High-Speed Optical Convolutional Neural Network Accelerator with 100 Gbaud EO-polymer/Si Hybrid Optical Modulator

Akito Shinya
Graduate School of Science and Technology for innovation
Tokushima University
Tokushima, Japan
akid3228@gmail.com

Koji Kida
Fuculty of Engineering and Design
Kagawa University
Takamatsu, Japan
kida.koji@kagawa-u.ac.jp

Hiromu Sato
Institute for Materials Chemistry and Engineering
Kyushu University
Kasuga, Japan
h_sato@cm.kyushu-u.ac.jp

Guo-Wei Lu
Division of Computer Engineering
The University of Aizu
Aizu, Japan
guoweilu@u-aizu.ac.jp

Shiyoshi Yokoyama
Institute for Materials Chemistry and Engineering
Kyushu University
Kasuga, Japan
s_yokoyama@cm.kyushu-u.ac.jp

Junichi Fujikata
Institute of Post-LED Photonics
Tokushima University
Tokushima, Japan
fujikata.junichi@tokushima-u.ac.jp

***Abstract*—Convolutional neural networks are a powerful category of artificial neural networks that can extract hierarchical features of biological data to significantly reduce parametric complexity and increase prediction accuracy. In this paper, we propose high-speed optical convolutional neural network accelerator with 100Gbaud EO-polymer/Si hybrid optical modulator and demonstrated proof of concept both in simulation and experiment.**

Keywords—Optical convolutional neural network, EO polymer, Hybrid optical modulator

I. INTRODUCTION

Convolutional neural networks (CNNs), stimulated by biological behavior, abstract the representation of input data in a raw form, and can predict its properties with great precision and parametric complexity. CNNs have applications in computer vision, natural language processing, and other fields. CNN are highly suitable neural networks in fields such as image processing and image recognition. Maintaining the integrity of the specifications. In recent years, image processing technology has attracted attention due to the development of AI technologies such as autonomous driving. Basically, neural networks are calculated using the GPU (CPU). Therefore, the calculation speed depends on the performance of the GPU. In recent years, modern technologies such as autonomous driving technology have grown rapidly. It means that the amount of computation required will be very large. As a result, there is a problem that the calculation time by the GPU and the power consumption of the computer increase. Optical neural networks (ONN) that use optical signal processing in parallel to solve this problem are attracting attention [1]. ONN has the potential to achieve ultra-high-speed computation that enables electrically processed parts such as interconnection by optical communication bands. Therefore, high-speed and low power optical modulation technique is essential for highly-efficient optical convolutional neural network accelerator.

In this study, we propose high-speed optical convolutional neural network accelerator with EO polymer/Si hybrid modulators [2-3] and demonstrate the proof of concept using MNIST data set both in simulation and experiment.

II. PRINCIPLE OF ONN

Figure 1 shows the CNN configuration. The input data vector is encoded as the intensity of the time symbol of the serial electrical waveform at a symbol ate of $1/\tau$ (baud) where τ is the symbol period. The convolutional kernel in the CNN is represented by a weight vector W of length R encoded into optical power via optical attenuators. A weighted replica is generated for the kernel wavelength channel. Input data weighted differently for each wavelength is synthesized. This results in convolution between the input data and W(weights) relative to the convolutional window. The calculation speed is given in 2 R/τ TOPS. Since the speed of this process varies depending on both the baud rate and the number of wavelengths, the number of wavelengths from the microcomb gives a lot of TOPS speeds. In addition, the length of the input data is theoretically unlimited, so data can be processed at very large scales. Simultaneous convolution with multiple kernels is achieved by preparing multiple weighted wavelengths for each kernel. Following multicast and distributed delay, the sub bands (kernels) are demultiplexed and detected separately, generating electronic waveforms for each kernel. As shown in Fig. 1, Kernel data is input at multiple wavelengths. By shifting them by one bit, the kernel can multiply-accumulate operations by combining data flowing at the same time.

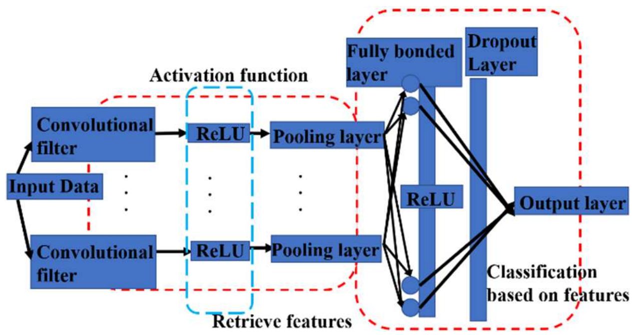

Fig.1 CNN configuration

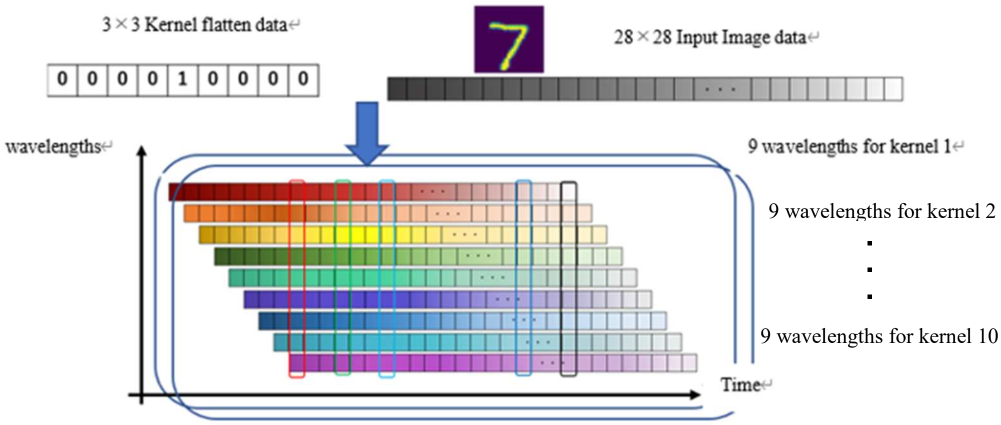

Fig 2 Optical CNN with 9 wavelength data sequence with 1 bit delay for each.

III. ONN CONFIGURATION OF OPTICAL TRANSMISSION IN SIMULATION

CNNs are mainly composed of convolutional filters, pooling layers, fully bonded layers, and dropout layers. In this study, the operation of the convolutional layer is performed with optical transmission system. Using the MNIST dataset, which is one of the machine learning methods, we recognized handwritten characters. An ONN circuit as shown in Fig.2 was established using an optical transmission system in simulation and experiment. With synthesizing the multiple wavelengths of optical data transmission, convolutional ONN was processed by synthesizing data with changing light intensity by ATT (Attenuator) which works as a kernel filter. Input CW laser power of 13dBm is input into Mach-Zehnder interferometer modulator. The main sources of transmission data noise are thermal noise, which is set to be about 1.3nV/$\sqrt{H_z}$ for the receiver circuit. Symbol rate in the optical circuit is 100Gbit/s. Image data of 784, which is a flattened format of "7", one of the MNIST datasets, is input to the MZM (Mach-Zehnder Modulator) by the transfer function and irradiated as optical data. This is performed with multiple ATT along the kernel value, which plays the same role as a multi-wavelength CNN. The input data is multiplicated by combining data at the same time by shifting one bit for each ATT value. After ONN processing, transmission image data with 10 Kernel matrixes were synthesized electrically by activation function of Sigmoid function.

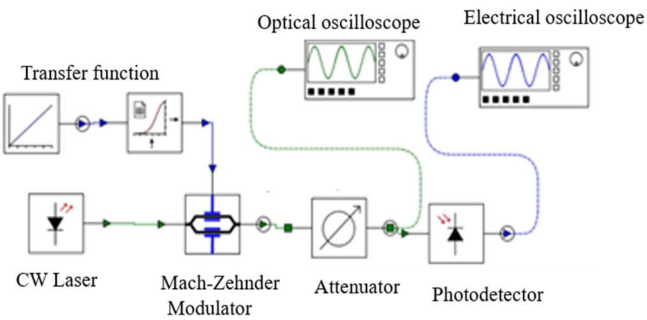

Fig. 3 Optical transmission CNN circuit system in simulation

IV. EO POLYMER/SI HYBRID MODULATOR

In order to realize high-performance EO polymer/Si hybrid optical modulator, we developed side-chain EO polymers with an ultra-high Tg up to 172℃, which are synthesized according to a modified procedure based on [4]. Thermal stability is the other crucial property for EO materials as it depicts the durability of materials under an operating temperature. We have achieved thermal stability up to 110℃ by designing the EO polymer materials with an optimized loading density of the chromophore and adamantyl. The synthesized EO polymer is used to fabricate the SPH modulator, which is fabricated in a travelling-wave (TW) and silicon-slot Mach-Zehnder interferometer (MZI) configurations, consisting of the silicon waveguide, side-chain EO polymer, and SiO_2/polymer cladding and electrodes. Figure 4(a), (b), (c), and (d) show the schematic structure, the picture of the fabricated SPH modulators using a silicon slot phase shifter, and the SEM image of the fabricated SPH modulator. With optimizing device structure, 1.2 Vπ has been obtained for a 1.5mm-long phase-shifter, which shows the fabricated SPH modulator is about 10 times

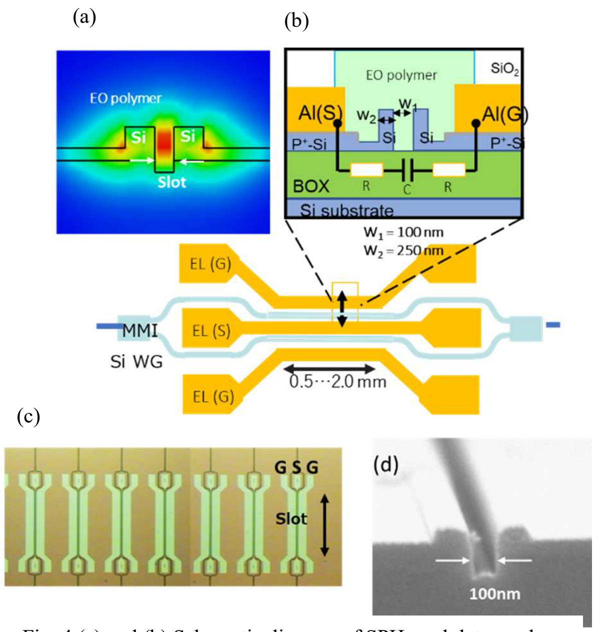

Fig. 4 (a) and (b) Schematic diagram of SPH modulator and optical field contour map. (c) Picture of the fabricated SPH modulators. (d) SEM image of the SPH modulator.

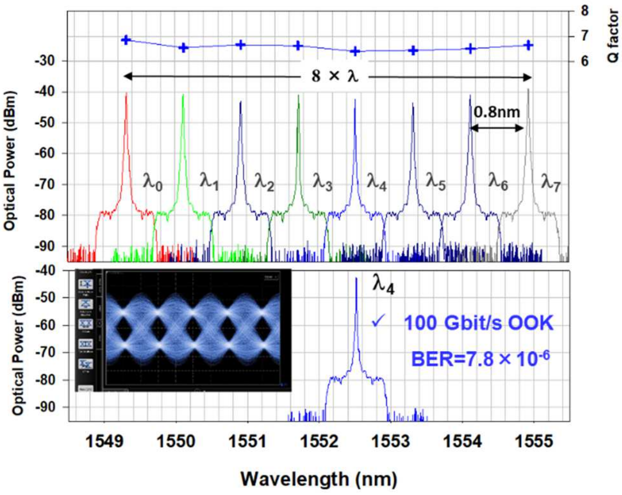

Fig. 5 Optical power and frequency at each wavelength.

more efficient compared with other types of the MZI optical modulators. We achieved enough low optical loss of 0.47dB.

Next, we studied 100Gbps OOK modulation with eight-wavelengths of a wavelength division multiplexing (WDM) filter using a TW EO polymer modulator. The wavelength spacing was 0.8nm. Figure 5 shows optical power spectrum through a WDM filter and an obtained 100Gbps OOK output waveform for λ_4 wavelength. Q factor of the eye diagram for each wavelength was more than 6.2 and around 7.8×10^{-6} bit error rate was obtained for each wavelength. Therefore, an EO polymer modulator is suitable to transmit the MNIST image data with ultra-high-speed signal.

In addition, our developed EO polymer material has very low optical loss less than 1dB with 1.5mm long phase-shifter for full C-band wavelengths, which shows the large number of high-speed image data can be processed in parallel by using the EO polymer/Si hybrid modulator.

V. SIMULATION RESULT OF ONN

To demonstrate the ONN process for MNIST data, optical transmission with 9 wavelength division multiplexing (WDM) system was simulated by optical circuit simulator. Optical transmission system is shown in Fig. 2. Figure 6 shows 10 kernel matrixes for CNN process [1]. Each kernel filter can extract the feature value of the image. After image data transmission with kernel filter of 9 ATTs, transmission data was synthesized electrically by Sigmoid function to minimize the MSE value between raw image data and processed image data. In this study, image data with 256 shades is normalized by maximum value of image data and Kernel filter value was defined by the optical ATT value set to be from -3dB to -10dB. To perform the multiply-accumulate operation, 1bit delay of 10ps was applied to 9 wavelengths image data such as the data with one wavelength is delayed by 10ps with another wavelength data.

Optical transmission data with a kernel filter of ATT was monitored by sampling oscilloscope. As for the system noise, thermal noise was set to be about $1.3\text{nV}/\sqrt{H_z}$. Image

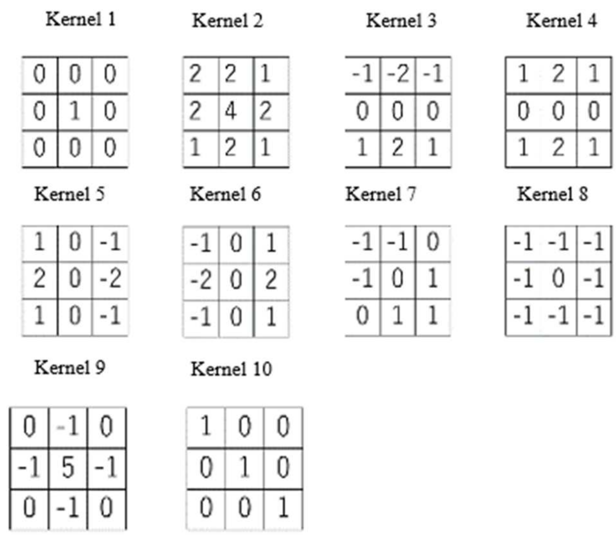

Fig 6. 10 kernel matrices for CNN process

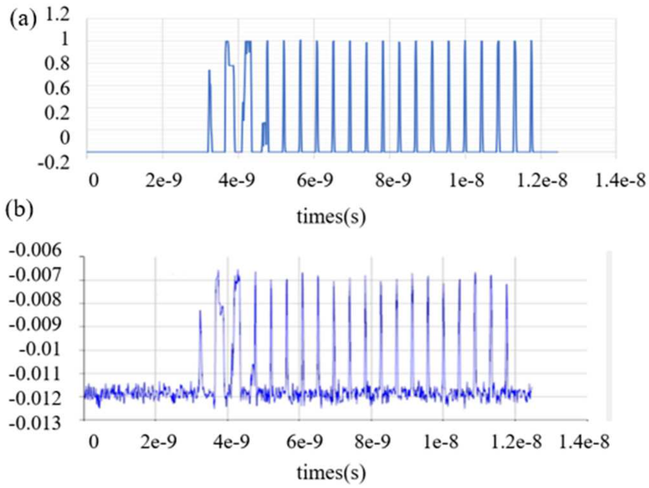

Fig 7. MNIST image data
(a) Input data waveform
(b) Photoreceiver output waveform

data is 28 × 28 matrix and was flattened to be 784 series data in line. Figure 7 shows the input data waveform at 100Gbps and photoreceiver output waveform through the ATT. Figure 8 shows the simulation result of 10 kernel CNN image data. With each kernel, feature value of image data could be extracted by ONN transmission system.

With synthesized 10 CNN date by 10 kernel matrices, MSE value between raw image data and the synthesized data is about 0.0864, which is enough small for image identification. In this simulation, ONN process speed is estimated to be 1.8 TOPS, which is expressed by $2 \times 9 \times$ 100Gbaud. With increase in the number of wavelengths, ONN process speed can be accelerated [1].

VI. EXPERIMENTAL RESULTS OF ONN

Finally, ONN process was demonstrated experimentally. Figure 8 shows the experimental set-up for ONN. Image data of 256 shades was signalized electrically by an arbitrary waveform generator (AWG). In the first

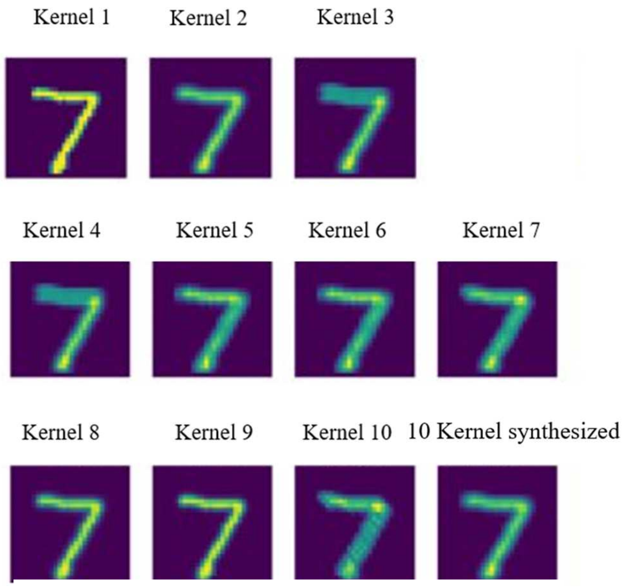

Fig. 8 Result of 10 kernels optical CNN data

experiment, image data transmission rate was set to be 4Gbaud to avoid the frequency bandwidth correction processing of the optical signal. Figure 10 shows the experimental data waveform both for the electrical input data and photoreceiver output data. These data are consistent with each other very much by tuning the voltage scale. Thermal noise was comparable to that of simulation. In this experiment, tunable laser source was applied with 100GHz free spectral range and 9 wavelengths in C-band wavelengths. With applying 10 kernel filter matrices with optical power attenuation, 0.72TOPS of high-speed data processing was demonstrated. Figure 11 (a) and (b) show optical transmission simulation and experimental results for MNIST of 7 character. MSE is 0.25 in experiment, which is larger than that of simulation, because electrical circuit in AWG has high-bandpass filter and low frequency portion of image data was filtered. By calibrating the electrical signal to be matched with MNIST data value with broader bandwidth, higher-recognition accuracy would be achieved.

VII. CONCLUSION

Convolutional neural networks inspired by biological brain systems are a powerful category of artificial neural networks that can extract hierarchical features of biological data to significantly reduce parametric complexity and increase prediction accuracy It is. They are very useful for machine learning tasks such as computer vision, speech recognition, image recognition, and medical diagnostics. Here, image convolution was performed using "7" 784-pixel MNIST data, which is one of the machine learning data, and image restoration was performed with high

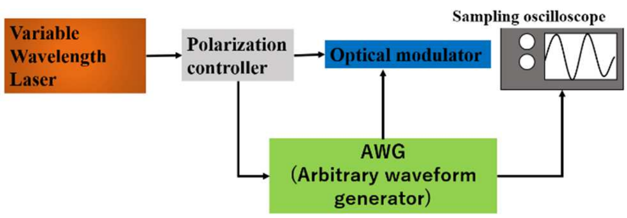

Fig. 9 Schematic of measurement system for ONN.

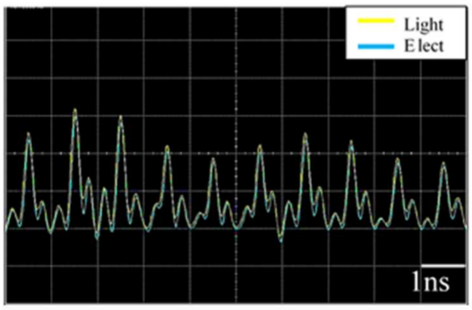

Fig. 10 Experimental data waveform for input electric signal and receiver output signal.

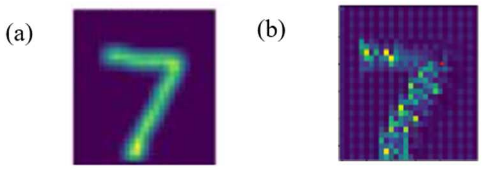

Fig 11. 10 Kernel sum CNN data
(a) Optical transmission simulation results.
(b) Experimental circuit processing results

accuracy. We demonstrated 0.72TOPS ONN process and it can be further increased with 100Gbaud by using the EO polymer/Si hybrid optical modulator.

ACKNOWLEDGMENT

These research results were partly obtained from the commissioned research of 02101 by National Institute of Information and Communications Technology (NICT), JAPAN.

REFERENCES

[1] Xingyuan Xu, Mengxi Tan, Bill Corcoran, Jiayang Wu, Andreas Boes, Thach G. Nguyen, Sai T. Chu, Brent E. Little, Damien G. Hicks, Roberto Morandotti, Arnan Mitchell, and David J. Moss, "11 TOPS photonic convolutional accelerator for optical neural networks," Nature 589 (7), pp. 44-51 (2021).

[2] Guo-Wei Lu, Jianxun Hong, Feng Qiu, Andrew M. Spring, Tsubasa Kashino, Juro Oshima, Masa-aki Ozawa, Hideyuki Nawata, and Shiyoshi Yokoyama, "High-temperature-resistant silicon-polymer hybrid modulator operating at up to 200 Gbit s^{-1} for energy-efficient datacentres and harsh-environment applications," Nat. Communications 11, 4224, pp. 1-9 (2020).

[3] Clements Kieninger, Christoph Fullner, Heiner Zwickel, Yasar Kutuvantavida, Juned N. Kemal, Carsten Eschenbaum, Delwin L. Elder, Larry R. Dalton, Wolfgang Freude, Sebastian Randel, and Christian Koos, "Silicon-organic hybrid (SOH) Mach-Zehnder modulators for 100 GBd PAM4 signaling with sub-1 dB phase-shifter loss," Opt. Express, vol. 28, p. 24693-24707 (2020).

[4] Piao, X. et al., "Nonlinear optical side-chain polymers post-functionalized with high-β chromophores exhibiting large electro-optic property," J. Polym. Sci. Part A 49, pp. 47–54 (2011).