

# Photonic Tensor Core with Low-insertion-loss, Non-volatile $\text{Ge}_2\text{Sb}_2\text{Te}_3\text{S}_2$ Intensity Modulators

Yuto Miyatake<sup>(1)</sup>, Rui Tang<sup>(1)</sup>, Kotaro Makino<sup>(2)</sup>, Junji Tominaga<sup>(2)</sup>, Noriyuki Miyata<sup>(2)</sup>, Makoto Okano<sup>(2)</sup>  
Kasidit Toprasertpong<sup>(1)</sup>, Shinichi Takagi<sup>(1)</sup>, and Mitsuru Takenaka<sup>(1)</sup>

<sup>(1)</sup> Department of Electrical Engineering and Information Systems, The University of Tokyo, Tokyo, Japan, [miyatake@mosfet.t.u-tokyo.ac.jp](mailto:miyatake@mosfet.t.u-tokyo.ac.jp)

<sup>(2)</sup> Device Technology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

**Abstract** We proposed a low-insertion-loss and non-volatile optical intensity modulator based on  $\text{Ge}_2\text{Sb}_2\text{Te}_3\text{S}_2$  (GSTS), a recently proposed widegap phase change material. We demonstrated matrix-vector multiplication using a new photonic tensor core with GSTS intensity modulators, enabling a high recognition accuracy of 97.34 % in CNN. ©2023 The Author(s)

## Introduction

Photonic tensor cores for matrix multiplication based on Si photonic integrated circuits (PICs) are expected to realize energy-efficient, large-bandwidth, and low-latency optical neural networks. For this purpose, various types of PICs have been proposed and demonstrated so far [1–4]. However, these previous PICs have suffered from matrix decomposition complexity or low fabrication tolerance. To overcome these limitations, we have recently proposed a novel photonic tensor core architecture that is tolerant to hardware error accumulation [5].

In this paper, we experimentally demonstrate the proposed photonic tensor core. For this purpose, we propose a new optical intensity modulator based on  $\text{Ge}_2\text{Sb}_2\text{Te}_3\text{S}_2$  (GSTS), which we have recently developed [6]. GSTS has a wider bandgap than conventional  $\text{Ge}_2\text{Sb}_2\text{Te}_5$  (GST) and features negligible optical absorption in its amorphous state and large optical absorption in its crystalline state at a 1550 nm wavelength. These features enable us to realize low-insertion-loss (IL) and non-volatile optical intensity modulators based on GSTS. Thanks to the low IL, the GSTS intensity modulator enables multi-level intensity modulation up to 24 levels for the wavelength range from 1520 nm to 1600 nm. Using the GSTS intensity modulators, we compose a photonic tensor core and successfully demonstrate  $4 \times 4$  matrix-vector multiplication. Based on the experimental results, we numerically implement a convolutional neural network (CNN) for handwritten digit recognition, in which convolutional operations are performed in the optical domain and demonstrate a high recognition accuracy of 97.34 %.

## Operation principle of photonic tensor core

Figure 1 shows the schematic of the proposed photonic tensor core for  $4 \times 4$  matrix-vector

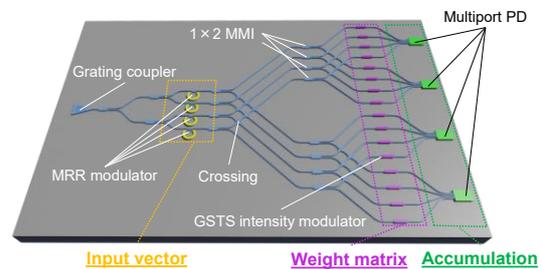


Fig. 1: Schematic of  $4 \times 4$  photonic tensor core.

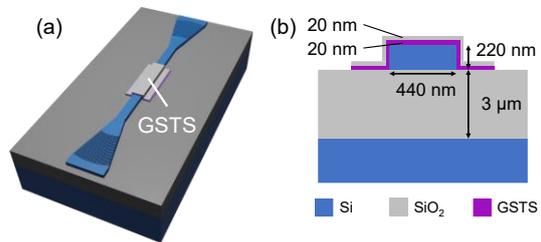
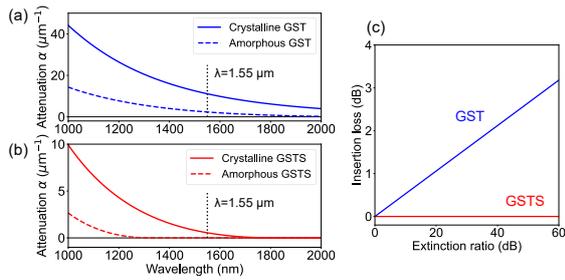
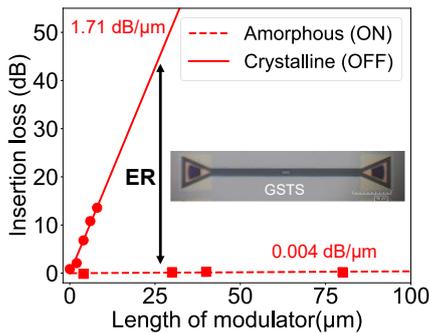


Fig. 2: (a) 3-D view and (b) cross-sectional view of GSTS intensity modulator.

multiplication. Input light from a grating coupler is split into four Si waveguides and injected into the input vector section where microring modulators are used to encode an input vector with optical intensity. Through waveguide couplers and waveguide crossings, the modulated vector signal is distributed to the weight matrix section where the intensity of the vector signal is again modulated according to each weight by non-volatile GSTS intensity modulators, enabling matrix-vector multiplication in the optical domain. Owing to the non-volatility of the GSTS intensity modulator, the weight matrix section does not consume power as long as the weight matrix is fixed. In the following accumulation section, the light is collected into multiport photodetectors to perform accumulation as photocurrent in the electrical domain. The photonic tensor core has an intrinsically small hardware error that does not



**Fig. 3:** Attenuation spectra of (a) GST and (b) GSTS. (c) Extinction ratio vs insertion loss of GST and GSTS intensity modulators at a 1550 nm wavelength.



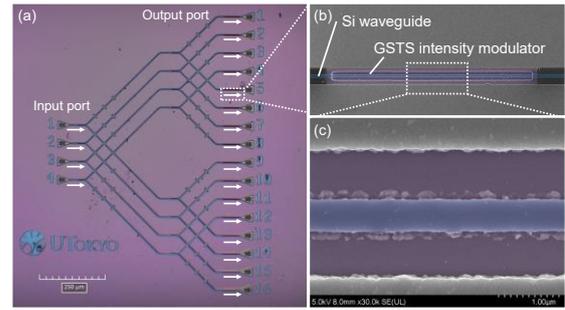
**Fig. 4:** Insertion loss vs length of GST and GSTS intensity modulators at a 1550 nm wavelength. The inset shows an optical microscopy image of the device.

increase with the device scale, being a promising approach to realize a high fidelity and high energy efficiency optical computing platform.

### GSTS intensity modulator

For the intensity modulators, we adopted GSTS, a new wide-gap PCM that we have recently proposed [6]. Figure 2(a) shows the schematic of the GSTS intensity modulator. A 20-nm-thick GSTS and  $\text{SiO}_2$  layers cover the top of a 220 nm-thick Si strip waveguide on a 3- $\mu\text{m}$ -thick buried oxide (BOX) layer as shown in Fig. 2(b).

The advantages of GSTS for the intensity modulators over other PCMs are as follows. GST the most widely used PCM, has non-negligible optical absorption at a 1550 nm wavelength in its amorphous state, resulting in the non-negligible IL of the intensity modulator.  $\text{Sb}_2\text{Se}_3$  [7] and  $\text{Sb}_2\text{S}_3$  [8] have almost no optical absorption in both amorphous and crystalline states at a 1550 nm wavelength, which is not suitable for intensity modulation.  $\text{Ge}_2\text{Sb}_2\text{Se}_4\text{Te}_1$  (GSST) [9] has no absorption in the amorphous state and large absorption in the crystalline state, suitable for intensity modulation, while the toxicity of Se is an issue. GSTS features both a transparent amorphous state and an absorptive crystalline state at a 1550 nm wavelength, like GSST. However, unlike GSST, it is non-toxic. These features are ideal for the realization of compact



**Fig. 5:** (a) Optical microscopy image of 4x4 photonic tensor core. (b), (c) SEM images of a GSTS intensity modulator.

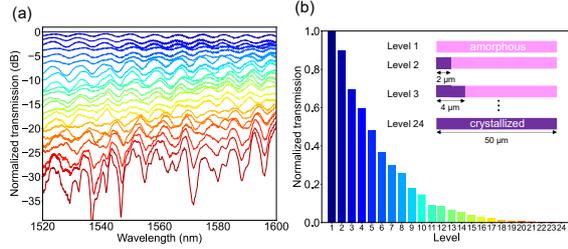
and low-IL intensity modulators.

Figures 3(a) and (b) show the attenuation coefficients of GST and GSTS, respectively. Although amorphous GST has optical absorption at a 1550 nm wavelength, that of GSTS at the wavelength is negligibly small. Figure 3(c) shows the relationship between the extinction ratio (ER) and IL of two PCM intensity modulators at a 1550 nm wavelength. Thanks to the negligibly small optical absorption of the amorphous GSTS, the trade-off between the IL and ER can be overcome, making the GSTS intensity modulator essentially superior to the GST intensity modulators.

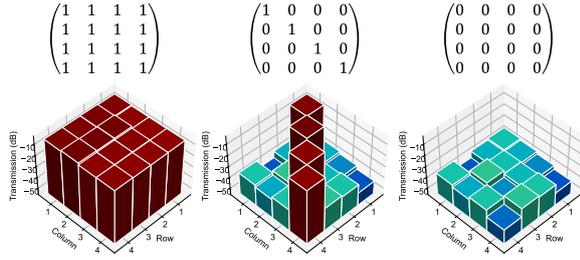
Next, we fabricated the GSTS modulator. After the formation of Si strip waveguides with grating couplers, a 400-nm-thick  $\text{SiO}_2$  cladding was deposited using plasma-enhanced chemical vapor deposition. 2- $\mu\text{m}$ -wide windows were opened in the cladding by electron-beam lithography, inductively coupled plasma reactive ion etching (ICP-RIE), and wet etching with BHF, which was followed by the deposition of a 20-nm-thick GSTS layer and a 20-nm- $\text{SiO}_2$  capping layer using RF sputtering. Lastly, the GSTS film and  $\text{SiO}_2$  film except for the device region were removed using ICP-RIE. Figure 4 shows the ILs of GSTS intensity modulators before and after the crystallization of GSTS at a 1550 nm wavelength. The inset shows an optical microscopy image of the fabricated device. The measured loss of the modulators with amorphous GSTS was 0.004 dB/ $\mu\text{m}$ , which was significantly smaller than that of the GST modulator estimated from Fig. 3(a) (0.2 dB/ $\mu\text{m}$ ). The loss with crystalline GSTS was 1.71 dB/ $\mu\text{m}$ . This result shows that GSTS is more suitable than GST for low-IL intensity modulators.

### Evaluation of photonics tensor core

Figure 5(a) shows an optical microscopy image of the fabricated 4x4 photonic tensor core. As a proof-of-concept demonstration, we implemented only the matrix-vector multiplication section. Figures 5(b) and (c) show scanning electron



**Fig. 6:** Multi-level modulation of GSTS intensity modulator. (a) Transmission spectra and (b) transmissions at a 1550 nm wavelength in the linear scale.



**Fig. 7:** Results of matrix-vector multiplication for the three different matrices at a 1550 nm wavelength.

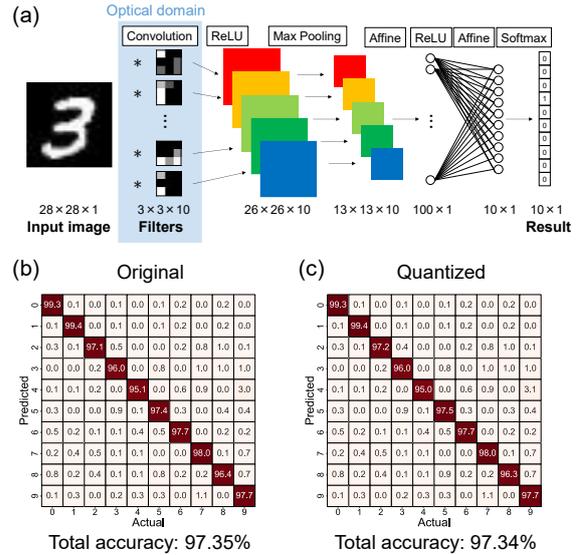
microscopy (SEM) images of a 50- $\mu\text{m}$ -long GSTS modulator. The estimated IL of the 50- $\mu\text{m}$ -long GSTS modulator is 0.2 dB.

To demonstrate the multi-level modulation of the GSTS intensity modulator, the as-deposited amorphous GSTS was crystallized by visible laser irradiation. By increasing the length of the crystallized region, 24-level modulation was achieved for the wavelength range from 1520 nm to 1600 nm as shown in Fig. 6(a) and (b). Note that for crystallization using visible laser irradiation, the spot size of the radiated laser is limited to  $\sim 1 \mu\text{m}$ . When we perform multi-level modulation based on spatially partial crystallization, as shown in the inset of Fig. 6(b), a longer modulator enables multi-level modulation with a larger number of levels. Therefore, the GSTS intensity modulator, which has a much smaller insertion loss than the GST intensity modulators, can be longer and modulated with a larger number of levels compared with the GST intensity modulators.

Figure 7 shows the result of  $4 \times 4$  matrix-vector multiplication for the three different matrices at a 1550 nm wavelength using the photonic tensor core. The ERs were greater than 26 dB for all the GSTS intensity modulators.

### Numerical analysis of CNN

Based on the experimental results, we numerically investigated handwritten digit recognition using CNN as shown in Fig. 8(a). An input image was convoluted with ten  $3 \times 3$  filters, which are assumed to be implemented in the optical domain. First, CNN's weights were determined through ex-situ training. Then, the



**Fig. 8:** (a) CNN for handwritten digit recognition. Confusion matrices of CNN with (b) original and (c) quantized weight.

weights of the five filters in the convolutional layer were quantized with nonuniform 24 levels in Fig. 6 to emulate matrix-vector multiplication using the presented photonic tensor core.

Figure 8(b)(c) shows the confusion matrix. The average total prediction accuracy of the emulated network with quantized weights was 97.34 %, which is almost identical to 97.35 % of the original (without quantization) network.

### Conclusions

we propose a new optical intensity modulator based on GSTS. The GSTS intensity modulator features lower IL than the GST intensity modulator. Thanks to the low IL, the GSTS intensity modulator enables multi-level intensity modulation up to 24 levels. With GSTS intensity modulators, we experimentally demonstrated  $4 \times 4$  matrix-vector multiplication using a photonic tensor core, enabling the high recognition accuracy of 97.34 % in CNN. Owing to the non-volatility of the GSTS intensity modulator, the weight matrix section does not consume power as long as the weight matrix is fixed. Therefore, the presented photonic tensor core is promising for energy-efficient in-memory photonic computing.

### Acknowledgements

This work was partly supported by JST-CREST (JPMJCR2004), JST-Mirai Program (JPMJMI20A1), JSPS KAKENHI (JP20H02198, JP21J20272, JP22K14298, and JP23H00172), and “Advanced Research Infrastructure for Materials and Nanotechnology in Japan (ARIM)” of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) (JPMXP1222UT1028).

## References

- [1] Y. Shen, N. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, Michael Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," *Nature Photon* vol. 11, pp. 441–446, 2017, DOI: 10.1038/nphoton.2017.93.
- [2] C. Ramey, "Silicon Photonics for Artificial Intelligence Acceleration: HotChips 32," 2020 IEEE Hot Chips 32 Symposium (HCS), pp. 1–26, 2020, DOI: 10.1109/HCS49909.2020.9220525.
- [3] A. N. Tait, A. X. Wu, T. F. Lima, E. Zhou, B. J. Shastri, M. A. Nahmias, and P. R. Prucnal, "Microring Weight Banks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, no. 6, pp. 312–325, 2016, DOI: 10.1109/JSTQE.2016.2573583.
- [4] S. Ohno, R. Tang, K. Toprasertpong, S. Takagi, and M. Takenaka, "Si Microring Resonator Crossbar Array for On-Chip Inference and Training of the Optical Neural Network," *ACS Photonics*, vol. 9, no. 8, pp. 2614–2622, 2022, DOI: 10.1021/acsp Photonics.1c01777.
- [5] R. Tang, M. Okano, K. Toprasertpong, S. Takagi, D. Englund, and M. Takenaka, "Two-layer integrated photonic architectures with multiphoton photodetectors for high-fidelity and energy-efficient matrix multiplications," *Optics Express*, vol. 30, no. 19, pp. 33940–33954, 2022, DOI: 10.1364/OE.457258.
- [6] Y. Miyatake, K. Makino, J. Tominaga, N. Miyata, T. Nakano, M. Okano, K. Toprasertpong, S. Takagi, and M. Takenaka, "Proposal of Low-Loss Non-Volatile Mid-Infrared Optical Phase Shifter Based on  $\text{Ge}_2\text{Sb}_2\text{Te}_3\text{S}_2$ ," in *IEEE Transactions on Electron Devices*, vol. 70, no. 4, pp. 2106–2112, 2023, DOI: 10.1109/TED.2023.3235865.
- [7] M. Delaney, I. Zeimpekis, D. Lawson, D. W. Hewak, and O. L. Muskens, "A New Family of Ultralow Loss Reversible Phase-Change Materials for Photonic Integrated Circuits:  $\text{Sb}_2\text{S}_3$  and  $\text{Sb}_2\text{Se}_3$ ," *Advanced Functional Materials*, vol. 30, no. 36, pp. 1–10, 2020, DOI: 10.1002/adfm.202002447.
- [8] Z. Fang, J. Zheng, A. Saxena, J. Whitehead, Y. Chen, and A. Majumdar, "Non-Volatile Reconfigurable Integrated Photonics Enabled by Broadband Low-Loss Phase Change Material," *Advanced Optical Materials*, vol. 9, no. 9, pp. 1–11, 2021, DOI: 10.1002/adom.202002049.
- [9] Y. Zhang, J. B. Chou, J. Li, H. Li, Q. Du, A. Yadav, S. Zhou, M. Y. Shalaginov, Z. Fang, H. Zhong, C. Roberts, P. Robinson, B. Bohlin, C. Ríos, H. Lin, T. Gu, J. Warner, V. Liberman, K. Richardson, and J. Hu, "Broadband transparent optical phase change materials for high-performance nonvolatile photonics," *Nature Communication*, vol. 10, no. 4279, pp. 1–9, 2019, DOI: 10.1038/s41467-019-12196-4.