This article has been accepted for publication in Proc. of SPIE Vol. 13903. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1117/12.3078430

Radix-expanded densely-spaced WDM technology for co-packaged optics switches

R. Palmer*[a], C. Antony[b], B. Zheng[c], M. P. G. Rombouts[c], J. Sarkis[a], N. Calabretta[c], P. D. Townsend[b], G. Talli[a], M. Kuschnerov[a]
[a]Huawei Heisenberg Research Center (Munich), Riesstr. 25, 80992 Munich, Germany;
[b]Tyndall National Institute (Ireland);
[c]Eindhoven University of Technology

ABSTRACT

Optical interconnects are a core technology for scaling AI/ML networks by linking NPU/GPUs directly or through ethernet switches with each other. With switch and link capacities doubling every 2 years, CPO (co-packaged optics) and OIO (optical Input/Output) connectivity solutions are emerging for tackling multiple challenges simultaneously: Reduction of power consumption and latency, and the increase of shoreline bandwidth density. In this work we explore the scaling limitations of DR-based (Direct Reach, parallel fiber output) CPO devices and demonstrate how densely-spaced WDM CPO variants can overcome this bottleneck, while simultaneously supporting a maximum switch radix, within the framework of Horizon Europe project ADOPTION. We demonstrate the first generation of (de-)multiplexer technology and their impact on transmission quality and network architecture.

Keywords: Co-packaged optics, port breakout, WDM, AWG, demultiplexing, PLC

1. INTRODUCTION

High-bandwidth optical interconnects are a core technology for scaling AI/ML networks. The number of GPUs in a training cluster can reach a hundred thousand for state-of-the-art clusters used for AI model development by major tech companies. GPU interconnects can be grouped in two domains: scale-up and scale-out. In the scale-up domain, a moderate number of GPUs is grouped inside a rack. Purely proprietary electrical switch/interconnect solutions are used to establish high-bandwidth interconnects between these GPUs. One example is the NVIDIA GB200 NVL72 system that combines 72 GPUs in the scale-up domain to a compute node. At 200 Gb/s, copper cables have a very limited reach of 1-3m, depending on whether passive or active copper cables are used. As compute power of GPUs increases, even electrical connections in the scale-up domain will approach their physical reach limits and optical interconnects will become essential. Interconnection between multiple compute nodes is realized in the scale-out domain, already relying on high-capacity network switches with optical I/Os. The capacity of these network switches has historically doubled every two years, see

Table 1. This is achieved by a doubling of the lane data rate and by an interleaved doubling of the number of channels every four years. It can be projected that the first 204.8 Tbps switches will be released in 2027, running at 224 Gbps per lane at 112 GBd PAM4 [1] with a total of 1024 lanes per switch. The combination of fast SERDES rates and large channel counts fueled the development of linear optical transceivers and co-packaged optics transceivers [2], [3]. Getting rid of electrical retiming reduces the power consumption by 50% as compared to a retimed (pluggable) module. CPO modules target shoreline bandwidth densities exceeding 1 Tbps/mm [4]. Such high shoreline bandwidth densities not only require ultra-compact transmitters and receivers, but also space efficient fiber-array interconnects. We will explore the fiber-array-unit (FAU)-related shoreline density limits for DR-based PIC's (with parallel fiber output) and derive when WDM has to be used for CPO for meeting the density requirements.

*robert.palmer@huawei.com

This article has been accepted for publication in Proc. of SPIE Vol. 13903. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1117/12.3078430

Table 1. Evolution and extrapolation of switch ASIC capacity and breakdown of associated lane rate and number of lanes.

Year of release	Switch capacity (Tbps)	Lane net rate (Gbps)	Number of lanes	Notes
2018	12.8	50	256	Broadcom Tomahawk 3
2020	25.6	100	256	Broadcom Tomahawk 4
2023	51.2	100	512	Broadcom Tomahawk 5
2025	102.4	200	512	Broadcom Tomahawk 6
* 2027	204.8	200	1024	* Extrapolated
* 2029	409.6	400	1024	* Extrapolated
* 2031	819.2	400	2048	* Extrapolated

2. SHORELINE BANDWIDTH DENSITY REQUIREMENTS

Due to the high lane rates of modern switch ASICs, the CPOs must be placed in close proximity to the switch. The RF attenuation is setting an upper limit to the maximum RF trace length. Consequently, there is a maximum number of CPOs that can surround the switch ASIC, see Figure 1, as the number of CPO modules and their respective width is determining the length of the shortest and of the longest RF traces. According to the "Next Generation CEI-224G Framework" (OIF-FD-CEI-224G-01.0) [5], today's RF package losses are roughly 2 dB/cm at 56 GHz. In an optimistic forecast, substrate loss might improve towards 1 dB/cm at 56 GHz and 2 dB/cm at 112 GHz. For keeping the roll-off at Nyquist frequency below 3 dB, this sets a trace length limit of 3 cm for 112 GBd PAM4 and 1.5 cm for 224 GBd PAM4.

We vary the number of CPO modules around the switch ASIC, as indicated in Figure 1, and study the impact of the CPO width on RF trace length. The resulting shoreline bandwidth density requirements of the CPO modules are summarized in Table 2. We can see that the next generation of 204.8 Tbps switches will require shoreline bandwidth densities exceeding 0.73 Tbps/mm. Assuming that the SERDES rate can be scaled by another factor of two in 2029, the required shoreline bandwidth densities will exceed 2.1 Tbps/mm and 4.2 Tbps/mm for 409.6 Tbps switches and 819.2 Tbps switches, respectively.

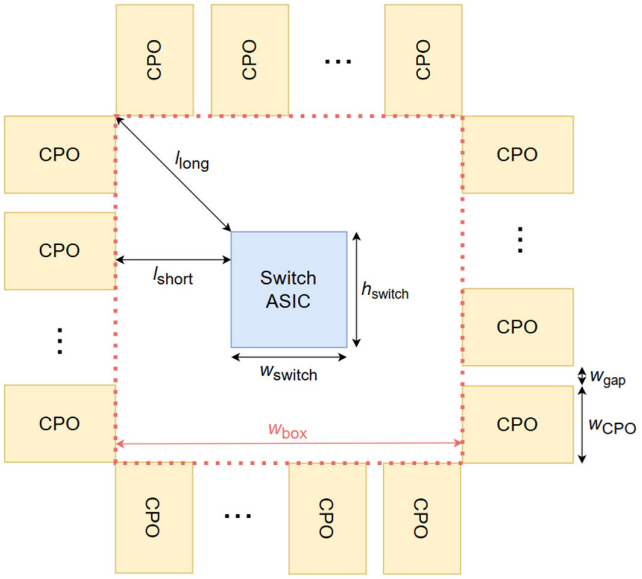

Figure 1. Derivation of shoreline bandwidth density requirements for CPO. The number of CPO modules and their respective width is determining the length of the shortest and the longest RF traces. Since the amount of tolerable RF loss for linear optics is limited, this sets a lower limit for the required shoreline bandwidth density.

This article has been accepted for publication in Proc. of SPIE Vol. 13903. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1117/12.3078430

Table 2. Shoreline bandwidth density requirements for CPO for supporting 204.8 Tbps and 409.6 Tbps switches. We assume switch ASIC dimensions of 33 mm x 26 mm (w_{switch} x h_{switch}), a 2 mm gap (w_{gap}) between the CPOs and RF substrate losses of 1 dB/cm at 56 GHz and 2 dB/cm at 112 GHz. The CPO width w_{CPO} is chosen such that the loss of the longest trace at the Nyquist frequency is 3 dB.

Switch capacity (Tbps)	Lane net rate (Gbps)	# of CPO modules	CPO width (mm)	Shortest RF path (mm)	Longest RF path (mm)	CPO capacity (Tbps)	Shoreline density (Tbps/mm)
204.8	200	8	34.9	19.4	30.0	25.60	0.73
204.8	200	12	22.6	19.4	30.0	17.07	0.76
204.8	200	16	16.45	19.4	30.0	12.80	0.78
409.6	400	8	24.2	8.7	15.0	51.20	2.12
409.6	400	12	15.48	8.7	15.0	34.13	2.20
409.6	400	16	11.1	8.7	15.0	25.60	2.31
819.2	400	8	24.2	8.7	15.0	102.40	4.23
819.2	400	12	15.48	8.7	15.0	68.27	4.41
819.2	400	16	11.1	8.7	15.0	51.20	4.61

In the following, we will assume that the density of the optical components and of the electrical IOs will not be the bottleneck for the shoreline density. This could be achieved e.g. by placing ring modulators and photo diodes on a 2D grid. Under this condition, the optical IO density will set a limit to the shoreline bandwidth density of the CPO. A standard compact fiber array unit (FAU) has a fiber pitch of 127 µm. To improve the density of the fiber interconnect, two strategies have been investigated in literature: 1) pitch-reduced FAUs based on reduced cladding fibers [8] or PLC pitch converters [9] to reduce the optical I/O pitch of the CPO device or 2) the per fiber capacity is increased through WDM [7], [10]. While the reduction of the fiber pitch may improve the I/O density by a factor of 2-3, the use of WDM has the potential to increase the density by a factor of 10 or beyond.

We study and compare different CPO IO approaches and derive the respective threshold for their shoreline bandwidth density. In particular, we study a fiber pitch of 127 µm and 84 µm, a 1-row FAU and a 2-row FAU, and single wavelength systems vs. 8-wavelength systems, and various combinations of these parameters. The resulting shoreline bandwidth density limit depends on the lane rate of the switch ASIC. A summary of the study if depicted in Table 3, and Table 4 for 200 Gbps per lane and 400 Gbps per lane, respectively. We can see that a single row, pitch-reduced fiber array unit is sufficient to reach the size and capacity requirements for 204.8 Tbps switches, but is incompatible with 409.6 Tbps switches. A 2D-fiber array can overcome this bottleneck and enable even up to 4.12 Tbps/mm when using reduced-cladding fibers. Meeting requirements for 819.2 Tbps switches in ~2031 would require even more fiber rows in the FAU. However, it is yet to be seen, if scaling the number of fiber rows is technically feasible. An increase of the number of fiber rows may increase the core concentricity error of the FAU, which would lead to an increase in variation of the fiber-chip coupling loss. The alternative architecture is to switch from parallel outputs to WDM. Table 4 illustrates that a system with 8 wavelengths can boost the shoreline density product to >15 Tbps/mm. Furthermore, the number of fibers of an 819.2 Tbps switch that need to be managed inside the switch housing may become problematic and a reduction of the number of fibers through WDM may be preferable.

This article has been accepted for publication in Proc. of SPIE Vol. 13903. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1117/12.3078430

Table 3. CPO variants operating at 200G net per lane for 204.8 Tbps switches. Here the FAU width is determining the shoreline bandwidth density. Various flavors of CPO are compared utilizing regular or pitch-reduced FAUs, dual fiber row FAUs, or WDM. The FAU bundles polarization maintaining fibers (PMF) together with regular single mode fibers (SMF).

# of fiber rows	Fiber pitch (µm)	Fiber ∅ (µm)	# of TX/RX fiber pairs	# of λ	# of SMF	# of PMF	FAU width (mm)	CPO capacity (Tbps)	Shoreline density (Tbps/mm)
1	127	125	64	1	128	8	18.27	12.8	0.70
1	84	80	64	1	128	8	12.42	12.8	1.03
2	127	125	128	1	256	16	18.27	25.6	1.40
2	84	80	128	1	256	16	12.42	25.6	2.06
1	127	125	32	8	64	8	10.14	51.2	5.05
1	84	80	32	8	64	8	7.04	51.2	7.27
1	127	125	64	8	128	16	19.29	102.4	5.31
1	84	80	64	8	128	16	13.09	102.4	7.82

Table 4. CPO variants operating at 400G net per lane for 409.6 Tbps and 819.2 Tbps switches. Here the FAU width is determining the shoreline bandwidth density. Various flavors of CPO are compared utilizing regular or pitch-reduced FAUs, dual fiber row FAUs, or WDM. The FAU bundles polarization maintaining fibers (PMF) together with regular single mode fibers (SMF).

# of fiber rows	Fiber pitch (µm)	Fiber ∅ (µm)	# of TX/RX fiber pairs	# of λ	# of SMF	# of PMF	FAU width (mm)	CPO capacity (Tbps)	Shoreline density (Tbps/mm)
1	127	125	64	1	128	8	18.27	25.6	1.40
1	84	80	64	1	128	8	12.42	25.6	2.06
2	127	125	128	1	256	16	18.27	51.2	2.80
2	84	80	128	1	256	16	12.42	51.2	4.12
1	127	125	32	8	64	8	10.14	102.4	10.10
1	84	80	32	8	64	8	7.044	102.4	14.54
1	127	125	64	8	128	16	19.29	204.8	10.62
1	84	80	64	8	128	16	13.09	204.8	15.64

3. OPTICAL PORT BREAKOUT CONCEPT

We have shown that WDM can significantly boost the shoreline bandwidth density of CPO modules, while simultaneously reducing the number of fiber IOs. The reduction of fibers, increases the capacity per TX/RX fiber pair, but at the same time it reduces the achievable radix of the switch. For example, an 819.2 Tbps switch may "only" have a radix of 256 instead of 2024. Networks architectures that benefit from a larger radix can utilize a so-called "optical port breakout device" to map the individual channels to discrete fibers. As illustrated in Figure 2(a), such a port breakout device can be unamplified or amplified, in case of insufficient margin in the link budget. In its simplest form, the optical port breakout module is a passive MUX/DEMUX, e.g., an arrayed waveguide grating (AWG). Such AWG can be paired with a semiconductor optical amplifier (SOA) to realize an amplified port breakout version. The channel spacing of the AWG must be chosen narrow enough, such that all channels are within the SOA bandwidth. At the same time, the 1dB bandwidth of the AWG must be wide enough to support 112 GBd or 224 GBd signals. Consequently, a 400 GHz grid, and a 300 GHz 1dB bandwidth have been chosen. The selected wavelength plan is depicted in Table 5. An amplified optical port breakout will add optical noise to the signal. The impact of the optical noise on the receiver sensitivity was studied with an analytical receiver model [6] implemented in Python and is illustrated in Figure 2(b). We find that the in-band OSNR of the signal after the SOA must remain above 33.7 dB and 36.7dB for keeping the RX sensitivity penalty below 0.5 dB for 112 GBd NRZ and for 224 GBd NRZ, respectively. For PAM4, the OSNR must be above 43.6 dB and 47.2 dB, for 112 GBd and 224 GBd, respectively. In DEMUX direction, assuming a typical SOA NF of 7 dB, the input power into the SOA must be above -14 dBm and above -4 dBm for 112 GBd NRZ and PAM4, respectively. In MUX direction, the signal is propagating

This article has been accepted for publication in Proc. of SPIE Vol. 13903. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1117/12.3078430

through the AWG before entering the SOA, hence, the AWG insertion loss (IL) has an impact on the output OSNR of the SOA. Assuming an AWG IL of ~3.5 dB, the required input power into the AWG (MUX-direction) must be above -10.5 dBm and above -0.5 dBm for NRZ and PAM4, respectively. For 224 GBd the input power requirements increase by 3 dB.

Crosstalk between individual WDM channels can further degrade the SNR of the signal. We approximate the integrated crosstalk of all neighboring channels as a Gaussian noise term that interferes with our channel under test and we use our analytical model to study the impact of the integrated crosstalk on the receiver sensitivity, see Figure 2(c). We find that the sensitivity penalty for PAM4 is below 0.2 dB if the integrated crosstalk is below -28.7 dB. For keeping the crosstalk related penalty below 0.5 dB, the integrated crosstalk from all channels must be below -17 dB for NRZ and below -25 dB for PAM4, respectively.

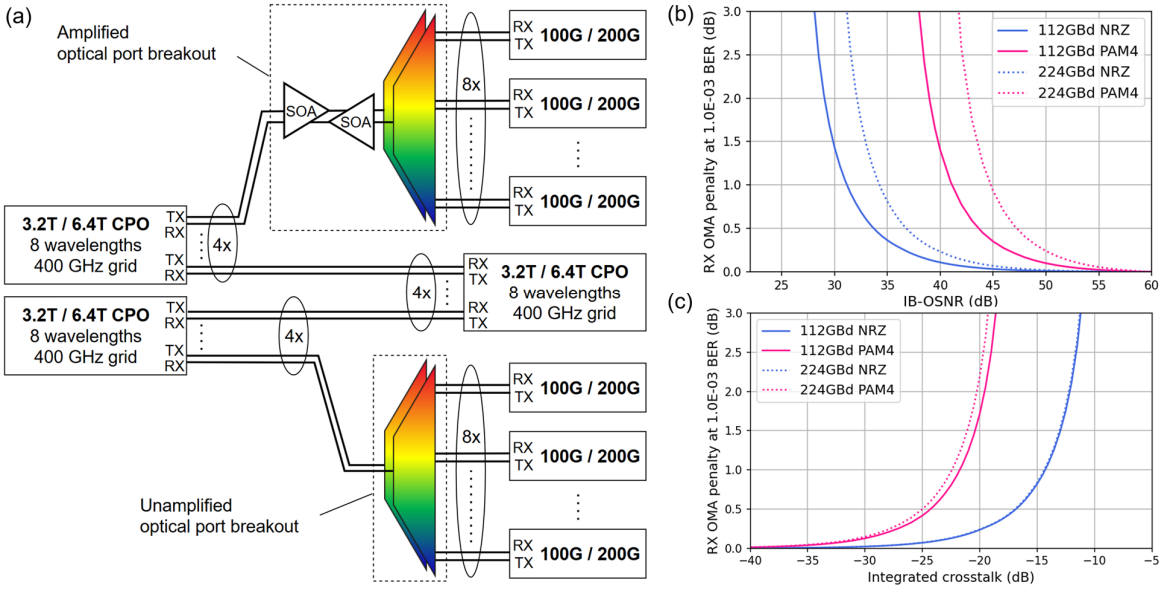

Figure 2. (a) Example configurations of the ADOPTION CPO modules. Each CPO has 4 TX/RX pairs, each pair is transmitting 1.6 Tbps per direction. This way, each CPO can connect to another CPO with 6.4 Tbps capacity, or to multiple CPO at a granularity of 1.6 Tbps. An increase of the CPO radix can be achieved by spectrally separating the individual channels with an amplified or unamplified port breakout module. (b) An amplified optical port breakout module is adding ASE noise to the signal, which is degrading the SNR. For keeping the OSNR-related penalty below 0.5 dB, an OSNR value must be above 33 dB for NRZ and 43 dB for PAM4, respectively. (c) For keeping the crosstalk related penalty below 0.2 dB, the integrated crosstalk from all channels must be below -20.8 dB for NRZ and below -28.7 dB for PAM4, respectively. For keeping the crosstalk related penalty below 0.5 dB, the integrated crosstalk from all channels must be below -17 dB for NRZ and below -25 dB for PAM4, respectively.

Table 5. ADOPTION wavelength plan. 8 wavelengths on a 400 GHz grid in agreement with DWDM-MSA [7].

Channel	Center frequency	Center wavelength
1	231.4 THz	1295.56 nm
2	231.0 THz	1297.80 nm
3	230.6 THz	1300.05 nm
4	230.2 THz	1302.31 nm
5	229.8 THz	1304.58 nm
6	229.4 THz	1306.85 nm
7	229.0 THz	1309.14 nm
8	228.6 THz	1311.43 nm

This article has been accepted for publication in Proc. of SPIE Vol. 13903. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1117/12.3078430

4. OPTICAL PORT BREAKOUT PERFORMANCE

An unamplified version of the port breakout module has been realized as a cascaded AWG in silica-on-silicon. The details of the component have been reported in [8]. The transmission spectra of the 8-port cascaded AWG with 400 GHz channel spacing is depicted in Figure 3. The wavelength plan of the AWG is provided in Table 5. The center frequencies of the AWG channels deviate by less than 40 GHz from the targeted values in in Table 5. We find an excellent 1 dB and 20 dB bandwidth of 307 GHz and 469 GHz, respectively. The rectangular passband shape can be expressed by its 1 dB to 20 dB bandwidth ratio of 0.65. The fiber-to-fiber IL of the pigtailed AWG is (3.1 ± 0.2) dB and the PDL is 0.1 dB. We evaluate the integrated crosstalk by summing the crosstalk of the 7 neighbor channels. The crosstalk of each neighbor is computed by integrating over 200 GHz or 325 GHz spans, for 112 GBd and 224 GBd, respectively, that are centered at the target frequencies as defined in Table 5. The spans include 75 GHz margin for thermal drift. We find an integrated crosstalk of -43 dB for 112 GBd and -37.7 dB for 224 GBd, which is sufficient for supporting both NRZ and PAM4. With such low crosstalk between the channels, such an AWG could also scale for (de-)multiplexing beyond 32 wavelength. A comparison between the achieved device performance and the targeted one is provided in Table 6.

We experimentally validate that the impact of crosstalk between the channels is indeed negligible by transmitting PAM4 data through the AWG. Due to restrictions of the experimental setup, the symbol rate of the channel under test was limited to 53 GBd. Four different experiments were performed as illustrated in Figure 4. Initially, the B2B performance is characterized according to the drawing in Figure 4(a). We find a receiver sensitivity at KP4-FEC threshold of -7.5dBm, see Figure 4(e). In experiments 2 and 3, we add a MUX AWG to the setup, see Figure 4(b), and a DEMUX AWG, see Figure 4(c). No filtering related penalty can be observed from this MUX and DEMUX AWG. Finally, we add neighboring channels to the MUX, see Figure 4(d). Due to limited HW availability, 4 of these neighboring channels were modulated with 25 GBd PAM4, while the other three neighboring channels were pure CW lasers. As expected, we find no penalty due to optical crosstalk from the AWG – the receiver sensitivity remained at -7.5 dBm also for the WDM scenario.

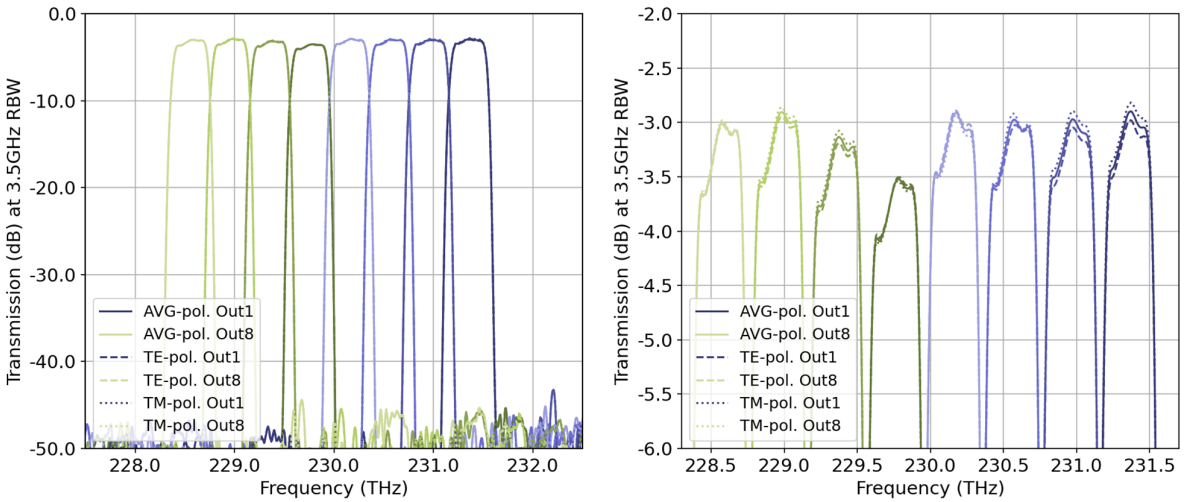

Figure 3. Transmission spectra of the 8 outputs of a 400 GHz AWG. The 1dB bandwidth is 307.1 GHz, the grid is 399.75 GHz, the IL is (3.1 ± 0.2) dB and the PDL is 0.1 dB. The integrated crosstalk for 112 GBd signals is -43 dB. For 224 GBd signals, the crosstalk increases to -37.7 dB.

This article has been accepted for publication in Proc. of SPIE Vol. 13903. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1117/12.3078430

Table 6. Comparison between achieved AWG device performance and targeted device performance. The crosstalk targets refer to 0.2 dB penalties.

Parameter	**Measured (112 GBd / 224 GBd)**	**112 GBd Target (NRZ / PAM4)**	**224 GBd Target (NRZ / PAM4)**
Insertion loss	(3.1 ± 0.2) dB	< 3.5 dB	< 3.5 dB
Polarization dependent loss	0.1 dB	-	-
1dB bandwidth	307.1 GHz	> 140 GHz	> 280 GHz
Grid	399.75 GHz	400 GHz	400 GHz
Integrated crosstalk of 7 neighboring channels	< -43.0 dB / -37.7 dB	< -20.8 dB / < -28.0 dB	< -20.8 dB / < -28.7 dB
(Extrapolated integrated crosstalk of 31 neighboring channels)	(< -36.5 dB / -31.2 dB)	(< -20.8 dB / < -28.0 dB)	(< -20.8 dB / < -28.7 dB)

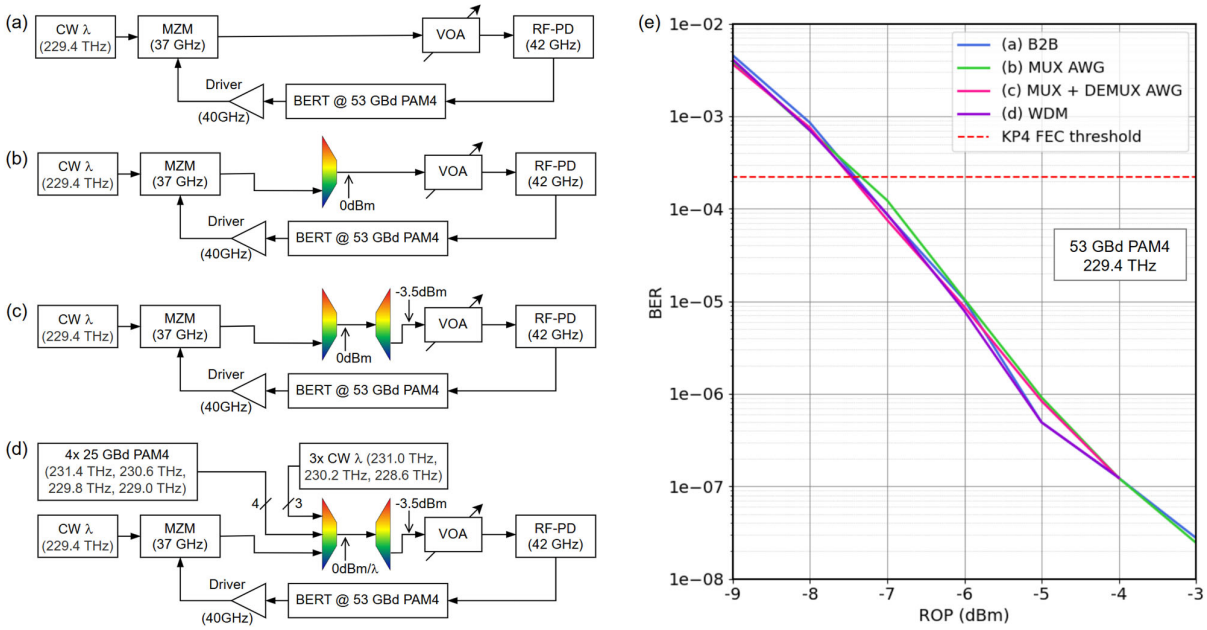

Figure 4. Experimental assessment of AWG-related performance penalty due to filtering and crosstalk with neighboring channels. (a) Experimental setup for B2B baseline performance assessment. (b) One AWG inserted. (c) Two AWGs inserted. (d) Channel under test + 4 modulated and 3 unmodulated aggressor channels combined by a first AWG. A second AWG is used to spectrally select only the channel under test for reception. (e) BER vs ROP for the 4 scenarios as illustrated in figures (a)-(d). Neither filtering related penalties, nor neighboring channel crosstalk-related penalties can be observed.

5. CONCLUSION

In this work we have discussed how CPO shoreline densities will need to approach 1 Tbps/mm in the 204.8 Tbps generation of switches with a lane rate of 200 Gbps and go beyond 2 Tbps/mm and 4 Tbps/mm in the 409.6 Tbps and 819.2 Tbps switches, respectively, assuming a lane rate of 400 Gbps. We have shown that WDM at a frequency grid of 400 GHz is a promising path for CPO to achieve shoreline density products beyond 15 Tbps/mm and for simplifying fiber management in the switch. In these WDM CPO architectures a high switch radix can be achieved by utilizing AWG-based optical port breakout modules. Such devices, realized on the silica-on-silicon platform, have the potential to (de-)multiplex up to 32

channels with negligible penalty from optical crosstalk between the channels, low insertion loss and very low polarization dependent loss. Our first generation of unamplified optical port breakout module has a flattop spectrum with 1 dB bandwidth of 307 GHz and 20 bandwidth of 469 GHz, while simultaneously featuring only 3.1 dB fiber-to-fiber IL and 0.1 dB PDL, and ultra-low crosstalk below -40 dB.

ACKNOWLEDGEMENTS

We acknowledge funding by the EU HORIZON 2021 project ADOPTION (grant agreement 101070178).

REFERENCES

[1] D. Ye, C. S. Geyik, Z. Qian, and K. Aygün, "Package Technology Enabling for 224 Gbps Electrical Signaling," in *2023 IEEE 32nd Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, Oct. 2023, pp. 1–3. doi: 10.1109/EPEPS58208.2023.10314950.

[2] S. Kannan *et al.*, "High Density Integration of Silicon Photonic Chiplets for 51.2T Co-packaged Optics," in *2024 IEEE 74th Electronic Components and Technology Conference (ECTC)*, May 2024, pp. 81–84. doi: 10.1109/ECTC51529.2024.00022.

[3] P. Ossieur *et al.*, "Integrated Photonics and Electronics for Optical Transceivers Supporting AI/ML Applications," *IEEE J. Sel. Top. Quantum Electron.*, vol. 31, no. 3: AI/ML Integrated Opto-electronics, pp. 1–16, May 2025, doi: 10.1109/JSTQE.2025.3539379.

[4] Y. Wang *et al.*, "Co-Designed Silicon Photonics Chip I/O for Energy-Efficient Petascale Connectivity," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 15, no. 8, pp. 1581–1591, Aug. 2025, doi: 10.1109/TCPMT.2024.3492189.

[5] C. Liu *et al.*, *Next Generation CEI-224G Framework*. [Online]. Available: https://www.oiforum.com/wp-content/uploads/OIF-FD-CEI-224G-01.0.pdf

[6] G. P. Agrawal, "Optical Receivers," in *Fiber-Optic Communication Systems*, John Wiley & Sons, Ltd, 2021, pp. 107–150. doi: https://doi.org/10.1002/9781119737391.ch4.

[7] *CW-WDM MSA Technical Specifications*. [Online]. Available: https://cw-wdm.org/

[8] R. Palmer *et al.*, "Polarization- and temperature-insensitive 8x400GHz O-band cascaded AWG for large radix data center and AI/ML networks," *submitted to Journal of Lightwave Technology*, doi: https://hdl.handle.net/10468/18349.