Optics Letters

Optical ReLU using membrane lasers for an all-optical neural network

Naoki Takahashi,[1,*] Weicheng Fang,[1] Ruihao Xue,[1] Sho Okada,[1] Yoshitaka Ohiso,[1] Tomohiro Amemiya,[1,2] and Nobuhiko Nishiyama[1,2]

[1] Department of Electrical and Electronic Engineering, Tokyo Institute of Technology, Tokyo 152-8552, Japan
[2] Institute of Innovative Research (IIR), Tokyo Institute of Technology, Tokyo 152-8552, Japan
*Corresponding author: takahashi.n.av@m.titech.ac.jp

Received 27 July 2022; revised 5 October 2022; accepted 7 October 2022; posted 10 October 2022; published 28 October 2022

In this study, we propose low power consumption, programmable on-chip optical nonlinear units (ONUs) for all-optical neural networks (all-ONNs). The proposed units were constructed using a III-V semiconductor membrane laser, and the nonlinearity of the laser was used as the activation function of a rectified linear unit (ReLU). By measuring the relationship of the output power and input light, we succeeded in obtaining the response as an activation function of the ReLU with low power consumption. With its low-power operation and high compatibility with silicon photonics, we believe that this is a very promising device for realizing the ReLU function in optical circuits.

https://doi.org/10.1364/OL.471603

Deep learning is a machine learning method that has been introduced in various fields, including image recognition, language translation, and speech recognition [1–4]. Deep learning is characterized by multilayer neural networks, and the development of this technology has been supported by recent improvements in computer processing and the accumulation of available data. The amount of computation required for neural networks is increasing exponentially each year, and various hardware architectures using graphics processing units (GPUs), application-specific integrated circuits (ASICs), and field programmable gate arrays (FPGAs) are being developed to meet the rapidly growing computational needs [5–8].

However, a completely different method using "optical" technology is being investigated as an approach that does not rely on conventional electrical methods. Ultra-high-speed processing using photonic integrated circuits can significantly improve power consumption because it generates no energy loss caused by the wiring resistance and capacitance in electronic circuits. Furthermore, parallel processing with various types of multiplexing can increase computation time. Therefore, optical neural networks (ONNs) are expected to be the next-generation computation and processing technology for deep learning [9,10].

In recent years, photonic integrated circuits based on silicon photonics technology have been proposed by various research groups to realize an ONN [11,12]. These are very promising in terms of their small footprint, low cost, and power-efficient data processing using mature complementary metal-oxide-semiconductor (CMOS)-compatible processes [13,14]. A chip-integrated ONN consists of optical interference units (OIUs) and an optical nonlinear unit (ONU), as shown in Fig. 1(a). The OIU can be realized relatively easily using a Mach–Zehnder interferometer (MZI) or a micro-ring resonator, which are widely used in silicon photonics [11,12,15]. However, nonlinear processing is often performed electrically after optical-electrical (OE) conversion, not meeting the requirements of ONNs in which all-optical processing without any electrical signal is desired. Several ONUs that use nonlinear optical effects in semiconductor optical amplifiers have also been reported [16–18]; however, a very large input power is required to produce nonlinear effects. Therefore, the realization of a simple ONU that can operate with low power consumption remains challenging.

This study proposes a low-power, programmable, on-chip ONU. Specifically, we demonstrated that the rectified linear unit (ReLU) activation function can be obtained using the nonlinearity of a membrane laser directly bonded on a silicon substrate [19–21]. Because this device can be driven with low power derived from the characteristics of membrane lasers, it is considered a promising activation function unit in an ONN consisting of a large parallel computing unit.

Figure 1(b) shows the schematic of the proposed optical ReLU. In the proposed structure, III-V semiconductor membrane components, such as lasers and waveguides, are integrated on a silicon on insulator (SOI) substrate. First, a certain level of current is injected into the laser to pump it to the vicinity of the threshold, and then the optical signal (excitation light) is fed to the laser through the waveguide. If the input optical intensity is sufficiently small, the laser cannot transition to the lasing state, and the output intensity will be approximately zero [Fig. 1(c)(i)]. As the input optical intensity is increased and the gain attributed to optical excitation exceeds the threshold value, the laser transitions to the lasing state, and the optical output responds linearly to the input light [Fig. 1(c)(ii)]. Because these output characteristics are very similar to the response of the ReLU activation function, they can be used as an element to obtain the nonlinearity required for the ONN.

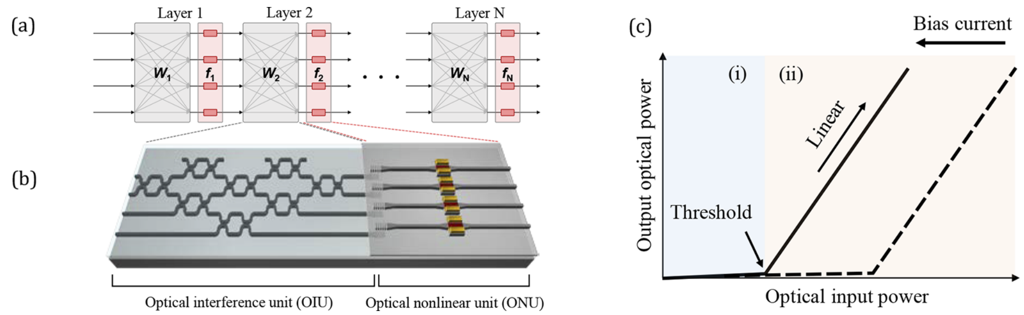

Fig. 1. (a) Block diagram of a feedforward neural network. (b) Schematic of the proposed all-ONN with membrane ONUs. (c) Nonlinear operation of the membrane laser depending on bias current: (i) below the threshold, and (ii) linear output above the threshold.

To achieve this nonlinearity, it is impractical to use a general semiconductor laser as the ONU for the following two reasons. First, the relatively long cavity length of the laser makes it difficult to excite the inside of the cavity uniformly with input light from one side, resulting in low sensitivity and weak nonlinearity of the input light. Second, when the OIU is composed of a silicon photonics-based MZI or micro-ring resonator, the ONU must also be silicon-based in terms of integration, whereas general lasers are InP-based.

Consequently, a membrane laser with a short cavity and ultra-low threshold characteristics was introduced as the ONU in this study. A schematic and scanning electron microscope (SEM) image of the membrane laser used in this study are shown in Figs. 2(a) and 2(b), respectively. The active layer consists of 1.55-µm band GaInAsP multiple quantum wells sandwiched by upper and lower InP cladding of approximately 100 nm, with a total thickness of 270 nm.

In this structure, the top and bottom of the III-V membrane are sandwiched between low refractive index materials (i.e., air and SiO_2), which results in a high refractive index difference in the vertical direction and strong optical confinement to the gain region. This allows lasing at a very low threshold current (sub-milliampere) with a short cavity (tens of micrometers). For current injection, a lateral injection structure was adopted, in which both sides of the active layer were buried with *n*-doped and *p*-doped InP through organometallic vapor-phase epitaxy (OMVPE) to form a p-i-n junction. The III-V membrane structure fabricated on the InP substrate was finally bonded to the silicon substrate using a direct bonding technique [22]. Furthermore, by introducing a distributed reflector (DR) structure, which consists of a surface grating to create distributed feedback (DFB) in the active section and a distributed Bragg reflector (DBR) in the back section, a lower threshold and high efficiency can be achieved.

Figure 2(c) shows the output characteristics of the membrane laser driven by the current. The fabricated device had a DR structure consisting of a 293-nm pitch DFB grating and 293-nm pitch DBR grating. The threshold current was 0.5 mA for a device with a DFB region length of 90 µm and a stripe width of 0.8 µm. By comparing the fiber-coupled output power with the

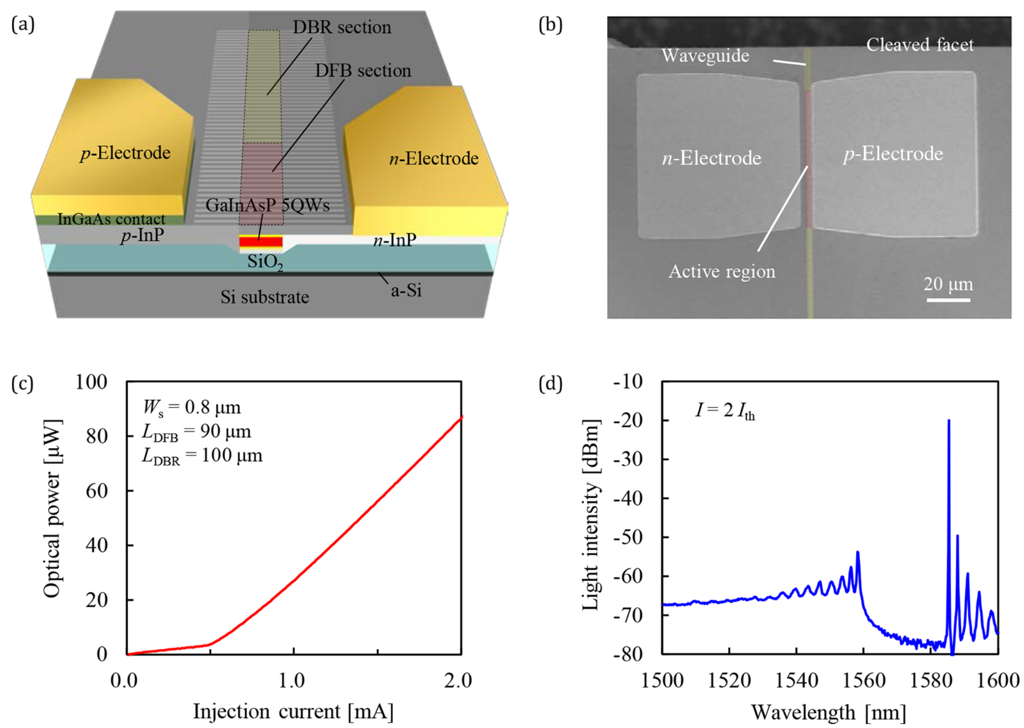

Fig. 2. (a) Schematic of the membrane DR laser. (b) SEM image of the fabricated membrane DR laser. (c) Output power versus injection current of the device. (d) Spectrum with an injection current of 1.0 mA ($2I_{th}$).

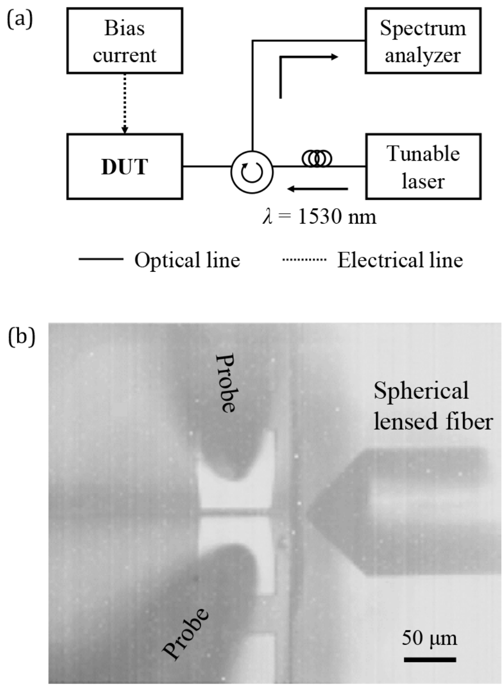

Fig. 3. (a) Schematic of the experimental setup. (b) Microscope image of the device under test (DUT).

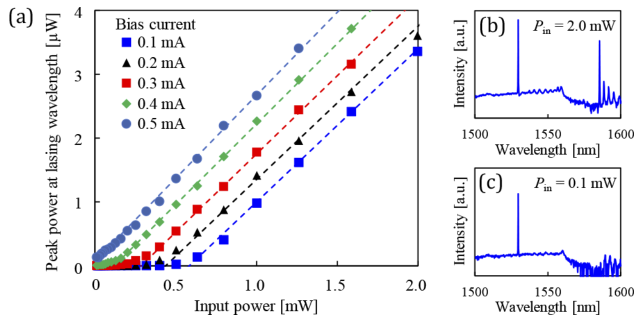

Fig. 4. (a) Response of the membrane nonlinear activation unit with various bias currents. Spectrum (b) before, and (c) after the threshold input power at a bias current of 0.2 mA.

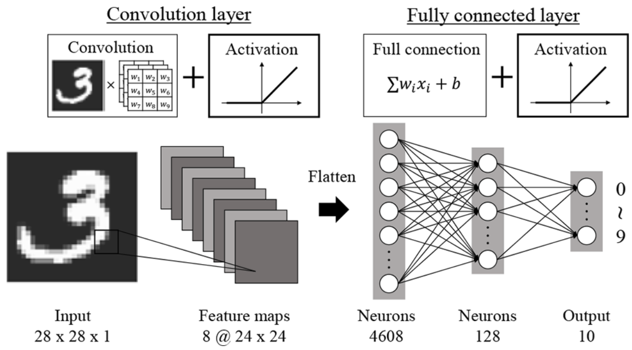

Fig. 5. Architecture of convolution neural network using proposed optical ReLU.

photodetector output, the coupling loss between the membrane laser and fiber was approximately 10 dB. Figure 2(d) shows the spectrum at a bias current of $I_b = 2I_{th}$ (1.0 mA), and a lasing wavelength of 1585 nm was confirmed.

Using the aforementioned device, we demonstrated that the membrane laser works as an optical ReLU [Fig. 1(b)]. Figure 3(a) shows the configuration of the measurement system. The excitation light from the tunable laser (i.e., the signal light from the OIU) was coupled to the cleaved facet of the waveguide by a lensed fiber, and the output from the membrane laser was observed using an analyzer through a circulator. The cavity length of the membrane laser used in this study was very small (90 μm), as shown in Fig. 3(b), which enabled uniform pumping inside the cavity by the injected light (signal light). To increase the pumping efficiency, the wavelength of the pumping light was set to be shorter (1530 nm) than the lasing wavelength of the membrane laser (1585 nm).

Figure 4(a) shows the output lasing light peak intensity as a function of input light. In this experiment, a bias current below the threshold was applied to the membrane laser to control the gain in the vicinity of the lasing condition. Here, when the bias current reaches the threshold current (0.5 mA), the carrier density inside the cavity satisfies the oscillation condition; hence, all the carriers excited by the 1530-nm input light contribute to stimulated emission, resulting in a linear optical output to the input light. However, when the bias current was below the threshold current (< 0.5 mA), the output was nonlinear in response to the input light. Until the carrier density inside the cavity reached the threshold value, the optical output was approximately zero, because the total loss exceeded the gain [Fig. 4(b)]. After the excited carrier density reached the threshold value, a linear optical output at the wavelength of 1585 nm was obtained from the stimulated emission of the laser [Fig. 4(c)]. For example, when a bias current of 0.1 mA (the bias voltage is 0.9 V at this current) is applied, a nonlinear characteristic of the ReLU with a threshold input optical power of 0.6 mW was successfully obtained. Note that the power inside the chip should be approximately 0.06 mW according to our previous experiment of the fiber coupling efficiency (10%). Output power is also ten times higher than the values of Fig. 4(a). This low coupling efficiency will not be a problem after integration of OIUs in the final form. The ReLU threshold point can be controlled by changing the bias current applied to the laser in advance. This implies that the bias term of the OIU can be addressed by the activation function side of the ReLU. In ONN circuits, OIUs consisting of MZIs or micro-ring cavities tend to be complex in general. Thus, it is advantageous to reduce the number of optical paths for the bias terms in OIUs.

The measured optical ReLU was applied to the convolutional neural network (CNN) for the Modified National Institute of Standards and Technology (MNIST) handwritten digit classification. After formulating the bias-dependent optical ReLU from the results of Fig. 4(a), we constructed the CNN shown in Fig. 5. The MNIST database contains 60,000 training images and 10,000 testing images, with small square 28×28 pixel grayscale images of handwritten single digits between 0 and 9. For this classification task, the batch size was set to 100 and the learning rate to 0.001, and convolution was performed with a kernel size of 5×5 and stride of 1. Table 1 shows the test accuracy of the network after training with various nonlinear activation functions. When the optical ReLU was applied to the neural network, the accuracy was 98.22%, which is comparable to the accuracy when the ideal ReLU was used for the same neural network. These results indicate that the optical ReLU obtained in this study is valid as a nonlinear activation function.

Table 1. Test Accuracy of CNN with Nonlinear Active Functions

Type	Activation Function	Epoch	Test Accuracy
1	Optical ReLU	10	98.22%
2	ReLU	10	98.32%
3	ELU	10	98.05%
4	Sigmoid	10	97.71%
5	Softplus	10	97.75%

In conclusion, this study proposes a method for realizing an optical ReLU using a III-V semiconductor membrane laser. We have demonstrated that a programmable ReLU function can be realized with low power consumption by using a membrane laser on silicon, which has a short cavity and ultra-low threshold. Consequently, we believe that the membrane laser is a promising device for realizing the ReLU function in optical circuits.

Funding. Japan Society for the Promotion of Science (19H02193, 20H02200, 21J14548); Core Research for Evolutional Science and Technology (JPMJCR15N6).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
2. D. Silver, A. Huang, and C. J. Maddison, *et al.*, Nature **529**, 484 (2016).
3. S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2017), pp. 131–135.
4. A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, Nature **542**, 115 (2017).
5. P. A. Merolla, J. V. Arthur, and R. A. Icaza, *et al.*, Science **345**, 668 (2014).
6. S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, Proc. IEEE **102**, 652 (2014).
7. C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, and J. Cong, in *Proceedings of the 2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays* (2015), pp. 161–170.
8. A. Graves, G. Wayne, and M. Reynolds, *et al.*, Nature **538**, 471 (2016).
9. Y. S. Abu-Mostafa and D. Psaltis, Sci. Am. **256**, 88 (1987).
10. D. Psaltis, D. Brady, X. -G. Gu, and S. Lin, Nature **343**, 325 (1990).
11. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, Nat. Photonics **11**, 441 (2017).
12. M. M. Pour Fard, I. A. D. Williamson, M. Edwards, K. Liu, S. Pai, B. Bartlett, M. Minkov, T. W. Hughes, S. Fan, and T. A. Nguyen, Opt. Express **28**, 12138 (2020).
13. M. A. Nahmias, B. J. Shastri, A. N. Tait, T. Ferreira de Lima, and P. R. Prucnal, Opt. Photonics News **29**, 34 (2018).
14. J. George, R. Amin, A. Mehrabian, J. Khurgin, T. El-Ghazawi, P. R. Prucnal, and V. J. Sorger, in *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF)*, OSA Technical Digest (online) (Optica Publishing Group, 2018), paper SpW4G.3.
15. A. N. Tait, T. Ferreira de Lima, M. A. Nahmias, H. B. Miller, H.-T. Peng, B. J. Shastri, and P. R. Prucnal, Phys. Rev. Appl. **11**, 064043 (2019).
16. G. Mourgias-Alexandris, A. Tsakyridis, N. Passalis, A. Tefas, K. Vyrsokinos, and N. Pleros, Opt. Express **27**, 9620 (2019).
17. J. Crnjanski, M. Krstić, A. Totović, N. Pleros, and D. Gvozdić, Opt. Lett. **46**, 2003 (2021).
18. Y. Huang, W. Wang, L. Qiao, X. Hu, and T. Chu, Opt. Lett. **47**, 1810 (2022).
19. T. Hiratani, D. Inoue, T. Tomiyasu, K. Fukuda, N. Nakamura, T. Amemiya, N. Nishiyama, and S. Arai, IEEE Photon. Technol. Lett. **29**, 1832 (2017).
20. T. Tomiyasu, D. Inoue, T. Hiratani, K. Fukuda, N. Nakamura, T. Uryu, T. Amemiya, N. Nishiyama, and S. Arai, Appl. Phys. Express **11**, 012704 (2018).
21. N. Takahashi, W. Fang, Y. Ohiso, T. Amemiya, and N. Nishiyama, J. Opt. Soc. Am. B **38**, 3340 (2021).
22. W. Fang, N. Takahashi, Y. Ohiso, T. Amemiya, and N. Nishiyama, IEEE J. Quantum Electron. **58**, 2000208 (2022).