

Supporting Information

MXene Nanoflakes Enabled All-Optical Nonlinear Activation Function for On-Chip Photonic Deep Neural Networks

*Adir Hazan, Barak Ratzker, Danzhen Zhang, Aviad Katiyi, Maxim Sokol, Yury Gogotsi and Alina Karabchevsky**

It is observed from Figure S1a that a dielectric to metallic crossover occurs at wavelengths larger than 1139 nm (summarised in Table S1). It is embodied in the real component ϵ_1 that becomes negative as the wavelength increases, indicating the onset of free carrier oscillations [s1]. Accordingly, $\text{Ti}_3\text{C}_2\text{T}_x$ is a plasmonic material at NIR spectra, where either surface plasmon resonance (SPR) or localized surface plasmon resonance (LSPR) could be excited at the MXene layered material interface. In addition, the crossover of ϵ_1 from positive to negative values exhibited film thickness dependence, namely, the wavelength in which ϵ_1 is negative value increases as the thickness of the film decreases. In addition, the imaginary component ϵ_2 for all four samples exhibited a peak around 820 nm, as shown in Figure S1b. The observed ϵ_2 for the MXene thin films incrementally increased at wavelengths longer than the resonant wavelength. Despite their quantitative differences, the spectral behavior of imaginary components of the dielectric functions is qualitatively similar at all four thicknesses. The changes in the relative permittivity's real and imaginary components for varying MXene thin films could be indicated by the thickness dependence of the medium's linear optical properties.

MXene films thickness measurements

Thicknesses of MXene films were measured using a profilometer (Stylus Veeco Dektak-8), and their surface uniformity was evaluated in terms of roughness, continuous region, and profile flatness. The nanometer-scale topography measurements of the $\text{Ti}_3\text{C}_2\text{T}_x$ thin films are shown in Figure S9 and summarised in Table S1. The step function fitting was deduced from the corresponding line profiles indicating the films' average thicknesses of approximately 50 nm, 66 nm, 72 nm, and 91 nm, where 0 height refers to the surface of the BK-7 substrate (i.e., the probe reference point on the substrate plane).

Table S1: Characterisation parameters of the $\text{Ti}_3\text{C}_2\text{T}_x$ (MXene) thin films.

$\text{Ti}_3\text{C}_2\text{T}_x$ Sample	Thickness (nm)	Surface roughness (nm)	ϵ_1 crossover wavelength (nm)
Sample #1	50	8	1347
Sample #2	66	12	1217
Sample #3	72	10	1210
Sample #4	91	15	1139

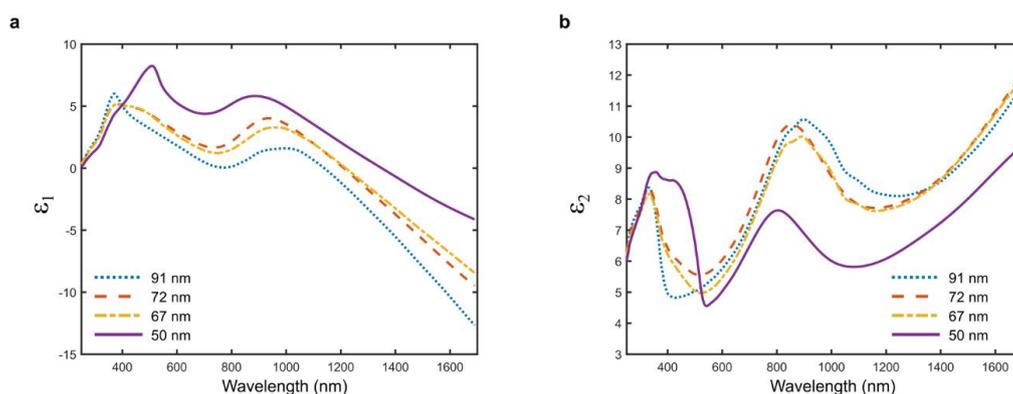


Figure S1. Dispersion of MXene thin films. **a)-b)** The complex permittivity $\tilde{\epsilon}(\lambda)$, with real ϵ_1 (**a**) and imaginary ϵ_2 (**b**) parts as a function of the wavelength for different film thicknesses: 50 nm (solid purple curve), 67 nm (yellow dash-dotted curve), 72 nm (orange dashed curve), and 91 nm (blue dotted curve).

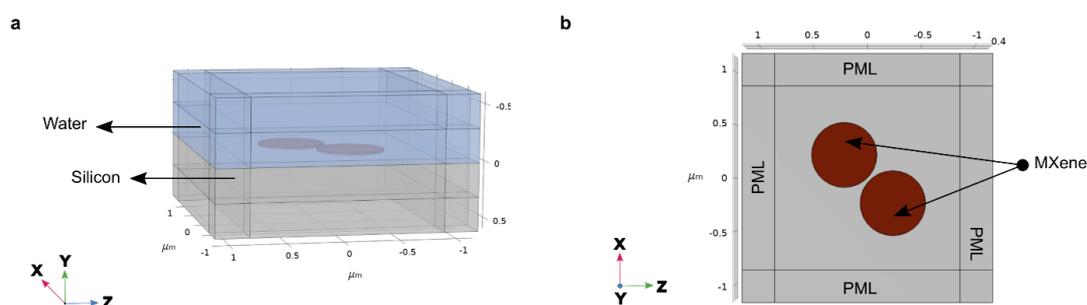


Figure S2. The nanodisks arrangement. **a)-b)** A unit cell including two nanodisks embedded in water (light blue medium) located on top of silicon (grey medium) surface (**a**) as simulated in the numerical model. Considering the light propagating along with the waveguide core (z-direction), with evanescent field components in the y-direction extending into the sample medium. The MXene nanodisks (red) are separated by 15 nm (**b**). Perfectly matched layers (PML).

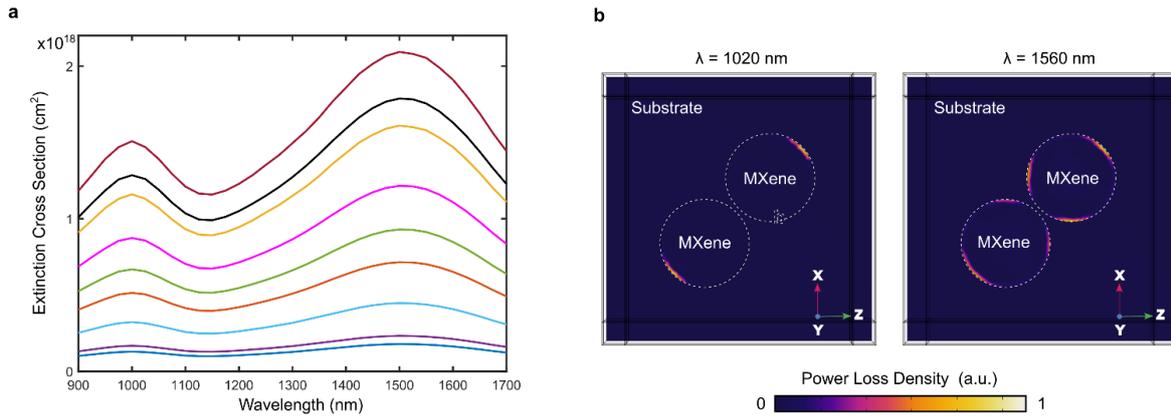


Figure S3. MXene on a substrate. **a)** Calculated results of extinction cross-section spectra of MXene nanodiscs atop the waveguide for different input powers. **b)** Normalised power loss density in a slice through the MXene nanodiscs (the silicon-water interface) at lower (1020 nm) and longer (1560 nm) wavelengths resonances.

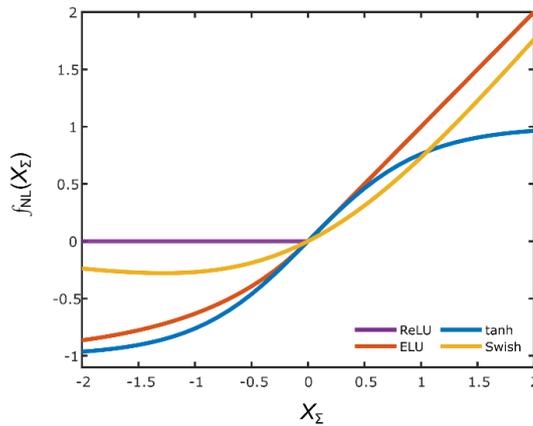


Figure S4. Different common software-based nonlinear activation functions. Represented by input/output relation: ReLU (purple), ELU (orange), tanh (blue), and Swish (yellow).

Despite there are no negative inputs (or weights) involved in the deep-learning task (corresponding to standard intensity encoding scheme), the proposed MXene-based AFs can efficiently work in such deep structured learning configurations. Hence, we have only plotted the regime where $P_{in} > 0$, because the difference between input signals with $P_{in} < 0$ and $P_{in} > 0$ is only reflected by their phases, that is $-P_{in} = e^{j\pi} P_{in}$. Therefore, $P_{out}(P_{in})$ is an odd function $P(-P_{in}) = -P_{out}(P_{in})$. In addition, the devices have intensity dependency, and changes in the optical signal phase do not affect the applied modulation. However, since the

photodetectors and the proposed all-optical configuration are not sensitive to the optical signal phase, $P_{\text{out}}(P_{\text{in}})$ need to be considered as an even function, that is $P(-P_{\text{in}}) = P_{\text{out}}(P_{\text{in}})$.

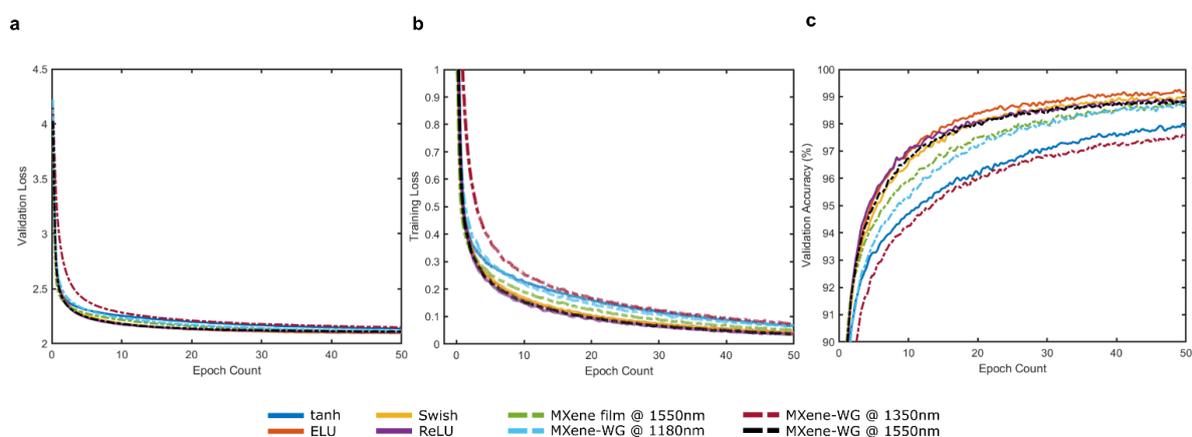


Figure S5. Fully connected network for handwritten MNIST digit classification. **a)-c)** Comparisons of loss as a function of epoch count during the training (**a**) and validation (**b**) processes and model accuracy (**c**), with proposed all-optical nonlinear activation functions considering MXene metasurface overlayer on waveguide and MXene thin film as compared to software-based nonlinear activation functions.

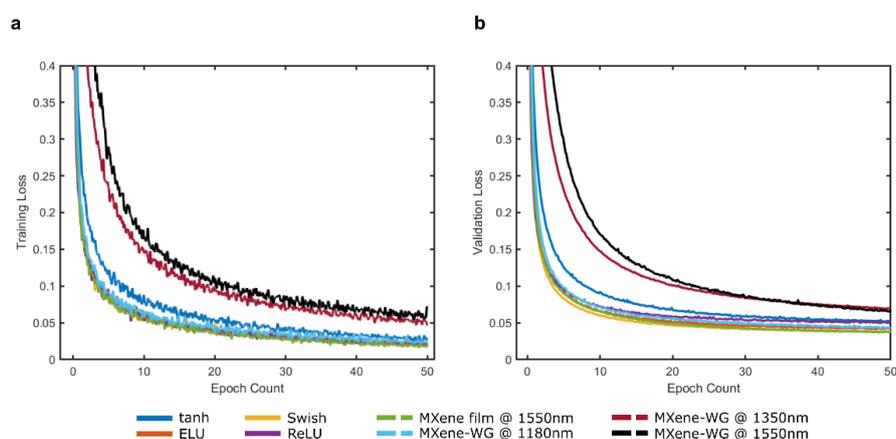


Figure S6. CNN network for handwritten MNIST digit classification. **a)-b)** Comparisons of loss as a function of epoch count during the training (**a**) and validation (**b**) processes, with proposed all-optical nonlinear activation functions considering a MXene metasurface overlayer on waveguide and a MXene thin film as compared to software-based nonlinear activation functions.

Although thermally controlled saturable absorber properties were utilized via free-space optics [s2-6], they were never implemented with MXene on a chip. We note that our architecture does not require thermal control.

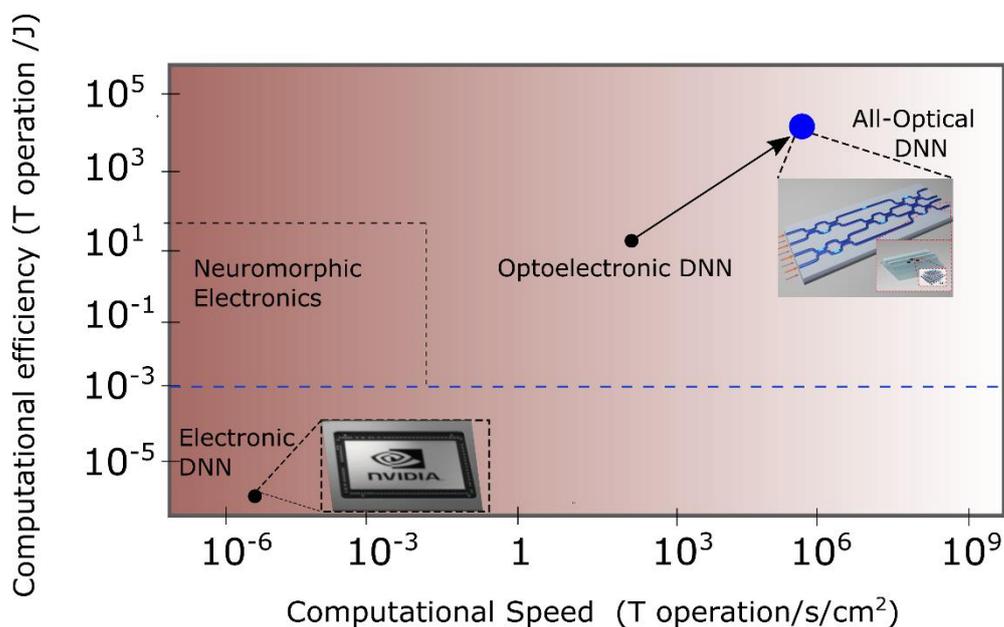


Figure S7. The power efficiency and processing speed dependence of various deep-structure learning architectures: conventional digital machines (e.g., GPUs), optoelectronic configuration, and the all-optical architecture consist of OIU and the proposed nonlinear activator.

These proposed configurations can serve as nonlinear units in which the nonlinear transfer function can be flexibly designed depending on the operating wavelength without modifying the physical structure. They operate at commonly used optical communication wavelength bands and can be adjusted by controlling a variable broadband light source via software. In addition, MXenes family materials have plenty of both optical and electronic properties. An optoelectronic material such as MXene may exhibit hysteresis in the neuron's nonlinear activation function for implementing complex photonic deep structured learning, where output is not a strict function of the corresponding input. The dependence of changing the MXene electronic properties on its optical response to incident light is a candidate to exhibit hysteretic behavior in the nonlinear transfer function, which could be a part of future study. The MXenes family also has many thermal properties that differ from their bulk counterparts [s7]. Hence, the dependence of changing the MXene thermal properties on its optical response to incident light is an additional candidate to exhibit the possibility of converting the nonlinear AFs into different shapes, which could be a part of future study.

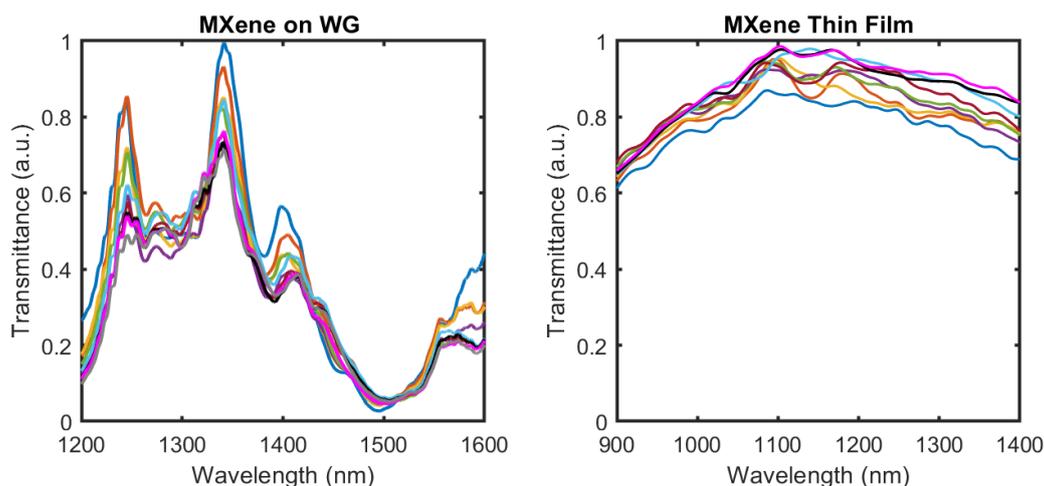


Figure S8. a-b) Zoom-in of normalized transmission measurements (divided by the measured maximum transmission value) dependence on input power signal varying from 6% to 96% for both devices: silicon rib waveguide covered with MXene flakes (**a**) and MXene thin films with 50 nm thickness on BK-7 substrate (**b**).

MXenes family materials have plenty of both optical and electronic properties. An optoelectronic material such as MXene may exhibit hysteresis in the neuron's nonlinear activation function for implementing complex photonic deep structured learning, where output is not a strict function of the corresponding input. The dependence of changing the MXene electronic properties on its optical response to incident light is a candidate to exhibit hysteretic behavior in the nonlinear transfer function, which could be a part of future study.

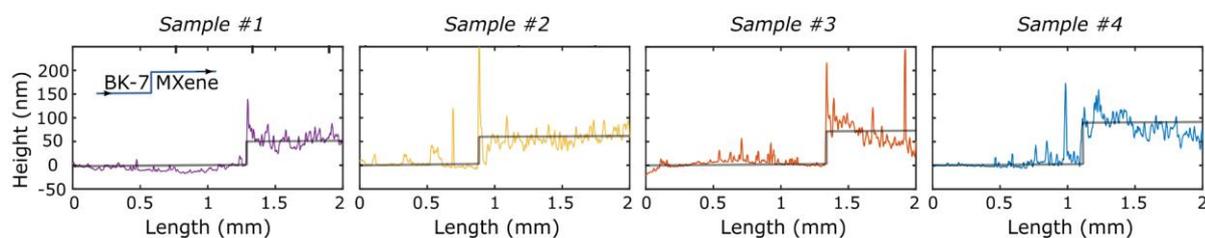


Figure S9. Surface profile measurements of $\text{Ti}_3\text{C}_2\text{T}_x$ MXene thin films. The average thickness (marked by the solid black line) and surface roughness of each sample was determined using a profilometer. Each sample consists of a spray-coated $\text{Ti}_3\text{C}_2\text{T}_x$ film on a BK-7 substrate with different thicknesses. The inset of the left subplot indicates the profile scanning direction for all samples, in which zero height indicates the substrate plane.

To better understand the benefits of using all-optical AF mechanisms that enable ultrafast and energy efficient photonic NNs, we estimated the computational performances with respect to optoelectronic and electronic approaches. Since the suggested all-optical AF allows bypassing the optical-to-electrical-to-optical conversion of current photonic NN implementations, the execution time of a photonic algorithm is, literally, the time that takes the light to propagate in matter through the photonic deep structured learning chip. Hence, it would be possible to estimate the all-optical NN computational performances considering: 1) the physical size of the photonic elemental devices that construct the OIU and ONLU and 2) the RI of the waveguide that constructs the neurons interconnections. The computational performance comparisons of the demonstrated all-optical architecture considering the state-of-the-art NNs technologies are summarised in Table S2.

The deep structured learning approaches' computing speed and energy efficiency can be estimated based on the time cost and power consumption of NNs implementations. This measures the number of total optical and/or electrical computational operations required to execute the NN linear and nonlinear processes per second and per watt-second, respectively. Considering a layer with N neurons containing an $N \times N$ weighting matrix, conventional digital computing will require $R_e = (2N - 1)N$ *electronic operations* to carry out N -dimensional matrix multiplication with more $R_e = N$ *electronic operations* corresponding to the nonlinear transformation. Thus, the total required *electronic operation* is $R_e = 2N^2$, and the corresponding processing rate can be expressed as $v_c = R_e \times 10^9 = 2N^2 \times 10^9$ operations/s, and efficiency of $E/R = 1$ pJ per floating point operations for an outstanding digital computing machine. These values can guarantee the estimation, considering 1-2 more significant orders compared to current digital computing machines (e.g., NVIDIA TITAN RTX aimed to be the best GPU now) with approximately power efficiency of 30-100 pJ per floating point operations and boost clock of 1.77 GHz.

Taking into account the optoelectronics computing approach [s8] employs the same OIU to emphasize the proposed all-optical activation's outcome of significantly improving computational performance. The OIU transformation is consistent with the proposed photonic NN building block chip as shown in Figure 1a and corresponds to matrix multiplication. In contrast, it contains an electronic-based nonlinear AF. Therefore, linear computing implements a layer with N neurons requiring $R_o = (2N - 1)N$ *optical operations* with more

$R_e = N$ electronic operations for realizing the nonlinear AF. Then, the processing rate, $v_c = f(R_o + R_e)$ operations/s, is limited by the optical-to-electronic conversation. In this architecture, the processing rate is the sum of the electrooptic components and the waveguide propagation delay of the light via the OIU formed by the mesh of MZIs. Assuming a conventional telecommunication ultrafast photodetector and electrooptic (EO) modulator with operation frequency 10's-80 GHz, such an optoelectronics system can beat the state-of-the-art electronic computing processes in which their clock rate is bounded to a few GHz. Using all-optical architecture allows a single photonic neuron to be integrated well within a few hundred of μm in length, leading to THz computation processing rate. This architecture does not suffer from electronic delays involved in the two other comparison architectures. Consider the physical size of MZI, approximately 100 μm in length, together with integrated photonic waveguide-based summation of 100 μm with RI of 3.48 at a wavelength of 1550 nm, and an estimated nonlinear activator which is less than 100 μm long in the case of MXene flakes overlayer on a silicon waveguide, or less than 50 nm in the case of MXene thin film. Therefore, the all-optical configuration would have a processing rate of $v_c = R_o \times 10^{11} = 2N^2 \times 2.9 \times 10^{11}$ operations/s for silicon waveguide with MXene flakes overlayer and $v_c = R_o \times 10^{11} = 2N^2 \times 4.3 \times 10^{11}$ operations/s for MXene thin films. The power consumption during linear and nonlinear operations is mainly defined by the optical power necessary for the nonlinear activator [s8]. Thus, the total power required for implementing forward propagation is estimated to be $R/P \approx 2N \times 0.86 \times 10^{14}$ in the case of MXene thin films (with $P_{AF} \approx 5 \text{ mW}$) and $R/P \approx 2N \times 58 \times 10^{14}$ in the case of the for silicon waveguide with MXene flakes overlayer (with $P_{AF} \approx 0.05 \text{ mW}$). Consequently, the proposed all-optical DNN approach may be far more potential by several orders of magnitude in terms of energy efficiency and processing speed than the conventional digital machines for standard and complex problem sizes. In general, as the DNN dimensions (the number of both layers and neurons) increase the more significant advantage of using photonic configuration.

Table S2: Comparison of different DNN technologies in terms of processing speed and power efficiency per operation and per second. All-optical configuration is formed by both OIU and ONLU. OIU and nonlinear electronic activator form optoelectronic configuration. Both linear and nonlinear operations are carried out via processing electrical signals in electronics-based DNNs. All the configurations consider $N = 4$.

<i>DNN Computing Architecture</i>	<i>Optical OPs / Electronic OPs</i>	<i>Efficiency (pJ/OPs)</i>	<i>Processing Rate (TOPs/s)</i>
All-Optical	∞	1.45	13.79
Optoelectronic	7	4×10^3	3.2×10^{-2}
Electronics	0	3.2×10^4	56.66×10^{-3}

In ref. [s9] the optical nonlinearity of lithium niobate waveguide was utilized that exploits the strong and instantaneous $\chi^{(2)}$ and tight spatial confinement of the waveguide modes to enhance the nonlinearity, however, such cumbersome in terms of fabrications periodically poled thin-film lithium niobate (PPLN) nanophotonic waveguide is not compatible with silicon photonics platform and fabs which is the main platform for the OLIU. In contrast, here we demonstrate an effect related to the imaginary part of the $\chi^{(3)}$ optical nonlinearity and field localisation effects in MXene nanoflakes overlay on silicon rib waveguide. The MXene-based configuration that we proposed, exhibits broadband optical responses in contrast to the ref. [s9]. In terms of technology readiness level: our devices are ready to be used and do not require additional components such as wavelength modulators or diabatic mode converters with advanced mode matching, for integrating with existing and developed optical NN architectures. Here the reconfigurable materials platform is proposed and so the all-optical nonlinear transform functions via the MXene-based devices' optical responses are achieved in the cases of 1) evanescent excitation on a chip, and 2) in free-space configurations.

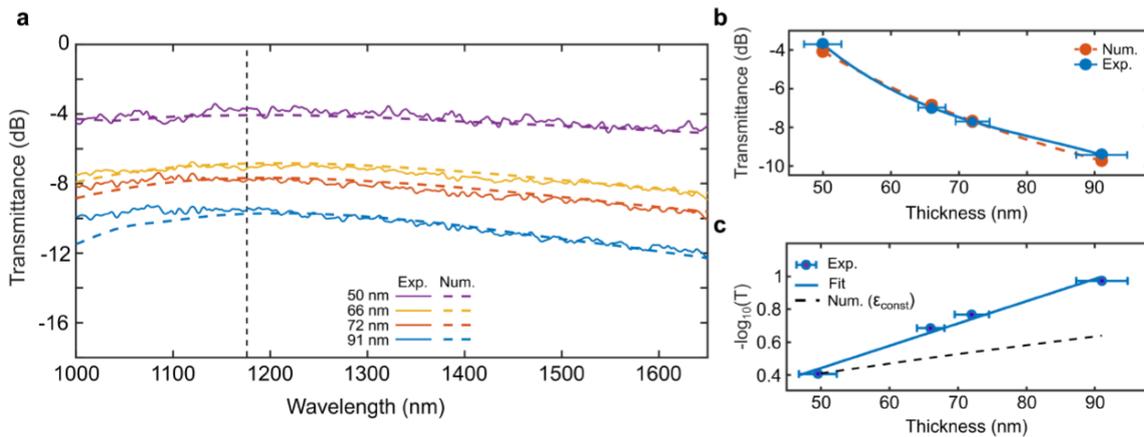


Figure S10. NIR transmittance spectra of MXene thin films. **a**, Experimentally obtained and numerically calculated spectral transmittance for $\text{Ti}_3\text{C}_2\text{T}_x$ thin films with various thicknesses (top to bottom) 50 nm, 66 nm, 72 nm, and 91 nm. **b**, Dependence of transmittance on MXene film thickness of the numerical (orange) and experimental (blue) results at a wavelength of 1180 nm. **c**, The logarithm of transmittance vs film thickness of MXene thin films (blue dots) and linear fit to the data

using the derived thickness (solid blue curve). The dashed black curve indicates the numerical $-\log_{10}(T)$ values with thickness-independent permittivity.

Since the transmittance increases as MXene film thickness decreases, a thinner film is more suitable to serve as an optical nonlinear AF for photonics deep structured learning architectures. Therefore, we focused on the thinnest MXene sample in this work.

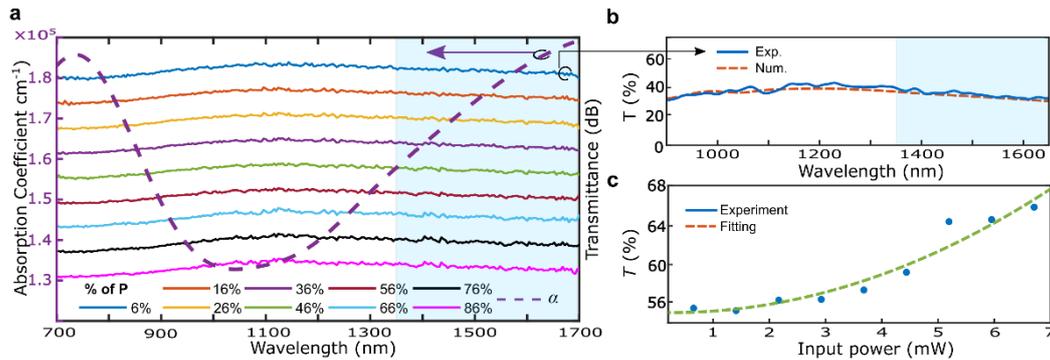


Figure S11. Experimental results of MXene thin film optical transfer function. **a**, The measured transmittance spectra for 50 nm thickness MXene thin film deposited on BK-7 substrate for different input power varying from 6% to 96% (from top to bottom). The measurement values related to the lowest input power are shown in subplot **(b)**, and the others are plotted with a step of -3dB. The purple dashed curve indicates the absorption coefficient for MXene thin film calculated from the spectroscopic ellipsometry measurement by its connection to the imaginary component of the RI as $\alpha(\lambda) = 4\pi \frac{\kappa(\lambda)}{\lambda}$. The blue area indicates the spectral region in which ϵ_1 becomes negative. **b**, The numerically calculated (dashed orange curve) and experimentally obtained (blue curve) transmittance response as a function of the wavelength for the lowest input power (6%). **c**, The measured nonlinear transmittance at a wavelength of 1236 nm as a function of incident power for $\text{Ti}_3\text{C}_2\text{T}_x$ 50 nm film and the corresponding calculated fitted curve.

References:

1. Dillon, A.D., Ghidui, M.J., Krick, A.L., Griggs, J., May, S.J., Gogotsi, Y., Barsoum, M.W., and Fafarman, A.T. (2016) Highly conductive optical quality solution-processed films of 2D titanium carbide. *Adv Funct Mater*, **26** (23), 4162–4168.
2. Feldmann, J., Youngblood, N., Wright, C.D., Bhaskaran, H., and Pernice, W.H.P. (2019) All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature*, **569** (7755), 208–214.

3. Moerland, P.D., Fiesler, E., and Saxena, I. (1996) Incorporation of liquid-crystal light valve nonlinearities in optical multilayer neural networks. *Appl Opt*, **35** (26), 5301–5307.
4. Hill, M.T., Frietman, E.E.E., de Waardt, H., Khoe, G., and Dorren, H.J.S. (2002) All fiber-optic neural network using coupled SOA based ring lasers. *IEEE Trans Neural Netw*, **13** (6), 1504–1513.
5. Huang, C., de Lima, T.F., Jha, A., Abbaslou, S., Shastri, BJ, and Prucnal, PR (2019) Giant enhancement in signal contrast using integrated all-optical nonlinear thresholder. *Optical Fiber Communication Conference*, M3E-2.
6. Mirek, R., Opala, A., Comaron, P., Furman, M., Król, M., Tyszka, K., Seredyński, B., Ballarini, D., Sanvitto, D., and Liew, T.C.H. (2021) Neuromorphic binarised polariton networks. *Nano Lett*, **21** (9), 3715–3720.
7. Zha, X.-H., Huang, Q., He, J., He, H., Zhai, J., Francisco, J.S., and Du, S. (2016) The thermal and electrical properties of the promising semiconductor MXene Hf₂CO₂. *Sci Rep*, **6** (1), 1–10.
8. Shen, Y., Harris, N.C., Skirlo, S., Prabhu, M., Baehr-Jones, T., Hochberg, M., Sun, X., Zhao, S., Larochelle, H., and Englund, D. (2017) Deep learning with coherent nanophotonic circuits. *Nat Photonics*, **11** (7), 441–446.
9. Li, GHY, Sekine, R., Nehra, R., Gray, R.M., Ledezma, L., Guo, Q., and Marandi, A. (2022) All-optical ultrafast ReLU function for energy-efficient nanophotonic deep learning. *Nanophotonics*.