Scalability Analysis of the SOA-based All-optical Deep Neural Network

B. Shi[1,2], B. Pan[1], N. Calabretta[1], and R. Stabile[1,2]

[1]Institute for Photonic Integration, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands
[2]Eindhoven AI Systems Institute, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands

In this work we propose a noise model to investigate the scaling of the SOA-based all-optical deep neural networks regarding the number of WDM inputs and the cascading layers. The model is validated experimentally by emulating the OSNR evolution of the all-optical neuron. The results show that our all-optical neuron structure can be interconnected to establish a 16-input/neuron 16-neuron/layer 10-layer all-optical neural network with minor accuracy degradation for image classification.

Introduction

Neuromorphic photonics is emerging as an alternative to electronics in data processing, exploiting the intrinsic parallelism of light [1], especially for the artificial neural networks (ANN), which require heavy parallel computing to extract meaningful information from a massive volume of data. Recently, we have demonstrated a monolithically integrated all-optical neuron based on semiconductor optical amplifiers, which is operated in both the linear regime and nonlinear regime to implement the linear weighted addition and wavelength converter based nonlinear function, resulting in two-order faster speed respect with the GPU [2]. An accurate noise model of the neural network for analyzing the input channel numbers as well as the OSNR degradation along the propagation is pivotal to investigate the scalability of these kinds of networks. The noise model for in-line amplification transmission [3] has been used in the communication system for many years, however, no model is available for more complex neurons based on wavelength division multiplexed (WDM) channels and on optical amplifiers working at different regimes. In this paper, we use the proposed model of the noise evolution for WDM-input SOA-based all-optical neural networks (AONNs). With the noise figure estimation [4] and the small-signal method [5], we demonstrated and simulated the noise evolution of SOA-based AONN by tuning the optical signal to noise ratio (OSNR) at the neuron input for different numbers of input channels. We further employ the AONN, with error induction, as a fully connected network after two convolutional layers, to investigate the performance of the all-optical neuron based fully connected network solving the image classification problem.

Experiments

The envisioned convolutional neural network embedded with an all-optical fully connected neural network is illustrated in Fig. 1a, with circle nodes representing one all-optical neuron. This neural network structure is also used in the simulation with the evaluation of the noised fully-connected all-optical neural network. Fig. 1b illustrates the experimental setup for assessing the performance of the SOA-based all-optical neuron (AON) and emulated its scalability. As shown in the gray box (the AON), the multiple inputs of the all-optical neuron are weighted and then summated at the SOA-WC and

converted to a single wavelength. The signal degradation defines the maximum number of input wavelengths, therefore, the number of neurons that can be used simultaneously in one layer, for a certain error level induced. The AON is demonstrated with off-shelf components for this primary analysis, but it can be integrated on-chip [2]. We use 4 WDM inputs operated at 1560.61, 1558.98 1557.36, and 1555.75 nm, which are modulated with 10 Gbaud/s up to 9-bit/sample amplitude modulation, generated by the arbitrary waveform generator. Then the input is sent to the AON after decorrelation and multiplexing. In the AON, the input is de-multiplexed and weighted by the SOAs linear amplification, which is calibrated on the previously trained weights, and controlled by an FPGA. The weighted signal is then multiplexed and fed to the optical nonlinear function, a nonlinear (NL) SOA-WC, which converts the summation of the weighted inputs to another WDM channel at 1554.13 nm, exploiting the cross-gain-modulation. To emulate the evolution of the optical signal propagation from layer to layer, a noise source is coupled to the input signal to control the input noise floor. We assess the filtered converted signal with an APD, and the time trace is recorded on an digital phosphor oscilloscope (DPO). The performance of the neuron is evaluated by calculating the normalized root mean square error (NRMSE) between the measured output and expectation time trace calculated with inputs and weighting factors.
To model the noise evolution of the neuron connection layer by layer, we consider the multiple-input and single converted output neuron as shown in Fig1.b. The noise factors for the spontaneous emission of the linear unit and NL-SOA wavelength converter are:

$$F_{lin,sp} = 2\rho_{ASE-eff}/h\nu_o\bar{G} - 2\rho_{sse-eff}/h\nu_o\,, \qquad F_{soa-wc,sp} = 2\rho_{ASE}/h\nu\eta G' \qquad (1)$$

Where $\bar{G}$, $\rho_{ASE\text{-}eff}$, $\rho_{sse\text{-}eff}$ is the effective gain, amplified spontaneous emission (ASE) noise density and the source spontaneous emission (sse) noise density, which can be obtained with spectrum at linear unit output. And G' is the effective gain of the NL-SOA. The ASE density at the converted output is: $\rho_{ASE} = \sum\eta_i\rho_{sse,i} + \sum\eta_i\rho_{ASE,i}/G$, considering WDM input with i-th channel source spontaneous emission $\rho_{sse,i}$ and conversion efficiency η_i. Then the error can be defined with the ratio of noise and signal:

$$NRMSE \propto Noise/Signal \propto M\sqrt{\rho_{ASE}/(\eta G P_{cw})},$$

with M as the input channel number and $\eta=\sum\eta_i$, calculated with the NL-SOA property.

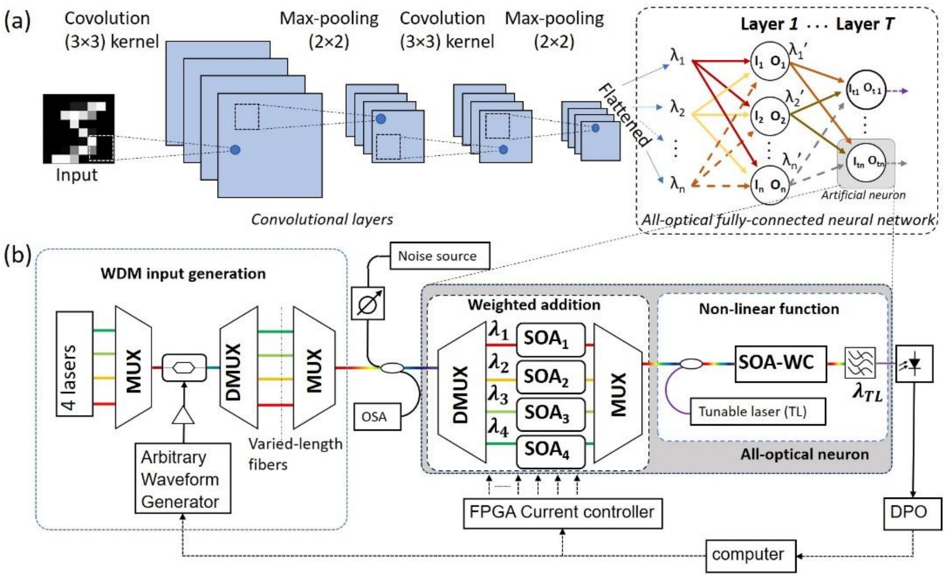

Fig. 1: (a) schematic of convolutional neural network with all-optical fully connected network. (b) experimental setup, with a gray box indicating the SOA-based all-optical neuron.

Results

To simulate the noise response of the all-optical neuron, we measure and define the parameter of the nonlinear SOA for the cross-gain modulation for modeling. In Fig. 2a, the squares and triangles depict the gain and amplified spontaneous emission (ASE) vs. total input power for NL-SOA (Kamelian, NL-SOA-L-C-FA). The SOA is set at 100 mA, for which the carrier lifetime is estimated as $\tau \approx 400$ ps. The simulation is carried out with unsaturated gain G_0 = 23 dB, saturated output power P_{sat} = 6 dB, normalized waveguide loss α' = 0.52, inversion parameter n_{sp} = 5.6, and b_{sp}=2.0, using methods in [4, 5].

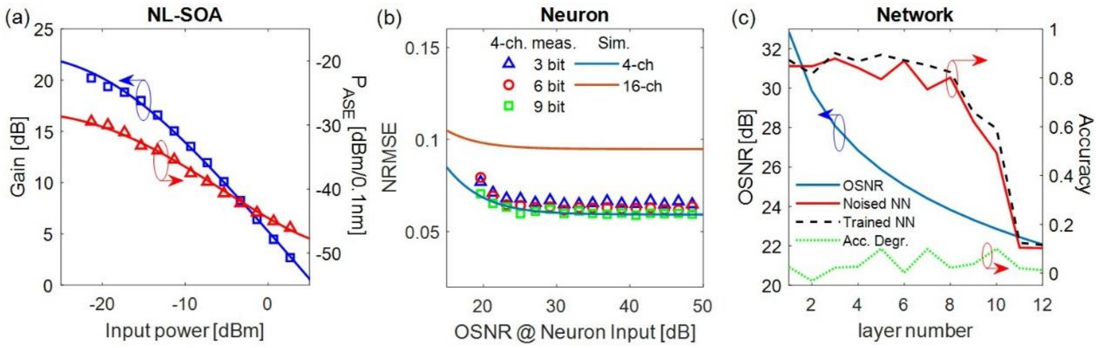

Fig. 2: (a) Gain (square) and ASE power (triangle) vs. input power of the SOA-WC, and the simulated results (lines). (b) Measured NRMSE for multi-level input neuron operation with tuning input OSNR, for 3, 6, and 9 bit/symbol per channel input and simulated neuron output error for 4- and 16-channel input. (c) OSNR vs. AONN layer number (blue) and image classification accuracy obtained by noised AONN evaluation (Red) and trained un-noised NN evaluation (black dashed line) and the accuracy degradation (green dashed line).

Fig. 2b plots the measured results for NRMSE vs. input OSNR for 3, 6, and 9 bit/symbol per channel 10 Gbaud/s input and 4-channel input for the neuron, which emulating 4×4 (M inputs × M neurons) neural network, with the signal OSNR changed related to the cascade of the layers. The experimental results fit the noise evolution model well, shown as the blue line in Fig. 2b. And the measurements show that by increasing the bit/symbol on multi-level modulation, the error slighly decreases. However, they all follow the same trend. We use the 9-bit/symbol at the following simulation since it enhances the throughput and maths to the noise model. We further use the model to analyze the performance of the SOA-based AONN by using a different input channel number. The red line in Fig. 2b plots the simulated NRMSE 16-ch neuron inputs, the error is greater than the 4-channel input case attributing to the decrease of the signal gain, but the error is still <0.1 [6]. This error evolution can be used to simulate a 16-input, 16-neuron neural network layer connecting to further layers, as present in Fig. 1a, as a fully-connected all-optical neural network after the convolutional layers. Based on the defined noise factors on the linear unit and the nonlinear unit, in euqation (1), and the cascade rule for the noise factor [3], we can estimate the output OSNR at the output of the all-optical neural network layers. The blue line in Fig. 2c shows the output OSNR vs. AONN layer numbers. And using error to OSNR relation obtained in Fig. 2b, we can evaluate the performance of the AONN by inducing the noise evolution in the neural network in solving the MNIST classification problem with TensorFlow. We use the size-reduced (28×28 pixels to 8×8 pixels) MNIST handwritten dataset to simulate the implementation of fully-connected AONN, with the convolutional neural network shown in Fig.1a, including two convolutional layers (4 (3×3) filters and followed by a (2×2) MaxPooling). We simulate the noised AONN with the error evolution with respect to the layer number and obtain the testing accuracy after 20 epochs of training. Fig. 2c red line presents the accuracy of the noised AONN when increasing the layer number. The black dashed line plots the accuracy of the trained neural network with the same structure of the AONN but realized on TensorFlow without error introduction. The descending of the curves is due to the limited training time of the network since a larger network requires more training epochs to reach the maximum accuracy, and the maximum accuracy can be further improved by choosing a better algorithm or better network structure. However, we could compare the noised AONN to the original trained neural network to show the degradation of the noised AONN implementation. By subtracting the red line from the dashed line, we obtained the green dashed line, which presents the accuracy degradation due to the noise. It shows that for layer numbers lower than 10, the accuracy is on average 4% lower than the neural network without noise. These results suggest that implementing a 16-channel 16-neuron 10-layer all-optical neural network is feasible with the proposed SOA-based all-optical neuron structure.

Conclusion

We emulated the scaling of the SOA-based all-optical deep neural networks in terms of the number of WDM inputs and the cascading layers with the aid of our noise model. We simulated the AONN with the noise evolution when increasing the layer number to solve an image classification problem. The result suggests the 16-input 16-neuron 10-layer all-optical neural network can be realized with the SOA-based all-optical neuron and with minor accuracy degradation.

Acknowledgment

This research work is financially supported by the Netherlands Organization of Scientific Research (NWO) under the Zwaartekracht programma, ‘Research Centre for Integrated Nanophotonics’.

References

[1] B. J. Shastri, et al., “Photonics for artificial intelligence and neuromorphic computing,” Nat. Photonics, vol. 15, no. 2, pp. 102–114, 2021.
[2] B. Shi, et al., “Lossless Monolithically Integrated Photonic InP Neuron for All-Optical Computation,” in OFC, 2020, p. W2A.12.
[3] M. J. Connelly, Semiconductor Optical Amplifiers. Boston: Kluwer Academic Publishers, 2002.
[4] A. Mecozzi, “Small-signal theory of wavelength converters based on cross-gain modulation in semiconductor optical amplifiers,” IEEE Photonics Technol. Lett., vol. 8, no. 11, pp. 1471–1473, Nov. 1996.
[5] K. Obermann et al., “Noise characteristics of semiconductor-optical amplifiers used for wavelength conversion via cross-gain and cross-phase modulation,” IEEE Photonics Technol. Lett., vol. 9, no. 3, pp. 312–314, 1997.
[6] B. Shi, et al., “Noise analysis of SOA-based All-optical Photonic Deep Neural Network with WDM input,” in Photonics in Switching and Computing, paper Tu3B.4.,2021.