cyclic-Frequency Arrayed-Waveguide Grating Router, http://www.ntt.co.jp/dic/phlab/eng/theme/2004/e2004_10_03.pdf, Oct. 2016.

[34] G. P. Agrawal, *Lightwave Technology: Telecommunication Systems*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

[35] G.957 : Optical interfaces for equipments and systems relating to the synchronous digital hierarchy, https://www.itu.int/rec/T-REC-G.957-200603-I/en, Jan. 2017.

[36] L. Beygi, E. Agrell, P. Johannisson, M. Karlsson, and H. Wymeersch, "A discrete-time model for uncompensated single-channel fiber-optical links," *IEEE Transactions on Communications*, vol. 60, no. 11, pp. 3440–3450, Nov. 2012.

[37] E. Säckinger, *Broadband Circuits for Optical Fiber Communication*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

[38] T. Benson, A. Akella, and D. A. Maltz, "Network traffic characteristics of data centers in the wild," in *Proc. The 10th ACM SIGCOMM Conference on Internet Measurement*, Nov. 2010, pp. 267–280.

[39] ITU-T G.975 : Forward error correction for submarine systems, https://www.itu.int/rec/T-REC-G.975-200010-I/en, Oct. 2016.

[40] W. J. E. Ebel and W. H. Tranter, "The performance of Reed-Solomon codes on a bursty-noise channel," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 298–306, Feb./Mar./Apr. 1995.

[41] Y. Yan, G. M. Saridis, Y. Shu, B. R. Rofoee, S. Yan, M. Arslan, T. Bradley, N. V. Wheeler, N. H.-L. Wong, F. Poletti, M. N. Petrovich, D. J. Richardson, S. Poole, G. Zervas, and D. Simeonidou, "All-optical programmable disaggregated data centre network realized by FPGA-based switch and interface card," *Journal of Lightwave Technology*, vol. 34, no. 8, pp. 1925–1932, Apr. 2016.