Vector–Matrix Multiplication at the Speed of Light

RESEARCHERS

Vahid Nikkhah, **Ali Pirmoradi**, **Farshid Ashtiani**, **Brian Edwards**, **Firooz Aflatouni** and **Nader Engheta** (engheta@seas.upenn.edu), University of Pennsylvania, Philadelphia, USA

REFERENCES

1. P. Ambs. Adv. Opt. Technol. **2010**, 372652 (2010).
2. N. Mohammadi Estakhri et al. Science **363**, 1333 (2019).
3. M. Camacho et al. Nat. Commun. **12**, 1466 (2021).
4. V. Nikkhah et al. Nat. Photonics **18**, 501 (2024).
5. A.Y. Piggott et. al. Nat. Photonics **9**, 374 (2015).

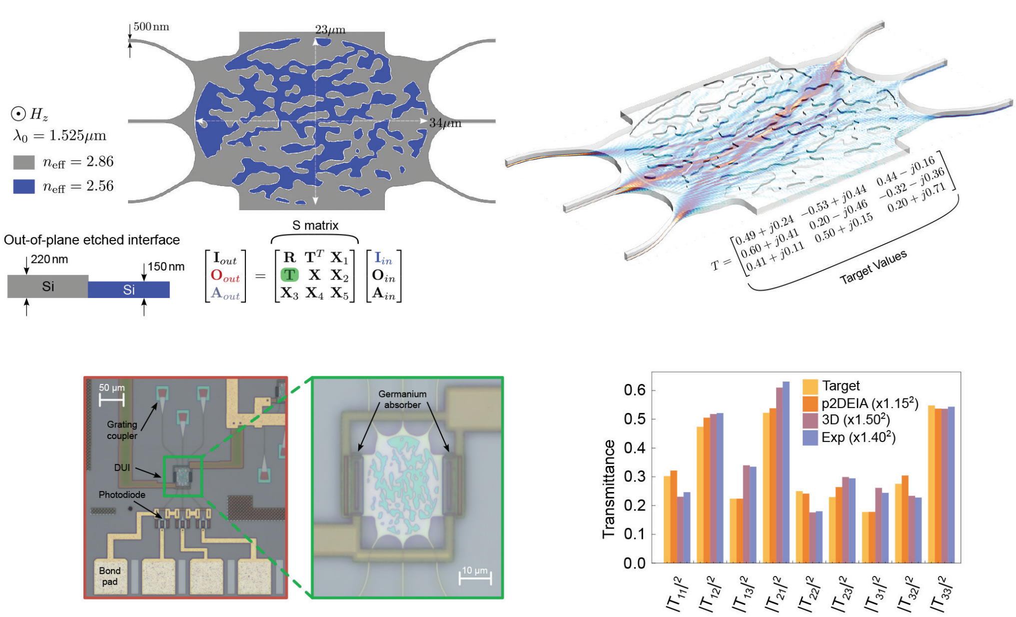
Artificial intelligence (AI) and machine learning (ML) have received considerable attention and growing interest. As generative AI and large language models such as ChatGPT find their way into daily life, the demand for ultrafast, low-power processing and handling of gigantic amounts of data become more pressing.

Vector–matrix multiplication is one of the key mathematical operations needed in such information processing and data handling. Since the speed of light is the fastest speed in the universe and photons do not interact with each other in linear media, research on photonic systems capable of relevant mathematical operations and computation has been actively pursued.[1–3] In recent work, we demonstrated how vector–matrix multiplication can be achieved with light propagation in silicon photonics platforms.[4]

Our structure consists of a 220-nm-thick silicon slab linked to several input and output single-mode waveguide ports. The complex-valued input $N\times 1$ vectors are encoded as the complex analog amplitudes of the dominant optical mode in the input waveguides. The main section of the silicon slab is designed to behave as the desired $N\times N$ matrix to be multiplied by the input vectors. This is achieved by selectively etching a thin, 70-nm layer in specific regions of the slab, creating low-effective-mode-index contrast between the etched and unetched areas and thus providing in-plane mixing of the signals coming from the N input waveguides.

We used the method of inverse design[5] on a 2D model of the structure based on the effective mode index to determine the shape, size and location of the 70-nm etched region in the silicon slab to provide the desired $N\times N$ matrix, while minimizing reflection. Our design was then fabricated in a foundry and tested, successfully achieving the desired transmittance.

Our method provides light-speed vector–matrix multiplication in a significantly smaller footprint compared with networks of Mach-Zehnder interferometers (MZIs) with long waveguides. Moreover, the low-index contrast makes our design less sensitive to aberration in the shape of etched regions—another advantage over MZI networks, where the long waveguides may be more prone to phase sensitivity. Our structures offer a mechanism for vector–matrix multiplication at the speed of light in tiny footprints and, we believe, provide a new low-power solution for the AI and ML paradigms. OPN



Top row: Top (left) and perspective (right) views of an inverse-designed structure that performs vector–matrix multiplication with light on a silicon photonic chip. Bottom row: Photograph of chip (left) and sample of experimental results (right).

Adapted from V. Nikkhah et al. Nat. Photonics **18**, 501 (2024).