Si Microring Weightbank for Photonic Signed Multiplication

Minkyu Cho, Yongjin Ji, *Woo-Young Choi
Department of Electrical and Electronic Engineering, Yonsei University, South Korea
**wchoi@yonsei.ac.kr*

Abstract: We demonstrate a new type of Si microring weightbank performing signed multiplication of $x \times w$. The PIC achieves 31 levels for inputs and weights, with effective resolutions of 6.51 and 5.97 bits, respectively.

1. Introduction

AI/ML computing requires massive matrix computation and, presently, this is achieved by hyper-parallel tensor cores such as graphics processing units (GPUs) and/or neural processing units (NPUs). The photonic tensor core platform is also receiving a large amount of attention due to the inherent parallelism light can provide. In particular, microring weightbanks provide full reconfigurability and compact footprint within the intensity-modulation/direct-detection (IM/DD) scheme. However, conventional IM/DD-based microring weightbanks have limitations in that complex time-sequential encoding schemes or additional modulators are required for representing signed parameters of input x and weight w needed for multiplication-and-accumulation (MAC) [1-3]. To overcome this limitation, we propose and experimentally demonstrate a Si photonic integrated circuit (PIC) featuring a bidirectional microring weightbank architecture. By leveraging bidirectional propagation, the PIC performs signed-input and signed-weight encoding directly without increasing the number of modulators. The experimental results show that the PIC achieves 31 distinct levels for signed multiplications of $x \times w$ resulting in the effective resolutions of 6.51 and 5.97 bits for inputs and weights, respectively.

2. Operating Principle

The proposed architecture performs signed multiplication of $x \times w$, by utilizing two microring resonators (MRs), each dedicated for x-encoding and w-encoding. The operation is governed by four variables, which are the light intensities f and g entering the 5:5 couplers and the through and drop port transmission coefficients T_w and D_w of the w-encoding MR. As shown in Figure 1, the multiplication is realized by injecting two continuous-wave (CW) lights at the same wavelength into opposite ends of a single waveguide. The CW light from the left grating coupler is modulated by both the x-encoding and w-encoding MRs and then detected by photodetectors (PDs) PD1 and PD4. Simultaneously, the CW light from the right grating coupler is modulated only by the w-encoding MR and is detected by PD2 and PD3.

As shown in Figure 1, S_{out} is determined by $I_1 - I_2 + I_3 - I_4$ which is proportional to $(f - g)(T_w - 2D_w)$. By mapping the parameters such that $x \propto (f - g)$ and $w \propto (T_w - 2D_w)$, the output S_{out} directly represents the signed multiplication of $x \times w$. The fundamental advantage of this bidirectional design is that it isolates the input encoding $(f - g)$ from the weight encoding $(T_w - 2D_w)$, allowing each to be controlled independently. This enables signed multiplication in the optical-domain without additional modulators or complex time-sequential encoding schemes.

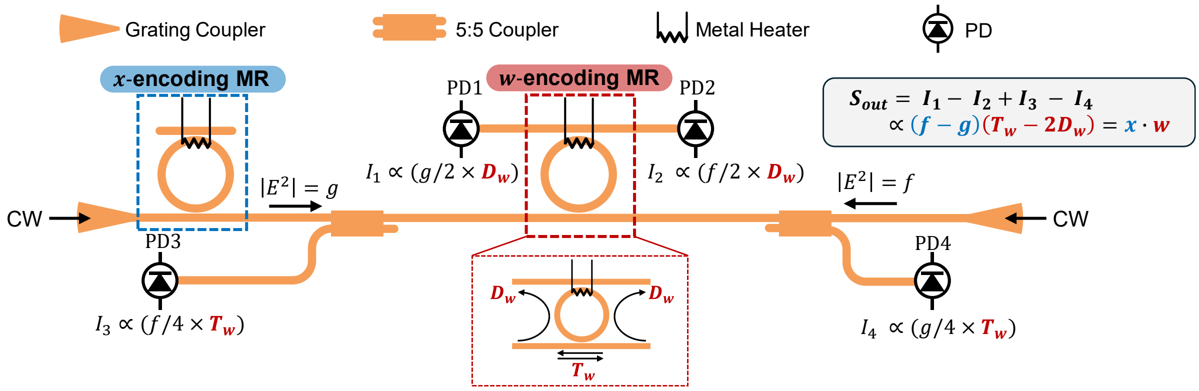

Figure 1. Bidirectional microring weightbank unit cell. The inset highlights the w-encoding MR, where T_w and D_w represent the through and drop coefficients of the microring, respectively.

3. PIC Design and Measurement

The proposed architecture is implemented with a passive Si PIC fabricated by the NanoSOI foundry. As shown in the microscope image of Figure 2(a), the chip features two MRs for x-encoding and w-encoding, each equipped with an integrated metal heater for thermal tuning. The experimental setup, illustrated in Figure 2(a), consists of a multi-channel power supply for electrical control, a CW laser source, and a custom PCB containing PDs and transimpedance amplifiers (TIAs). The optical output signals are converted into voltages by the PD-TIA and captured by a data acquisition (DAQ) unit for digital processing.

The experimental procedure follows a systematic calibration and mapping process. First, the heater voltages for both MRs are swept to characterize the initial transmission responses. To compensate for variations in coupling strengths at the grating couplers, the measured signal is calibrated in the digital domain. Subsequently, a Look-Up Tables (LUT)s for x and w are constructed to map the heater voltages to the desired parameter values of x and w. Before utilizing these LUTs, the thermal crosstalk between the two integrated heaters is modeled and compensated to ensure independent control of each MR during the encoding process.

Figures 2(b) and 2(c) show the multiplication results of $x \times w$ obtained by sequentially sweeping the parameters using the generated LUTs. To verify the independent controllability of each MR, a two-dimensional loop sweep is performed. Figure 2(b) presents the results with x as the sweep variable and w as the stepped parameter, while Figure 2(c) shows the results with their roles interchanged. The consistent linearity in both cases confirms that each MR can be controlled independently without mutual interference. To further evaluate the system's precision, Figure 2(d) presents error histograms derived from 300 random sweeps. The error standard deviations (σ) were measured at 0.022 for x-sweeps (with w held constant) and 0.032 for w-sweeps (with x held constant). These noise characteristics, verified through actual operation across 31 distinct levels, correspond to effective resolutions of 6.51 bits for x and 5.97 bits for w, respectively.

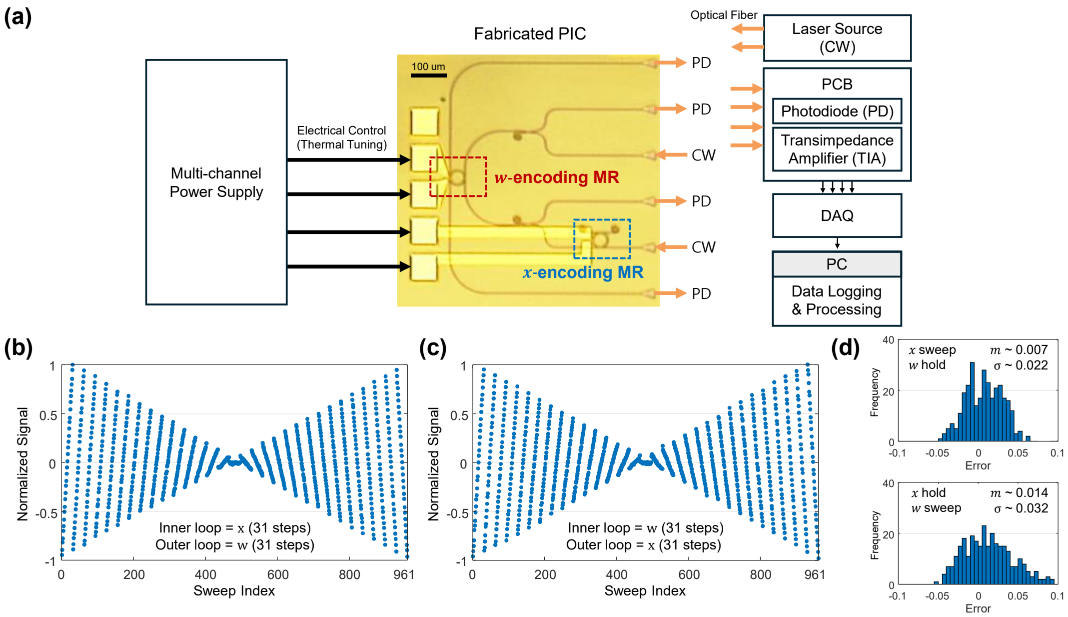

Figure 2. (a) Experimental setup and fabricated PIC. (b)-(c) 2D LUT sweep results for x and w. (d) Error histograms for 300 random sweeps of x and w, where m and σ denote the mean and standard deviation, respectively.

4. Conclusions

We demonstrated a bidirectional microring weightbank for signed encoding within the IM/DD framework. This bidirectional architecture enables direct optical-domain processing of signed values, eliminating the need for auxiliary modulators. The system successfully distinguished 31 levels, achieving effective resolutions of 6.51 bits for inputs and 5.97 bits for weights. This hardware-efficient design can provide a compact platform for high-speed signed operations in next-generation photonic neural networks.

References

[1] J. M. Tait et al., "Neuromorphic photonic networks using silicon photonic weight banks," *IEEE J. Sel. Top. Quantum Electron.* 23(6), 1–15 (2017).

[2] F. Ashtiani et al., "An on-chip photonic trained deep neural network," *Nature* 606, 501–506 (2022).

[3] X. Xu et al., "11 TOPS photonic convolutional accelerator for optical neural networks," *Nature* 589, 44–51 (2021).

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00455038).