仿神经

On-Chip Optical Neuromorphic Computing

Yichen Shen,[1†]**, Scott Skirlo,**[1]**, Nicholas C. Harris,**[2]**, Dirk Englund,**[2] **and Marin Soljačić**[1]

[1]*Department of Physics, Massachusetts Institute of Technology,Cambridge, MA 02139, USA*

[2]*Research Laboratory for Electronics, Massachusetts Institute of Technology,Cambridge, MA 02139, USA*

[†]*ycshen@mit.edu*

Abstract: We propose an on-chip nanophotonic system that do the neural network computing all in optical domain. Our system is able to give equivalent 相等的，相当的 learning performance, while potentially achieve 3 orders of magnitude faster speed than conventional electronic neural nets.

OCIS codes: 200.4260 (Neural Networks), 200.4700 (Optical Neural Systems), 130.3120 (Integrated Optical Device)

The brain, unlike the von Neumann processors found in conventional computers, is very power efficient, extremely effective at certain computing tasks, and highly adaptable to novel situations and environments. Artificial Neural Network (ANNW) is an algorithm that mimic the signal processing architecture in the brain, and has recently received an explosion of interests [1]. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics.

Nowadays, both the quantity and also the size of data files are growing at a rapid speed, therefore computing speed and the power efficiency is the key on evaluating the performance of any machine learning algorithm. In real life applications, artificial neural networks can contain up to millions of units in each hidden layers. The shear number of units make forward propagation the rate limiting step in many applications. Many efforts have been made to increase the computing speed of artificial neural networks. Among these methods, neuromorphic computing, where electronic devices are 调整 tweaked to have optimized architecture for the kinds of computing for neural networks [2], is one of the most heavily invested techniques. Photonic platforms offer an alternative approach to microelectronics. The high speeds, high bandwidth, and low cross-talk achievable in photonics are very well suited for an ultrafast ANNW processor. In addition, the high wall-plug 墙上的插座 efficiencies of photonic devices may allow such implementations to match or 使黯然失色 eclipse equivalent electronic systems in low energy usage. With the recent advances in quantum optical devices and on-chip nanophotonic circuit fabrication, we reasoned it is possible to design a viable on-chip optical neural network (ONNW) architecture. In this work, we first propose that conventional neural networks architecture can be entirely and equivalently represented by on-chip optical components. We then experimentally demonstrate the on-chip optical computing of the weighted sum part in our ONNW, with the help of which we show that our optical neural networks are able to give similar accuracy performance on a standard evaluation dataset compared with result from conventional neural networks. In the last part, we show that in certain condition optical neural nets can be made at least 3 orders of magnitude faster in forward propagation than conventional neural nets.

Artificial neural network architecture contains an input layer, at least one hidden layers, and an output layer. In each layer, information propagate through neural network via linear combination (e.g. matrix multiplication) followed by nonlinear activation function applied to the result from linear combination. In training an artificial neural network model, data are fed into the input layer, and output is calculated through the forward propagation step. Then the parameters are optimized through the back propagation procedure. The general architecture of our optical neural network (ONNW) computing is depicted in Fig. 1A. It is mainly composed of three optical processing units:

通过MZI实现酉矩阵的乘法

1. **Optical Interference Unit.** Which is used to perform arbitrary *unitary* 酉矩阵 matrix multiplication on the input optical signal. The unitary matrix can be obtained using a network of Mach-Zehnder interferometers, as shown in Fig. 1D,E. Mathematically, it can be rigorously proved that any arbitrary unitary matrix can be represented by the network of Mach-Zehnder interferometers [3] .

通过光放大单元，将酉矩阵转换为任意矩阵

2. **Optical Amplification Unit.** Which is used to generalize the unitary matrix to arbitrary matrix operation. In general, any arbitrary matrix can be generated using optical interference and linear amplification through SVD decomposition [4].

3. **Optical Nonlinearity Unit.** Which is used to apply the nonlinear activation function. In modern optics, luckily, nonlinearity is a very common phenomena. Many materials respond to external light signals in a nonlinear way with respect to light intensity. Here we propose one of the most commonly used optical nonlinearities – saturable absorption in our system. The saturable absorber nonlinear function can be modeled as $\sigma\tau_s I_0 = \frac{1}{2}\frac{\ln(T_m/T_0)}{1-T_m}$ [5], as shown in Fig. 1C.

With these three units, in principle, our ONNW architecture can do computation in a way that is mathematically equivalent to the traditional artificial neural networks.

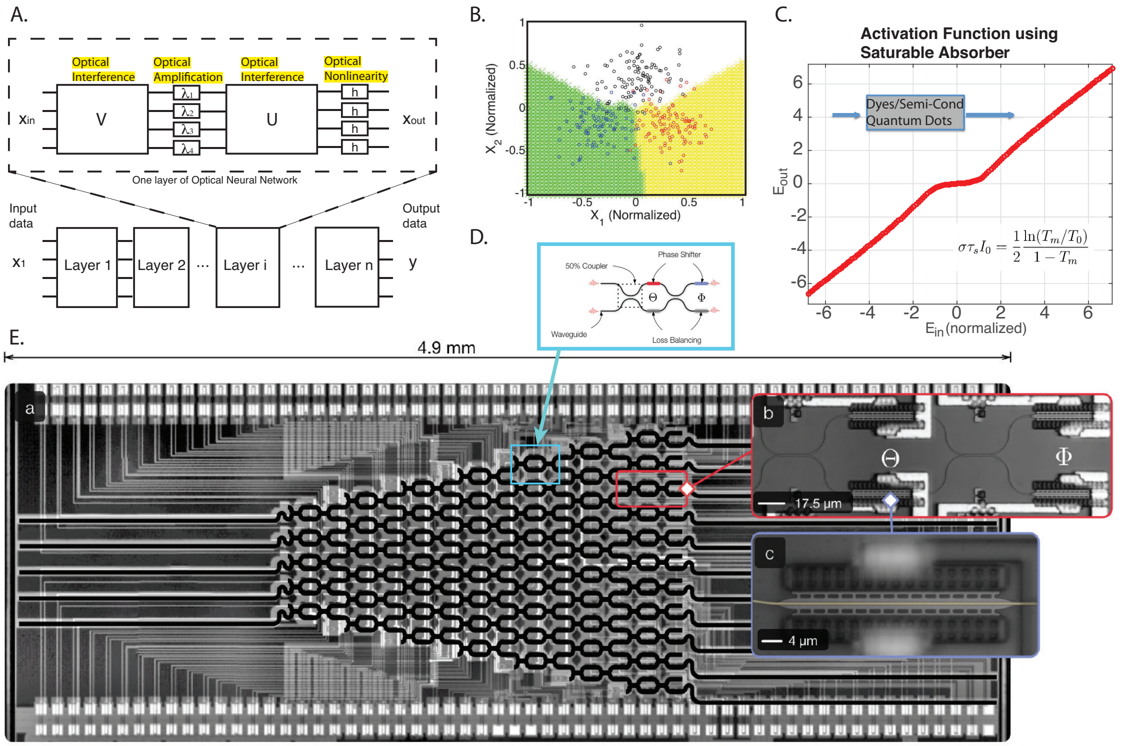

Fig. 1: (A) General Architecture of Optical Neural Network (ONNW) (B) Decision boundary for a simple 2 dimensional, 3 classes classification problem trained on our Optical Neural Network System. Three categories of data are labeled by different colors, areas are also labeled by different colors based on predictions (C) The optical response of a nonlinear optical system (saturable abosrption) that perform nonlinear calculation. Inset: Schematic illustration of the saturable absorption system. (D). Schematic illustration of a single Mach-Zehnder Interferometer. (E) Microscope image of an experimentally fabricated 5 × 5 unit on chip optical interference unit.

As a proof of concept, we apply a similar procedure to conventional ANN, to train an optical neural network structure (choosing saturable absorber as the activation function) with N input units, 1 hidden layer, and M output units. Batch forward propagation and backpropagation are used to optimize the parameters. Fig. 1B shows the classification result for the N=2, M=3 case (8 % error rate). We are also able to get less than 10% error rate for standard MNIST handwriting dataset (N=786, M=6), which is similar to the result from conventional neural net result.

Optical neural networks have the intrinsic advantage of speed. Once all parameter has been trained and implemented on the nanophotonic system, the forward propagation computing is done by light, which is only limited by the physical size and the pulse bandwidth (THz). So in principle it can be 3 orders of magnitude faster than electronic neural net.

References

1. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature **521**, 436–444 (2015).
2. J. Misra and I. Saha, “Artificial neural networks in hardware: A survey of two decades of progress,” Neurocomputing **74**, 239–255 (2010).
3. M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, “Experimental realization of any discrete unitary operator,” Phys. Rev. Lett. **73**, 58–61 (1994).
4. G. Golub and W. Kahan, “Calculating the singular values and pseudo-inverse of a matrix,” Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis **2**, 205–224 (1965).
5. A. Selden, “Pulse transmission through a saturable absorber,” British Journal of Applied Physics **18**, 743 (1967).