An integrated optical neural network chip based on Mach-Zehnder interferometers

Xu Zhao, Zhenming Yu, Benlin Liu, Yu Li, Hongwei Chen, Minghua Chen
Beijing National Research Center for Information Science and Technology, Beijing, China
Department of Electrical Engineering, Tsinghua University, Beijing, 100084, China
email address: chenmh@tsinghua.edu.cn

***Abstract*—We propose an integrated optical neural network chip based on Mach-Zehnder interferometers, which can replace analog to digital conversion device and solve energy consumption problems in large-scale fiber-optic communication systems.**

***Keywords*—optical neural network, silicon based integration, optical communication, MZI**

I. Introduction

In face of the demand for massive information processing and transmission in the future, traditional digital electrical signal processing faces a dual challenge of bandwidth and energy consumption[1,2]. As an analog signal processing technology, the main advantage of optical signal processing is reduction of required energy consumption in high-speed, large bandwidth and high sampling rate signal processing tasks. As the sampling rate of the signal increases, the energy consumption of the digital signal processing system will show an explosive increase[3]. Therefore, a neural network photonic on-chip system based on silicon-based integration is proposed. This system is combined with digital signal processor units, giving full play to the advantages of high efficiency, large bandwidth and low power consumption of photonic signal processing, and alleviates the pressure of ADC and DSP of optical communication receiver.

II. System Design

The schematic diagram of the system is shown in Fig. 1. in which different colors are used to distinguish optical and electrical components.

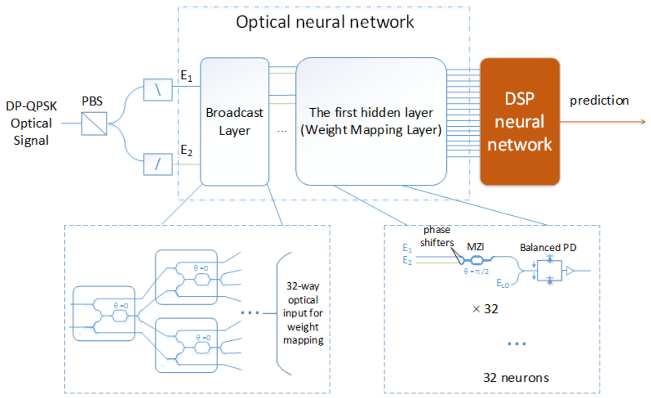

Fig. 1. Schematic diagram of optical neural network based on MZIs.

The neural network framework we choose is a binary neural network with a weight of +1 or -1 and the activation function is a symbolic decision function, which means that any real number are mapped to +1 or -1 by its symbol. This binarization makes it easy to implement on hardware, and as long as the number of trainings is sufficient, the performance of neural network is not significantly reduced compared to the ordinary neural network[4]. The weight of binarization is easily obtained by adjusting the Mach-Zehnder interferometer(MZI) parameter mapping. We alter MZI transmission matrix by setting the voltages on both the internal and external phase shifters of each MZI[5,6], and a variety of optical neural network auxiliary structures containing broadcast layers and weight-mapping layers can be realized according to this principle.

After the DP-QPSK signal being separated by the PBS, the separated signals are adjusted to the same polarization direction to fit in the MZI system. The basic unit of the broadcast layer is the splitter and MZI, realizing the process of copying two input optical signals up and down to form four output signals after which optical power is halved. The basic unit of the broadcast layer forms a tree structure by layer superposition, and generates an input of 2^5=32 hidden layer units, that is, 64 input optical signals. The basic unit of the weight mapping layer consists of an MZI and two balanced photo detectors followed by an 180 degree optical mixer. The phase modulation of the input light is realized by adjusting the lengths of the two input arms of the MZI, and the correspondence between the phase and the hidden layer weight is realized through the balance detection. By setting the internal phase of MZI in each neuron of weight mapping layer to $\pi/2$, each MZI realizes the function of adding two ways of signal together. By adjusting the lengths of the two input arms of the MZI, phases of two input optical signals($\boldsymbol{E_1}$ and $\boldsymbol{E_2}$) are set to certain numbers, namely α_1 and α_2. Given that

$$\boldsymbol{E_1} = |E_1|e^{j\alpha_1} \tag{1}$$

$$\boldsymbol{E_2} = |E_2|e^{j\alpha_2} \tag{2}$$

then we get the optical signal of MZI upper output arm $\boldsymbol{E}$ as

$$\boldsymbol{E} = \kappa(|E_1|e^{j\alpha_1} + |E_2|e^{j\alpha_2})\, e^{j\varphi} \tag{3}$$

where both κ and φ are constants. Given that the local optical oscillator signal is

$$\boldsymbol{E}_{LO} = |E_{LO}|e^{j\alpha_{LO}} \tag{4}$$

the output power signal of the j-th balanced photo detector unit is shown in (5):

$$P_j = 2\kappa e^{j\varphi}|E_{LO}|[|E_1|\cos(\alpha_1 - \alpha_{LO}) + |E_2|\cos(\alpha_2 - \alpha_{LO})] \tag{5}$$

By setting α_1 and α_2 to $\pi/4$, $3\pi/4$, $5\pi/4$ or $7\pi/4$, the value of P changes in limited cases; given that the input is two optical signals($\boldsymbol{E}_1$ and $\boldsymbol{E}_2$) after depolarization multiplexing, the transfer function $F_j(\boldsymbol{E}_1,\boldsymbol{E}_2)$ for the j-th weight mapping unit is shown in (6):

$$P_j = F_j(\boldsymbol{E}_1,\boldsymbol{E}_2) = \lambda_{1j}I_1 + \lambda_{2j}Q_1 + \lambda_{3j}I_2 + \lambda_{4j}Q_2 \tag{6}$$

where λ_{ij} (i=1,2,3,4; j=1,2, … 32) is either +1 or -1, and each λ_{ij} is one unit of the first hidden layer's weigh matrix; I_1, Q_1, I_2, Q_2 are the real and imaginary parts of the two-way polarized optical signal IQ modulated, also used for data training in simulation.

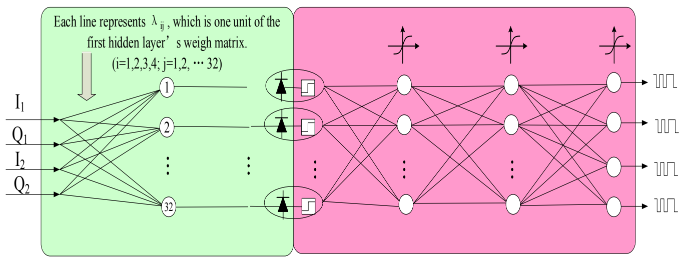

Fig. 2. Schematic diagram of optical neural network for simulation, different colors are used to distinguish optical and electrical parts.

It is learned by calculation that when the number of neurons in the hidden layer is 32, the number of MZIs in the broadcast layer is 63; the weight mapping layer MZI number is 32, and each MZI is set to implement the same direction addition function.

III. Simulation Result

Our simulation verified neural network structure consists of one input layer(4 neurons), 7 hidden layers(32 neurons each) and one output layer(4 neurons). The input is the real and imaginary parts of the two-way polarized optical signal IQ modulated and the output is the prediction of the input position by the neural network, as the structure shown in Fig. 2. The training data are the real and imaginary parts of the two DP-QPSK received signals separated by polarizing beam splitter in the optical fiber communication system and the targets are the real and imaginary parts of the 100Gbps DP-QPSK received signals in the optical fiber communication system. The influence of the channel on the transmitted signal is shown by the constellation map. For optical signals with Gaussian white noise, simulation comparison is also performed according to different optical signal-to-noise ratios. The basis for measuring the quality of the neural network model is the comparison between the accuracy of the verification set of the constellation after the nearest neighbor decision and the true decision result recovered by the neural network.

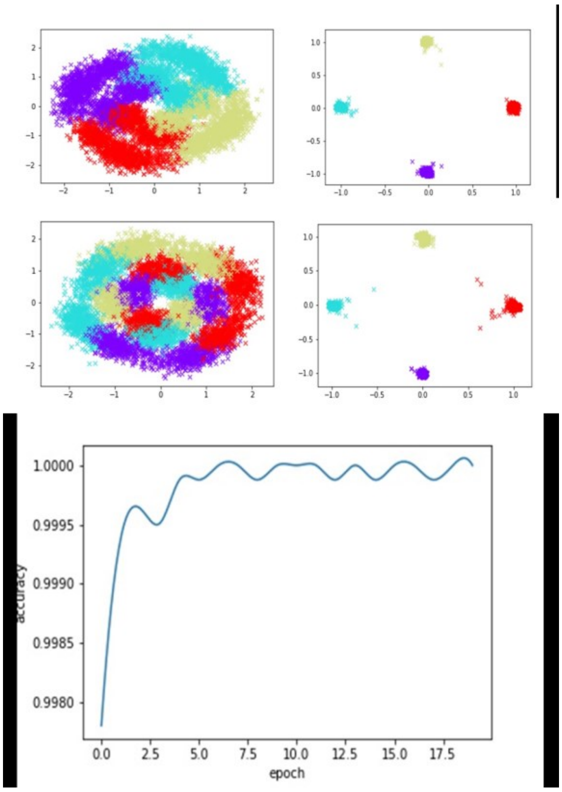

Fig. 3. Constellation map before/after the neural network system and trend of verification set accuracy during iteration of training(20dB OSNR).

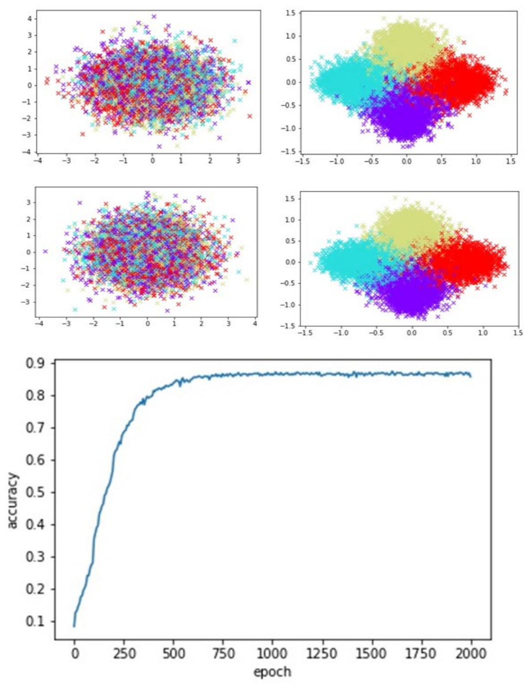

Fig. 4. Constellation map before/after the neural network system and trend of verification set accuracy during iteration of training(5dB OSNR).

For high OSNR(20dB) data, the training converges very quickly, and convergence can be achieved with only a few iterations, with an accuracy of almost 100%. For low OSNR(5dB) data, the training takes a long time and costs more than 1,000 iterations to reach convergence, and the accuracy is close to 90%.

IV. Conclusion

We demonstrate an integrated optical neural network chip design based on Mach-Zehnder interferometer which can replace analog to digital conversion device and solve energy consumption problems in large-scale fiber-optic communication systems. The optical part of this system contains a broadcast layer and a weight-mapping layer, realizing IQ demodulation of 100Gbps DP-QPSK signal and giving full play to the advantages of large bandwidth and high speed of optical signal processing system.

Acknowledgment

This work has been supported by National Natural Science Foundation of China under contract 61771285.

References

[1] Barroso L A, Hölzle U. The case for energy-proportional computing[J]. Computer, 2007 (12): 33-37.

[2] Andrae A S G, Edler T. On global electricity usage of communication technology: trends to 2030[J]. Challenges, 2015, 6(1): 117-157.

[3] Rivet, F., Deval, Y., Begueret, J. B., Dallet, D., Cathelin, P., & Belot, D. (2010). The experimental demonstration of a SASP-based full software radio receiver. IEEE Journal of Solid-State Circuits, 45(5), 979-988.

[4] Courbariaux, M., Hubara, I., Soudry, D., El-Yaniv, R., & Bengio, Y. (2016). Binarized neural networks: Training deep neural networks with weights and activations constrained to+ 1 or-1. arXiv preprint arXiv:1602.02830.

[5] Shen, Y., Harris, N. C., Skirlo, S., Prabhu, M., Baehr-Jones, T., Hochberg, M., & Soljačić, M. (2017). Deep learning with coherent nanophotonic circuits. Nature Photonics, 11(7), 441.

[6] Miller D A B. Perfect optics with imperfect components[J]. Optica, 2015, 2(8): 747-750.