A single photonic perceptron based on a soliton crystal microcomb for scalable, high speed, optical neural networks

SUBMISSION DATE / POSTED DATE

03-03-2020 / 11-08-2020

CITATION

Moss, David; xu, xingyuan; tan, mengxi; Wu, Jiayang; Morandotti, Roberto (2020): A single photonic perceptron based on a soliton crystal microcomb for scalable, high speed, optical neural networks. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.11925225.v3

DOI

10.36227/techrxiv.11925225.v3

Supplementary Materials for "**Photonic perceptron based on a Kerr microcomb for high-speed, scalable, optical neural networks**"

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

[1]Optical Sciences Centre, Swinburne University of Technology, Hawthorn, VIC 3122, Australia

[2]Department of Electrical and Computer Systems Engineering, Monash University, Clayton, 3800 VIC, Australia

[3]School of Engineering, RMIT University, Melbourne, VIC 3001, Australia

[4]Department of Physics and Material Science, City University of Hong Kong, Tat Chee Avenue, Hong Kong, China.

[5]Xi'an Institute of Optics and Precision Mechanics Precision Mechanics of CAS, Xi'an, China.

[6]INRS-Énergie, Matériaux et Télécommunications, 1650 Boulevard Lionel-Boulet, Varennes, Québec, J3X 1S2, Canada.

[7]Bioinformatics Division, Walter & Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia

*Correspondence to: dmoss@swin.edu.au

Soliton crystal microcomb generation

Critical for micro-combs is the ability to phase-lock the frequency comb modes, and many oscillation states have been explored to achieve this, including feedback-stabilized Kerr combs [s1], dark solitons [s2] and dissipative Kerr solitons (DKS) [s3]. While many of these approaches have enabled breakthroughs [27], all (particularly DKS and dark soliton states described by the Lugiato-Lefever equation [27]) require sophisticated feedback systems and complex dynamic pumping schemes to initiate and sustain [s1, s3]. Here, we employ a new and powerful class of micro-comb based on what have been termed "soliton crystals" that are generated from a fundamentally different process and which offer significantly improved simplicity compared to DKS states. They are naturally formed in micro-cavities that display the appropriate form of mode crossings, without the need for the complex dynamic tuning mechanisms that DKS require. They were termed 'soliton crystals', due to their crystal-like profile in the angular domain in micro-ring resonators [36, 37]. To generate coherent micro-combs, a CW pump laser (Yenista Tunics – 100S-HP) was employed, with the power amplified to 30dBm by an optical amplifier (Pritel PMFA-37) and the wavelength subsequently swept from blue to red. The acquired soliton crystals optical spectra are shown in Fig. S1. We note that when locking the pump wavelength to the resonance of the MRR, the stability of the microcomb can be further enhanced that could even serve as frequency standards [22].

Datasets and pre-training

The datasets we employed was from MNIST (Modified National Institute of Standards and Technology) handwritten digit database [s4] and part of the publicly available Wisconsin Breast Cancer dataset [s5]. The datasets of recognition tasks were first separated as training sets and test sets. The training sets were used for the offline training with the Back Propagation algorithm [s6], performed on an electronic computer using Matlab^{TM}, to acquire pre-trained weights and bias. The test sets were tested with both the ONN and an electronic computer for comparison. We note that, since the number of training sets is sufficiently large compared with the number of synaptic connections, the cross validation process was not employed in this work—and in any case, it could be performed offline before the pre-training.

Experimental setup

During the experiment, the 7×7 gray scale data of the handwritten digit figures were first converted into a one dimension array $\mathbf{X}=[x(1), x(2), \ldots, x(49)]$ by assembling each column head-to-tail. Then a 49-symbol waveform was generated and coded with the intensities at each time slot in proportion to the values of $\mathbf{X}$ at corresponding sequences, thus the input data $\mathbf{X}$ were multiplexed in the time domain. The 49-symbol waveform was generated by an arbitrary waveform generator (Keysight M8195A), which supported a sample rate of 65 Giga-Samples/s and an analog bandwidth of up to 25 GHz. To acquire stepwise waveforms for the input nodes, we used 5 sample points at 59.421642 Giga Samples/s to form each single symbol of the input waveform, which also matched the progressive time delay τ (84 ps) of the dispersive buffer.

The optical power of the 49 microcomb lines was shaped according to the intensity of pre-trained neuron weights $\mathbf{W}=[w(1), w(2), \ldots, w(49)]$. We shaped the comb lines' power with a programmable optical spectral shaper using liquid crystal on silicon techniques (Finisar WaveShaper 4000S), which could dynamically reconfigure the ONN connections within 500 ms with a resolution of 1 GHz. Two stages of programmable optical spectral shapers were employed for a larger loss dynamic range. The first WaveShaper was used to flatten the microcomb, while the second one was used to achieve pre-trained neuron weights. A feedback loop was used to enhance the shaping accuracy, where the comb lines' power after shaping was measured

by an optical spectrum analyser (Anritsu MS9710C) and compared with the pre-trained weights to generate an error signal for the calibration of the WaveShapers' loss characteristics.

Then the waveform was multicast via intensity modulation on the wavelength channels established by shaped comb lines, such that the optical signal at the kth (k=1, 2, …, 49) channel was $w(k)\cdot\mathbf{X}$. Next the optical signals at all wavelength channels passed through 13-km dispersive single mode fibre and obtained a time delay of $(49 - k)\times\tau$ for the kth channel, and τ was measured to be 84 ps. Thus, the optical signals were progressively shifted in the time domain. The optical signal after the single mode fibre was converted to the electrical domain by a photodetector (Finisar VPDV2120), and the waveform was then measured by a high-speed oscilloscope (Keysight DSOZ504A). The sampled output of the photodetector was added to the bias symbol and rescaled in intensity by the reference symbol to extract the recovered output of the ONN and locate the hyper-plane (a trained subspace in the high-dimension space of the input data, which serves as a decision boundary that separates different classes of data).

We note here that the accuracy of the ONN predictions was experimentally limited by the performance of the arbitrary waveform generator, which introduced errors to the symbols' intensities and thus deteriorated the correctness of the matrix multiplication. This can be addressed by using an arbitrary waveform generator with a larger analog bandwidth, or a higher sampling rate. Addressing this issue would result in higher levels of correctness than reported here.

Speed calculation of the scaled ONN

Following our definition of throughput and latency introduced in the manuscript, the overall throughput of the deep ONN is roughly the product of each hidden layer's speed and the number of hidden layers, although we note that rigorous and accurate calculation of the throughput is only possible with specific configurations of the network.

Here is a simple example of calculation (this example is just to show the calculations of throughput and latency, the actual performance in terms of prediction accuracies is not the focus of our discussion here): the input waveform/layer is the same as the demonstrated perceptron (49×1 vector at 11.9 Giga Baud with 8-bit resolution, τ= 84ps), the network has a hidden layer that each has 7 fully connected neurons, and an output layer that has 10 fully connected neurons (to match with the number of categories for digits from 0 to 9). As a result, 343 (49×7) and 70 (7×10) wavelengths would be needed in the hidden and output layer, respectively. This can be achieved by using smaller FSR microcombs such as 25GHz across the wide optical band (the C + L bands already reach >11THz wide).

In the hidden layer, each initial electrical output waveform (right after the photodetection and before the digital signal processing) corresponds to the output of a single neuron and has a duration of $(49\times2-1)\times84\text{ps}=8.148$ ns. Only one time slot of each group of symbols represents the result of matrix multiplication between the input vector and the weight vector that constitutes of 49×2=98 floating point operations. As a result, the throughput of each neuron is given as 98/8.148=12.0275 Giga-FLOPS. Since different neurons are multiplexed in both the spatial and wavelength domain and detected in parallel, the total throughput of the hidden layer would be 12.0275×7=84.1925 Giga-FLOPS.

In the output layer, the generated electrical waveform of each neuron has a duration of $(7\times2-1)\times84\text{ps}=1.092$ ns. Only one time slot of each group of symbols represents the result of matrix multiplication between the input vector (sampled and re-multiplexed waveform from the hidden layer) and the weight vector that constitutes of 7×2=14 floating point operations, thus the throughput would be 14/1.092=12.8205 Giga-FLOPS for each neuron and the total throughput of output layer would be 12.8205×10=128.205 Giga-FLOPS. As such, the total peak throughput of the network would be 84.1925+128.205 =212.3975 Giga-FLOPS. In addition, the latency of the overall network is the sum of each layer's latency, which mainly

comes from the dispersive optical buffer and the electrical sampling and multiplexing module. We assume the latency to be 200 ps for the buffer in integrated forms and to be twice of the waveform duration for the re-sampling unit (2×8.148 ns and 2×1.092 ns for the hidden and output layer, respectively), the total latency of the example network would roughly be 18.68 ns. We note that the latency is just a very rough estimation showing how to calculate or measure the performance of our approach, the practical calculations of the latency are subject to more detailed parameters.

The speed of the network has the potential to reach 10 Tera-FLOPS, determined as follows. With 20 layers, each layer featuring 20 neurons and a modulation rate of 25 Giga baud, the overall throughput should be around 20×20×25=10 tera-FLOPS, according to the discussion in the above section. With 8-bit resolution, the total potential throughput in terms of bit rate could reach 10×8=80 Tbps. We note that other widely used techniques in telecommunications such as polarization multiplexing and coherent modulation formats could also potentially boost the computing speed of the proposed neuron network in this work.

Table S1 shows the performance matrices of state-of-art ONNs. We note that it is difficult to directly compare different kinds of ONNs, since on one hand, there are no universal and specific definitions of ONN's parameters. On the other hand, the operation principles of existing ONNs are quite different and have their unique advantages. As such, here we highlight the decent advances of existing works and focus on the speed parameters, including the *latency* and *throughput*, to reflect our ONN's advantages in this aspect.

Table S1 Performance comparison of state-of-the-art ONNs

Approach\Parameter		Compatibility with digital electronics	Latency	Throughput speed per unit	
				FLOPS	bits/s
Diffraction devices [10]		—	< 10 ns	—	—
Integrated couplers [4]		—	< 0.1 ns	—	—
Reservoir computing [13]		Yes	< 1 μs	17.6 G	—
Spike computing [16]		Yes	< 1 μs	8 G	8 G
Spike computing [17]		—	< 0.1 μs	—	—
This work	Demonstrated single Perceptron	Yes	64 μs	11.9 G	95.2 G
	Designed deep ONN	Yes	>18.68 ns	>10 T	>80 T

"—" denotes the corresponding parameter is either not demonstrated or not indicated in the work.

References of the supplementary

[s1] P. Del'Haye *et al.*, Optical frequency comb generation from a monolithic microresonator. Nature **450**, 1214 (2007).

[s2] X. Xue *et al.*, Mode-locked dark pulse Kerr combs in normal-dispersion microresonators. Nature Photonics **9**, 594–600 (2015).

[s3] T. Herr *et al.*, Temporal solitons in optical microresonators. Nature Photonics **8**, 145–152 (2013).

[s4] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**, 2278-2324 (1998).

[s5] O. L. Mangasarian, W. N. Street, W. H. Wolberg, Breast cancer diagnosis and prognosis via linear programming. *Operations Research* **43**, 570-577 (1995).

[s6] C. M. Bishop, Neural networks for pattern recognition. (Oxford university press, 1995).

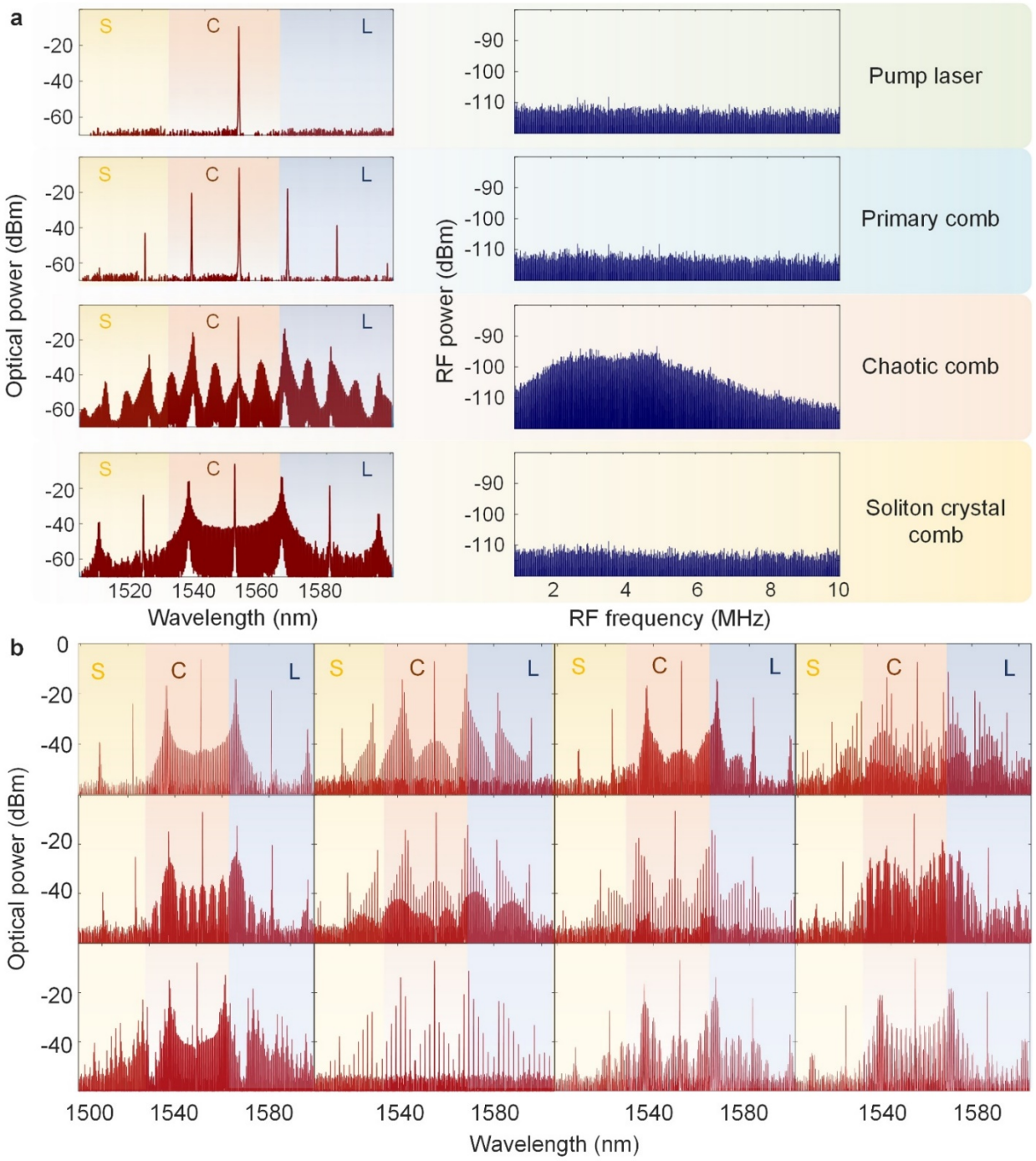

Fig. S1 | a, Different states and measured RF intensity noise of the microcomb. **b,** Measured low intensity noise states.

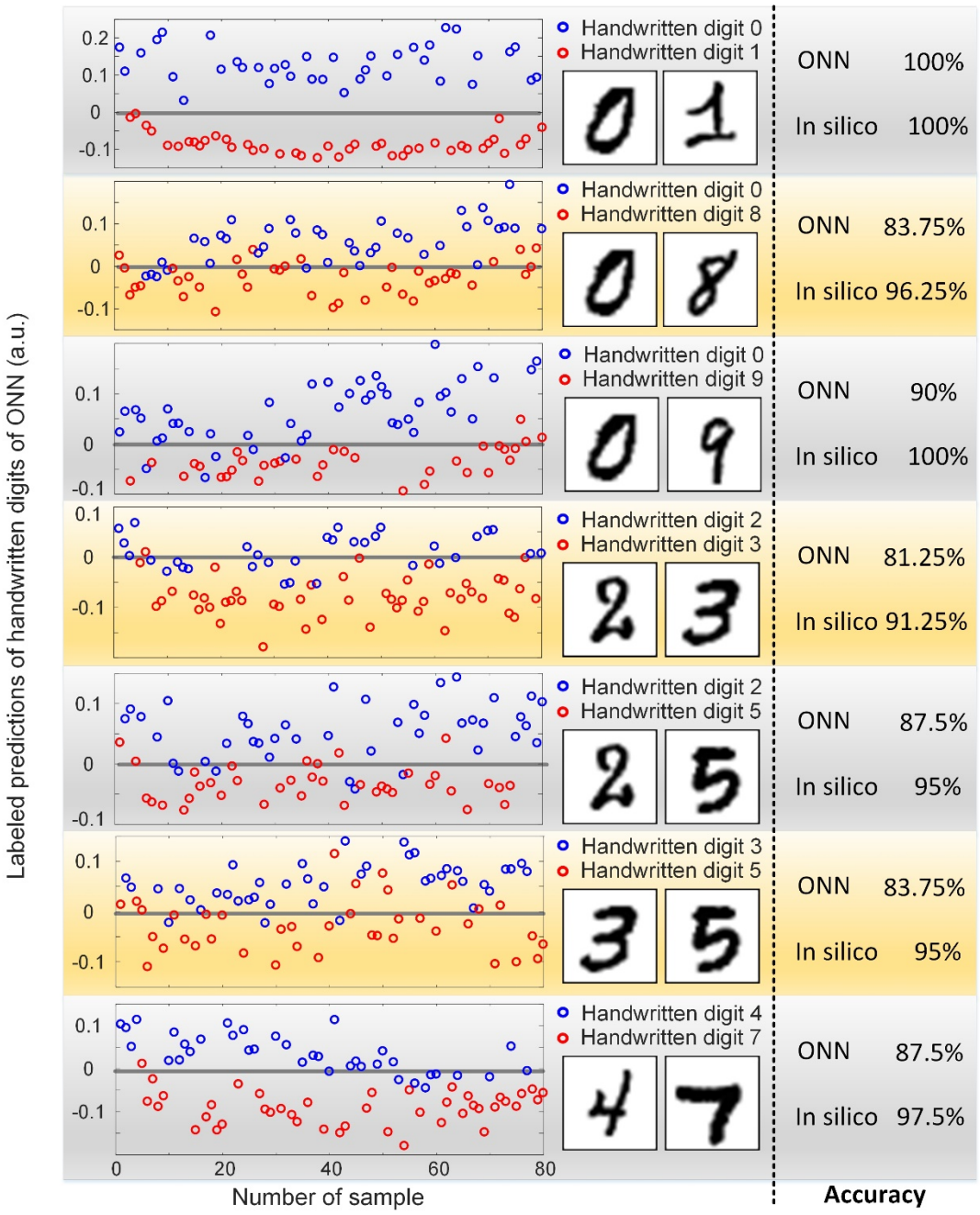

Fig. S2 | ONN predictions of handwritten digits labeled according to their correct answers.

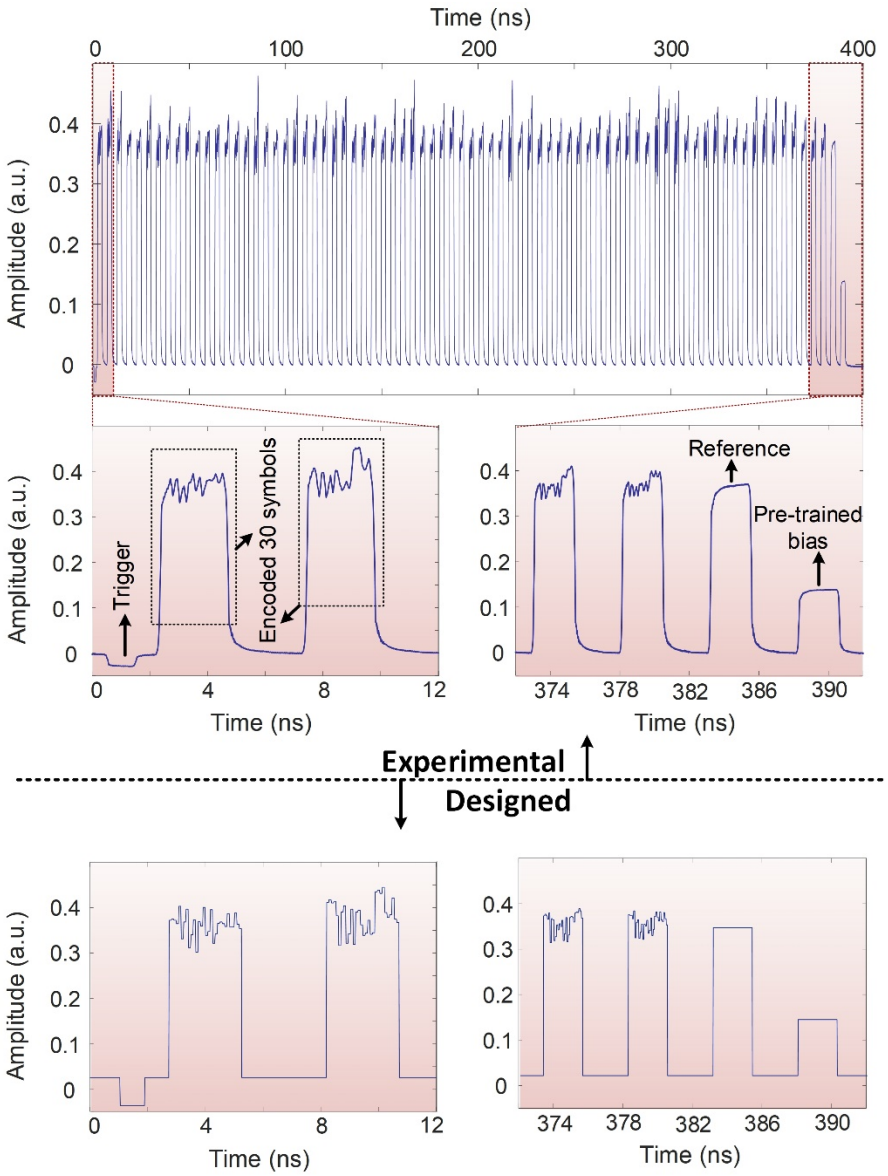

Fig. S3 | Time-domain multiplexed input layer of cancer diagnosis test. Generated 11.9 Giga-baud data stream of the encoded 75 sets of features showing 30-symbol encoded data for each set and 3 symbols padded for post measurement, including a trigger symbol to trigger the oscilloscope, a reference symbol to calibrate the reference level, and a bias symbol encoded with the pre-trained bias to locate the decision boundary.

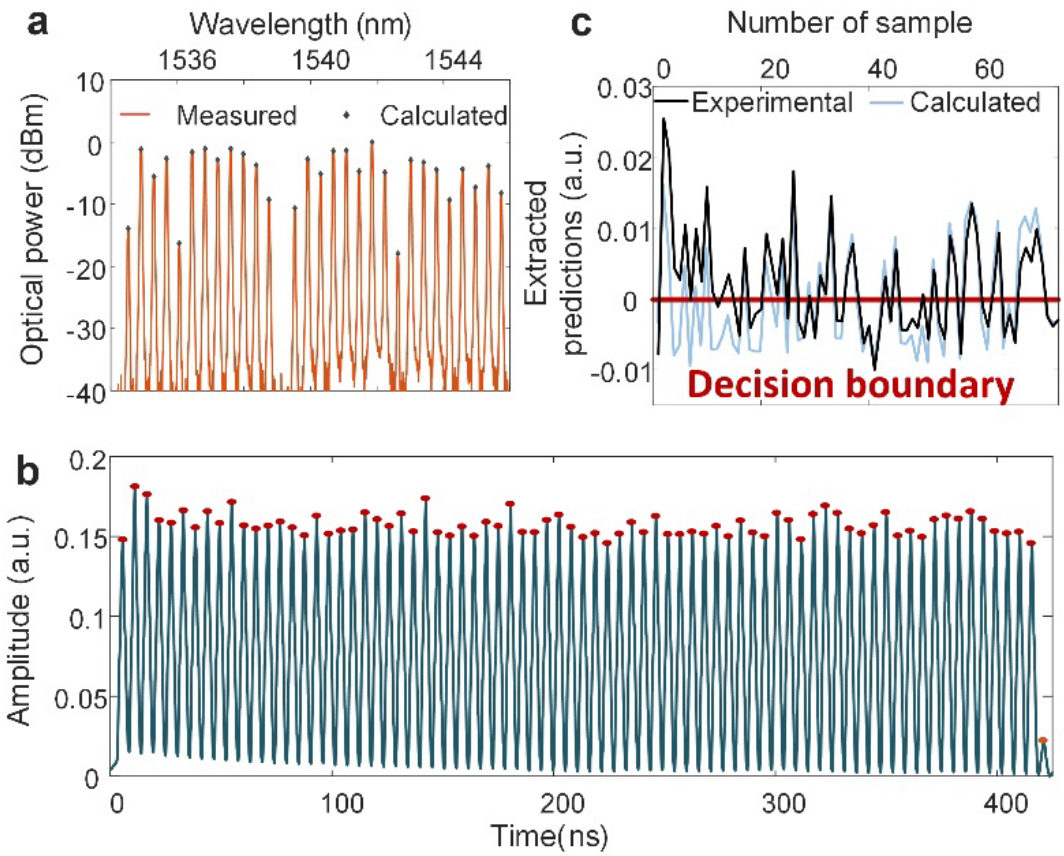

Fig. S4 | Experimental recognition of cancer diagnosis. a, Optical spectrum of the shaped (soliton crystal) micro-comb measured by an optical spectrum analyser. **b,** Measured and sampled output waveform from the photodetector. **c,** Recovered ONN predictions X×W+b acquired by rescaling the sampled results via the reference symbol, and the hyper-plane X×W+b=0 (black line).