Chip

Research Article

PHOTONICS

DOI: 10.1016/j.chip.2023.100061

On-chip optical matrix-vector multiplier based on mode division multiplexing

Qiaolv Ling[1,2], Penghui Dong[1,2], Yayan Chu[3], Xiaowen Dong[3], Jingye Chen[1,2], Daoxin Dai[1,2] & Yaocheng Shi[1,2,*]

[1]*State Key Laboratory for Modern Optical Instrumentation, Centre for Optical and Electromagnetic Research, College of Optical Science and Engineering, International Research Center for Advanced Photonics, Zhejiang University, Zijingang Campus, Hangzhou 310058, China* [2]*Ningbo Research Institute, Zhejiang University, Ningbo 315100, China* [3]*Huawei Technologies Co., Ltd, Shenzhen 518000, China*

E-mail: yaocheng@zju.edu.cn (Yaocheng Shi)

Cite as: Ling, Q. et al. On-chip optical matrix-vector multiplier based on mode division multiplexing. *Chip* **2**, 100061 (2023).
https://doi.org/10.1016/j.chip.2023.100061

Received: 6 July 2023

Accepted: 7 August 2023

Published online: 14 August 2023

A matrix-vector multiplication (MVM) optical signal processor based on mode division multiplexing (MDM) was proposed and demonstrated in the current work, which is composed of a mode multiplexer, a multimode beam splitter, a mode demultiplexer, a modulator array and combiners. In addition, the characteristics of MDM obviate the need for multiple wavelengths and therefore multiple laser light sources are unneeded, which greatly reduces the complexity and cost. A 4 × 4 MDM-MVM was realized on a standard silicon-on-insulator (SOI) platform. Combined with the off-chip light source and photodetectors (PDs), 4-level modulation has been demonstrated, and each level of the output signal could represent 2 bits of information.

Keywords: Photonic computing, Silicon photonics, Matrix vector multiplier, Mode division multiplexing

INTRODUCTION

With the proliferation of ultra-high-speed mobile networks and device numbers connected to the Internet, people's expectations for computing performance have been increased by leaps and bounds. The processors under the traditional von Neumann architecture are constrained by energy consumption and bandwidth, which leads to the inability to carry the massive data processing demands. On the other hand, the emergence of artificial intelligence, especially artificial neural networks, poses great challenges to the slowing down of Moore's Law. Endowed with the inherent characteristics of high speed, low energy consumption and parallel processing, photonics is expected to achieve higher-performance information processing and even solve the complex algorithm problems that cannot be solved by microelectronic processors cooperating with electronics[1,2]. It is worth noting that the training process of neural networks heavily relies on massive matrix operations during the feed-forward and backpropagation stages[3–5], which determines the computation time and energy consumption of many workloads. The optical matrix-vector multiplier (MVM) has the potential to greatly reduce the power consumption of information transmission, increase the operating speed of the device, as well as expand the information transmission capacity[6,7]. Therefore, MVM is expected to be viable candidate for high-performance deep neural network acceleration hardware. In addition, many discrete numerical operations are based on the matrix so that MVM can perform various related operations, such as vector inner product operations, discrete Fourier operations[8] and convolution[9].

Initially, researchers implemented optical multipliers in free space[10–12]. However, the traditional optical MVM adopts detachable components, which is not conducive to the scaling and integration of the overall structure. The development of photonic integration technology has made it possible to implement multipliers on-chip. The on-chip multipliers can be divided into three types according to the principle and the realization structure: wavelength division multiplexing (WDM) type, multiplane light conversion (MPLC) type, and Mach-Zehnder interferometer (MZI) type[5]. In a WDM-MVM, wavelengths are used to differentiate signals input in parallel. In addition to the initial 4 × 4 microring multiplier demonstration[13], multipliers that can implement negative modulation matrix loading[14–17] and non-volatile loading of modulation elements[18–21] have also been demonstrated. The MPLC-MVM can load the matrix, which is decomposed into a series of programmable unitary diagonal matrices and unitary diffraction matrices[22]. Tang et al. experimentally demonstrated a ten-port single optical processor[23]. In the MZI-MVM, it is also necessary to utilize the triangular[24] or rectangular decomposition algorithm[25] to convert an ordinary matrix into the multiplication of several unitary matrices and diagonal matrices[26]. MZI-MVMs are often adopted to accelerate the optical neural networks (ONN) linear calculation part[27–30].

WDM-MVM usually requires multiple lasers to support multiple wavelengths. Modulation element loading requires various algorithms in MPLC-MVM and MZI-MVM. In order to reduce the cost and system complexity, in the current work, an on-chip optical signal processor was proposed to perform matrix-vector multiplication based on mode division multiplexing (MDM), which consists of a mode multiplexer, a multimode beam splitter, a mode demultiplexer, MZI modulators, and a power combiner. To the best of our knowledge, it is the first time that the MDM has been introduced to realize MVMs. As a proof-of-concept, a 4 × 4 MDM-

 DOI: 10.1016/j.chip.2023.100061

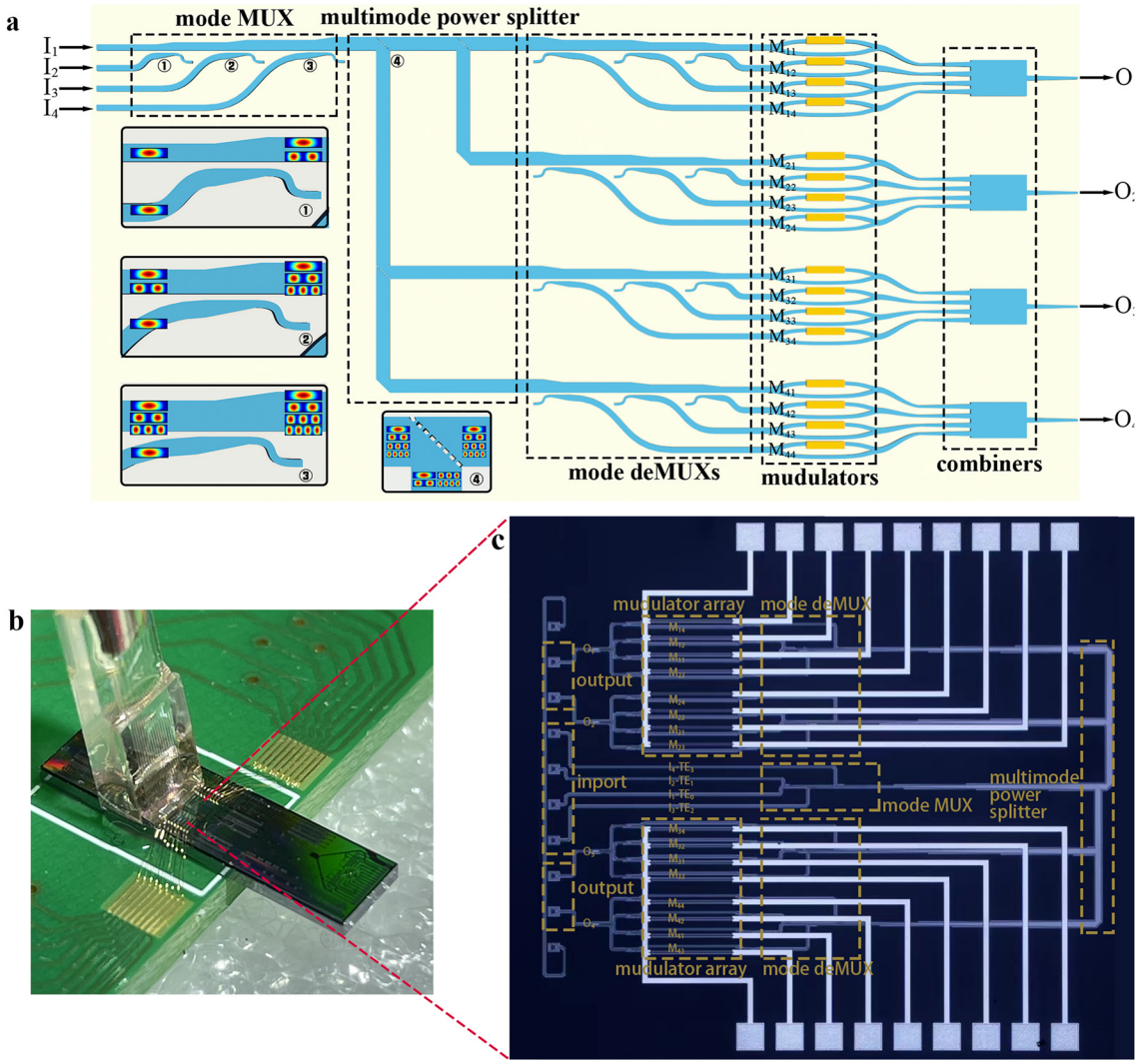

Fig. 1 | a, Schematic of the on-chip optical matrix-vector multiplier. The system can implement the matrix-vector multiplication of $\mathbf{M} \cdot \mathbf{I} = \mathbf{O}$, in which the matrix **M** is represented by the transmittances of the 4 × 4 MZI modulator matrix, the vector I is represented by the optical power of the 4 × 1 off-chip light source array, and the result vector O is represented by the optical powers detected by the 4 × 1 photodetector array (MUX: multiplexer). **b,** Image of the packaged chip. **c,** Microscope image of the fabricated device.

MVM was demonstrated on silicon-on-insulator (SOI) platform with 4-level modulation.

PRINCIPLE AND DESIGN

Matrix-vector multiplication can be described by the following formulas:

$$\mathbf{O} = \mathbf{M} \cdot \mathbf{I} = \begin{bmatrix} O_1 \\ O_2 \\ \vdots \\ O_m \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{m1} & M_{m2} & \cdots & M_{mn} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_n \end{bmatrix} \quad (1)$$

$$O_i = \sum_{j=1}^{n} M_{ij} I_j, (i = 1, 2, \ldots, m) \quad (2)$$

where, **I** is an n-dimensional input vector; **M** is an $m \times n$ modulation matrix; **O** is an m-dimensional result vector; I_j, M_{ij}, and O_i are elements of **I**, **M**, **O**, respectively. The proposed architecture for performing the multiplication of a 4 × 4 matrix **M** and a 4 × 1 vector **I** is shown in Fig. 1a. Four incoherent fundamental mode (TE_0) light signals were input at the ports I_1–I_4. The elements of vector **I** were sequentially loaded on these four signals, which can be achieved by modulating the signal intensity. Each input vector element corresponds to a different mode (TE_0–TE_3), combined to the bus waveguide by the asymmetric directional coupler (ADC) based mode multiplexer[31,32]. Later, a 1 × 4 multimode power beam splitter was utilized to divide the light into four even parts, containing all the four modes. Each part of the multimode signal was sequentially demultiplexed into four TE_0 lights by the four-channel mode de-multiplexer. Then the intensities of all 16 channels were modulated by the 4 × 4 modulator array loaded with matrix **M**. All the 4 × 4 multiplication processes were performed simultaneously. Finally, the four TE_0 signals of each channel were output to the single-mode waveguide through the 4 × 1 beam combiner, realizing the addition operation. Four optical powers detected by the off-chip PD array represent the elements of the result vector **O**.

In the MDM-MVM, the mode multiplexer and demultiplexer are realized by ADCs based on mode evolution, as shown in Fig. 2a. The TE_0 light is input at waveguide A. In the counter-tapered coupling region, a particular high-order mode is coupled into the bus waveguide B, while other

DOI: 10.1016/j.chip.2023.100061

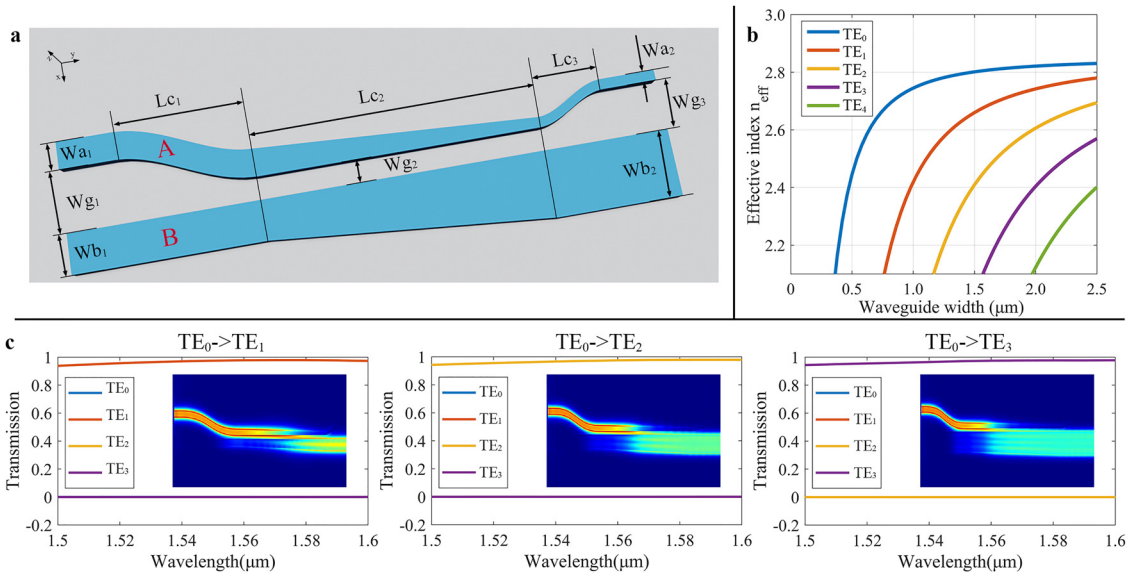

Fig. 2 | a, Schematic of the mode converter based on ADC. **b,** Calculated effective indices of eigen-modes of 220 nm-thick SOI strip waveguide. **c,** Calculated conversion efficiencies and field distributions of the three designed ADCs.

Table 1 | The parameters of the designed ADCs. (Unit: μm).

	Wa_1	Wa_2	Wb_1	Wb_2	Lc_1	Lc_2	Lc_3	Wg_1	Wg_2	Wg_3
TE_0- > TE_1	0.30	0.18	0.38	0.62	20.0	30.0	5.0	1.20	0.20	0.50
TE_0- > TE_2	0.30	0.18	0.80	1.06	20.0	40.0	10.0	1.20	0.20	0.50
TE_0- > TE_3	0.30	0.18	1.20	1.40	20.0	50.0	15.0	1.20	0.20	0.90

lower-order modes are transmitted in waveguide B without interference. In order to achieve high mode conversion efficiency, the counter-tapered coupling region should be carefully designed following the phase-matching condition. A 220 nm-thick strip waveguide was chosen to construct the device. Under such a geometric structure, calculation of the effective indices for different modes of the strip waveguides with different widths at 1.55 μm wavelength was conducted, as shown in Fig. 2b. Assuming that the multiplexer is to implement conversion coupling from the TE_0 mode to the TE_i mode. The width at the input of waveguide A (Wa_1) needs to support the TE_0 mode, and the width at the output (Wa_2) is required to be small enough so as to reduce the coupling of waveguide B to waveguide A. The width at the input of waveguide B (Wb_1) must support the transmission of the original modes (TE_0–TE_{i-1}). Wb_2 needs to be greater than the minimum width corresponding to TE_i, while at the same time smaller than the minimum width for exciting TE_{i+1} to avoid conversion of the TE_0 mode into a higher-order mode. Parameters such as lengths and gaps were optimized with the adoption of the finite-difference time-domain (FDTD) method. The optimized parameters of the three multiplexers are shown in Table 1. Fig. 2c shows the calculated field distribution and the mode conversion rate.

The structure of multimode waveguide crossing with an obliquely embedded subwavelength grating (SWG)[33] (Fig. 3a) was employed as the multimode power splitter. Due to the thin-film interference effect, the SWG reflector can split the multimode incident beam into two equal beams simultaneously. Since the effective refractive index of the SWG can be adjusted by changing the duty cycle (f_{swg}), the splitting ratio can be adjusted arbitrarily with a large bandwidth. In consideration of the fabrication feasibility, the SWG pitch Λ_{swg} and SWG width w_{swg} were chosen to be 500 nm and 200 nm, respectively. According to the simulation, when the multimode waveguide width was set as 15 μm, and the f_{swg} as 0.4, a splitting ratio of about 50 : 50 was achieved. The uniformity of the splitting ratio of each mode and the crosstalk between modes are shown in Fig. 3b. In order to reduce the difficulty and complexity of the fabrication process, the 1 × 4 beam splitter is realized with the adoption of three 50 : 50 beam splitters cascaded in a parallel structure (Fig. 3c).

The modulator is realized with the MZI based on thermal tuning, forming a modulation matrix to verify the feasibility of the MVM. The electrode in the heating area is titanium nitride (TiN), the length of the heating arm is about 300 μm, the resistance is about 1.8 kΩ, and the width of the heating waveguide is 500 nm. The beam combiner is achieved by a 4 × 1 MMI beam combiner. Assuming that the input power of each Input Port is equal, the power output from input port 1, port 2, port 3, and port 4 to output port accounts for 23.69%, 24.48%, 24.48%, and 23.69%, respectively.

CHARACTERIZATION

Fig. 1b, c show the fabricated and wire bonded SOI photonic chip and the microscope image of the MDM-MVM device, respectively. The input is the TE_0 light, marked as I_j-TE_{j-1} (j = 1, 2, 3, 4) according to the modes multiplexed into the bus waveguide, and the modulation MZI array can be marked as M_{ij} in the same way. The corresponding output port is O_i. It should be noted that, in order to improve the compactness of the device and reduce the footprint, 90° corner bends are adopted to change the transmission direction of the optical path[34].

The experimental setup for the device characterization is shown in Fig. 4. To avoid disturbing coherence of different input ports in MMI, an amplified spontaneous emission source (ASE) was adopted as the light source. The ASE was connected to a commercial-grade fiber optic splitter, which splits the original light into four parts as input vector. It is evi-

DOI: 10.1016/j.chip.2023.100061

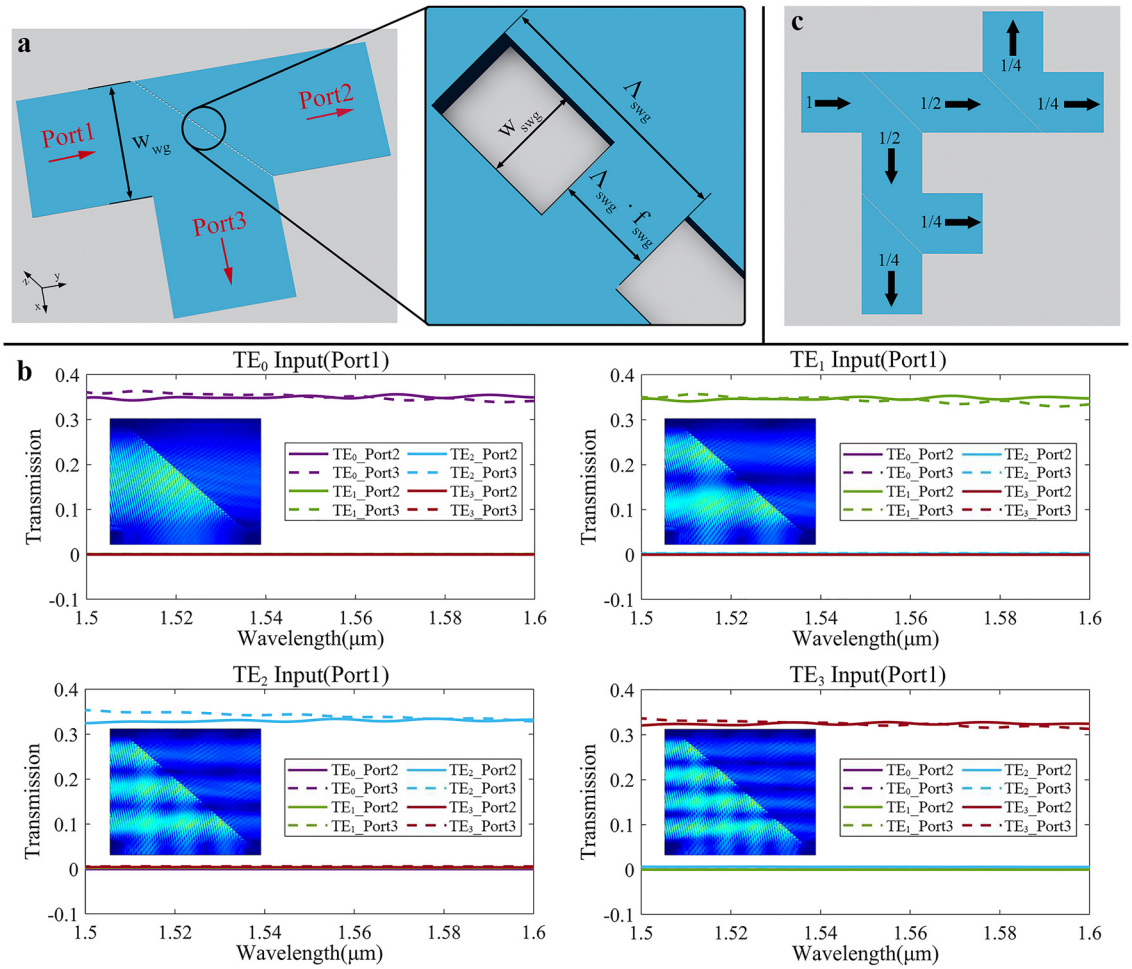

Fig. 3 | a, Schematic of multimode beam splitter based on SWG, and the enlarged top view of the SWG transflector with some key parameters labeled. **b,** The calculated transmission spectra and mode crosstalk for the multimode power splitters of inputting TE_0–TE_3 modes, with the calculated light propagation profiles. **c,** Schematic of the 1 × 4 parallel beam splitter.

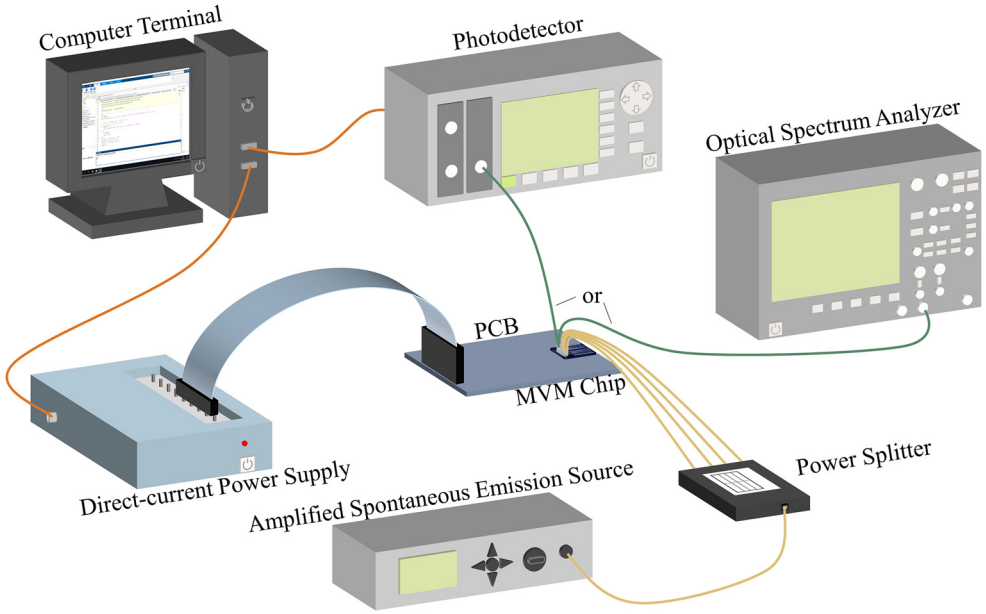

Fig. 4 | Experimental setup for the static and dynamic response characterization of the optical MDM-MVM.

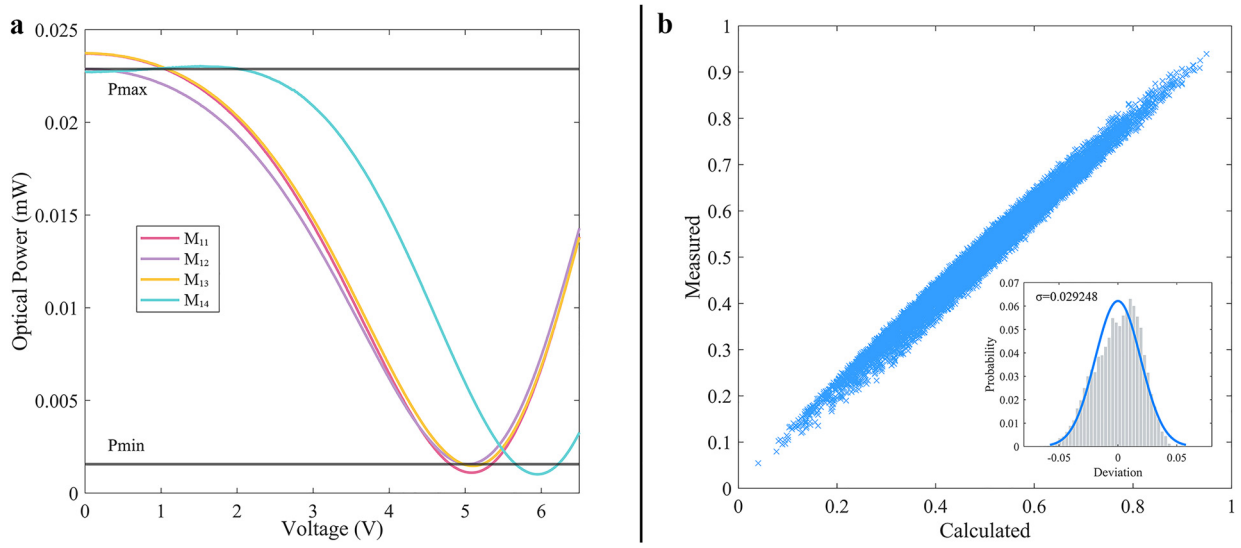

Fig. 5 | a, Optical power of four modulators corresponding to the first row of **M** when a voltage is applied from 0 to 6.5 V. P_{max} and P_{min} represent the maximum and minimum values of the overlap between the dynamic ranges of these four modulators. **b,** Scatter plot for MAC accuracy measurement with an input vector. The inset is a residual error distribution histogram.

dent that modulators can be connected after the beam splitter to load different input vectors. To generate the modulated signals, a multi-channel direct-current power supply was utilized, which generates sixteen parallel electrical drive signals. Each set of four signals represents the ith row of the modulation matrix **M** and is applied in parallel to the MZI modulator corresponding to O_i. The light on the chip is emitted from the coupling grating, which can be measured by an optical spectrum analyzer (OSA) to evaluate the static spectral response of the MVM. The chip can also be connected to PDs so as to convert the inner product optical power into an electrical signal.

In the modulation matrix section, analysis on the computational accuracy of the MVM is a pivotal aspect. The factors affecting accuracy originate from various aspects, some of which can be eliminated. For instance, the imbalance in the splitting ratio between different ports of the splitter can be addressed by setting different power detection thresholds for each output port O_i of the multiplier. Additionally, the losses of the splitter and mode multiplexer for different modes, along with the differences in the dynamic range of the MZIs, can lead to variations in the optical power of the same level due to the different mode channels through which the light passes. This can be compensated by normalizing the four MZIs corresponding to the same output port. This means that each MZI's modulation range is constrained within the overall minimum and maximum optical power (*P*min and *P*max) for the four paths, as depicted in Fig. 5a. Within this modulation range, four random numbers were selected to form a modulation vector, which is later subjected to multiplication and addition (MAC) with a fixed input vector. The measured results were compared against the expected analytically calculated multiplication result. The results of 10,000 calculations were scaled to the range [0, 1] and plotted together with the corresponding histogram in Fig. 5b, showing a standard deviation of 0.029. Based on the equation $bit\ depth = \log_2(1/(3\sigma * 2))$[20,35], a modulation resolution of 2 bits can be derived.

With the modulation accuracy determined, the "level-voltage" lookup table can be obtained for each MZI. Based on the interval of $(P\max - P\min)/(2^2-1)$ as the step size, each MZI is configured with corresponding intensity levels and modulation voltages for each level. By traversing all the modulation vectors, the overlap range of output signal ranges for different result elements can be examined for validating the MVM's capability so as to support 4-levels modulation. This approach guarantees the accuracy and consistency of the performance of MVM across all the possible modulation combinations. The measurement results demonstrate that the fabricated MVM can effectively support modulation with the precision of 2 bits, which further confirms the above calculation. This implies that there are totally $(2^2-1)\times 4 + 1 = 13$ possible elements in the result vector. Fig. 6 illustrates the waveforms of the four drive signals, the responses of the dynamic characteristics of each modulator, and the final result of the vector-vector multiplication.

DISCUSSION

In the characterization system mentioned above, factors that affect the modulation bit resolution also arise from other connected devices, such as the fluctuation of optical power from the light source, the accuracy of voltage loading from the multi-channel voltage source, as well as the stability of the PD. In order to estimate the best bit resolution of the MVM, the general model of finite precision analysis proposed by Li et al[35] can be utilized, which is based on the assumption that the external devices are ideal and focusing only on the SOI chip. In this case, the extinction ratio (ER) of the modulator is a crucial factor. By inputting the **I** of 1010 to the MVM and loading 0101 or 1111 signals on each row of the modulation matrix (M_i), the ER of the MVM can be obtained, as is shown in Fig. 7, which varies with wavelength. The utilization of ASE results in an effective ER of the multiplier, which is the average value within the wavelength range of the light source. When employing a non-coherent single-wavelength light source, the ER of the multiplier can be fixed at a higher value specific to that wavelength, which can reach up to 15 dB. Based on the formula $1/\mathrm{ER} < 0.5/(2^{\text{bit depth}} - 1)$derived from the model, the MVM can support a precision of 4 bits. To further enhance the accuracy of MDM-MVM, certain components can be optimized within the current structure. For instance, during testing, it was observed that the power fluctuations resulting from thermal crosstalk between modulators account for approximately 2.15% of the MZI's dynamic range, making it a main contributing factor to the overall deviation. In future work, thermo-

DOI: 10.1016/j.chip.2023.100061

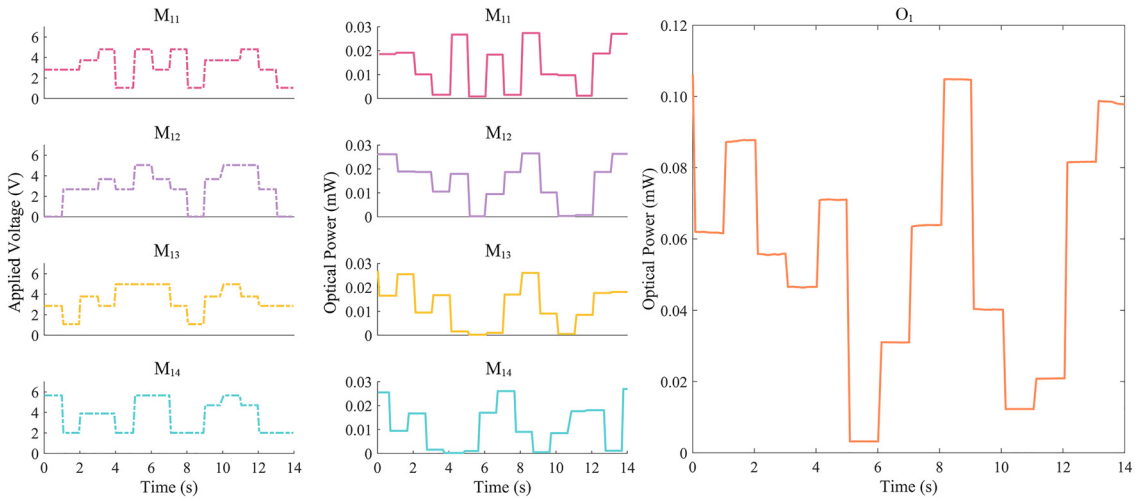

Fig. 6 | Waveforms of the driving voltages and response of each modulator and MVM. Vector **I** is fixed. Voltages are applied to the MZI modulator array, representing the elements in the first row of matrix **M**.

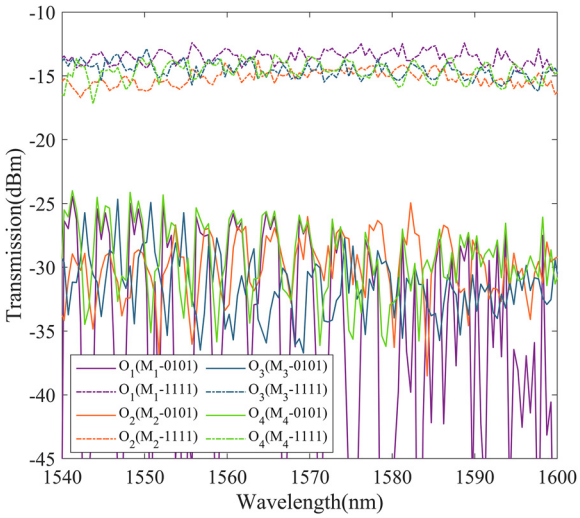

Fig. 7 | Transmission spectra obtained by loading different M_i values with input vector I as 1010.

electric cooler (TEC) can be implemented to reduce the thermal crosstalk. The modulator can also be replaced, such as by using a GeSi waveguide electro-absorption modulator[36,37]. The GeSi modulator can mitigate the error caused by the heat conduction of a certain heater to the adjacent heater. It can also effectively increase the modulation rate, thereby increasing the calculation rate.

To achieve larger-scale multiplication in this architecture, it is necessary to evaluate the optical losses introduced by each system component. Based on individual component characterization, the average loss of the 4-mode multiplexer/demultiplexer is –0.44 dB, the 1 × 4 splitter is –9.80 dB, and the 4 × 1 coupler is –6.26 dB. The MVM principle dictates that the loss of the 1 × 4 splitter should be –6 dB, and the losses of other components should be as close to 0 as possible. It is worth noting that the discrepancy between the actual loss and the ideal loss for the coupler is relatively large, which hinders the scalability of the multiplier. It is the structure of the MMI that leads to the fact that only $1/2^{\text{port number}}$ of light can enter the output port for each input port. To address the problem above, the approach of utilizing MDM and replacing the MMI with a mode multiplexer can be adopted. This multiplexer combines the four TE_0 signals into a multimode bus waveguide, enabling highly efficient and low-loss beam combining. This approach could improve the signal-to-noise ratio and reduce the power requirements of the light source. Furthermore, edge couplers[38,39]are essential for extracting signals from the chip. Alternatively, on-chip integration of a PD array[40,41] allows for direct electrical signal output. In these two signal output schemes, the difference between the coupling efficiency or conversion efficiency of different modes needs particular attention to ensure the accuracy of the results.

To expand the calculation dimension of MDM-MVM, one approach is to continue increasing the number of the multiplexed modes. It has been reported that 11-mode MDM has been achieved[42]. This suggests that, with minimal changes to the existing architecture, the number of multiplexable modes can be increased by adding ADCs for each mode, increasing the number of parallel layers in the splitter's SWG, increasing the number of MZIs, and utilizing an 8 × 1 MMI for an 8-channel multiplier. However, as this method exhibits a clear upper limit to the increase in dimensionality, another approach that can be taken into consideration is to integrate WDM into the architecture. Except the wavelength-sensitive beam combiners, the other components mentioned above are broadband devices that can be directly used in combination with WDM devices. This allows each input element to be characterized by the intensity of light at a specific wavelength and particular mode. The loading of modulation matrix elements can be achieved by either adding wavelength demultiplexer in front of the existing modulator array or replacing the array with a wavelength-selective modulator array[16,43]. In the MDM & WDM-MVM architecture, one of the primary limiting factors has become the losses. Assuming that the multiplier scale is 2^n, based on the power of the light source used and the current measured losses of the components, it can be estimated that the output optical power of the system can be expressed as: $P_{\text{output}} = P_{\text{ASE}} \pm \text{Loss}_{\text{off-chip splitter}} \pm \text{Loss}_{\text{grating}} \times 2 \pm \text{Loss}_{\text{mode MUX \& deMUX}} \pm \text{Loss}_{\text{multimode splitter}} \pm \text{Loss}_{\text{MZI}} \pm \text{Loss}_{\text{MMI}} = 19 - 3n - 5.87 \times 2 - 0.44 - 4.90n - 0.73 - 3.13n$. In order to ensure that elements 0 to 1 are distinguishable, the ER of the MZI also needs to be considered in P_{output}. Based on the minimum detectable power of the PD, it can be deduced that the maximum value of n is 6, which means the current achievable upper limit for the dimensionality is 64.

Compared with other MVM optical signal processors, the MDM-MVM in the current work is endowed with the advantage of having less stringent requirements on the light source. It does not necessitate multiple lasers or microcombs to support multiple wavelengths. Additionally,

Table 2 | Comparison of different photonic matrix multiplier.

	Structures	Dimension	Resolution	Matrix loading	Light source
Our work	MDM-MVM	4	2-bit	One-one	ASE
Ref. [13]	WDM-MVM	4	Binary	One-one	Four tunable lasers
Ref. [28]	MZI-MVM	4	8-bit	Algorithm-aided	Laser
Ref. [36]	WDM-MVM	2	5.1-bit	One-one	Two DFB lasers
Ref. [30]	MZI-MVM	6	Supplied by 16-bit current	Algorithm-aided	Coherent laser
Ref. [20]	WDM-MVM	16	5-bit	One-one	Optical frequency comb
Ref. [17]	WDM-MVM	4	9-bit	One-one	DFB pumped microcomb

MDM-MVM possesses a direct execution capability, which eliminates the requirement of intricate algorithms, owing to the one-to-one mapping relationship between modes and matrix elements. The independence between different channels also makes it easier to analyze and subsequently improve crosstalk and losses in the system. The specific comparisons are presented in Table 2.

CONCLUSIONS

In the current work, a novel matrix-vector multiplier based on mode division multiplexing was introduced and demonstrated. The core components of the processor include a mode multiplexer, a multimode beam splitter, a mode demultiplexer, a modulator array, and combiners. Efficient 4-level modulation was achieved with the adoption of an off-chip light source and PDs, with each level of the output signal encoding 2 bits of valuable information. The utilization of MDM obviated the necessity for multiple laser light sources and reduced both the system complexity and overall cost. Lasers and detectors can be further integrated on the same chip, resulting in a fully integrated on-chip optical matrix-vector multiplier.

METHODS

The proposed MDM-MVM was fabricated on an SOI platform with a 220-nm-thick top silicon layer and a 2-μm-thick buried silicon dioxide layer. 248-nm DUV photolithography defines the patterns and inductively coupled plasma reactive ion etching (ICP-RIE) was employed to form the silicon waveguides. Subsequently, a 1-μm thick silica thin film was deposited as the upper cladding. After the deposition, 100-nm thick strip TiN micro heaters were fabricated on top of the heating arm waveguide. Finally, aluminum was deposited and etched to form the metal wires and pads.

REFERENCES

1. Shastri, B. J. et al. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **15**, 102–114 (2021). https://doi.org/10.1038/s41566-020-00754-y.
2. Xu, X. et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021). https://doi.org/10.1038/s41586-020-03063-0.
3. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015). https://doi.org/10.1038/nature14539.
4. Sarle, W. S. Neural networks and statistical models. In *Proceedings of the 19th Annual SAS Users Group International Conference*, 1538–1550 (CiNii, 1994). https://cir.nii.ac.jp/crid/1570854175364916608.
5. Zhou, H. et al. Photonic matrix multiplication lights up photonic accelerator and beyond. *Light: Sci. Appl.* **11**, 30 (2022). https://doi.org/10.1038/s41377-022-00717-8.
6. Kitayama, K.-I. et al. Novel frontier of photonics for data processing—photonic accelerator. *APL Photonics* **4**, 090901 (2019). https://doi.org/10.1063/1.5108912.
7. Zhou, Z., Xu, P. & Dong, X. Computing on silicon photonic platform. *Chin. J. Lasers* **47**, 9–23 (2020). https://doi.org/10.3788/CJL202047.0600001.
8. Zhou, Y., Cao, W., Liu, L., Agaian, S. & Chen, C. L. P. Fast Fourier transform using matrix decomposition. *Inf. Sci.* **291**, 172–183 (2015). https://doi.org/10.1016/j.ins.2014.08.022.
9. Vasudevan, A., Anderson, A. & Gregg, D. Parallel multi channel convolution using general matrix multiplication. In *2017 IEEE 28th International Conference on Application-specific Systems, Architectures and Processors (ASAP)*, 19–24 (IEEE, 2017). https://doi.org/10.1109/ASAP.2017.7995254.
10. Athale, R. A. & Collins, W. C. Optical matrix–matrix multiplier based on outer product decomposition. *Appl. Opt.* **21**, 2089–2090 (1982). https://doi.org/10.1364/AO.21.002089.
11. Habiby, S. F. & Collins, S. A. Implementation of a fast digital optical matrix–vector multiplier using a holographic look-up table and residue arithmetic. *Appl. Opt.* **26**, 4639–4652 (1987). https://doi.org/10.1364/AO.26.004639.
12. Tamir, D. E., Shaked, N. T., Wilson, P. J. & Dolev, S. High-speed and low-power electro-optical DSP coprocessor. *J. Opt. Soc. Am. A* **26**, A11–A20 (2009). https://doi.org/10.1364/JOSAA.26.000A11.
13. Yang, L., Ji, R., Zhang, L., Ding, J. & Xu, Q. On-chip CMOS-compatible optical signal processor. *Opt. Express* **20**, 13560–13565 (2012). https://doi.org/10.1364/OE.20.013560.
14. Tait, A. N., Nahmias, M. A., Shastri, B. J. & Prucnal, P. R. Broadcast and weight: an integrated network for scalable photonic spike processing. *J. Lightwave Technol.* **32**, 3427–3439 (2014). https://opg.optica.org/jlt/abstract.cfm?uri=jlt-32-21-3427.
15. Tait, A. N. et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2014). https://doi.org/10.1038/s41598-017-07754-z.
16. Luan, E. et al. Towards a high-density photonic tensor core enabled by intensity-modulated microrings and photonic wire bonding. *Sci. Rep.* **13**, 1260 (2023). https://doi.org/10.1038/s41598-023-27724-y.
17. Bai, B. et al. Microcomb-based integrated photonic processing unit. *Nat. Commun.* **14**, 66 (2023). https://doi.org/10.1038/s41467-022-35506-9.
18. Ríos, C. et al. In-memory computing on a photonic platform. *Sci. adv.* **5**, eaau5759 (2019). https://doi.org/10.1126/sciadv.aau5759.
19. Feldmann, J., Youngblood, N., Wright, C. D., Bhaskaran, H. & Pernice, W. H. P. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019). https://doi.org/10.1038/s41586-019-1157-8.
20. Feldmann, J. et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021). https://doi.org/10.1038/s41586-020-03070-1.
21. Zhou, W. et al. In-memory photonic dot-product engine with electrically programmable weight banks. Preprint at https://doi.org/10.48550/arXiv.2304.14302 (2023).
22. Tang, R., Tanemura, T. & Nakano, Y. Integrated reconfigurable unitary optical mode converter using MMI couplers. *IEEE Photonics Technol. Lett.* **29**, 971–974 (2017). https://doi.org/10.1109/LPT.2017.2700619.
23. Tang, R., Tanomura, R., Tanemura, T. & Nakano, Y. Ten-port unitary optical processor on a silicon photonic chip. *ACS Photonics* **8**, 2074–2080 (2021). https://doi.org/10.1021/acsphotonics.1c00419.
24. Reck, M., Zeilinger, A., Bernstein, H. J. & Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73**, 58–61 (1994). https://doi.org/10.1103/PhysRevLett.73.58.
25. Clements, W. R., Humphreys, P. C., Metcalf, B. J., Kolthammer, W. S. & Walmsley, I. A. Optimal design for universal multiport interferometers. *Optica* **3**, 1460–1465 (2016). https://doi.org/10.1364/OPTICA.3.001460.
26. Harris, N. C. et al. Linear programmable nanophotonic processors. *Optica* **5**, 1623–1631 (2018). https://doi.org/10.1364/OPTICA.5.001623.
27. Hughes, T. W., Minkov, M., Shi, Y. & Fan, S. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica* **5**, 864–871 (2018). https://doi.org/10.1364/OPTICA.5.000864.
28. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017). https://doi.org/10.1038/nphoton.2017.93.
29. Bogaerts, W. et al. Programmable photonic circuits. *Nature* **586**, 207–216 (2020). https://doi.org/10.1038/s41586-020-2764-0.
30. Zhang, H. et al. An optical neural chip for implementing complex-valued neural network. *Nat. Commun.* **12**, 457 (2021). https://doi.org/10.1038/s41467-020-20719-7.
31. Dai, D. et al. 10-channel mode (de)multiplexer with dual polarizations. *Laser Photonics Rev.* **12**, 1700109 (2018). https://doi.org/10.1002/lpor.201700109.

32. Wang, J. et al. Broadband and fabrication-tolerant on-chip scalable mode-division multiplexing based on mode-evolution counter-tapered couplers. *Opt. Lett.* **40**, 1956–1959 (2015). https://doi.org/10.1364/OL.40.001956.
33. Xu, H., Dai, D. & Shi, Y. Ultra-broadband on-chip multimode power splitter with an arbitrary splitting ratio. *OSA Contin.* **3**, 1212–1221 (2020). https://doi.org/10.1364/OSAC.396024.
34. Wang, Y. & Dai, D. Multimode silicon photonic waveguide corner-bend. *Opt. Express* **28**, 9062–9071 (2020). https://doi.org/10.1364/OE.387978.
35. Li, C., Zhang, X., Li, J., Fang, T. & Dong, X. The challenges of modern computing and new opportunities for optics. *PhotoniX* **2**, 20 (2021). https://doi.org/10.1186/s43074-021-00042-0.
36. Shen, W., Du, J., Xiong, J., Ma, L. & He, Z. Silicon-integrated dual-mode fiber-to-chip edge coupler for 2 × 100 Gbps/lambda MDM optical interconnection. *Opt. Express* **28**, 33254–33262 (2020). https://doi.org/10.1364/OE.408700.
37. Cao, X., Li, K., Wan, Y. & Wang, J. Efficient mode coupling between a few-mode fiber and multi-mode photonic chip with low crosstalk. *Opt. Express* **30**, 22637–22648 (2022). https://doi.org/10.1364/OE.457549.
38. Li, G. et al. Improving CMOS-compatible germanium photodetectors. *Opt. Express* **20**, 26345–26350 (2012). https://doi.org/10.1364/OE.20.026345.
39. Liow, T.-Y. et al. Silicon modulators and germanium photodetectors on SOI: monolithic integration, compatibility, and performance optimization. *IEEE J. Sel. Top. Quantum Electron.* **16**, 307–315 (2010). https://doi.org/10.1109/JSTQE.2009.2028657.
40. He, Y. et al. Silicon high-order mode (de)multiplexer on single polarization. *J. Lightwave Technol.* **36**, 5746–5753 (2018). https://opg.optica.org/jlt/abstract.cfm?uri=jlt-36-24-5746.
41. Xu, Q., Schmidt, B., Pradhan, S. & Lipson, M. Micrometre-scale silicon electro-optic modulator. *Nature* **435**, 325–327 (2005). https://doi.org/10.1038/nature03569.
42. Tait, A. N. et al. Feedback control for microring weight banks. *Opt. Express* **26**, 26422–26443 (2018). https://doi.org/10.1364/OE.26.026422.
43. Srinivasan, S. A. et al. 50 Gb/s C-band GeSi waveguide electro-absorption modulator. In *Optical Fiber Communication Conference (OFC), Tu3D.7* (Optica, 2016). https://doi.org/10.1364/OFC.2016.Tu3D.7.

MISCELLANEA

Funding This work was supported by National Major Research and Development Program (2021YFB2801703), National Natural Science Foundation of China (62135011 & 62105286), "Pioneer" and "Leading Goose" R & D Program of Zhejiang (2022C01103), and the Fundamental Research Funds for the Central Universities.

Declaration of Competing Interests The authors declare no competing interests.

© 2023 The Author(s). Published by Elsevier B.V. on behalf of Shanghai Jiao Tong University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)