nature communications

Article

https://doi.org/10.1038/s41467-024-54776-z

Programmable nonlinear optical neuromorphic computing with bare 2D material MoS_2

Received: 27 March 2024

Accepted: 21 November 2024

Published online: 27 November 2024

Check for updates

Lei Tong[1,2,3,9], Yali Bi[4,5,9], Yilun Wang[1,9], Kai Peng[6], Xinyu Huang[1], Wei Ju[1,4], Zhuiri Peng[1], Zheng Li[1], Langlang Xu[1], Runfeng Lin[1], Xiangxiang Yu[1], Wenhao Shi[1], Hui Yu[3], Huajun Sun[1,2], Kanhao Xue[1,2], Qiang He[1,2], Ming Tang[1], Jianbin Xu[3], Xinliang Zhang[1], Jinshui Miao[7], Deep Jariwala[8], Wei Bao[6] ✉, Xiangshui Miao[1,2] ✉, Ping Wang[1,4] ✉ & Lei Ye[1,2,7] ✉

Nonlinear optical responses in two-dimensional (2D) materials can build free-space optical neuromorphic computing systems. Ensuring the high performance and the tunability of the system is essential to encode diverse functions. However, common strategies, including the integration of external electrode arrays or photonic structures with 2D materials, and barely patterned 2D materials, exhibit a contradiction between performance and tunability. Because the unique band dispersions of 2D materials can provide hidden paths to boost nonlinear responses independently, here we introduced a new free-space optical computing concept within a bare molybdenum disulfide array. This system can preserve high modulation performance with fast speed, low energy consumption, and high signal-to-noise ratio. Due to the freedom from the restrictions of fixed photonic structures, the tunability is also enhanced through the synergistic encodings of the 2D cells and the excitation pulses. The computing mechanism of transition from two-photon absorption to synergistic excited states absorption intrinsically improved the modulation capability of nonlinear optical responses, revealed from the relative transmittance modulated by a pump-probe-control strategy. Optical artificial neural network (ANN) and digital processing were demonstrated, revealing the feasibility of the free-space optical computing based on bare 2D materials toward neuromorphic applications.

Two-dimensional (2D) materials exhibit a plethora of nonlinear optical responses[1–3], such as saturable absorption[4,5], stimulated emission[6,7], and high harmonic generation[8–13]. Especially, due to the Coulomb interactions confined in the atom-thin plane, the nonlinear responses of 2D materials are sensitive to external modulations[1–3,8–15]. This is desired for optical neuromorphic computing[16–21], enabling the characteristics of intrinsic tunability, ultrafast, power-efficient, and parallelism[6,8,10,20,22–24]. The modulation strategies are mainly conducted in guided space by integrating 2D materials onto waveguides[15,25–27], ring resonators[28,29], and fibers[30]. The embedding of 2D materials in free-space computing systems can provide new flexibilities to design neuromorphic functions[14,22,23,31–33]. However, the modulation performance in free space is limited[8–14,34], because of the limited light-matter interaction length along the out-of-plane direction of 2D materials[8,10,11,34]. In

A full list of affiliations appears at the end of the paper. ✉e-mail: baow2@rpi.edu; miaoxs@hust.edu.cn; p_wang@hust.edu.cn; leiye@hust.edu.cn

addition, the requirement of reconfigurable encoding procedures for different functions in free-space computing systems is another important challenge for stable and effective tunability[22–24,31–33].

Fabricating electrode array onto 2D materials and applying electric fields can modulate nonlinear signals by changing the oscillating strength of excitons or breaking the crystal inversion symmetry[8,9]; the electronic system shows high tunability but also sacrifices the speed to ~ ns[8,9,12,13]. Free-space photonic structures, including nanocavities and metamaterials, can modulate signals at high speed and efficiency[10,11,15,34,35]. However, the spectral bandwidth and modulation depth are restricted by their resonant nature, limiting the tunability range to fit with neural dynamics[10,14,34]. Except for the above contradiction, the complex fabrication and large-scale integration are also challenging for the coupling of 2D materials with external structures. Therefore, using bare 2D materials independently is envisioned as a promising route to develop new free-space optical computing systems. Recently, related devices, such as spatial light modulators[36] and meta-lenses[37] have been demonstrated with purely patterned 2D materials and have realized neural networks. However, the tunability of the system is still limited by the fixed geometric structures[38]. To support computing functions with high performance and tunability, directly encoding optical excitation fields to intrinsically excite various nonlinear responses rather than rely on geometric structures is a solution[4,8,10]. However, compared to the above strategies, the limited signal-to-noise ratio (SNR) is a congenital restriction. Therefore, building a free-space optical neuromorphic computing system based on 2D materials remains a challenge (Fig. 1a).

In this work, we proposed a new free-space optical neuromorphic computing concept to conduct signal processing and computing within a bare 2D molybdenum disulfide (MoS_2) array. This system shows an improved modulation performance with intrinsic fast speed, low energy consumption, and enhanced SNR of > 16 dB while simultaneously ensuring tunability through the synergistic encodings of the 2D cells and the excitation pulses. The key mechanism is the intrinsic transition from two-photon absorption (TPA)[39] to synergistic excited states absorption (SESA)[40], revealed from the modulated relative transmittance (ΔT) of MoS_2 by a pump-probe-control scheme. After synergistic excitation, the ΔT is enhanced because of the much higher density of states (DOS) for the final excited states[40]. In addition, the ΔT shows strong thickness-dependence, due to the generation of ionized electron-hole pairs assisted by Coulomb screening. The relaxation of ΔT is prolonged, mainly attributed to the thermalization of ionized electron-hole pairs[40,41]. Then, the control pulse delay, pulse power, and flake thickness can encode different computing functions to demonstrate the modulated performance and tunability. Here, an artificial neural network (ANN) was developed by encoding the control time delay into 20 weight levels. The optical ANN showed a maximum operating time of ~ 5.47 ps in one cell, and the operation energy consumption for one multiplication calculation was between ~ 5.12 – 9.12 pJ. After training, the experimental testing agreed with the simulated results. In addition, digital processing, such as optical logic gates and a

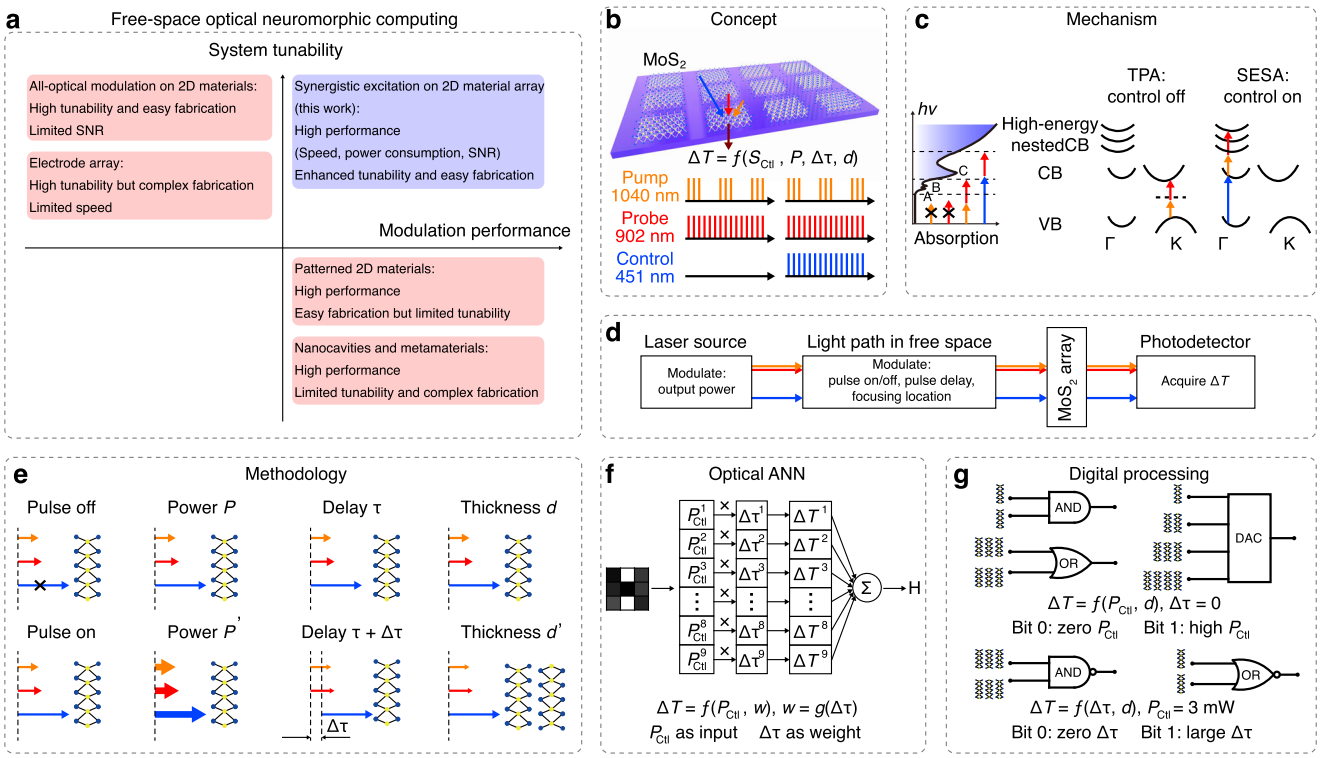

Fig. 1 | Free-space optical neuromorphic computing concept based on a pure 2D MoS_2 array. a Contradiction between performance and tunability of the system for different strategies. The proposed system shows an improved modulation performance and simultaneously ensures the tunability of the system. **b** Schematic of the computing concept based on a pure MoS_2 flake array. Within the pump-probe-control setup (bottom), the relative transmittance (ΔT) is effectively enhanced and modulated, assisting the design of computing functions. **c** Schematic of the computing mechanism. The synergistic transition mainly involves five specific bands: the VB maximum and CB minimum near the K points, the parallel VB and CB between the K and Γ points, and the high-energy nested CBs near the Γ point. The ΔT with the control pulse off is dominated by TPA at the K points. The ΔT with the control pulse on is dominated by SESA between the K and Γ points. **d** Simplified schematic of the computing system. **e** Schematic of the computing methodology. Computing functionalities can be realized by modulating the pulse on/off state, pulse power, pulse delay, and encoding the flake thickness. **f** Schematic of optical ANN. The ΔT is a function of control power and control time delay. The control power in each cell is the input signal, and the control time delay is encoded into weight. The cell thicknesses are identical. **g** Schematic of digital processing functions. Optical AND, OR, NAND, NOR gates, and DAC are realized by encoding the control pulse power and delay into input bits and choosing appropriate cell thickness as input ports.

digital-to-analog converter (DAC) was demonstrated to show the tunability and adaptability to different functions. The performance modulation in MoS_2 is stable and reliable. The pulse spot size determines the equivalent cell size, showing the potential for a higher cell density. More importantly, the methodology and concept apply to various 2D materials. Therefore, this free-space optical neuromorphic computing concept based on pure 2D materials can provide a new approach to program neuromorphic applications.

Results

Free-space computing concept based on bare 2D MoS_2 flake array

The system concept is based on a pure MoS_2 flake array without electrode arrays, complex nanocavities, or meta-structures, therefore simultaneously avoiding the complex fabrication and restrictions of these structures (Fig. 1b). To improve the modulation of nonlinear responses based on the synergistic transitions, we designed a pump-probe-control experimental setup[4] (Supplementary Fig. 1a) based on the band structure of MoS_2 (Fig. 1c). In theory, the A and B excitons near the band edge are residing at the K points of the Brillouin zone[40]. The C peak is determined by transitions between K and Γ points, distributed in a circular region around the Γ point where the conduction band (CB) and valence band (VB) are nearly parallel[40]. This particular band alignment generates larger joint DOS for much higher absorption near the C peak, and more substantial many-body effects can occur under external excitation[40]. Here, A and B excitons are in the same state by neglecting the spin-orbit coupling. Therefore, the synergistic excitation mechanism mainly involves five specific bands in the momentum space: the VB maximum and CB minimum near the K points, the parallel VB-CB between the K and Γ points, and the high-energy nested CBs near the Γ point. Then we can determine the pump pulse (1040 nm, 220 fs), probe pulse (902 nm, 100 fs), and control pulse (451 nm, 100 fs) to modulate the ΔT of MoS_2. The superimposed light field exhibited an effective full width at half maximum of ~156 nm. The pump (control) time delay is defined as the excitation time difference of the pump (control) pulse relative to the probe pulse, modulated mechanically by two time-delay lines. Because the pump and probe photon energies were smaller than the bandgap, the pump-probe response without the control pulse is attributed to TPA. The control pulse photon energy is near the C peak to effectively promote transitions in the parallel VB-CB bands. The synergistic excitation under the control pulse can generate excited states in the high-energy nested CBs, where the much higher DOS can enhance the ΔT, therefore, intrinsically improving the modulation capability. Within this mechanism, we can build a free-space computing system (Fig. 1d), in which the output power of each pulse is controlled within the laser source, the pulse on/off state and delay are controlled within the light path, and the MoS_2 cell is selected by modulating the focusing location. As a result, the ΔT responses, as a function of control pulse state, pulse power, pulse delay, and flake thickness, can be used to encode different computing functions (Fig. 1e), as will be discussed later. Because our mechanism and function encoding scheme are mainly determined by the absorption in MoS_2 rather than the geometry of the MoS_2 array, here inverse design is not needed. In the experimental setups (Supplementary Fig. 1b), the ΔT responses were measured by a photodiode, and we used the original data for the following discussions, with the signal unit in counts per second (cts/s).

Modulation performance

The ΔT responses were characterized to reveal the key mechanism and the modulation performance. The experimental setups are summarized in the methods. It should be noted that the nonlinear characteristics are sample-independent. The on/off state of the control pulse was switched in situ to directly enhance the ΔT response. The ΔT was near zero (pump absorption ~1%) when the control pulse was off, while an enhanced ΔT (pump absorption ~15%) was detected when the control pulse was on (Fig. 2a, sample A), indicating the synergistic excitation nature different from saturable absorption[1,4]. The on/off state of the control pulse can realize an intensity modulator with an enhanced SNR of > 16 dB, which is determined by the higher DOS in the high-energy nested CBs. With the increase of the control power, the ΔT dynamics under scanned pump time delay were enhanced monotonically (Fig. 2b, Supplementary Fig. 2, Sample B, ΔT mappings in Supplementary Figs. 3 and 4), indicating the enhancement is mainly attributed to the control pulse. The non-shifted peak and identical relaxation lifetime reveal that the control power did not alter the final excited states. Quantitatively, the ΔT showed a linear relationship with power for all three pulses (Fig. 2c, sample C, ΔT mappings in Supplementary Figs. 5 to 7). The signal followed $\Delta T \propto P_{\mathrm{Pump}} \times P_{\mathrm{Probe}} \times P_{\mathrm{Ctl}}$, matching the TPA nature when the control pulse was off, and excluding the stimulated emission mechanism when the control pulse was on. This linear relationship with the pulse power can encode power as input and ΔT as output for computing. It should be noted that higher pulse power can lead to saturable absorption. The pulse power should be limited to keep the above linear relationship and ensure accuracy based on the proposed computing methods. Besides, due to the synergistic effects, the ΔT dynamics under scanned control time delay showed a prolonged relaxation to over ~5 ps (Fig. 2d), which can be feasible for pulse delay-encoded applications. This prolonged relaxation is also revealed from ΔT mappings in Supplementary Figs. 8 and 9.

The ΔT response under scanned pump time delay was analyzed for regions with different thicknesses (Sample D, regions marked in Supplementary Figs. 10 and 11). Regardless of the thickness, the ΔT peak was blue shifted to the same pump time delay of 48 fs when the control pulse was on, indicating all regions share the same excitation path (Supplementary Fig. 12). After summarizing the thickness distributions for all regions in Fig. 2e (atomic force microscopy (AFM) image in Supplementary Fig. 13), we focused on the ΔT-thickness correlation at a pump time delay of 82 fs. The ΔT showed weak thickness-dependence without the control pulse (Fig. 2f and Supplementary Fig. 14) due to the limited changes in DOS for TPA[42]. With the control pulse included, the ΔT decreased monotonically with the increase of thickness when the thickness was below ~70 nm, while the ΔT enhancement was quenched for thicker locations over ~70 nm (Fig. 2g and Supplementary Fig. 15), showing the screening of the control pulse. The enhanced ΔT is attributed to the much higher DOS for SESA[40]. As the thickness increases, the high DOS near the nested CBs contributes to stronger Coulomb screening[41,43]. This screening effect makes more hot carriers generate ionized electron-hole pairs and relax through intraband scattering, leading to the observed lower ΔT[41,42]. Over the thickness limit of ~70 nm, the intraband scattering dominates and leads to the observed quenching effect. The thermalization of ionized electron-hole pairs can contribute to the prolonged relaxation lifetime[41,43] (detailed explanation in Supplementary Note 1). It should be noted that computing functions in this study are encoded with cell thickness below ~70 nm, where SESA dominates. A similar thickness dependence is revealed in the gain distribution $G = \Delta T_{\mathrm{control\ on}} / \Delta T_{\mathrm{control\ off}}$ (Supplementary Fig. 16). Hence, the improved modulation performance relies on enhanced excitation and prolonged relaxation. This strong thickness dependence is also evidence that the nonlinear process in MoS_2 is dominated by the interaction with the in-plane electric field component of the light field[44].

Optical ANN in the system

Then, this system can realize an optical ANN, and using bare 2D materials for ANN was rarely estimated. The optical system performed input encoding, weight encoding, vector-matrix multiplication, and output detection. The weight updating values during back-propagation, and output summation and activation were calculated

https://doi.org/10.1038/s41467-024-54776-z

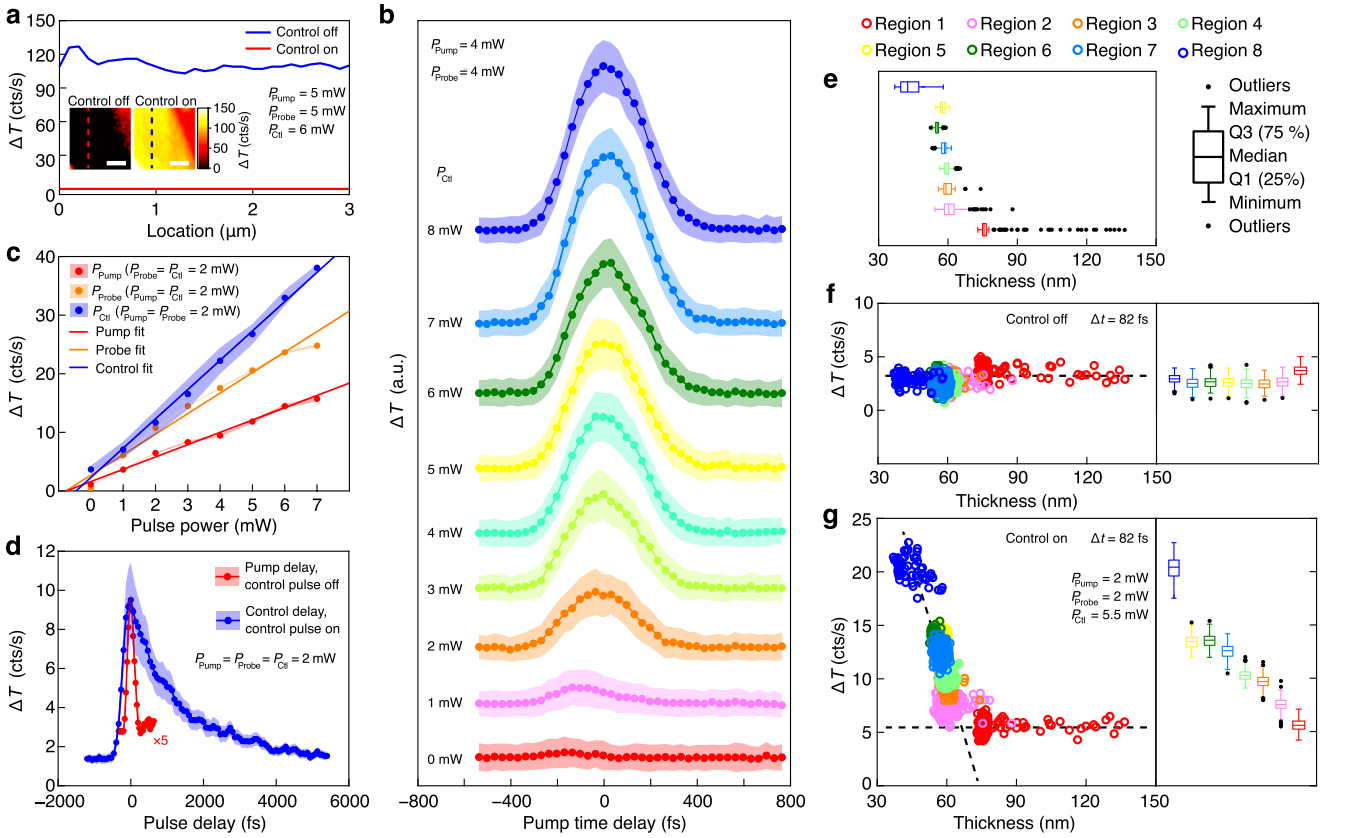

Fig. 2 | Modulation performance. a The ΔT enhancement (pump absorption ~15%) is controlled by the on/off state of the control pulse. The ΔT distribution is extracted along the dashed lines in the inset mappings. Scale bars, 5 μm. **b** ΔT dynamics vs. pump time delay under different control powers in sample B. The shaded area shows the standard deviation of ΔT acquisition. **c** The linear dependence of ΔT with power for three pulses in sample C. The solid line shows the linear fitted results, and the shaded area shows the standard deviation of ΔT acquisition. **d** The red line shows the fast decay with the pump time delay when the control pulse was off. The blue line shows the prolonged decay with the control time delay when the control pulse was on. The shaded area shows the standard deviation of ΔT acquisition. **e** The thickness distribution for eight selected regions. **f**, **g** The thickness-ΔT correlation at a pump time delay of 82 fs, with the control pulse off (**f**) and on (**g**). The right panels show the ΔT distribution in each region.

with a digital computer. Next, we focus on the performance of the optical system. Here, the output ΔT for each cell was a function of control power and control time delay, where the control power functioned as the input signal, and the control time delay marked as the signal delay Δτ was encoded as the weight (Fig. 1f). Therefore, the output ΔT quantifies the calculation of input × weight in each MoS_2 cell. We used 9 cells with the same thickness for ANN computing (optical image in Supplementary Fig. 13h), so the output ΔT in each cell can be added directly. The pump time delay was zero, and pump-probe powers were 2 mW as fixed parameters during computing. ΔT responses were measured for each cell in sequence. Under different P_{Ctl} from 1 to 8 mW, the Δτ was continuously modulated from 5.47 to 0 ps in the first half cycle, and then increased to 5.47 ps in the second half cycle, to monitor the ΔT responses in 8 cycles (Fig. 3a and Supplementary Fig. 17a). This continuous Δτ was set by mechanically moving the delay line on the light path of the control pulse. A larger difference in light path length leads to a larger Δτ. From the results, the response is stable from cycle to cycle, and the symmetric characteristic is beneficial for weight updates. Besides, MoS_2 is stable at ambient conditions, the laser power is below the saturation threshold and induces no damage to the material, and the light-matter interaction shows an enhanced SNR, thus, the ANN computing performance can be reliable in the long term. Under fixed P_{Ctl} of 8 mW, we switched the Δτ between 0 ps and 5.47 ps periodically and measured the ΔT responses in 50 cycles, the system can maintain stable endurance performance (Supplementary Fig. 17b). Because the ΔT relaxation is generally fitted with exponential equations of Δτ [45], the logarithm of ΔT was applied to encode the normalized weight ($\ln(\Delta T) \propto w$), leading to an approximately linear relationship between weight and Δτ. Based on the experimental results in Fig. 3a, the Δτ in the range of 0–5.47 ps was encoded into 20 weight levels (Fig. 3b), which also determined the maximum operation time of ~5.47 ps in one cell. For the potentiation and depression process, multilevel ΔT states were measured for the encoded Δτ values under different P_{Ctl}, in which adjacent ΔT states can be distinguished clearly (Fig. 3c, d and Supplementary Fig. 17c–j).

Based on the encoded multilevel ΔT states, a single-layer ANN with 36 synapses was applied for pattern recognition, with SoftMax as the activation function and cross-entropy as the cost function (detailed ANN setup can be found in Supplementary Note 2)[18]. A custom training dataset consisted of 3 × 3 patterns for letters H, U, S, and T, encoded by P_{Ctl} for each pixel (Supplementary Fig. 18). To evaluate the noise robustness, the Gaussian noise with a mean value of zero and standard deviation of 0.2 mW is included in the training dataset. Each pattern was transformed into a 1 × 9 input vector, and after calculating the vector-by-matrix multiplication in the pure MoS_2 array, the output of the network was a 1 × 4 vector $[\Delta T_1 \quad \Delta T_2 \quad \Delta T_3 \quad \Delta T_4]$, and the maximum component in the output vector could acquire the recognition result. In the off-chip training process, the initial Δτ components were set to zero, which were then updated to modulate the output vector in 30 epochs (Supplementary Fig. 19), and the corresponding weights were summarized in Supplementary Fig. 20. The average output vector components for the true label H, U, S, and T in each epoch were calculated in Supplementary Fig. 21a. By comparing the

https://doi.org/10.1038/s41467-024-54776-z

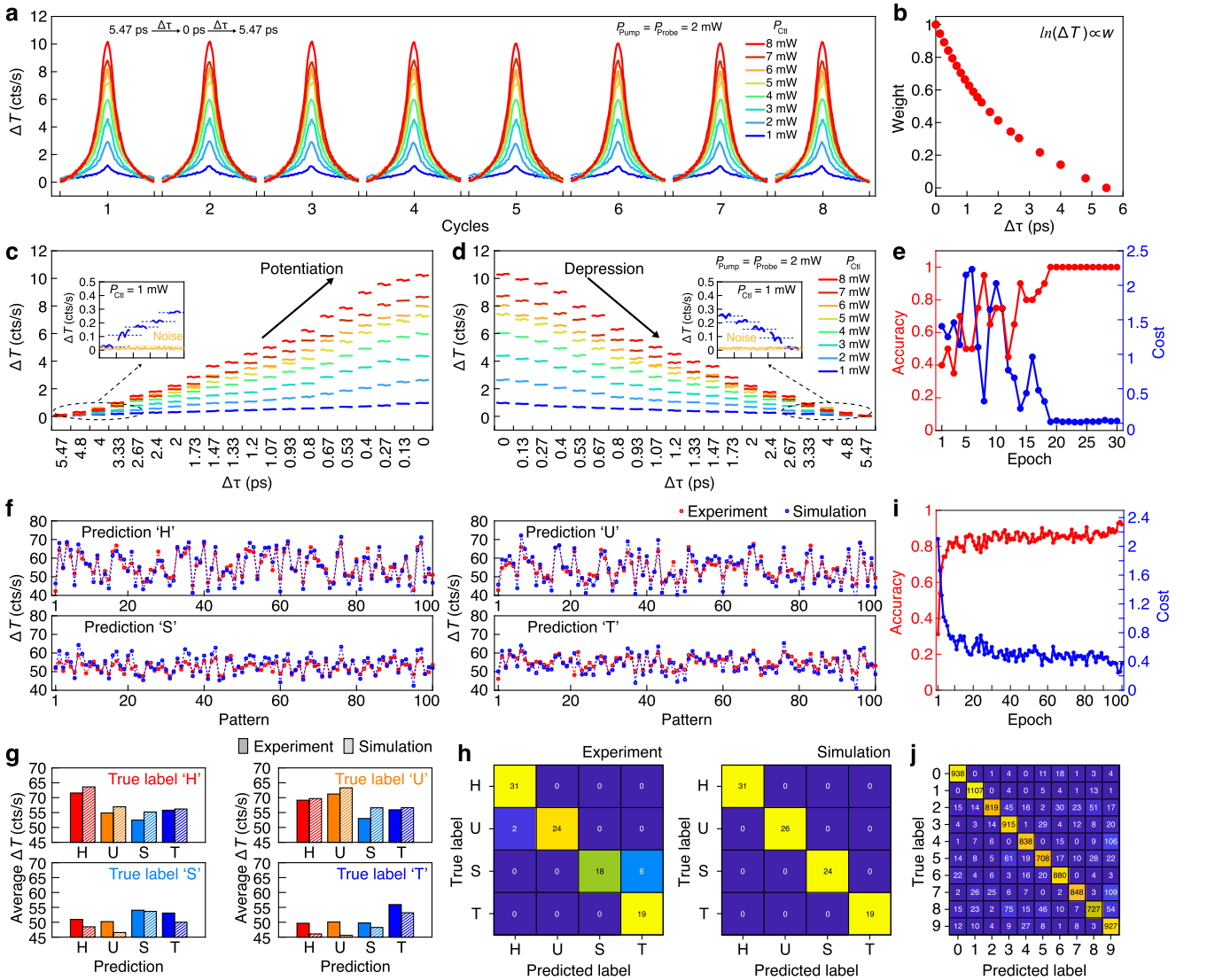

Fig. 3 | Optical ANN in the system. a The stable and symmetric Δ*T* potentiation and depression in 8 cycles under different control powers. In each cycle, the control time delay was continuously decreased from 5.47 to 0 ps in the first half cycle and then increased to 5.47 ps in the second half cycle. **b** The approximately linear relationship between weight and Δτ. Based on the experimental results in (**a**), we encoded Δτ in the range of 0 to 5.47 ps into 20 weight levels. **c**, **d** The multilevel Δ*T* states for the potentiation (**c**) and depression (**d**) process under different control powers. The inset shows the minimum 5 Δ*T* states under P_{Ctl} = 1 mW, indicating adjacent Δ*T* states can be distinguished. **e** The training accuracy and cost in 30 epochs for a single layer ANN to classify a custom training dataset. **f** The experimental and simulated output vector components for the custom testing dataset. **g** The average output vector components for the true label H, U, S, and T in both experiment and simulation. **h** The testing confusion matrix in experiment and simulation for the custom testing dataset. **i** The training accuracy and cost in 100 epochs for a single layer ANN to classify the MNIST training dataset. **j** The testing confusion matrix in simulation for the MNIST testing dataset.

true labels and the predictions (Supplementary Fig. 21b), the training accuracy reached 100%, and the cost decreased to 0.03 (Fig. 3e). To evaluate the recognition performance, predictions for a testing dataset were carried out in both experiments and simulations. Four components in the output vector were summarized in Fig. 3f for the 100 testing patterns, and the average output vector for the true label H, U, S, and T was calculated in Fig. 3g, showing the experiment results were in accordance with the simulation (testing details in Supplementary Fig. 22). The testing accuracy was 100% and 92% in simulation and experiment, respectively. The confusion matrix for the testing results in the experiment and simulation is shown in Fig. 3h. We emphasize that the signal-detecting error exhibited less influence on the experimental testing accuracy since the SNR is large enough to be distinguished by our system. In addition, the off-chip trained weight values were used for the experimental testing. It is normal there exists a deviation from actual on-chip trained weight values, and this is also a reason for the lower testing accuracy in the experiment. It should be noted that the testing procedure was directly carried out based on the detected Δ*T* signals, which included an intrinsic noise level of ~5% and the experimental results can indicate the noise robustness. Besides, we also used the MNIST dataset to further evaluate the applicability of our system for optical neuromorphic computing in simulation (details in Supplementary Note 3). We built a one-layer neural network following the system setup. The training accuracy and cost in 100 epochs are shown in Fig. 3i. After testing, the testing accuracy is 87.07%, and the testing confusion matrix is shown in Fig. 3j. Therefore, we conclude that the system can nicely perform classification within an optical ANN. Based on the experimental setup, the operation energy consumption for one multiplication calculation with the three pulses after Δ*T* accumulation was between ~5.12–9.12 pJ. The performance is

compared to technologies implementing electronic or photonic schemes in Supplementary Table 1.

Digital processing in the system

Through the synergistic encodings of the 2D cells and the excitation pulses, digital processing functions were designed by implementing the ΔT-thickness correlation directly. The correlation ensures a higher SNR at relatively weak illumination, which is beyond the reach of the conventional pump-probe setup. In addition, etching the thickness of 2D flakes is simpler compared to the fabrication of external electronic circuits or photonic structures. As schematically shown in Fig. 1g, the control pulse power and control time delay encode the input bits, and by choosing cells with appropriate thickness as input ports, the output ΔT is summarized externally to encode different functions. The pump time delay was fixed at 0. Here we chose 6 cells in a patterned MoS_2 array with different thicknesses (optical image in Supplementary Fig. 13g). The thickness dependence for each cell was measured at first, under the pump and control time delays of zero and the pump-probe powers of $P_{Pump} = P_{Probe} = 2$ mW. When the control power was increased from 0 mW to 7 mW, the ΔT mappings of each cell were summarized in Fig. 4a, showing the highest SNR of 129.4. The average ΔT under increasing control power was calculated in Fig. 4b, with the gradient $\frac{d\Delta T}{dP_{Ctl}}$ determined by the cell thickness. Four logic functions (AND, OR, NAND, and NOR) were designed with different input encoding schemes and selected cells (Fig. 4d), and the truth table is summarized in Fig. 4c. The output threshold was encoded at 60 cts/s. When the control pulse delay was fixed at 0 ps, and the input bit 1 (bit 0) was encoded by a control pulse power of 3 mW (0 mW), AND (OR) logic can be achieved by using cell 3 and 4 (cell 5 and 6). The output ΔT of cells 3 and 4 are both below the threshold, and the summed ΔT is above the threshold, encoding, AND logic (Fig. 4e). The output ΔT of cells 5 and 6 are both above the threshold, therefore, encoding OR logic (Fig. 4e). When the control pulse power was fixed at 3 mW, and the input bit 1 (bit 0) was encoded by a control pulse delay of 4 ps (0 ps), NAND (NOR) logic can be achieved by using cell 5 and 6 (cell 3 and 4). A larger delay suppresses the output ΔT, therefore, the summed ΔT of cells 5 and 6 is above the threshold when at least one input bit is 0, achieving NAND logic (Fig. 4e). The summed ΔT of cells 3 and 4 is above the threshold when both input bits are 0, achieving NOR logic (Fig. 4e). In addition, based on the same encoding scheme of AND/OR logic, a 4-bit optical DAC was demonstrated by choosing cells 1, 3, 4, and 6 as the input ports, and the summed ΔT output directly converted each 4-bit digit into an analog value, as shown in Fig. 4f.

Discussion

In the free-space optical neuromorphic computing concept, the synergistic transition is the key to function designs, and this requires sample-independent ΔT modulation, which is evaluated in more samples (Supplementary Figs. 23–26). The results also indicate that the computing mechanism and setup are reproducible and reliable. Except for the general results for digital processing and neuromorphic computing, special phenomena at the flake edge and wrinkles are also discussed in Supplementary Figs. 27, 28, and Supplementary Note 4, 5, and a deeper understanding of ΔT responses in MoS_2 will enrich new computing mechanisms.

The free-space optical neuromorphic computing system based on a pure 2D cell array enables computing functions with fast speed, low power consumption, large SNR, and simplified architecture, relieving the restrictions of peripheral electronic circuits, complex nanocavities, and meta-structures. In addition, the synergistic encodings of the 2D cells and the excitation pulses also show higher tunability to program different computing functions. Besides, the excited states in 2D materials can be transferred into electronic signals in situ, assisting the matching between optical and electronic modules for further integration in free space. As a physics-aware neural network platform, free-space optical neuromorphic computing in 2D cell arrays is feasible for analog signal processing in the real world. For example, combined with the excellent photon detection characteristics of 2D materials, this platform can function as a smart sensor to detect and pre-process optical information simultaneously.

For array fabrication, thickness-controllable assembly of 2D films can be realized through bottom-up CVD growth[46] and top-down etching[47]. Recent progress in the robotic assembly also opens a new way to fabricate an array of 2D flakes[48]. In the free-space neuromorphic computing system with controllable cell thicknesses, a large-scale array can be fabricated by patterning large-area 2D film directly. The

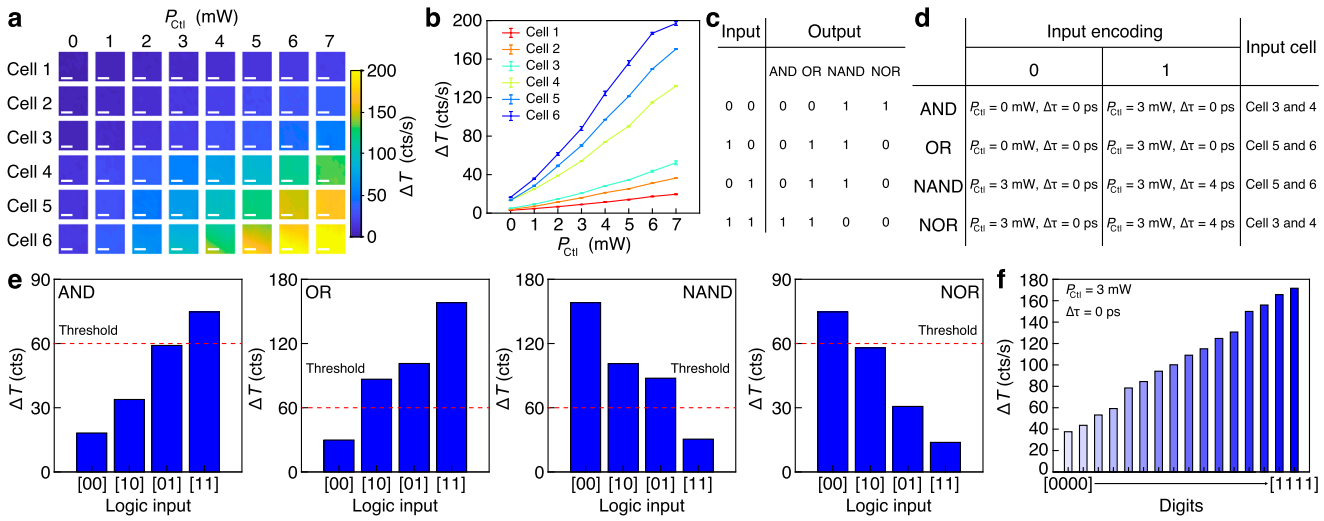

Fig. 4 | Digital processing in the system. a ΔT mappings under different control powers in 6 cells with different thicknesses. Scale bars, 100 nm. **b** The average ΔT under different control powers in 6 cells, the gradient $\frac{d\Delta T}{dP_{Ctl}}$ is determined by the cell thickness. The control pulse delay was 0 for (**a**) and (**b**). **c** Truth table of logic functions. **d** The input signal encodings and selected input cells for different logic functions. For AND and OR logic, the control pulse delay is 0, and input bits are encoded by the control pulse power. For NAND and NOR logic, the control pulse power is 0, and input bits are encoded by the control pulse delay. **e** The output ΔT of the system can achieve different logic functions based on the setup scheme. The red dashed line indicates the defined ΔT threshold. **f** Based on the same encoding scheme of AND and OR logic, the output ΔT in cells 1, 3, 4, and 6 can form an optical 4-bit DAC.

unit size can be narrowed down to the laser spot size[4]. Thus, a higher cell integration density can be achieved. The large-scale high-density 2D flake array can be inserted on the light path of free-space optical systems directly for computing. For functionalization, vector-by-matrix multiplication was carried out by scanning the laser focusing spot over the whole array, this setup compromises the intrinsic fast speed of our concept, but the pump-probe-control process is still fast[4]. The solution is integrating an array of laser sources to modulate and read out signals simultaneously, then we can fully apply the fast speed and high parallelism. Except for multilayer MoS_2, the intrinsic nonlinear responses in many 2D materials are advantageous[1–3], and more efforts are needed to develop new free-space applications in 2D materials. This work focuses on multilayer MoS_2 because the band structure of MoS_2 can match the laser source parameters to realize the best modulation performance. By matching the laser photon energy with the band structure of the selected 2D material, other 2D materials are also applicable in our system setup under the same modulation methodology and the same mechanism. Although integrated devices on waveguides, ring resonators, and fibers are not the topic in this work, the key mechanism of synergistic transitions and the compatibility of 2D materials with these structures will also enrich optical neuromorphic applications. For example, a 2D flake array can be capped above a waveguide array, this structure functions as an array of semiconductor optical amplifiers (SOA). The three pulses as input signals can synergistically modulate the gain at each SOA node, and then computation can be realized with the amplified signals on waveguides. Based on the discussions in data processing and computing applications, the proposed concept can enlighten the integration of optical communication and neuromorphic computing systems.

Methods

1. Sample preparation

Bulk MoS_2 crystals (hq-Graphene) were mechanically exfoliated with Nitto tape, and few-layer samples were attached to a SiO_2/Si 300 nm/500 μm wafer. We selected MoS_2 samples with different thicknesses under an optical microscope. Then the selected samples were attached to PVA (Aladin) films at 45 °C and transferred onto glass slides at 70 °C. The PVA film was removed in deionized water for ~ 30 min. The AFM mapping for sample D was acquired by the Digital TNSTRUments/Veeco Dimension 3000 system under tapping mode. MoS_2 arrays were fabricated by etching exfoliated MoS_2 flakes with Ar/SF_6 (20 sccm/20 sccm) plasma directly. Because the exfoliated MoS_2 flake generally has nonuniform thicknesses in different locations, each cell in the array can have different thickness after patterning. We chose cells with appropriate thicknesses for measurements.

2. Pump-probe-control strategy

The schematic of the pump-probe-control system is shown in Supplementary Fig. 1a. The dual-output femtosecond laser source (InSight DeepSee, Spectra-Physics, Newport) was operated at the frequency of 80 MHz. The pump (1040 nm, 220 fs) pulse was modulated by an acousto-optic modulator (AOM, 1205C-1, Isomet) at a frequency of 2.23 MHz. A time-delay line (TD) after AOM was used to modulate the pump time delay. Both the pump and probe (902 nm, 100 fs) pulses were combined by a dichroic mirror (DMSP1000, Thorlabs) collinearly, and filtered by a polarization-maintaining single-mode optical fiber (PM-SM-F) (LPC-08-1060-6/125-P-3.8-20AC-40-3AF, OZ Optics) within a length of 0.3 m to keep the transverse Gaussian distribution. The control pulse was frequency-doubled by the probe pulse through a beta-barium borate crystal (BBOC) (SHG@900 nm, Union Optic). The control pulse was filtered by a 25 μm pinhole (H) and a quarter-wave plate (QWP). The three pulses with matched Gaussian distribution were controlled and scanned by a galvanometer scanning lens (SL) and tube lens (TL). The objectives (OBs) with high N.A. (N.A. 1.49, UAPON 100XOTIRF, and N. A. 1.2 UPLSAPO 60XW, Olympus) were used to focus the pulses on the MoS_2 samples. The pump pulse signal was transmitted through an oil condenser (OC) (N.A. 1.4, U-ACC, Olympus) and three bandpass filters (BPF) (ZET800/200 bp, ET980sp-2P8, Chroma Technology). Then the transmitted photons of the probe beam were detected by a Si PIN photodiode (S3994-01, Hamamatsu) at the frequency of 2.23 MHz. We used the original data for all discussions, with the signal unit in counts per second (cts/s). A high-speed lock-in amplifier (HF2 LI, Zurich Instruments) was implemented to demodulate the pump-probe signal for sensitive heterodyne detection. The intensity mapping results were acquired by an acquisition card (PCI 6251, National Instrument) with a frequency of 1.25 MHz.

Before carrying out measurements within the MoS_2 array, the whole system was calibrated to synchronize the three incident beams. The illumination time was synchronized by calibrating the initial location of the TDs. Then we can acquire the accurate pump time delay and control time delay. The spatial location of each beam was calibrated by dichroic mirrors, then we can accurately synchronize the coupling of the three beams at the same focusing spot.

Here, we selected the pump and probe wavelengths based on the band structure of MoS_2. The pump and probe photon energies were below the bandgap ($E_{pump}<E_{gap}$, $E_{probe}<E_{gap}$), thus, using a single pulse alone cannot modulate the relative transmittance (ΔT) signals. The $E_{pump}+E_{probe}$ was selected to excite carriers near the band edge, thus the ΔT modulation was relatively low. The control photon energy was much higher than the band gap ($E_{ctl}>E_{gap}$) and near the high-energy nested conduction bands to assist exciting carriers in a high density-of-states (near C peak in Fig. 1b). Therefore, the ΔT signals can be significantly enhanced and modulated due to the synergistic excitation mechanism. Compared to photonic structures, no resonant excitation is required. This computing setup and function encoding schemes are mainly determined by the modulation of absorption in MoS_2, rather than the geometry of the MoS_2 array. Therefore, our system can realize various computing by simply changing the laser pulse setup. This is different from the design of meta-structures or other optical structures, whose function is related to the geometry structure and inverse design is necessary. Here, inverse design is not needed for our scheme.

3. Setups of ΔT modulation

Four modulation methods of control pulse state, pulse power, pulse delay, and flake thickness, are proposed to encode different functions. The experimental setup is schematically explained in Supplementary Fig. 1b. For control pulse state switching, a baffler is used on the light path of the control pulse to control the transmission (on state) or blockage (off state). The pulse power is modulated by controlling the laser source output. The delay time is modulated by mechanically moving the delay line on the light path of the control pulse. For thickness control, the focusing spot of the laser pulse is turned to select a different cell with a different thickness. Here, we used the exfoliated MoS_2 flake with nonuniform thickness for array patterning. Therefore, each cell in the array has a different thickness. Appropriate cells are selected in the same means to encode computing functions.

The key mechanism of synergistic transitions from two-photon absorption (TPA) to synergistic excited states absorption (SESA) is characterized in Fig. 2. Figure 2a shows a large ΔT enhancement by controlling the on/off state of the control pulse. In this setup, the pump and probe power is $P_{Pump}=P_{Probe}=5$ mW, the on-state power of the control pulse is $P_{Ctl}=5.5$ mW, the off-state power of the control pulse is zero, and the pump and control time delay are zero. Figure 2b shows the ΔT vs. pump time delay under different control power to reveal the signal enhancement is mainly attributed to the control pulse. In this setup, the pump and probe power are $P_{Pump}=P_{Probe}=4$ mW, the control power is increased from 0 mW to 8 mW, the control time delay is zero. Figure 2c shows the linearly enhanced ΔT vs. pulse power for three pulses. In this setup, the pump and control time delay is zero, and only one pulse power is

monotonically increased from 0 mW to 7 mW, while the other two pulse powers are fixed at 2 mW. Figure 2d shows the ΔT vs. pulse delay to reveal the prolonged relaxation. In this setup, the pump, probe, and control power are $P_{Pump} = P_{Probe} = P_{Ctl} = 2$ mW. When the control pulse is off, the red dots show the fast decay of ΔT vs. pump time delay, here the control time delay is zero. When the control pulse is on, the blue dots show the prolonged relaxation of ΔT vs. control time delay, here the pump time delay is zero. Under synergistic excitation, we also noted the response started at an earlier negative control time delay of ~ − 534 fs. Figure 2f, g shows the ΔT vs. thickness under a pump time delay of 82 fs. In this setup, the pump, probe, and control power are $P_{Pump} = P_{Probe} = 2$ mW, $P_{Ctl} = 5.5$ mW, and the control time delay is zero.

Data availability

All data are available in the main text or the supplementary materials. The data on which the figures are constructed and from which the conclusions are drawn are available in an online open dataset archived on Zenodo (DOI: 10.5281/zenodo.13984751).

Code availability

All codes are available from the corresponding authors upon request.

References

1. Liu, X., Guo, Q. & Qiu, J. Emerging low-dimensional materials for nonlinear optics and ultrafast photonics. *Adv. Mater.* **29**, 1605886 (2017).
2. Autere, A. et al. Nonlinear optics with 2D layered materials. *Adv. Mater.* **30**, 1705963 (2018).
3. Yu, S., Wu, X., Wang, Y., Guo, X. & Tong, L. 2D materials for optical modulation: challenges and opportunities. *Adv. Mater.* **29**, 1606128 (2017).
4. Bi, Y. et al. Far-field transient absorption nanoscopy with sub-50 nm optical super-resolution. *Optica* **7**, 1402–1407 (2020).
5. Yu, J. et al. Giant nonlinear optical activity in two-dimensional palladium diselenide. *Nat. Commun.* **12**, 1083 (2021).
6. Paik, E. Y. et al. Interlayer exciton laser of extended spatial coherence in atomically thin heterostructures. *Nature* **576**, 80–84 (2019).
7. Sung, J. et al. Room-temperature continuous-wave indirect-bandgap transition lasing in an ultra-thin WS_2 disk. *Nat. Photon.* **16**, 792–797 (2022).
8. Klimmer, S. et al. All-optical polarization and amplitude modulation of second-harmonic generation in atomically thin semiconductors. *Nat. Photon.* **15**, 837–842 (2021).
9. Wang, Y. et al. Direct electrical modulation of second-order optical susceptibility via phase transitions. *Nat. Electron.* **4**, 725–730 (2021).
10. Calafell, I. A. et al. Giant enhancement of third-harmonic generation in graphene-metal heterostructures. *Nat. Nanotechnol.* **16**, 318–324 (2021).
11. Hong, H. et al. Giant enhancement of optical nonlinearity in two-dimensional materials by multiphoton-excitation resonance energy transfer from quantum dots. *Nat. Photon.* **15**, 510–515 (2021).
12. Cha, S. et al. Gate-tunable quantum pathways of high harmonic generation in graphene. *Nat. Commun.* **13**, 6630 (2022).
13. Jiang, T. et al. Gate-tunable third-order nonlinear optical response of massless Dirac fermions in graphene. *Nat. Photon.* **12**, 430–436 (2018).
14. You, J. W., Bongu, S. R., Bao, Q. & Panoiu, N. C. Nonlinear optical properties and applications of 2D materials: Theoretical and experimental aspects. *Nanophotonics* **8**, 63–97 (2019).
15. Xia, F., Wang, H., Xiao, D., Dubey, M. & Ramasubramaniam, A. Two-dimensional material nanophotonics. *Nat. Photon.* **8**, 899–907 (2014).
16. Shastri, B. J. et al. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photon.* **15**, 102–114 (2021).
17. Genty, G. et al. Machine learning and applications in ultrafast photonics. *Nat. Photon.* **15**, 91–101 (2021).
18. Tong, L. et al. 2D materials–based homogeneous transistor-memory architecture for neuromorphic hardware. *Science* **373**, 1353–1358 (2021).
19. Lee, S., Peng, R., Wu, C. & Li, M. Programmable black phosphorus image sensor for broadband optoelectronic edge computing. *Nat. Commun.* **13**, 1485 (2022).
20. Mennel, L. et al. Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020).
21. Jang, H. et al. An atomically thin optoelectronic machine vision processor. *Adv. Mater.* **32**, 2002431 (2020).
22. Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
23. Wright, L. G. et al. Deep physical neural networks trained with backpropagation. *Nature* **601**, 549–555 (2022).
24. Wang, T. et al. An optical neural network using less than 1 photon per multiplication. *Nat. Commun.* **13**, 123 (2022).
25. Flöry, N. et al. Waveguide-integrated van der Waals heterostructure photodetector at telecom wavelengths with high speed and high responsivity. *Nat. Nanotechnol.* **15**, 118–124 (2020).
26. Ono, M. et al. Ultrafast and energy-efficient all-optical switching with graphene-loaded deep-subwavelength plasmonic waveguides. *Nat. Photon.* **14**, 37–43 (2020).
27. Wang, Y. et al. Ultrahigh-speed graphene-based optical coherent receiver. *Nat. Commun.* **12**, 5076 (2021).
28. Datta, I. et al. Low-loss composite photonic platform based on 2D semiconductor monolayers. *Nat. Photon.* **14**, 256–262 (2020).
29. He, J. et al. Low-loss integrated nanophotonic circuits with layered semiconductor materials. *Nano Lett.* **21**, 2709–2718 (2021).
30. Zuo, Y. et al. Optical fibres with embedded two-dimensional materials for ultrahigh nonlinearity. *Nat. Nanotechnol.* **15**, 987–991 (2020).
31. Teğin, U., Yıldırım, M., Oğuz, I., Moser, C. & Psaltis, D. Scalable optical learning operator. *Nat. Comput. Sci.* **1**, 542–549 (2021).
32. Mengu, D. et al. At the intersection of optics and deep learning: statistical inference, computing, and inverse design. *Adv. Opt. Photon.* **14**, 209–290 (2022).
33. Cordaro, A. et al. Solving integral equations in free space with inverse-designed ultrathin optical metagratings. *Nat. Nanotechnol.* **18**, 365–372 (2023).
34. Lee, Y. et al. Boosting quantum yields in two-dimensional semiconductors via proximal metal plates. *Nat. Commun.* **12**, 7095 (2021).
35. Liu, X. et al. Nonlinear valley phonon scattering under the strong coupling regime. *Nat. Mater.* **20**, 1210–1215 (2021).
36. Löchner, F. J. et al. Controlling second-harmonic diffraction by nano-patterning MoS_2 monolayers. *Opt. Express.* **27**, 35475–35484 (2019).
37. van de Groep, J. et al. Exciton resonance tuning of an atomically thin lens. *Nat. Photon.* **14**, 426–430 (2020).
38. Chen, M. K., Liu, X., Sun, Y. & Tsai, D. P. Artificial intelligence in meta-optics. *Chem. Rev.* **122**, 15356–15413 (2022).
39. Das, S., Wang, Y., Dai, Y., Li, S. & Sun, Z. Ultrafast transient sub-bandgap absorption of monolayer MoS_2. *Light Sci. Appl.* **10**, 27 (2021).
40. Wang, L. et al. Slow cooling and efficient extraction of C-exciton hot carriers in MoS_2 monolayer. *Nat. Commun.* **8**, 13906 (2017).
41. Wang, Y., Nie, Z. & Wang, F. Modulation of photocarrier relaxation dynamics in two-dimensional semiconductors. *Light Sci. Appl.* **9**, 192 (2020).
42. Mak, K. F. et al. Tightly bound trions in monolayer MoS_2. *Nat. Mater.* **12**, 207–211 (2013).

43. Versteegh, M. A. M., Kuis, T., Stoof, H. T. C. & Dijkhuis, J. I. Ultrafast screening and carrier dynamics in ZnO: Theory and experiment. *Phys. Rev. B* **84**, 035207 (2011).
44. Latini, S., Olsen, T. & Thygesen, K. S. Excitons in van der Waals heterostructures: the important role of dielectric screening. *Phys. Rev. B* **92**, 245123 (2015).
45. Yang, Z. et al. MXene-based broadband ultrafast nonlinear activator for optical computing. *Adv. Optical Mater.* **10**, 2200714 (2022).
46. Wang, X. et al. Controlled synthesis of highly crystalline MoS_2 flakes by chemical vapor deposition. *J. Am. Chem. Soc.* **135**, 5304–5307 (2013).
47. Lin, T. et al. Controlled layer-by-layer etching of MoS_2. *ACS Appl. Mater. Interfaces* **7**, 15892–15897 (2015).
48. Mannix, A. J. et al. Robotic four-dimensional pixel assembly of van der Waals solids. *Nat. Nanotechnol.* **17**, 361–366 (2022).

Acknowledgements

This work was supported by the National Key Research and Development Plan of China (grants 2019YFB2205100, 2021YFB3601200), the National Natural Science Foundation of China (grants 62222404, 61974050, 61927817, 61931010, 62075076), Innovation Fund of the Wuhan National Laboratory for Optoelectronics, RGC Postdoctoral Fellowship Scheme of Research Grants Council of Hong Kong (PDFS2223-4S06), Open Fund of State Key Laboratory of Infrared Physics, China Postdoctoral Science Foundation (BX20220123, 2022M720051). W.B. and K.P. acknowledge support from the National Science Foundation (award no. DMR-2143041).

Author contributions

L.T., Y.B., Y.W., X.M., P.W., and L.Y. conceived the idea and designed the experiments. L.T., Y.B., and Y.W. carried out exfoliated sample preparation. L.T., Y.B., W.J., and X.H. carried out array fabrication. L.T., Y.B., and W.J. performed experiments. Y.B. and P.W. carried out the construction of the imaging system. X.H., Z.P., and Y.B. conducted AFM measurements. Z.L., R.L., and S.C. conducted digital processing measurements. L.T., L.X., W.S., and H.Y. conducted optical ANN simulations. L.T. and W.J. conducted optical ANN testing experiments. D.J. provided critical suggestions for ANN. L.T. and Y.B. analyzed the TPA and SESA mechanisms. X.Z., P.W., and L.Y. helped with mechanism discussions. Y.B., X.H., and Z.P. helped with crystal edge response analyses. Y.B., Z.L., R.L., and S.C. helped with wrinkle response analyses. Y.W. and X.Z. helped with optical intensity modulator analyses. L.T. and Y.B. carried out the data analyses. K.P. and W.B. provided valuable insight and critical suggestions. L.T., W.B., P.W., and L.Y. wrote the initial manuscript. X.Y., H.S., K.X., Q.H., M.T., J.X., and J.M. provided useful comments during the manuscript preparation. All authors participated in revising the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-54776-z.

Correspondence and requests for materials should be addressed to Wei Bao, Xiangshui Miao, Ping Wang or Lei Ye.

Peer review information *Nature Communications* thanks Sunkook Kim, who co-reviewed with Muhammad Naqi, Ugur Tegin, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]School of Integrated Circuits, Huazhong University of Science and Technology, Wuhan, Hubei, China. [2]Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, Hubei, China. [3]Department of Electronic Engineering, Materials Science and Technology Research Center, The Chinese University of Hong Kong, Hong Kong, China. [4]Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, Hubei, China. [5]Anhui Province Key Laboratory of Measuring Theory and Precision Instrument, School of Instrument Science and Opto-Electronics Engineering, Hefei University of Technology, Hefei, China. [6]Department of Materials Science and Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA. [7]State Key Laboratory of Infrared Physics, Shanghai Institute of Technical Physics Chinese Academy of Sciences, Shanghai, China. [8]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA. [9]These authors contributed equally: Lei Tong, Yali Bi, Yilun Wang. ✉e-mail: baow2@rpi.edu; miaoxs@hust.edu.cn; p_wang@hust.edu.cn; leiye@hust.edu.cn