# Optical Neuromorphic Processor at 11 TeraOPs/s based on Kerr Soliton Crystal Micro-combs

Mengxi Tan,[1] Xingyuan Xu,[2] Jiayang Wu,[1] Andreas Boes,[3] Bill Corcoran,[2] Thach G. Nguyen,[3] Sai T. Chu,[4] Brent E. Little,[5] Damien G. Hicks,[1,6] Roberto Morandotti,[7] Arnan Mitchell,[3] and David J. Moss[1,*]

[1]Optical Sciences Centre, Swinburne University of Technology, Hawthorn, VIC, Australia; [2]Department of Electrical and Computer Systems Engineering, Monash University, Clayton, VIC, Australia; [3]School of Engineering, RMIT University, Melbourne, VIC, Australia; [4]Department of Physics, City University of Hong Kong, Hong Kong, China. [5]Xi'an Institute of Optics and Precision Mechanics Precision Mechanics of CAS, Xi'an, China. [6]Bioinformatics Division, Walter & Eliza Hall Institute of Medical Research, Parkville, VIC, Australia. [7]INRS-Énergie, Matériaux et Télécommunications, 1650 Boulevard Lionel-Boulet, Varennes, Québec, Canada. *dmoss@swin.edu.au.

**Abstract:** We demonstrate a universal optical vector convolutional accelerator operating at 11 Tera-OPS, generating convolutions of images of 250,000 pixels with 8-bit resolution for 10 kernels simultaneously. We use the same hardware to form a deep optical CNN with ten output neurons, achieving successful recognition of full 10 digits with 88% accuracy. Our approach is scalable and trainable for applications to unmanned vehicle and real-time video recognition.

Convolutional neural networks (CNNs) can abstract the representations of input data in their raw form, and predict their properties with both unprecedented accuracy and greatly reduced parametric complexity [1, 2], and have been applied to computer vision, natural language processing and other areas [3]. Optical neural networks (ONNs) [4-10] are promising next-generation neuromorphic computers for ultra-high computing speeds enabled by the >10 THz wide telecom band. They avoid limitations of reading and storing data, known as the von Neumann bottleneck [11]. While Significant progress has been made in highly parallel, high-speed and trainable ONNs, processing large-scale data, needed for computer vision tasks, remains challenging because ONNs are fully connected structures with their input scale determined by hardware parallelism. This leads to tradeoffs in network scale and footprint. ONNs have not achieved the extreme computing speeds that analog photonics are capable of.

Here, by interleaving wavelength, temporal, and spatial dimensions with an integrated Kerr microcomb, we demonstrate an optical convolution accelerator with a vector computing speed of 11.322 Tera-OPS/s (TOPS) and use it to process 250,000 pixel images with 10 convolution kernels at 3.8 TOPs [7]. Our convolution accelerator (CA) is dynamically reconfigurable and scalable, serving as both a CA front-end as well as an optically deep CNN with fully connected neurons, with the same hardware. We use the deep CNN to achieve recognition of the full ten handwritten digits (0-9) with an accuracy of 88%. Our accelerator is stand alone and universal — fully compatible with electrical and optical interfaces, as a universal ultrahigh bandwidth data compressing front end for neuromorphic hardware bringing massive-data machine learning for real-time ultrahigh bandwidth data within reach.

Figure 1 shows the photonic matrix CA. The input data vector **X** is encoded as the intensity of temporal symbols in a serial electrical waveform at a symbol rate $1/\tau$ (baud), where $\tau$ is the symbol period. The convolution kernel is represented by a weight vector **W** of length $R$ that is encoded in the optical power of the microcomb lines through spectral shaping performed by a Waveshaper. The temporal waveform **X** is multi-cast onto the kernel wavelength channels via electro-optical modulation, generating replicas weighted by **W**. Next the optical waveform is transmitted through a dispersive delay with a delay step (between wavelengths) equal to the symbol duration of **X**, achieving time and wavelength interleaving. Finally, the delayed and weighted replicas are summed via high speed photodetection so that each time slot yields a convolution between **X** and **W** for a given convolution window, or receptive field. The convolution window effectively slides at the modulation speed matching the baud rate of **X**. Each output symbol is the result of $R$ multiply-and-accumulate operations, with the computing speed given by $2R/\tau$ OPS. Since the speed of this process scales with both the baud rate and number of wavelengths, it can be dramatically boosted into the TOP regime with the massively parallel wavelength

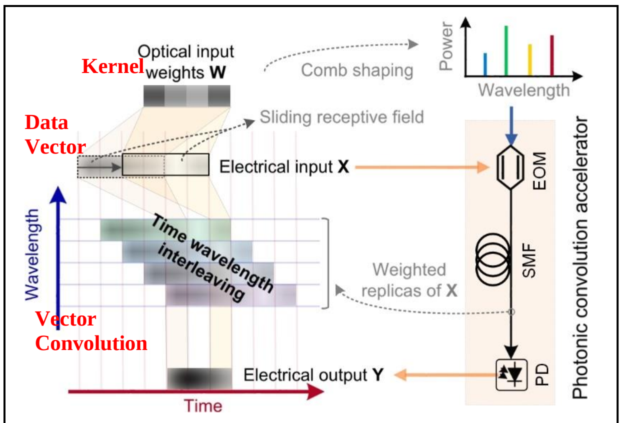


**Figure 1 | Operation principle of the Tera-FLOPS photonic convolution accelerator**. EOM: electro-optical Mach-Zehnder modulator. SMF: standard single mode fibre for telecommunications. PD: photodetector.

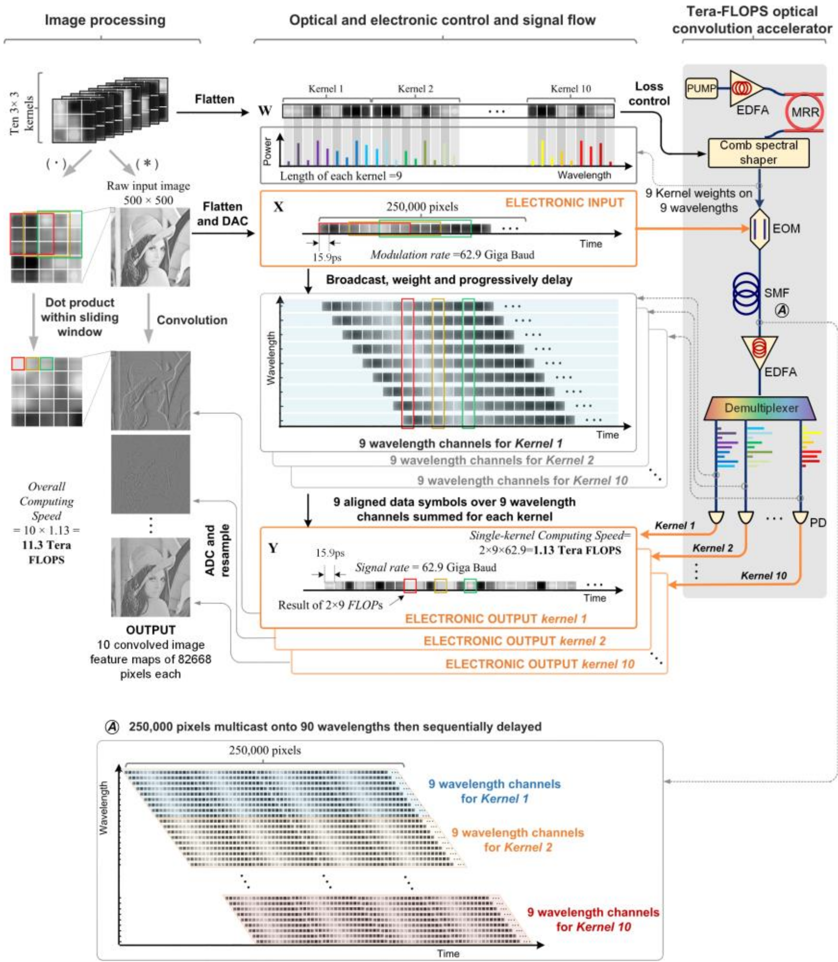


**Figure 2 | Image processing.** The optical and electronic control and signal flow (middle panel), and the corresponding processing flow of the raw input image (left panel). CW pump: continues-wave pump laser. PC: polarization controller. EDFA: erbium doped fibre amplifier. MRR: micro-ring resonator. EOM: electro-optical Mach-Zehnder modulator. SMF: standard single mode fibre for telecommunications. PD: photodetector.

channels of the microcomb source. The length of the input data **X** is unlimited so that the convolution accelerator can process data with an arbitrarily large scale—the only limitation being the capability of the external electronics and the number of wavelengths (for speed).

For matrix operations, the matrix is flattened into a vector, determining the sliding convolution window's stride and effective matrix computing speed. Our flattening method resulted in an effective reduction (overhead) in matrix computing speed that scales inversely with the kernel size (a 3x3 kernel yields an overhead 1/3). This can be avoided to produce convolutions with a symmetric stride and no speed overhead, but is not necessary for most applications.

Figure 2 shows the experimental setup for the full matrix convolutional accelerator that we use to process a classic 500×500 face image. The system performs 10 simultaneous convolutions with ten 3×3 kernels to achieve distinctive image processing functions. The weight matrices for all kernels were flattened into a composite kernel vector **W** containing all 90 weights (10 kernels with 3x3=9 weights each), which were then encoded onto the optical power of 90 microcomb lines by an optical spectral shaper, each kernel occupying a band of 9 wavelengths. The wavelengths were supplied by a coherent soliton crystal microcomb in a micro-ring resonator [7, 12-16], radius = 592 µm, FSR spacing ~ 48.9 GHz, with a bandwidth of ~ 36 nm for 90 wavelengths over the C-band.

The raw 500×500 input face image was flattened electronically into a vector **X** and encoded as the intensities of 250,000 temporal symbols with a resolution of 8 bits/symbol (limited by the electronic arbitrary waveform generator (AWG)), to form the electrical input waveform via a high-speed electrical DAC converter, at a data rate of 62.9 Giga Baud (time-slot τ =15.9 ps) (Fig. 4b). The waveform duration was 3.975µs for each image corresponding to a processing rate for all ten kernels of > 1/3.975µs, or 0.25 million ultra-large-scale images per second. The input waveform **X** was then multi-cast onto the 90 shaped comb lines via electro-optical modulation, yielding replicas weighted by the kernel vector **W**. Following this, the waveform was then transmitted through a ~2.2 km length of standard single mode fibre having a dispersion of ~17ps/nm/km. The fibre length was carefully chosen to induce a relative temporal shift in the weighted replicas with a progressive delay step of 15.9 ps between adjacent wavelength channels. This delay exactly matched the duration of each input data symbol τ, which effectively resulted in time and wavelength interleaving for all ten kernels.

The 90 wavelengths were then de-multiplexed into 10 sub-bands of 9 wavelengths, each sub-band corresponding to a kernel, and separately detected by 10 high speed photodetectors. The detection process effectively summed the aligned symbols of the replicas (the electrical output waveform of one of the kernels (*kernel 4*) is shown in [7] Fig. 4c). The 10 electrical waveforms were converted into digital signals via ADCs and resampled so that each time slot of each of the waveforms corresponded to the dot product between one of the convolutional kernel matrices and the input image within a sliding window (i.e., receptive field). This effectively achieved convolutions between the 10 kernels and the raw input image. The resulting waveforms thus yielded the 10 feature maps containing the extracted hierarchical features of the input image (Fig. 4d and Supplementary in [7]).

The convolutional vector accelerator makes full use of time, wavelength, and spatial multiplexing, where the convolution window effectively slides across the input vector **X** at a speed equal to the modulation baud-rate — 62.9 Giga Symbols/s. Each output symbol is the result of 9 (the length of each kernel) multiply-and-accumulate operations, thus the core vector computing speed (i.e., *throughput*) of each kernel is 2×9×62.9 = 1.13 TOPS. For ten kernels computed in parallel the overall computing speed of the CA is therefore 1.13×10 =11.3 TOPS, or 11.321×8=90.568 tera-bits per second (Tb/s). This speed is > 500 x higher than the fastest speed of ONNs reported to date. The convolutional accelerator is fully and dynamically reconfigurable and scalable with the same hardware system. We used the accelerator to sequentially form both a frontend convolution processor as well as a fully connected layer, together yielding an optical deep CNN. We applied the CNN full 10 (0-9) handwritten digit image recognition.

For the optical CNN the fully connected layer had ten neurons, each corresponding to one of the ten categories of handwritten digits from 0 to 9, with the synaptic weights represented by a 72×10 weight matrix $\mathbf{W}_{FC}^{(l)}$ (ie., ten 72×1 column vectors) for the $l$th neuron ($l \in [1, 10]$) – with the number of comb lines (72) matching the length of the flattened feature map vector $\mathbf{X}_{FC}$. The shaped optical spectrum at the $l$th port had an optical power distribution proportional to the weight vector $\mathbf{W}_{FC}^{(l)}$, thus serving as the equivalent optical input of the $l$th neuron. After being multicast onto the 72 wavelengths and progressively delayed, the optical signal was weighted and demultiplexed with a single Waveshaper into 10 spatial output ports — each corresponding to a neuron. Finally, the different node/neuron outputs were obtained by sampling the 73$^{rd}$ symbol of the convolved results. The final output of the optical CNN was represented by the intensities of the output neurons, where the highest intensity for each tested image corresponded to the predicted category. Supervised network training was performed offline electronically.

We tested 500 8-bit 30 × 30 resolution images of the handwritten digit dataset with the deep optical CNN, achieving an accuracy of 88%. The computing speed of the VCA front end of the deep optical CNN was 2×75×11.9 =1.785 TOPS, or 14.3 Terabits/s. The computing speed of the fully connected layer was 119.8 GigaOPS. The waveform duration was 30×30×84ps=75.6ns for each image, and so the convolutional layer processed images at the rate of 1/75.6ns = 13.2 million handwritten digit images per second. Handwritten digit recognition, although a benchmark test in digital hardware, is still (for full 10 digit (0 - 9) recognition) beyond analog reconfigurable ONNs. Digit recognition requires a large number of parallel paths for fully-connected networks (e.g., a hidden layer with 10 neurons requires 9000 physical paths), which poses a huge challenge for current nanofabrication techniques. Our CNN represents the first reconfigurable and integrable ONN capable not only of performing high level complex tasks such as full handwritten digit recognition, but at ultrahigh TOP speeds. The optical *latency* of 0.11 μs of the dispersive fibre spool did not affect the operational speed, and can be eliminated (< 200 ps) with integrated highly dispersive devices such as photonic crystals or chirped Bragg gratings [17].

In summary, we demonstrate an optical convolutional accelerator operating at 11.3 TOPS and use it to perform convolutions on face images with 250,000 8-bit resolution pixels. We then form an optical deep learning CNN to achieve recognition of handwritten digit images. Our network is capable of recognizing and processing large-scale data and images at ultra-high computing speeds for real-time massive-data machine learning tasks [18,19].

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
2. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. Commun. ACM **60**, 84-90 (2017).
3. Yao, P. et al. Fully hardware-implemented memristor convolutional neural network. *Nature* **577**, 641–646 (2020).
4. Shen, Y. et al., "Deep learning with coherent nanophotonic circuits", *Nature Photonics* **11**, 441 (2018).
5. Lin, X. et al., "All-optical machine learning using diffractive deep neural networks", *Science* **361**, 1004-1008 (2018).
6. Feldmann, J., et al., "All-optical spiking neurosynaptic networks with self-learning capabilities", *Nature* **569**, 208-214 (2019).
7. X. Xu, et al., "11 TOPs photonic convolutional accelerator for optical neural networks", *Nature* **589** 44-51 (2021).
8. X.Xu et al., "Photonic perceptron based on a Kerr microcomb for scalable high speed optical neural networks", Laser & Phot. Rev. **14** (8) 2000070 (2020).
9. G. Wetzstein et al., "Inference in artificial intelligence with deep optics and photonics", *Nature* **588** (7836), 39-47 (2020).
10. J. Feldmann, *et al*., "Parallel convolutional processing using an integrated photonic tensor core," *Nature* **589** (7840), 52-58 (2021).
11. Ambrogio, S. *et al.* Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **558**, 60 (2018).
12. M. Tan et al., "RF and microwave photonic temporal signal processing with Kerr micro-combs", Advances in Physics: X **6** (1), 1838946 (2021).
13. D.J. Moss, et al.,"New CMOS-compatible platforms based on silicon nitride and Hydex for nonlinear optics", *Nat. Photon.***7**, (8), 597 (2013).
14. B. Corcoran *et al.*, "Ultra-dense optical data transmission over standard fiber with a single chip source", *Nature Comm.* **11** Article: 2568 (2020).
15. J. Wu et al., "RF Photonics: An Optical Microcombs' Perspective," IEEE J. Sel. Top. Quantum Electron. **24**, no. 4, 6101020, Jul-Aug. (2018).
16. X Xu et al., "Microcomb-based photonic RF signal processing", IEEE Photonics Technology Letters **31,** (23), 1854-1857 (2020).
17. Sahin, E. et al., "Large scalable dispersion engineering using cladding-modulated Bragg gratings on a silicon chip", Applied Physics Letters **110**, 161113 (2017).
18. Esteva, A. et al., "Dermatologist-level classification of skin cancer with deep neural networks", *Nature* **542**, 115-118 (2017).
19. Capper, D. et al., "DNA methylation-based classification of central nervous system tumours", *Nature* **555**, 469 (2017).