Supporting Information

Integrated multi-operand optical neurons for scalable and hardware-efficient deep learning

Chenghao Feng1,2 , Jiaqi Gu2,3, Hanqing Zhu2, Shupeng Ning1, Rongxing Tang1, May Hlaing2, Jason Midkiff2, Sourabh Jain1, David Z. Pan2, Ray T. Chen2,4,\*

1Microelectronics Research Center, The University of Texas at Austin, Austin, Texas 78758, USA.

2Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78705, USA.

3School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287,

USA

4Omega Optics, Inc., 8500 Shoal Creek Blvd., Bldg. 4, Suite 200, Austin, TX 78757, USA.

**Supplementary Note 1: MOON training algorithms.**

To train the MOON weights while being aware of hardware noises, we adopt a hardware-aware training method where the weights are quantized mapped to the measured MOMZI responses. For each neural network layer, no matter Linear or Convolution, we map it to a general matrix multiplication operation. For each length- vector product , we map and to the real chip measurement results during forward propagation,

 Eq. (S1)

where is the -bit quantization function, is the look-up tanble (LUT) with 2kb entries that mapps weights and inputs to the real chip measurement results, and is a fitted proxy of that enables gradient backpropagation to and . Parameters are regressed from the curve fitting.

During backpropagation, the gradients are as follows:

 Eq. (S2)

We also inject Gaussian noises to MOMZI output to improve the noise robustness of the model. Output deviation is input-dependent based on Fig. 5(b), i.e., . In this way, we fully consider physical chip responses during training to close the gap of simulation and on-chip deployment accuracy.

**Supplementary Note 2: Parameter tables for the delay, propagation loss, and footprint estimation**

**Table. S1.** Device parameters used in our performance estimation based on AIM photonics’ PDK1.

|  |  |  |
| --- | --- | --- |
| Optical component | Length (μm) | Insertion loss (dB) |
| High-speed EO MZI () | 1600 | 3 |
| High-speed plasmonic EO MZI2 | ~220a | 11.2 |
| Low-speed TO MZI () | 550 | 1 |
| Microring-based filter | 16 | 0.25 |

1. Based on the layout picture of the MZI in the reference (~200 . The high-speed phase shifter part is only 15 in length, so the size of the modulator can be further optimized with more compact directional couplers.

**Table. S2.** Parameters to calculate the performance of -op MOMZI-PTC.

|  |  |
| --- | --- |
| Parameters | Values |
|   |  |
|  |  |
|  | 10 (1.5 after scaling) |
|  |  |
|  |   |
|  |  |
|  | ( |
|  |  |
|  |  |

**Supplementary Note 3: Energy efficiency.**

The power consumptionof *n*-input, *m*-output MOMZI-PTC for computing is contributed by lasers, weight configuration, and conversion between electrons and photons, which is obtained by:

 Eq. (S3)

The parameter table for modeling Eq. (S1) is provided in Table. S3. In Eq. (S3), is the laser power. represents the energy consumption of modulators, and includes the power consumption for photodetection, amplification, and analog-to-digital conversion. is the operating speed of modulators, as determined by the total delay of the MOMZI-ONN. In the MOMZI-ONN, is the static power to tune the MOMZI to a bias point. Using thermal phase shifters for bias tuning, . By utilizing energy-efficient active optical components based on nano-opto-electro-mechanical systems or phase change materials3,4, we can eliminate the power consumption for phase maintenance. To carrier-depletion-based silicon MZI modulators, can achieve ~146 fJ/bit. Using energy-efficient plasmonic-on-silicon modulators , fJ/bit. One DAC’s energy consumption can be estimated by 5:

 Eq. (S4)

where is the DAC figure of merit, is the DAC resolution, is the sampling frequency, is the bit rate. In our estimation, fJ/step in a 7-nm microprocessor6, bit, .

The propagation loss, the photodetectors' minimum detectable power, and the outputs' precision dictate the laser power. The total laser power can then be calculated by the following equation7:

 (Eq. S5)

where ℎ𝜈 is the photon energy at 1.55 μm, is the wall-plug efficiency of the laser8, is the number of wavelengths used in the MOMZI-PTC. The precision of output signals is bits. is the capacitance of the photodetector while is the operating voltage. Note that in some zero-biased energy-efficient photodetectors9,10, is the baud rate of the intput signals. In this work, we choose Baud/s and use a GSPS ADC for reading the output.

**Table. S3.** Energy consumption of a *k*-point MOMZI-PTC – modeling parameters

|  |  |
| --- | --- |
| Expression | Value |
|  | ~146 fJ/bit11 (0.1 fJ/bit after scaling)2 |
|  |  (PDK)0 after scaling |
|  5 |  fJ/step 6 |
|  7 |  |
| (10 GSPS)  | 39 mW/channel 12 |
|  | 0.52 fJ/level13 (1.3 mW/channel)after scaling |

References

1. E. Timurdogan et al., “APSUNY Process Design Kit (PDKv3.0): O, C and L Band Silicon Photonics Component Libraries on 300mm Wafers,” 2019 Optical Fiber Communications Conference and Exhibition (OFC), 1–3, OSA (2019) [doi:10.1364/ofc.2019.tu2a.1].

2. W. Heni et al., “Plasmonic IQ modulators with attojoule per bit electrical energy consumption,” Nature Communications **10**, 1–8 (2019) [doi:10.1038/s41467-019-09724-7].

3. L. Midolo, A. Schliesser, and A. Fiore, “Nano-opto-electro-mechanical systems,” Nature Nanotechnology **13**(1), 11–18, Springer US (2018) [doi:10.1038/s41565-017-0039-1].

4. M. Wuttig, H. Bhaskaran, and T. Taubner, “Phase-change materials for non-volatile photonic applications,” Nature Photonics **11**(8), 465–476, Nature Publishing Group (2017) [doi:10.1038/nphoton.2017.126].

5. B. S. G. Pillai et al., “End-to-end energy modeling and analysis of long-haul coherent transmission systems,” Journal of Lightwave Technology **32**(18), 3093–3111, Institute of Electrical and Electronics Engineers Inc. (2014) [doi:10.1109/JLT.2014.2331086].

6. C. Huang et al., “A silicon photonic–electronic neural network for fibre nonlinearity compensation,” Nature Electronics **4**(11), 837–844, Nature Publishing Group (2021) [doi:10.1038/s41928-021-00661-2].

7. M. A. Nahmias et al., “Photonic Multiply-Accumulate Operations for Neural Networks,” IEEE Journal of Selected Topics in Quantum Electronics **26**(1), 1–18, IEEE (2020) [doi:10.1109/JSTQE.2019.2941485].

8. H. Wang et al., “High-Power Wide-Bandwidth 1.55-μm Directly Modulated DFB Lasers for Free Space Optical Communications,” in 2019 Conference on Lasers and Electro-Optics, CLEO 2019 - Proceedings, pp. JTu2A-72 (2019) [doi:10.23919/CLEO.2019.8750482].

9. L. Vivien et al., “Zero-bias 40Gbit/s germanium waveguide photodetector on silicon,” Optics Express **20**(2), 1096 (2012) [doi:10.1364/OE.20.001096].

10. T. M. Photodiodes et al., “High-Speed Evanescently-Coupled Waveguide,” 6827–6832 (2020).

11. J. Ding et al., “Ultra-low-power carrier-depletion Mach-Zehnder silicon optical modulator,” Opt. Express, OE **20**(7), 7081–7087, Optica Publishing Group (2012) [doi:10.1364/OE.20.007081].

12. “ADC (Analog-to-Digital converters) – Alphacore,” <https://www.alphacoreinc.com/adc-analog-to-digital-converters/> (accessed 30 August 2021).

13. C. Li et al., “Analog content-addressable memories with memristors,” 1, Nat Commun **11**(1), 1638, Nature Publishing Group (2020) [doi:10.1038/s41467-020-15254-4].