



Programmable low-power consumption all-optical nonlinear activation functions using a micro-ring resonator with phase-change materials

ZILING FU,¹ ZHI WANG,^{1,*} PETER BIENSTMAN,² RUI JIANG,¹ 
JIAN WANG,¹ AND CHONGQING WU¹

¹*Institute of Optical Information, Key Laboratory of Luminescence and Optical Information, Ministry of Education, School of Physical Science and Engineering, Beijing Jiaotong University, Beijing 100044, China*

²*Photonic Research Group, Ghent University – IMEC, Ghent 9052, Belgium*

*zhiwang@bjtu.edu.cn

Abstract: A programmable hardware implementation of all-optical nonlinear activation functions for different scenarios and applications in all-optical neural networks is essential. We demonstrate a programmable, low-loss all-optical activation function device based on a silicon micro-ring resonator loaded with phase change materials. Four different nonlinear activation functions of Relu, ELU, Softplus and radial basis functions are implemented for incident signal light of the same wavelength. The maximum power consumption required to switch between the four different nonlinear activation functions in calculation is only 1.748 nJ. The simulation of classification of hand-written digit images also shows that they can perform well as alternative nonlinear activation functions. The device we design can serve as nonlinear units in photonic neural networks, while its nonlinear transfer function can be flexibly programmed to optimize the performance of different neuromorphic tasks.

© 2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

With the rapid development of information technology such as big data, cloud computing, smart terminals, and global data traffic is growing geometrically. Therefore, in the era of artificial intelligence(AI), traditional architecture computing systems face serious challenges in terms of energy efficiency and volume, which are limited by Moore's law [1,2]. Therefore, AI technologies represented by neural networks are rapidly developing toward achieving high speed and low power consumption [3].

Photonics has the inherent characteristics of high speed and massive parallelism with low energy consumption, thus, it has attracted attention as a promising candidate [4]. Indeed, optical neuromorphic computing has been experimentally validated in photonic integrated circuits [5–10] and free-space optics [11–14].

Optical neurons are one of the key technologies in optical neuromorphic computing. The nonlinear activation function(NLAF), one of the perceptrons [15] of the optical neuron, is crucial to the training and decision mapping processes of the network. Compared to those in electrical neurons, all-optical NLAFs are not yet mature [16]. Optical devices can have a superlarge bandwidth and low power consumption. Therefore, photonics provides advantages in connectivity and matrix multiplication over electronics. A multitude of photonic devices exhibit nonlinear transfer functions that resemble neuron-like or gate-like transfer functions; however, a non-linear response alone is not sufficient for a photonic device to act as a neuron. Photonic neurons must be capable of reacting to multiple optical inputs (fan-in), applying a nonlinearity and producing an optical output suitable to drive other like photonic neurons (cascadability). Optical devices face fundamental challenges in satisfying these requirements in particular [3].

In recent years, all-optical nonlinear activation functions based on interference between the dipoles of the plasmon oscillation in the metal nanoparticles and the exciton transition in the Quantum dot [17], photonic crystal Fano lasers [18], the traditional non-volatile of phase change materials (PCMs) [8] and the volatile switching of PCMs excited with a free space femtosecond laser pulse [19] have been proposed. These are all ultra-fast, compact, on-chip solutions for neuromorphic photonic computing.

Usually, for different AI applications, the activation function needs to be selected according to the specific task [16]. In addition, the proper activation function affects the overall average test accuracy [20]. Experimental results show that radial basis functions (RBF) in support vector machines [21], ReLU in deep learning networks for a 50-hour English Broadcast News task [22], ELU in different vision datasets [23], and Softplus in deep learning networks for phone recognition tasks [24], significantly outperform some other activation functions. Therefore, it is essential to achieve programmability of the all-optical NLAFs.

At present, tuning the bias pulse energy injected in the Periodically poled thin-film lithium niobate (PPLN) nanophotonic waveguide [25] can be used to implement common used variants of the Relu function. The cavity paired with the tuning biases on the interferometers [26] and varying the wavelength of light input to the racetrack resonator with a span of Ge/Si hybrid waveguide [27] provide programmability among different kinds of activation functions. However, the programmable all-optical nonlinear activation function has yet to be optimized in terms of power consumption.

In this paper, we propose a programmable, low-loss all-optical activation function device based on a silicon micro-ring resonator loaded with PCMs. The NLAf relies on the nonlinear properties of the silicon micro-ring resonator, which are due to thermal and free-carrier-related nonlinearities. Programmability is achieved by loading the $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (GST) PCM on the micro-ring resonator in four different intermediate states (refractive indexes) between the crystalline and amorphous states. Four different NLAFs of the Relu, ELU, Softplus and RBF are implemented for incident signal pulses at the same wavelength. The non-volatility of GST is used to maintain the four nonlinear activation functions without any extra power consumption. The maximum power consumption required to switch between the four different NLAFs is only 1.748 nJ. Finally, we simulate benchmark machine learning tasks using our all-optical NLAFs with accuracy higher than 94.8% in the task of classification of the MNIST handwritten digital image dataset, benchmark MNIST handwritten digit classification task, which demonstrates the prospect of our scheme for future applications in all-optical neural networks.

2. Principle and design

2.1. Coupled mode theory for GST-loaded silicon micro-ring

We design an add-drop micro-ring resonator with a radius of $10\ \mu\text{m}$, and a $0.5\ \mu\text{m}$ long GST film is loaded as shown in Fig. 1. NLAFs are proposed for the TE mode signal light. When the signal light is injected into the micro-ring, the two-photon absorption (TPA) effect in the micro-ring generates free carriers, which will cause free carrier absorption (FCA) and free carrier dispersion (FCD). In addition, the TPA and FCA effects induce a thermo-optical (TO) effect. The FCD effect causes the resonant wavelength to blue-shift within the time scale of a few nanoseconds, whereas the time scale of the red-shift of the resonance wavelength caused by the thermo-optic effect is in the tens of nanoseconds [28]. Although the micro-ring resonator exhibits self-pulsation with both FCD and TO effects, it cannot be used as an optical NLAf device in our design when both effects are non-stationary.

We modeled the optical propagation in the micro-ring resonator using a nonlinear coupled-mode theory approach based on [29–32], with the inclusion of contributions from GST and an additional straight waveguide coupled to the micro-ring resonator, i.e. GST changes the refractive index of the Si/GST hybrid waveguide and introduces extra loss, and the straight waveguide

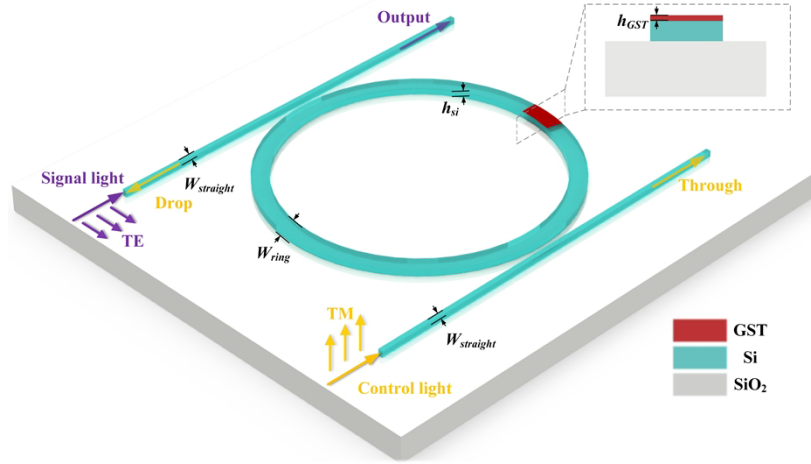


Fig. 1. Perspective view of the GST-loaded add-drop micro-ring resonator. The simulation design parameters are: $W_{\text{straight}} = 350$ nm, $W_{\text{ring}} = 590$ nm, gap = 630 nm, $h_{\text{si}} = 220$ nm, and $h_{\text{GST}} = 10$ nm.

induces extra coupling loss in our updated model. The coupled ordinary differential equations are expressed in Eq. (1), where u is the temporal evolution of the intracavity field, N is the free-carrier density, and ΔT is the temperature change in the GST-loaded micro-ring resonator.

$$\begin{aligned} \frac{\partial u}{\partial t} &= \left[i\delta\omega_{nl_hy} + i(\omega_{r_hy} - \omega) - \frac{\gamma_{\text{loss_hy}}}{2} \right] u + \kappa S_{in}, \\ \frac{\partial N}{\partial t} &= -\frac{N}{\tau_{fc}} + \frac{\beta_{\text{Si}}c^2}{2\hbar\omega V_{\text{FCA}}^2 n_{g_hy}^2} |u|^4, \\ \frac{\partial \Delta T}{\partial t} &= -\frac{\Delta T}{\tau_{th}} + \frac{\gamma_{\text{abs_hy}} |u|^2}{\rho_{\text{Si}} c_p \text{Si} V_m}, \end{aligned} \quad (1)$$

where S_{in} is the amplitude of the signal light (signal light power $P_{in} = |S_{in}|^2$), ω is the frequency of the signal light, and ω_{r_hy} is the resonance frequency of the hybrid GST/Si micro-ring cavity. The remaining parameters are listed in Table 1. $\gamma_{\text{loss_hy}}$ and $\gamma_{\text{abs_hy}}$ represent the total and absorption losses in the hybrid GST/Si micro-ring cavity, respectively.

$$\gamma_{\text{loss_hy}} = 2\gamma_{\text{coup}} + \gamma_{\text{rad}} + \gamma_{\text{abs_hy}} \quad (2)$$

where we have introduced the coupling loss into the waveguide γ_{coup} (with $\kappa = i\sqrt{\gamma_{\text{coup}}}$) and the radiation loss γ_{rad} . In the hybrid GST on silicon micro-ring we have absorption by linear surface absorption, TPA and FCA:

$$\gamma_{\text{abs_hy}} = \gamma_{\text{abs,lin}} + \frac{\beta_{\text{Si}}c^2}{n_{g_hy}^2 V_{\text{TPA}}} |u|^2 + \frac{\sigma_{\text{Si}}c}{n_{g_hy}} N \quad (3)$$

In the Silicon On Insulator (SOI) $\eta_{\text{lin}} = \gamma_{\text{abs,lin}} / (\gamma_{\text{rad}} + \gamma_{\text{abs,lin}}) \approx 0.4$ [33,34], $\gamma_{\text{rad}} + \gamma_{\text{abs,lin}} = c\alpha_{\text{ring_hy}}/n_{g_hy}$, $\alpha_{\text{ring_hy}}$, and n_{g_hy} are the total loss and group index of the hybrid GST on silicon micro-ring cavity.

$$\begin{aligned} \alpha_{\text{ring_hy}} &= (1 - \zeta)\alpha_{\text{Si_ring}} + \zeta\alpha_{\text{Si_GST}} \\ n_{g_hy} &= (1 - \zeta)n_{g_si} + \zeta n_{g_PCM} \end{aligned} \quad (4)$$

where $\zeta = L_{\text{PCM}}/2\pi R$, $\alpha_{\text{Si_GST}}$ and n_{g_PCM} depend on the state of the GST film.

Table 1. Parameters describing the properties of the micro-ring resonator and their values used in the simulations

Parameter	Description	Value
β_{Si}	TPA coefficient	$8.4 \times 10^{-12} \text{m} \cdot \text{W}^{-1}$
dn_{Si}/dT	thermal coefficient	$1.86 \times 10^{-4} \text{K}^{-1}$
dn_{Si}/dN	FCD coefficient	$-1.73 \times 10^{-27} \text{m}^3$
σ_{Si}	FCA absorption cross section	10^{-21}m^2
ρ_{Si}	density of silicon	$2.33 \times 10^3 \text{kg} \cdot \text{m}^{-3}$
$c_{p,Si}$	thermal capacity	$700 \text{J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$
γ_{coup}	coupling loss	$2.52 \times 10^9 \text{s}^{-1}$
τ_{th}	thermal relaxation time	65ns
τ_{jc}	carrier relaxation time	5.3ns
V_{th}	thermal effective volume	$3.19 \times 10^{-18} \text{m}^3$
V_{TPA}	TPA effective volume	$2.59 \times 10^{-18} \text{m}^3$
V_{FCA}	FCA effective volume	$2.36 \times 10^{-18} \text{m}^3$
$n_{g,Si}$	group index of the silicon waveguide	4.2
α_{Si_ring}	the total micro-ring loss	0.16cm^{-1}

Both TO and FCD effects cause a significant shift in the resonance frequency $\delta\omega_{nl_hy}$, whereas the shift caused by the Kerr-effect is negligible. Using first-order perturbation theory, this gives:

$$\delta\omega_{nl_hy} = \frac{\omega_{r_hy}}{n_{g_hy}} \left(\frac{dn_{si}}{dN} N + \frac{dn_{si}}{dT} \Delta T \right) \quad (5)$$

Equation (1) has steady-state solutions when $\partial u/\partial t = 0$, $\partial N/\partial t = 0$, $\partial \Delta T/\partial t = 0$. The corresponding linear matrix M is obtained by adding small perturbations to the stable results and substituting the updated parameters into the normalized differential equations with omitting higher-order terms. Then, a 4×4 eigenmatrix M is obtained by normalization according to the method described in [26]. Thus, we can find a stable fixed point (i.e., after a small perturbation, the system relaxes back to the same point) that is suitable for NLAFs if the real parts of all four eigenvalues are negative.

2.2. Coupled mode theory for GST-loaded silicon micro-ring

We aim to ensure that the micro-ring is at a stable fixed point when GST is in different states. We analyzed the corresponding relation between the signal light power and intracavity energy in Fig. 2(a) as well as the real and the imaginary parts of the four eigenvalues of the M matrix in Fig. 2(b)–(e) when the crystallization fraction of GST is 50–80%. Figure 2(b)–(e) shows that the real parts of all four eigenvalues are negative and that the micro-ring is at a stable fixed point when the GST is in one of these four states. Only two solid and dashed lines are shown in Fig. 2(b)–(e) because two of the four eigenvalues are conjugate to each other.

Moreover, we also analyzed the relationship between signal light wavelength detuning and stability, as shown in Fig. 2(f). When the crystallization fraction of the GST loaded on the micro-ring is 50%, the relationship between the output and input powers of the micro-ring is determined with the signal light wavelength detuning of 100–250 pm, separately. The red dots correspond to the unstable fixed point, which appears only when the signal light wavelength is far from the resonant wavelength. Therefore, there is no effect on our optical NLAFs subsequently.

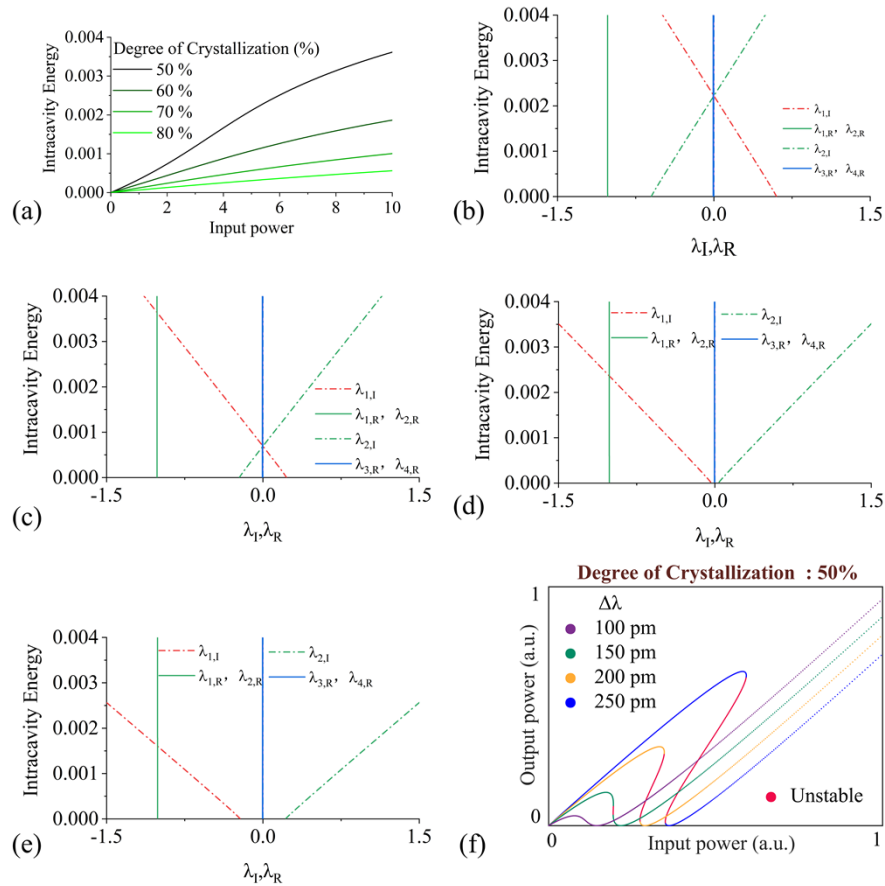


Fig. 2. (a) Steady-state response of the intracavity energy versus the signal light power when the crystallization fraction of the GST is 50, 60, 70 and 80%. (b)(c)(d)(e) Corresponding real (λ_R , solid lines) and imaginary (λ_I , dot-dashed lines) parts of the four eigenvalues of the M matrix, relatively to the abscissa for each corresponding intracavity energy. (f) Result of the output and input powers when the incident light wavelength detuning from the resonance of the micro-ring is 100, 150, 200 and 250 pm at a 50% crystallization fraction of the loaded-GST, respectively.

2.3. Implementation of programmability

While the nonlinearity results primarily from the power-dependent nonlinear phase change due to the free-carrier and TO effects, the change in the state of the loaded GST is equally important to achieve the programmability of the NLA. From Fig. 3(a)–(d), when the wavelength of the incident signal light is 1549.38 nm, four different optical NLAFs, RBF, Relu, Softplus, and ELU, can be generated between the output power and the input power as the crystallization fraction of the loaded GST on the micro-ring increases from 50 to 80%. There is good agreement between the ideal activation function (dotted red line) and the device response (solid blue line). The switching between the different nonlinear activation functions is determined by the state of the loaded-GST. The initial state of the loaded-GST is crystalline, which is modulated to crystalline fraction of 80% by optical pulses in TM mode.

The increase in the loaded-GST crystallization fraction implies an increase in the refractive index and loss, and the resonant wavelength is red-shifted.

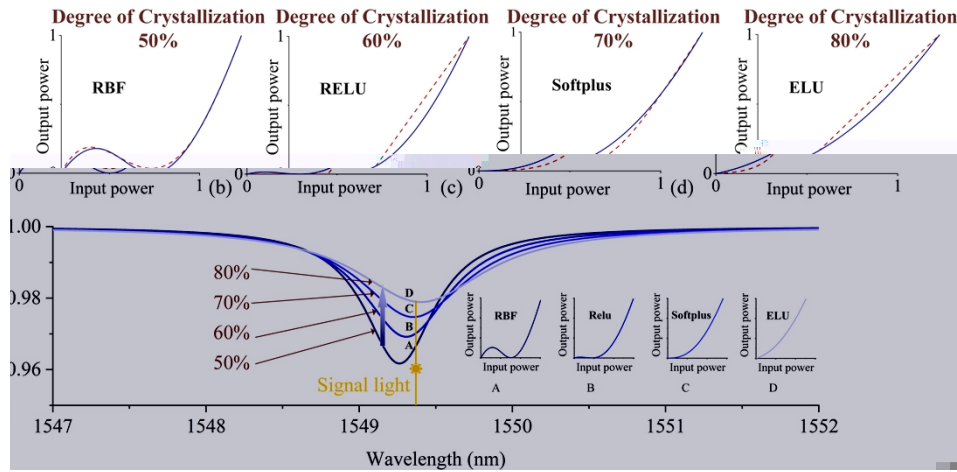


Fig. 3. (a)-(d) NLAFs at different states of the loaded GST. (dotted red: the ideal activation function, solid blue: the device response). (e) Sketch of the transmission evolution with incremental crystallization fraction of the loaded-GST for different NLAFs.

The state of the loaded-GST is determined by the control light with a wavelength of 1546.9 nm in the TM mode. Switching among different crystallization fractions of the GST can be realized by changing the power and duration of the injected optical pulse [35]. Thus, it is possible to achieve reversible switching among different activation functions.

Figure 3(e) shows the evolution of the transmission with respect to the input light power under different states of the loaded GST. It can be observed that the resonant wavelength of the micro-ring is red-shifted with an increase in the loaded-GST crystallization fraction, while the wavelength detuning is at a maximum of 50 pm. This further proves that our NLAFs work at a stable fixed point.

When the crystallization fraction of the loaded-GST is 50%, the incident signal light wavelength is longer than and relatively far from the resonant wavelength, with a detuning of -50 pm. Thus, as the signal light power increases, the output power undergoes a drop followed by a linear increase, resulting in an RBF, as shown in inset A of Fig. 3(e).

As the GST crystallization fraction increases to 60%, the red shift of the resonant wavelength causes the signal light wavelength to be closer to the resonant wavelength, even though it is still longer than the resonant wavelength, with a detuning of -23.4 pm. Thus, there is a mechanism whereby the output power remains almost zero as the input optical power increases, and then begins to rise as the input power continues to increase. This makes the Relu function, shown in inset B of Fig. 3(e), feasible.

After the loaded GST is further crystallized up to 70%, the resonant wavelength is red-shifted to a point shorter than the signal light wavelength. At this point, the resonant wavelength is very close to the signal light wavelength, and the detuning is 4.3 pm. As the input optical power increases, the output power increases very slowly resulting in an almost-zero initial output optical power, and then it increases linearly with the input power; thus, the Softplus function is generated, as shown in inset C of Fig. 3(e).

Finally, when the GST crystallization fraction is 80%, the resonant wavelength is further red-shifted towards a wavelength shorter than that of the signal light, and the detuning is -33.3 pm. The output power always increases with the input power; therefore, the ELU function can be obtained, as shown in inset D of Fig. 3(e).

In addition, to allow the NLAf to work stably for a long time, we need to ensure that the signal light does not change the state of the loaded GST. Referring to [36], the detected optical power range in the experiment is generally between -2 and -6 dBm; consequently, based on the modulation depth in Fig. 3(e) we conclude that the signal light power coupled into the Si/GST hybrid waveguide is less than -3 dBm. Thus, the power of the signal light does not change the loaded-GST state or stability of the optical NLAfs.

2.4. Simulated in benchmark MNIST handwritten digit classification

To validate the applicability of our NLAfs, we performed classification simulations on the MNIST dataset using (a), (b), (c) and (d) in Fig. 3, which are abbreviated as ONAF- RBF, ONAF-Relu, ONAF-Softplus and ONAF- ELU. In particular, we use a rational function to fit our discrete data in Fig. 3 and ensure that the fitting curve passes through the original points; only the positive part of the function is considered.

We simulated a three-layer fully connected neural network and studied its accuracy in a benchmark MNIST handwritten digit classification task, as illustrated in Fig. 4(a). Each input image in the MNIST dataset is of 28×28 pixels. To implement the classification of the MNIST dataset, raw images with a size of 28×28 pixels were first flattened into one-dimensional arrays. Then, 784 input pixels are fed into the three-layer network, and the output elements are normalized to represent probabilities from digit 0 to 9.

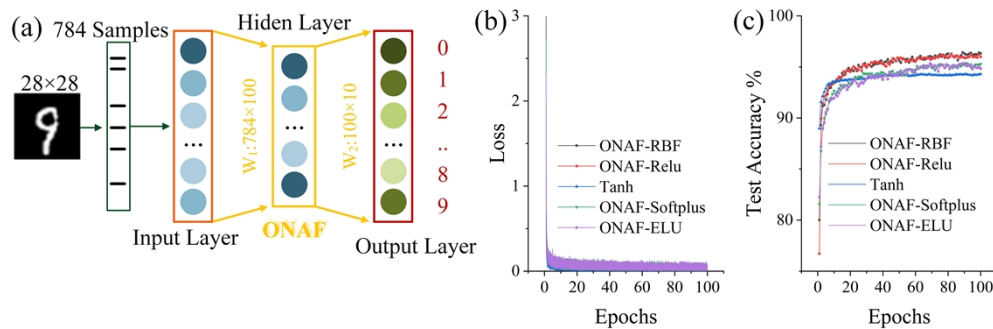


Fig. 4. (a) Schematic illustration of handwritten digit image recognition using neural network. (b) Cross-entropy loss of training dataset. (c) Accuracy of test dataset.

Networks with ONAF- RBF, ONAF- Relu, ONAF- Softplus and ONAF- ELU show a good performance in benchmark MNIST handwritten digit classification task, with accuracy of 96%, 96.4%, 95.3% and 94.8%, respectively. The cross-entropy loss of the training dataset during the training process is shown in Fig. 4(b), and the test accuracy in Fig. 4(c). Meanwhile, we adopt Tanh activation function to verify the programmability and the efficiency. Besides this classification example, there are other applications where such activation functions are routinely used for artificial neural network tasks.

3. Discussion

3.1. Control light

GST has a high refractive index contrast between its amorphous and crystalline states. We can induce a slight change in the resonant wavelength of the hybrid GST on silicon micro-ring using the intermediate crystallographic states of the GST, that is, states with a mixture of crystalline and amorphous regions.

Therefore, we used a control light to manipulate the state of the GST. The reason for chose the TM mode for the control light because the electric field distribution is more concentrated on

the upper and lower sides of the waveguide than in the TE mode, as shown in Fig. 5(a) and (b); thereby the GST overlaps with the optical field over a larger area and absorbs more optical power. This enables a lower power consumption for switching between different NLAFs.

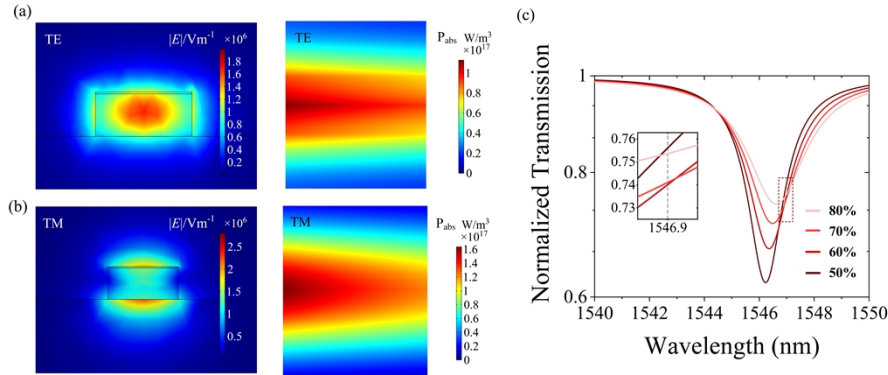


Fig. 5. Simulated (a) TE and (b) TM mode optical profiles (left) and optical absorption (right) of the waveguide with GST on top in the crystalline state. (c) Transmission spectra of the micro-ring for GST crystallization fraction from 50% to 80%. Inset: Enlarged view of the transmission around 1549.5 nm.

However, when the state of the GST changes, the resonant wavelength of the micro-ring and the optical power coupled into the cavity also change as well. Accordingly, we simulated the transmission spectrum of the micro-ring when the GST crystallization fraction is from 50% to 80%, as shown in Fig. 5(c). When the control light wavelength is 1546.9 nm, the difference of the power coupled into the micro-ring does not exceed 2%, as shown in the inset of Fig. 5(c). As a result, for the control light at 1546.9 nm, the energy coupled into the cavity is approximately 25% at different crystallization fractions of GST.

3.2. Power consumption

We analyzed the power consumption for switching among the four optical NLAFs at a control light wavelength of 1546.9 nm. Since the crystalline GST has a larger thermal conductivity and the crystallization process takes a longer time [37], we conclude that the highest power consumption for switching between the four optical NLAFs is corresponds to the switching from RBF to ELU (degree of crystallization from 50 to 80%), whereas the lowest power consumption is switching from ELU to Softplus (degree of crystallization from 80 to 70%).

We then performed a similar analysis referring to [38] for the switching process between the two states. The power consumption required for state switching is determined by both the incident optical pulse duration and pulse power. Due to the TO effect the relaxation time is in the nanosecond regime [39] and due to the limitation in the ablation temperature of the GST, we separately chose $P_1 = 107$ mW, $t_1 = 1$ ns and $P_1 = 107$ mW, $t_1 = 1$ ns, $P_2 = 10$ mW, $t_2 = 29$ ns optical pulses to realize the crystallization and amorphization processes, respectively, as shown in Fig. (6). The maximum and minimum power consumption required to switch between the optical NLAFs are 1.7488 nJ and 0.428 nJ, respectively. A major advantage of our device is its non-volatility: no additional power supply is required to maintain the state.

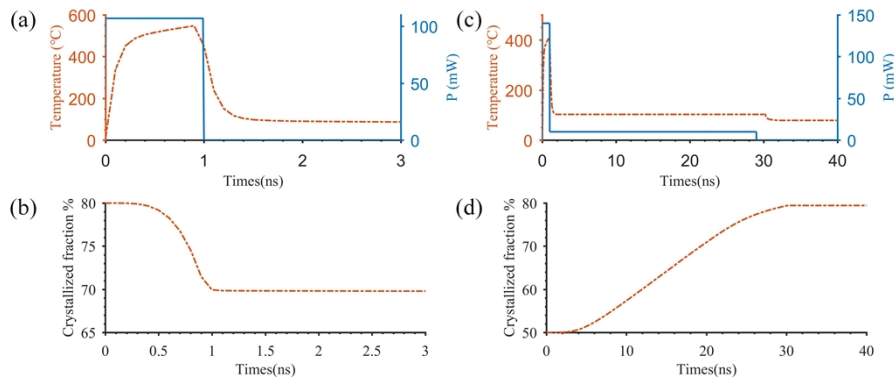


Fig. 6. (a) (b) Control light injection into a previously 80% crystalline loaded-GST/Si hybrid waveguide (a)The average temperature of the GST film. (b) Crystal fraction change of the GST film. (c) (d) Control light injection into a previously 50% crystalline loaded-GST/Si hybrid waveguide (c)Average temperature of the GST film. (d) Crystal fraction change of the GST film.

4. Conclusions

We designed a programmable, low-loss all-optical activation function device based on a silicon micro-ring resonator loaded with PCMs. The NLAF relied on the nonlinear properties of the silicon micro-ring resonator. Programmability was achieved by configuring the state of the GST loaded on the micro-ring resonator. Four different nonlinear activation functions, Relu, ELU, Softplus, and RBF, were implemented for the same incident signal light. The maximum power consumption required to switch between the four different NLAFs was only 1.748 nJ. Simulation of the classification of handwritten digit images also showed that they performed well as alternative NLAFs. Because of the non-volatility of GST, each implementation of the network after determining the NLAF does not need to be reconfigured and consumes almost no energy, thereby achieving a genuinely low-power programmable all-optical NLAF. This demonstrates the potential of the proposed scheme for future applications in all-optical neural networks.

Funding. National Key Research and Development Program of China (2021YFB2900700); Beijing Municipal Natural Science Foundation (L201021).

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. M. M. Waldrop, "The chips are down for Moore's law," *Nature* **530**(7589), 144–147 (2016).
2. K. Kitayama, M. Notomi, M. Naruse, K. Inoue, S. Kawakami, and A. Uchida, "Novel frontier of photonics for data processing—Photonic accelerator," *APL Photonics* **4**(9), 090901 (2019).
3. B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, "Photonics for artificial intelligence and neuromorphic computing," *Nat. Photonics* **15**(2), 102–114 (2021).
4. G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. B. Miller, and D. Psaltis, "Inference in artificial intelligence with deep optics and photonics," *Nature* **588**(7836), 39–47 (2020).
5. X. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T. G. Nguyen, S. T. Chu, B. E. Little, D. G. Hicks, R. Morandotti, A. Mitchell, and D. J. Moss, "11 TOPS photonic convolutional accelerator for optical neural networks," *Nature* **589**(7840), 44–51 (2021).
6. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, and D. Englund, "Deep learning with coherent nanophotonic circuits," *Nat. Photonics* **11**(7), 441–446 (2017).
7. J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, and H. Bhaskaran, "Parallel convolutional processing using an integrated photonic tensor core," *Nature* **589**(7840), 52–58 (2021).

8. J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature* **569**(7755), 208–214 (2019).
9. Z. Cheng, C. Ríos, W. H. P. Pernice, C. D. Wright, and H. Bhaskaran, "On-chip photonic synapse," *Sci. Adv.* **3**(9), e1700160 (2017).
10. J. Feldmann, M. Stegmaier, N. Gruhler, C. Ríos, H. Bhaskaran, C. D. Wright, and W. H. P. Pernice, "Calculating with light using a chip-scale all-optical abacus," *Nat. Commun.* **8**(1), 1256 (2017).
11. T. Zhou, X. Lin, J. Wu, Y. Chen, H. Xie, Y. Li, J. Fan, H. Wu, L. Fang, and Q. Dai, "Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit," *Nat. Photonics* **15**(5), 367–373 (2021).
12. C. Qian, X. Lin, X. Lin, J. Xu, Y. Sun, E. Li, B. Zhang, and H. Chen, "Performing optical logic operations by a diffractive neural network," *Light: Sci. Appl.* **9**(1), 59 (2020).
13. Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, and S. Du, "All-optical neural network with nonlinear activation functions," *Optica* **6**(9), 1132–1137 (2019).
14. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," *Science* **361**(6406), 1004–1008 (2018).
15. X. Xu, M. Tan, B. Corcoran, J. Wu, T. G. Nguyen, A. Boes, S. T. Chu, B. E. Little, R. Morandotti, A. Mitchell, D. G. Hicks, and D. J. Moss, "Photonic Perceptron Based on a Kerr Microcomb for High-Speed, Scalable, Optical Neural Networks," *Laser Photonics Rev.* **14**(10), 2000070 (2020).
16. H. Zhou, J. Dong, J. Cheng, W. Dong, C. Huang, Y. Shen, Q. Zhang, M. Gu, C. Qian, H. Chen, Z. Ruan, and X. Zhang, "Photonic matrix multiplication lights up photonic accelerator and beyond," *Light: Sci. Appl.* **11**(1), 30 (2022).
17. M. Miscuglio, A. Mehrabian, Z. Hu, S. I. Azzam, J. George, A. V. Kildishev, M. Pelton, and V. J. Sorger, "All-optical nonlinear activation function for photonic neural networks [Invited]," *Opt. Mater. Express* **8**(12), 3851 (2018).
18. T. S. Rasmussen, Y. Yu, and J. Mork, "All-optical non-linear activation function for neuromorphic photonic computing using semiconductor Fano lasers," *Opt. Lett.* **45**(14), 3844 (2020).
19. T. Y. Teo, X. Ma, E. Pastor, H. Wang, J. K. George, J. K. W. Yang, S. Wall, M. Miscuglio, R. E. Simpson, and V. J. Sorger, "Programmable chalcogenide-based all-optical deep neural networks," *Nanophotonics* **11**(17), 4073–4088 (2022).
20. M. Zhao, S. Zhong, X. Fu, B. Tang, S. Dong, and M. Pecht, "Deep Residual Networks With Adaptively Parametric Rectifier Linear Units for Fault Diagnosis," *IEEE Trans. Ind. Electron.* **68**(3), 2587–2597 (2021).
21. C. Cortes and V. Vapnik, "Support-vector networks," *Mach Learn* **20**(3), 273–297 (1995).
22. G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (2013), pp. 8609–8613.
23. D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," *arXiv*, arXiv:1511.07289 (2015).
24. H. Zheng, Z. Yang, W. Liu, J. Liang, and Y. Li, "Improving deep neural networks using softplus units," in *2015 International Joint Conference on Neural Networks (IJCNN)* (2015), pp. 1–4.
25. G. H. Y. Li, R. Sekine, R. Nehra, R. M. Gray, L. Ledezma, Q. Guo, and A. Marandi, "All-optical ultrafast ReLU function for energy-efficient nanophotonic deep learning," *Nanophotonics* (2022).
26. A. Jha, C. Huang, and P. R. Prucnal, "Reconfigurable all-optical nonlinear activation functions for neuromorphic photonics," *Opt. Lett.* **45**(17), 4819 (2020).
27. B. Wu, H. Li, W. Tong, J. Dong, and X. Zhang, "Low-threshold all-optical nonlinear activation function based on a Ge/Si hybrid structure in a microring resonator," *Opt. Mater. Express* **12**(3), 970 (2022).
28. T. Van Vaerenbergh, "All-optical spiking neurons integrated on a photonic chip," Ghent University (2014).
29. T. V. Vaerenbergh, M. Fiers, P. Mechet, T. Spuesens, R. Kumar, G. Morthier, B. Schrauwen, J. Dambre, and P. Bienstman, "Cascadable excitability in microrings," *Opt. Express* **20**(18), 20292–20308 (2012).
30. Z. Wang, Q. Li, Z. Fu, A. Katumba, F. D. Coarer, D. Rontani, M. Sciamanna, and P. Bienstman, "Threshold plasticity of hybrid Si-VO₂ microring resonators," in *Optical Fiber Communication Conference (OFC)* (OSA, 2020), p. Th2A.26.
31. L. Zhang, Y. Fei, T. Cao, Y. Cao, Q. Xu, and S. Chen, "Multibistability and self-pulsation in nonlinear high-Q silicon microring resonators considering thermo-optical effect," *Phys. Rev. A* **87**(5), 053805 (2013).
32. A. Luginan, S. García-Cuevas Carrillo, C. D. Wright, and P. Bienstman, "Rigorous dynamic model of a silicon ring resonator with phase change material for a neuromorphic node," *Opt. Express* **30**(14), 25177 (2022).
33. P. E. Barclay, K. Srinivasan, and O. Painter, "Nonlinear response of silicon photonic crystal microresonators excited via an integrated waveguide and fiber taper," *Opt. Express* **13**(3), 801–820 (2005).
34. G. Priem, P. Dumon, W. Bogaerts, D. V. Thourhout, G. Morthier, and R. Baets, "Optical bistability and pulsating behaviour in Silicon-On-Insulator ring resonator structures," *Opt. Express* **13**(23), 9623–9628 (2005).
35. Z. Cheng, C. Ríos, N. Youngblood, C. D. Wright, W. H. P. Pernice, and H. Bhaskaran, "Device-Level Photonic Memories and Logic Applications Using Phase-Change Materials," *Adv. Mater.* **30**(32), 1802435 (2018).
36. H. Zhang, L. Zhou, J. Xu, N. Wang, H. Hu, L. Lu, B. M. A. Rahman, and J. Chen, "Nonvolatile waveguide transmission tuning with electrically-driven ultra-small GST phase-change material," *Sci. Bull.* **64**(11), 782–789 (2019).

37. C. Ríos, M. Stegmaier, P. Hosseini, D. Wang, T. Scherer, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "Integrated all-photonic non-volatile multi-level memory," *Nat. Photonics* **9**(11), 725–732 (2015).
38. Z. Fu, Z. Wang, H. Wang, R. Jiang, L. Liu, C. Wu, and J. Wang, "Thermal dynamics of phase switching process of an SOI rib waveguide covered with a Ge₂Sb₂Te₅ phase change material film," *Opt. Mater.* **124**, 112046 (2022).
39. M. Stegmaier, C. Ríos, H. Bhaskaran, C. D. Wright, and W. H. P. Pernice, "Nonvolatile All-Optical 1 × 2 Switch for Chipscale Photonic Networks," *Adv. Opt. Mater.* **5**(1), 1600346 (2017).