Photonic parallel convolutional processing based on spectrum slicing on silicon photonics platform

Zong-Ming Chang
Department of Photonics
National Sun Yat-sen University
Kaohsiung, Taiwan
tyuio9856@gmail.com

Po-Hsiang Huang
Department of Photonics
National Sun Yat-sen University
Kaohsiung, Taiwan
alphonsehuang@gmail.com

Pao-Chia Shih
Department of Photonics
National Sun Yat-sen University
Kaohsiung, Taiwan
y0958698@gmail.com

Chewn-Pu Jou
Taiwan Semiconductor Manufacturing Company
Hsinchu, Taiwan
cpjou@tsmc.com

Chua-Chin Wang
Department of Electrical Engineering,
National Sun Yat-sen University
Kaohsiung, Taiwan
ccwang@mail.ee.nsysu.edu.tw

Yung-Jr Hung
Department of Photonics
National Sun Yat-sen University
Kaohsiung, Taiwan
yungjr@mail.nsysu.edu.tw

2025 IEEE Silicon Photonics Conference (SiPhotonics) | 979-8-3315-0618-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/SiPhotonics64386.2025.10985669

***Abstract*—We have demonstrated a photonic CNN system employing spectrum slicing on a broadband light source and testing with MNIST handwritten characters dataset. The standard deviation of the photonic multiply-accumulate operations is 0.0417, resulting in the bit precision of 4-bit, with a recognition accuracy reaching 94.4%.**

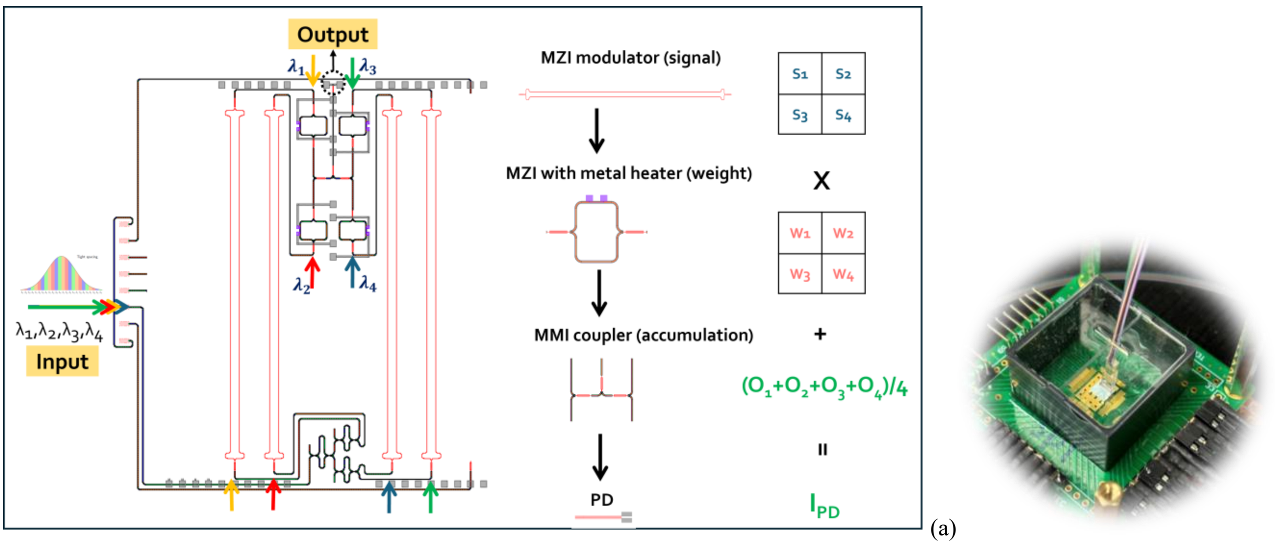
Keywords—Convolutional neural network, broadband light source, wavelength (de)multiplexer, silicon photonics

I. Introduction

Convolutional neural networks (CNNs) have achieved remarkable progress in image classification through extracting feature maps from thousands of images [1]. In the CNN architecture, convolution layers occupy most of computing time and resources due to high computation complexity of matrices multiply accumulate operation. To keep up with the exponentially growing need for processing capability, photonic convolutional processing is believed to be a key to hardware-based artificial intelligence accelerators. Photonic processors access a wide optical bandwidth by exploiting wavelength division multiplexing and eliminate capacitive delay and charge/discharge energy dissipation [2-4]. Most of the proposed system architectures for photonic CNN employed require distinct coherent light sources. For example, a broadcast-and-weight protocol based on cascaded micro-ring resonator (MRRs) arrays for convolution operation on an N-dimensional vector requires N distinct coherent light sources at different wavelengths [5]. It also requires precise phase control over a substantial number of MRRs. Recently, a photonic CNN system was proposed to boost computing parallelism by taking advantage of partially coherent light [6]. This approach eliminates the need for precise control of numerous phase shifters or MRRs and eases the requirements for stringent feedback control and thermal management. However, this partially coherent light approach requires cm-to-m long optical delay lines that tremendously reduces the computation density. In this work, we propose and demonstrate a photonic CNN system employing spectrum slicing on a broadband light source (named SS-PNN) for MNIST handwritten digits dataset classification. This SS-PNN approach addresses the phase and thermal issues in coherent solutions without significantly increasing the device footprint.

II. Design and Experiments

Spectrum-sliced incoherent light source was firstly utilized in wavelength-division-multiplexed passive optical networks (WDM PONs) for supporting the explosive growth of data



(a) (b)

Fig. 1 (a) Schematic diagram and operation principle of a SS-PNN system. (b) Photograph of fiber-pigtailed silicon photonic processor.

traffic. The same idea is utilized here to implement a photonic processor. Our photonic processor implements an on-chip matrix multiplication engine capable of performing parallel MAC operation, as shown in Fig. 1(a). A broadband light source such as super-luminescent diode (SLD) could support all MAC operations without requiring multiple single-wavelength semiconductor lasers or a comb laser. Spectrum-slicing is achieved automatically at the on-chip optical interleaver based on cascaded Mach-Zehnder interferometers (MZIs). Spectrum-sliced light would have multiple peaks separated by the free-spectral-range of the optical interleaver, as shown in Fig. 2(a). We utilized all optical peaks for each kernel to reduce the difference in the optical intensities, which is important to warrant correct MAC operation. The dispersion penalty, if any within very short optical path on chip, was suppressed by using the SLD operating in 1.3-µm wavelength region. The photonic processor equips with high-speed Mach-Zehnder modulators for signal inputs, low-speed micro-heaters in balanced MZIs for weight inputs, and an output Ge photodiode. Optical multiplexing is conducted by serially connecting two optical modulators while the accumulation is achieved by three multimode interference (MMI) couplers. This photonic processor chip was fabricated in a CMOS foundry and occupies an area of 6.25 mm^2.

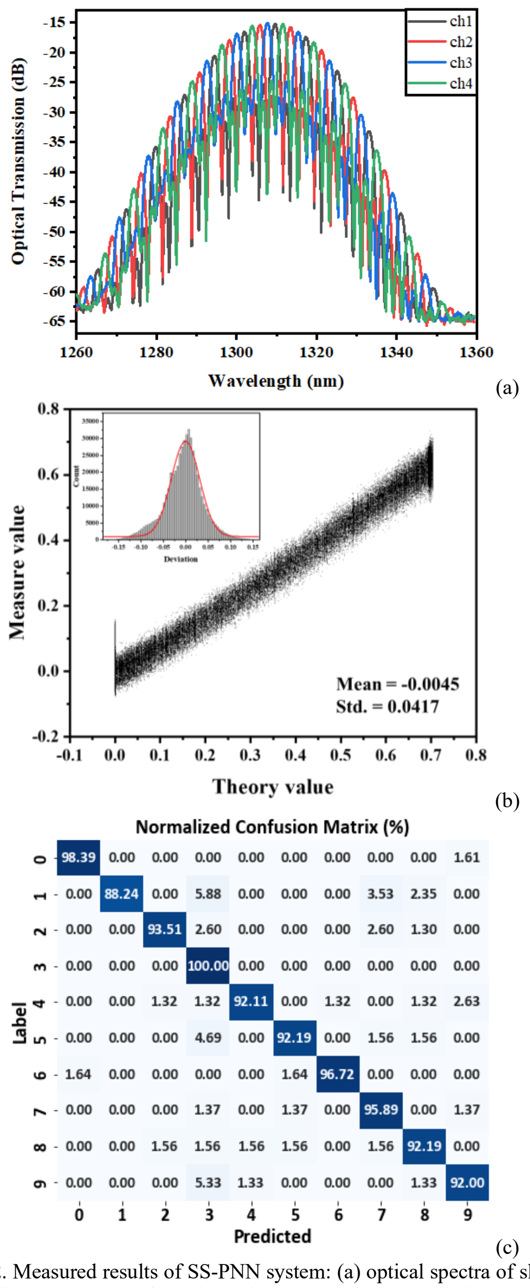


Fig. 2. Measured results of SS-PNN system: (a) optical spectra of sliced SLD light source into four channels, (b) quantitative statistical analysis of convolution accuracy, and (c) The confusion matrix of recognizing 710 characters in the MNIST database.

We successfully implemented MNIST handwritten characters dataset classification on SS-PNN system. The CNN model used in our experiments was pretrained by TensorFlow to obtain the weight values. During the convolution operations, we input the image signals through the MZI modulators and the weight signals through the MZIs with metal heater. Then utilizing MMI couplers to complete the multiply-accumulate operation, resulting in an optically computed feature map. These feature maps will be collected and processed by a computer for the following fully-connected processes. We observed some deviations between the theoretical and experimentally obtained feature maps. To quantify the accuracy of photonic convolution process, we perform a statistical analysis on the errors of the first 710 characters as shown in Fig. 2(b). In this case, the standard deviation is 0.0417, resulting in a bit precision of 4-bit. Furthermore, we present the inference results of these 710 characters in the form of a confusion matrix, as shown in Fig. 2(c). The accuracy of the experiment on these 710 characters can reach 94.4%, while the theoretical accuracy of this CNN model (with 10,000 digits) is 96.95%. There remains about 2% deviation between the theoretical and experimental accuracy. This discrepancy could potentially be reduced by increasing the number of characters or minimizing computational errors to bring the results closer to the theoretical values.

III. Conclusions

We propose and demonstrate, for the first time, an SS-PNN system for classifying the MNIST handwritten characters dataset. By incorporating a SLD with WDM, this approach effectively reduces the number of required light sources and eases the precision needed for channel wavelength alignment. Experimental results demonstrate that the SS-PNN computed values have a mean error of -0.0045 compared to theoretical results, with a standard deviation of 0.0417, achieving 4-bit precision. The system also attained a recognition accuracy of 94.4% for 710 MNIST character images.

Acknowledgments

The authors would like to acknowledge photonic layout support by Synopsys Inc. and chip fabrication support by TSMC. This work was supported in part by the Joint Development Project between TSMC and NSYSU.

References

[1] Y. Lecun et al., *Proc. IEEE* **86**, 2278 (1998).
[2] T. Fu et al., *Nat. Commun.* **14**, 70 (2023).
[3] B. Bai et al., *Nat. Commun.* **14**, 66 (2023).
[4] J. Feldmann et al., *Nature* **589**, 52-58 (2021).
[5] A. N. Tait et al., *J. Light. Technol.* **32**, 4029-4041 (2014).
[6] B. Dong et al., *Nature* **632**, 55-62 (2024).