Silicon Photonic Hopfield-like Electro-optical Recurrent Network for Time-series Data Processing and Recognition

Guangwei Cong, Noritsugu Yamamoto, Rai Kou, Yuriko Maegami, Morifumi Ohno, and Koji Yamada
Platform Photonics Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Onogawa 16-1, Tsukuba, Ibaraki 305-8569, Japan
gw-cong@aist.go.jp

Abstract: We propose and experimentally demonstrate a Hopfield-like electro-optical recurrent network based on silicon photonic circuits for processing time-series data to extract feature vectors just by one-time sampling, which can offer robust and simple waveform recognition. © 2022 The Author(s)

1. Introduction

Digital processors are facing energy and speed bottleneck with transistor technology approaching the physical limit. On the other hand, in future highly interconnected society, large amounts of analog signals need to be processed and recognized in high speed, low latency, and low power. It is becoming more and more challenging to fulfill this requirement for traditional digital processors. To analyze time series signals by digital processors, the signals need to be sampled to digital ones with very high resolution, which causes large latency and consumes a lot of energy. Recently, novel computing systems utilizing optics/photonics have been demonstrated to accelerate signal processing and artificial intelligent inference, such as optical neural network [1,2] and reservoir computing [3,4]. Electrical-photonic heterogeneous architecture could be one of promising solutions for future processors [5]. In this work, we propose a Hopfield-like electro-optical recurrent network by adding an optical-electrical (OE) feedback to our previous projection-based silicon photonic classifier circuits [6] and demonstrate a prototype of this architecture by using home-made feedback circuits. This architecture does not need continuous and fine sampling for temporal waveforms and can offer robust recognition just by one-time data sampling at the end of signal. We verify the role of OE feedback in recognizing the time-series Mel Frequency Cepstral Coefficients (MFCC) [7] of Japanese vowels. This work could enable a high-efficient OE hybrid hardware for directly analyzing temporal analog signal.

2. Architecture and Principle

Figure 1(a) shows the proposed OE recurrent network architecture, in which the photonic circuits is using our previously reported projection-based photonic classifier [6], which consists of Mach-Zehnder interferometers (MZI) and phase shifters (PS). The optical outputs are converted to electrical signals by photodetectors (PD) and the four electrical signals are feedbacked to four MZIs via amplifiers (Amp.), forming OE loops. The MZI provides EO nonlinearity. The analog data is input to a MZI with a continuous-wave laser input, and the signals detected by PDs

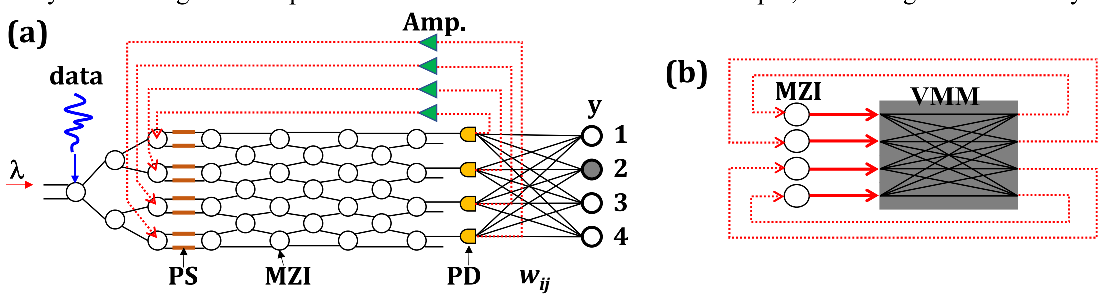

Fig. 1. (a) Silicon photonic OE recurrent network architecture. (b) Continuous Hopfield-like recurrent network principle.

are connected to the result y by using the readout coefficient w_{ij} which is done by electronics. The OE loop follows the continuous Hopfield-like recurrent network principle [8] as shown in Fig. 1(b). The output of each MZI (works as neurons) is feedbacked to other three MZIs and itself, depending on the state of the vector-matrix multiplication (VMM) which is implemented by the MZI mesh. The Hopfield-like recurrent network can be trained to an associative memory mode corresponding to the input waveforms. This current architecture can be trained to remember four waveforms which are marked by a four-element feature vector of the optical output at the end of signal input. For some simple tasks, this feature vector can directly give out the results of classification, while for complex tasks, this vector can be used as the input for a linear transformation to give out the results like that done in photonic reservoir computing [3,4]. This Hopfield-like network offers robust waveform recognition due to the memory effect. Even though the input waveform deviates from its learned one, it can be correctly recognized.

For easy understanding, we perform recognition for bit sequences by simulation to explain the principle. The input data has a unit of π and the feedback amplifier's gain coefficient (convert optical powers to phase shift) is set to 1.5π. The loop delay can be a tunable parameter and one bit delay is assumed. Here, no readout coefficients are used to understand the recurrent architecture itself. We consider four patterns: all-one (11…11), all-zero (00…00), half-zero/half-one (00…11...), half-one/half-zero (11…00...), with the bit lengths of 2, 4, 8, and 16. Firstly, Fig. 2(a) shows the OE loop induced responses of all ports for the 16 bit all-one pattern under the initial random phase distribution (before training). The network can self-consistently converges to a steady state after 4-5 OE loops when continuously input the same signal. Secondly, we train the network to distinguish above four patterns by just sampling the optical output at the last bit. We use the forward propagation training algorithm here [6]. Each pattern will have an output power distribution (feature vector) at four PDs, and we use the index of maximum-power position to mark each pattern. Obviously, if without OE loops, only by sampling the last output, it is no way to distinguish all-one and half-zero/half-one, and all-zero and half-one/half-zero as well, because their last bits are same. However, with OE loops, we can distinguish these patterns with 100% accuracy up to the 8-bit length as seen in Fig. 2(b). If the length of continuous identical bits is much longer than the response time given in Fig. 2(a), it will fail to distinguish the patterns because the influence from long-term history information will become negligible, as shown by the 16-bit case in Fig. 2(b). The results of temporal responses for 4-bit patterns after training are shown in Figs. 2(c)-2(f). For each pattern, the system maps the maximum power to the respective ports of 1, 2, 3, 4 at the last bit (indicated by colored arrows), indicating the pattern separation. Finally, we verify the recognition robustness by deviating the bit pattern to analog ones by 20% to the maximum. Using the learned model from training for Figs. 2(c)-2(f), we input these deviated patterns for verification and obtain the results in Figs. 2(g)-2(j). Comparing Figs. 2(g)-2(j) to Figs. 2(c)-2(f), respectively, we notice that the deviated analog patterns can still be correctly recognized. The Hopfield-like recurrent network can be trained to build an associative mode to memorize four patterns. Even though the input deviates from its training data, the learned network state tries to recover its output to that of its closest training data, offering robust pattern recognition.

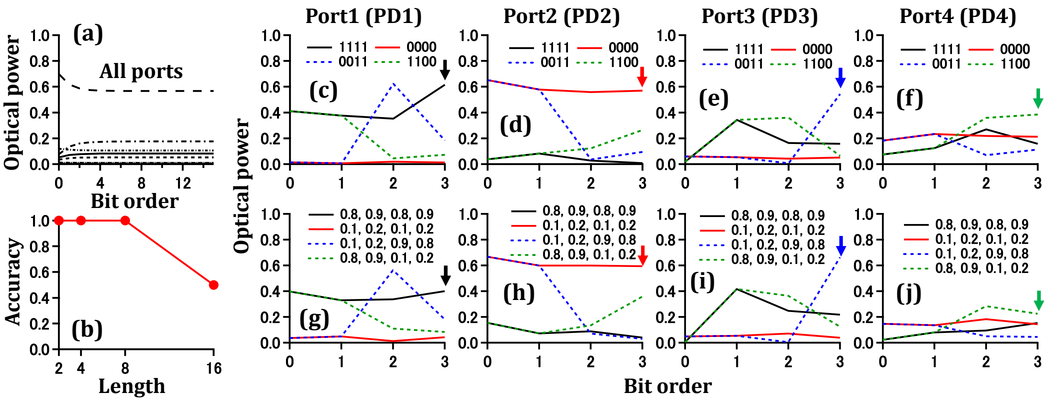

Fig. 2. (a) OE loop induced optical responses of all eight ports for an input of 16 bits of one under random phase distribution (before training). (b) Classification accuracy in relation of bit lengths for the 4 patterns: all-one, all-zero, half-zero/half-one, half-one/half-zero. (c)-(f) Temporal responses of four 4 bits pattern at four ports (corresponding to the PD number in the up-to-down order) after training. (g)-(j) Verified temporal responses of four analog signals deviated from the trained digital ones by 20% to the maximum.

3. Experimental Results of Japanese Vowel Recognition

To demonstrate the prototype of the architecture in Fig. 1(a), we built up the experimental setup in Fig. 3(a). The on-chip photonic circuit in Fig. 1(a) is the silicon photonic chip in Fig. 3(a) which was explained in [6]. The feedback circuit is home made on a bread board, in which the signals from PD are combined with those from direct-current (DC) sources by operational amplifiers. The combined signals are feedbacked to four MZIs. We control all phase shifters by two 40-channel DC sources and use an arbitrary function generator (AFG) and oscilloscope (OSC) for data input and result readout, respectively. Since the fiber and cable are long, the OE loop time is about 20 μs. This time is related to signal coding and can be greatly shorten if adopting an integrated form. The current setup is just for proof-of-concept demonstration and ideally it should be realized with on-chip PD integrated and feedback circuits packaged. The laser was set to the 1530 nm wavelength and TE polarization. The forward propagation training algorithm was adopted to train the chip in this experiment and the parameters were same as those used in [6].

Next, we perform Japanese vowel classification and recognition using the setup in Fig. 3(a) and readout coefficients (off chip) in Fig. 1(a). The whole system has 62 training parameters (46 on chip + 16 off chip) in total. The PD output was sampled only one time after finishing data input and normalized before linear transformation. The output layer used the Softmax normalization. Since the raw voice data contain lots of redundant information [1], it is rare to directly use the raw waveforms for recognition tasks in traditional recurrent neural networks (RNN) [7]. The traditional RNN usually takes the preprocessed data such as MFCC coefficients [7] as the network input. We treat Japanese vowels similarly to obtain the time dependent MFCC coefficients for each vowel and input them into the MZI after being coded by AFG. We use the Japanese vowel dataset in [9] and perform recognition for four vowels ("ha", "hi", "hu", "he") as seen in Fig. 3(b). For each vowel, four MFCC components of 11 time-sections are extracted for the colored parts in raw waveforms and flattened to the temporal data. These data are recoded to ~250 kHz by AFG, thus, the OE loop has a ~5-bit delay. For comparison, simulation was also done by assuming a 5-bit delay. For simulation, we used the vowels spoken by 28 persons from 21 to 56 years old. Thus, there are 118 vowels in total, 80 used for the train set and 38 for the test set. For experiment, we used 64 vowels (40 for the train and 24 for the test) to save the training time (limited by data input and readout). Fig. 3(c) shows the simulated result of recognition accuracy. Without OE loops (only using the last data point for training), the accuracy is about 56.3%. It can be enhanced by 20-30% to 86.3% for the train and 78.1% for the test if using OE loops. We performed two times training experiments and obtained the accuracies of about 75.0-82.5% and 66.7-70.8% for the train and test sets, respectively, as shown in Fig. 3(d), with only 62 parameters. The experimental accuracies are several percent lower than the simulated ones, which may be related to the system stability due to a long loop and jumper cables. The current experimental setup can be further improved in future from various aspects such as stability, amplifier gain tuning, integrated circuits, and packaging to shorten the loop time for high-speed waveform processing.

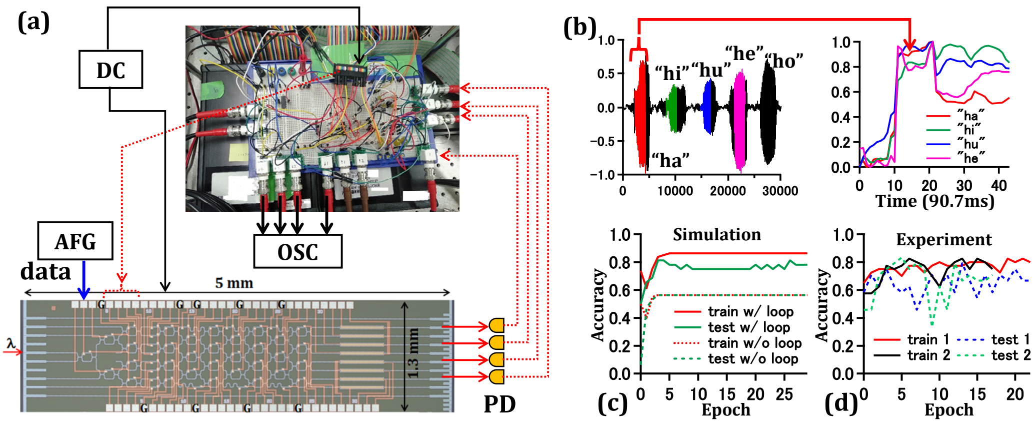

Fig. 3. (a) Experimental setup of the OE recurrent network using home-made feedback circuits. (b) An example of raw vowel waveform in the dataset [9] and the extracted temporal MFCC coefficients for each vowel. (c) Simulated training curves of accuracy with and without OE loop. (d) Experimental training curves of accuracy using the setup in (a) which takes temporal MFCC waveforms in (b) as the input analog data.

4. Summary

We proposed a Hopfield-like OE recurrent network and demonstrated its prototype using silicon photonic circuits. Using this prototype, we experimentally demonstrated recognition for four Japanese vowels by using MFCC as the input analog data, evidencing the role of OE recurrent network in recognizing time-series signal.

Acknowledgment

This work was supported by Japan Science and Technology Agency (JST), CREST Grant Number JPMJCR21C3.

[1] Y. Shen, et al., "Deep learning with coherent nanophotonic circuit," Nat. Photon. **11,** 441-446 (2017).
[2] X. Y. Xu, et al., "11 TOPS photonic convolutional accelerator for optical neural networks," Nature **589,** 44-51 (2021).
[3] K. Vandoorne, et al., "Experimental demonstration of reservoir computing on a silicon photonics chip," Nat. Commun. **5,** 3541 (2014).
[4] M. Nakajima, et al., "Scalable reservoir computing on coherent linear photonic processor," Communications Physics **4,** 20 (2021).
[5] N. Harris, "Passage—A Wafer-Scale, Programmable Photonic Communication Substrate," in Conference of HOT CHIPS 2022.
[6] G. Cong, et al., "On-chip bacterial foraging training in silicon photonic circuits for projection-enabled nonlinear classification," Nat. Commun. **13,** 3261 (2022).
[7] E. Rejaibi, et al., "MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech," Biomedical Signal Processing and Control **71**, 103107 (2022).
[8] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," The Proceedings of the National Academy of Sciences (PNAS) **79,** 2554-2558 (1982).
[9] Toshisada Deguchi, Ghen Ohyama and Hideki Kasuya (2011): Vowel Database: Five Japanese Vowels of Males, Females, and Children Along with Relevant Physical Data (JVPD). Speech Resources Consortium, National Institute of Informatics. https://doi.org/10.32130/src.JVPD.