Supplemental document accompanying submission to *Optics Express*

**Title:** High-speed serial deep learning through temporal optical neurons

**Authors:** Ming Li, Zhixing Lin, Shuqian Sun, Jose Azana, Wei Li

**Submitted:** 2/24/2021 8:33:37 PM
High-speed serial deep learning through temporal optical neurons: supplemental document

ZHIXING LIN,1,2,3 SHUQIAN SUN1,2,3 JOSÉ AZAÑA,4 WEI LI1,2,3 AND MING LI,1,2,3,*

1State Key Laboratory on Integrated Optoelectronics, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China
2School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
3Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100190, China
4Institut National de la Recherche Scientifique – Centre Énergie, Matériaux et Télécommunications (INRS-EMT), Montreal, Quebec, Canada.
*ml@semi.ac.cn

1. Optical field analysis of serial ONN

1.1 Forward propagation

We begin this analysis at the input part of the serial ONN. The light source we use in this ONN strategy is a pulse train emitted by an actively mode-locked laser whose complex output field function can be written as:

\[ E_{MM}(t) = \sum_i E_p(t - IT_i) e^{-j\phi_0}, \]  (1)

where \( T \) is the repetition period of the pulse train, \( \phi_0 \) is the constant initial phase over time, and \( E_p(t) \) is the expression of a single pulse:

\[ E_p(t) = \exp\left(-\frac{(t-T/2)^2}{2c_{lw}^2}\right), \]  (2)

where \( c_{lw} \) is the parameter controlling the temporal linewidth of the pulse, such that \( E_p(t) = 0 \) when \(|t| > T\). The object or waveform to be processed by the neural network \( S_{obj}(t) \) is sampled by this pulse train. This way, we obtain the input waveform to the serial ONN:

\[ E_0(t) = E_{in}(t) = S_{obj}(t) \cdot \sum_i E_p(t - iT_i) e^{-j\phi_0}, \]  (3)

For each layer of the serial ONN, the layer under consideration receives the output of the last layer \( E_{l-1} \), processes it and exports its processed waveform \( E_l \) to the next layer (the input of layer 1 is \( E_0 \)). The processing contains two parts, as shown in Fig. 2a, i.e., temporal phase modulation and dispersion. The input of layer \( l \) firstly experiences phase modulation, so we obtain:

\[ E_{t\phi}(t) = E_{l-1}(t) \cdot \exp(-j\phi_l(t)), \]  (4)

where \( \phi_l(t) = \sum_i \phi_{i\parallel} \cdot \text{ones}(t - iT_i) \), is defined as a step function of which each period \( T \) has one or more step heights \( \phi_{i\parallel} \), each confined within \( 2\pi \). All these step heights are the parameters to be trained. After the phase modulation, the wave will go through a dispersive medium with a specific dispersion value. For convenience, we model the operation of the dispersive medium in the frequency domain. In particular, the spectral transfer function of the dispersive medium can be expressed as follows:

\[ H(\omega) = \exp\left(i\Phi(\omega^2/2)\right), \]  (5)

where \( |\Phi| \) is the first-order dispersion coefficient and \( \omega \) is the angular frequency. By multiplying the expression after phase modulation with the spectral transfer function and performing the inverse Fourier
transform, we get the following expression for the waveform at the output of layer $l$:

$$E_l(t) = \mathcal{F}^{-1}\{H(\omega) \cdot \mathcal{F}[E_{l1}(t)]\} = \mathcal{F}^{-1}(H(\omega) \cdot \mathcal{F}[E_{l-1}(t) \cdot \varphi(t)])$$

(6)

The output of the serial ONN system is detected by a photodetector, which is only sensitive to optical intensity. Specifically, the intensity of the final output of the serial ONN, i.e., the output of layer $L$ (assuming that this ONN has $L$ layers), can be written as:

$$I_{\text{out}}(t) = |E_L(t)|^2.$$

(7)

This is the profile that is used to calculate the loss function of the whole network in the backward propagation section of the system.

Fig. S1 | The evolution of cost (red) and accuracy (blue) of the training process for the analog signals’ classification problem considered in the main text.

1.2 Backward propagation

The loss function is used as the indicator of the resemblance between the waveforms obtained at the output of the network and the ideal target outputs (“labels”). For this purpose, the corresponding mean square error (MSE) is calculated:

$$e(\varphi) = \frac{1}{\Delta t} \int \left[ I_{\text{out}}^{1/2}(t) - I_{\text{label}}^{1/2}(t) \right]^2 dt,$$

(8)

where $I_{\text{label}}(t)$ is the “label” output corresponding to a specific “feature” input. Because the independent variable $t$ is integrated, the MSE is independent of time and instead, it depends on the trainable variable $\varphi(t)$ — the modulation phase at each layer. So the MSE will change as the phase of each layer — $\varphi$ — is modified.

In the back-propagation algorithm, the gradient $g = \nabla_{\varphi} e(\varphi)$ — the partial derivative between the MSE and the trainable variables — is calculated to update the variables. It determines the direction in which the modulation phase is adjusted in the backward propagation. Using this differential parameter, the modulation phase is consecutively updated to minimize the MSE as follows:

$$\varphi_{l,k} = \varphi_{l,k-1} - \alpha \cdot g_{k-1},$$

(9)

where $k$ is the step of training, $\alpha$ is the learning rate, and $g_{k-1}$ is the gradient at step $k - 1$. Beside the stochastic gradient rule as the update rule used in our models, there are many other rules that might be
alternatively considered to perform this task.

2. Training details

2.1 Training model details
We constructed the SONN models with Python and Tensorflow-GPU on a computer with two 12-core CPUs and a 2080Ti GPU, and spend approximately 5.5 minutes on training for the classification of analog and digital signals every 1000 epochs. The batch size is 6. The number of neurons at each layer is fixed by the number of phase levels within each period of the temporal phase modulation profile. In this paper, the duration of each phase level is set to half the repetition period of the mode-locked laser, leading to a number of neurons equal to 62 and 48 for the SONNs used for analog signal classification and digital signal classification, respectively. The SONN model with four layers is designed according to the analysis given in Supplementary Section 1. It is crucial to rearrange the Fourier transform (FT) operations to ensure that the zero-frequency component is shifted to the center of the obtained spectral pattern, i.e., to implement what is usually referred to as an ‘fftshift’, to directly match the profile of the spectral transfer function of the dispersive media, \( H(\omega) = \exp(j\Phi_\omega\omega^2/2) \). However, no function in the Tensorflow module could support this operation. Fortunately, an alternative method operating in the time domain, i.e., inverting every other sampling value of the field function before the FT operation, was used here to mimic the fftshift function. The number of FT points is 210 for every pulse period. Eight periods of zero-amplitude were introduced on each side of the signal under analysis to be sampled by the optical pulse train. The dispersions used in the simulated SOONs were \( 2D_T, -2D_T, 2D_T, -2D_T \) for the analog signal classification case, and \( 3D_T, -3D_T, 3D_T, -3D_T \) for the digital signal classification problem \((D_T\) is the dispersion value that satisfies the condition of the first-order temporal Talbot effect). We exploited an Adam optimization algorithm to minimize the loss function, thus maximizing the prediction rate. To obtain an initial point with which the model can achieve a better accuracy rate, several rounds of training were performed to find the best starting point.

2.2 Classification of analog signals
Concerning the training for the analog signals’ classifier described in the main text, this was carried out with 1000 epochs, and the evolution profiles of the cost and accuracy functions during the training process are presented in Fig. S1. This figure confirms that the cost function follows the desired stable evolution. The fluctuation of the accuracy trend in early epochs is due to the large initial learning rate, which is set to decay at a factor of 0.3 every 200 epochs. The ripple of the cost evolution relates to the sensibility of the final output of this serial ONN model to the phase of each layer. We anticipate that the fluctuations in the accuracy curve could be reduced using a smaller learning rate, but this would translate into an increased convergence time.
The trained modulation phase profiles of the four layers are shown in Fig. S2. They are restricted within the range between 0 and $2\pi$ by using a sigmoid function. Because the periodic pulses will shift half of the period over time by the integer temporal Talbot effect, the phase profile is shifted by half of the phase period to prevent the pulses from meeting with the phase jump points in the even-numbered layers. The data rate of the modulation phase is set to be twice the optical pulse repetition rate, i.e., 10 gigabits per second, which is still within the available range of a state-of-the-art electronic AWG (Keysight, M8194A, 120 GSa/s simultaneously for four channels\(^1\)). A lower data rate of the phase...
profile would translate into the presence of an increased amount of noise around the main target peak in the output waveform, an issue that is further discussed in Supplementary Section 3.

2.3 Classification of digital signals

The evolution of the cost and accuracy functions, as well as the resulting modulation phase profiles, during the training process in the digital signals’ classification problem considered in the main text are shown in Fig. S3 and Fig. S4, respectively. In Fig. S3, we can observe a significant fluctuation of the accuracy profile; we select an epoch of 800 as the training result by checking the outcome after every epoch such that to get a high success rate. Here, we select a training point for which the accuracy is not stable since the accuracy gets lower when it is more stable.

Fig. S4: The trained modulation phase profiles for layers 1, 2, 3 and 4 (red, green, blue, and purple lines, respectively) that are obtained for the considered digital signals’ classification problem.

3. Considerations for the experimental configuration

We demonstrated a pseudo-3-layer SONN rather than a four-layer one due to laboratory limitations (in regard to the number of channels of the electrical signal generators available in our lab). The optical pulse train was generated by an actively mode-locked laser with a clock provided by an analog signal generator (ASG, Keysight, N5183B), and was filtered out-of-cavity by an optical bandpass filter (OBPF, Yenista, XTA-50/W) with a wavelength bandwidth of ~1 nm in order to match the desired specifications for the optical pulses in the network. Although the ASG, the arbitrary waveform generators (AWGs, Tektronix, AWG70001A) and the parallel bit error ratio tester (PBERT, Agilent, 81250) were synchronized through a 10-MHz reference signal emitted by the ASG, this synchronization was insufficient and in particular, the relative movement between the signals generated from these instruments could be visualized in a real-time oscilloscope (OSC, Tektronix, DPO73304D). To mitigate the effects of this limited synchronization, we firstly fixed the sampling rate of the AWGs to be $24.112199996 \times 10^{9}$ GS s$^{-1}$ (corresponding to two samples per one phase step). Subsequently, through
observation of the relative movement on the real-time OSC, we finely tuned the frequencies of the ASG and the PBER T. Finally, the three signals got synchronized when the frequencies of the ASG and the PBER T were set to 12.05609998950 GHz. This procedure helped to reduce the timing jitter significantly among these generated signals, though there was still some remaining jitter even after the application of the procedure. As for the alignment between the phase modulation signals and the event-carrying pulses, an optical tunable delay line (OTDL) and several fixed delay lines (each introducing a delay several times longer than the maximal tuning span of the OTDL) were used. The relative position of the phase profile and the pulses in the second layer did not affect the output waveforms, and this only had an effect on the relative position of the resulting waveform over time. Therefore, our efforts focused on aligning the pulses with the phase modulation profile in the third layer only. Besides, the two phase-modulators (EOSpace) employed in this experiment have the same $V_{\pi}$ of 4.5 V. The phase modulation signals were amplified by broadband low-noise electric amplifiers (SHF, L806A and SHF, S807C) to an appropriate $V_{pp}$. The integer Talbot effect undergone by the original unmodulated optical pulse train through the dispersive lines in the network was verified beforehand by using an optical sampling oscilloscope (EXFO, PSO-102).

As mentioned in the main text, the demonstrated experimental serial ONN platform was designed according to the equipment available to the team at the time of this work. Because there were only two AWG channels available in our labs, which could support the implementation of two layers of the network with phase modulation, we planned for the realization of a 2-layer serial ONN model. The limited depth of the model affects the capability of classification, such that this scheme can distinguish successfully two different patterns only. Moreover, the performance of the ONN with two phase-modulation layers only is still far from satisfactory, as shown in Fig. S5. In this figure, the trained classification results are shown in Fig. S5 a—f and Fig. S5 g—l for ‘u’ & ‘c’ and ‘a’ & ‘s’, respectively. Not only the success rate is relatively low, say 84.4% for the classification of ‘u’ & ‘c’ and 64.4% for that of ‘a’ & ‘s’, but additionally, the trained output waveforms lack the desired high contrast between the pulse used to mark the classification result and the neighboring pulses, as shown in Fig. S5 c, f, i and l. Fortunately, we find that an extra layer with no phase modulation can greatly improve the performance of the ONN. In particular, through this strategy, we achieve a far higher contrast in the output waveforms, as shown in Fig. 4 in the main text, and the success rate is increased to 100%. This performance improvement is partly attributed to the fact that the layer with no phase modulation can be regarded as a layer with a constant phase modulation and therefore, this scheme actually implements a network with a larger depth, leading to the observed improvements in the feature extraction process.
In regards to the specific dispersion value to be used at each layer, we have concluded that an optimal performance is achieved when choosing a dispersion that satisfies a temporal Talbot condition for the Fig. S5| a, d, j, ‘Feature’ waveforms of four English letters—‘u’, ‘c’, ‘c’ and ‘s’—in the form of binary ASCII code. Each bit is represented by the power of each pulse. b, e, h, k, Ideal ‘label’ waveforms corresponding to a, d, g, and j. c, f, i, l, Trained output waveforms corresponding to a, d, g, and j.

Fig. S6| The evolution of cost (red) and accuracy (blue) of the training process for the considered analog signals’ classification problem when the dispersion of each layer is set to 2D_T + 200 ps/nm (D_T is the dispersion value satisfying the first-order integer temporal Talbot effect).

4. Selection of dispersion value
In regards to the specific dispersion value to be used at each layer, we have concluded that an optimal performance is achieved when choosing a dispersion that satisfies a temporal Talbot condition for the
input optical pulse train. In general, the dispersion value at each layer will affect the performance (classification accuracy) of the serial ONN. As seen in Fig. S6, the accuracy of the considered deep learning system (for the analog signals’ classification problem described in the main text) is significantly deteriorated when the applied dispersion does not satisfy a Talbot condition. In the example shown herein (with a deviation of 1.57% over the closest Talbot dispersion value), the accuracy can only reach a maximum of 79.2% even after many rounds of training (>5000). Also, we did several other simulations with different dispersion deviations, and got similar results. In sharp contrast, the accuracy of the ONN when the dispersion is set to satisfy a Talbot condition at every layer can reach 100%, as shown in Fig. S1. We attribute the improved performance obtained under Talbot conditions to the fact that operation under these conditions ensures a more uniform re-distribution of the energy of each dispersed pulse along the time domain, ensuring an optimal interaction (i.e., coherent addition) between consecutive temporal neurons at the positions of the corresponding temporal pulses.

5. Demonstration of classification for time-consecutive patterns (objects)

![Fig. S7](image_url) The result of classification for single events and consecutive patterns using the same-signed dispersion of $D_T$ across the neural network. The first, second and third row represent the input patterns to be analyzed, the target ideal output waveforms and the simulated output waveforms, respectively.
As mentioned in the main text, the proposed serial ONN is well suited for the analysis (e.g., classification) of different time-consecutive patterns or objects but in this case, the scheme needs to be carefully designed. The main consideration is that the dispersion in the system may induce undesired interference effects among the processed waveforms corresponding to different consecutive input patterns, negatively affecting the overall performance of the network. The trained phase profiles for a given input object should still provide the predicted performance when the same object is repeated along the time domain. However, a significant deterioration of the performance may be produced when different patterns are intended to be analyzed in a consecutive fashion, see results in Fig. S7. This is so because in this case, unintended interactions may be introduced between the waveform under analysis and adjacent waveforms that differ from those considered in the training process. In the example shown in Fig. S7c, the two different patterns in Figs. S7a and b are analyzed in a consecutive fashion with a gap in between them of 10× the input pulse repetition period (5 periods on each side) and assuming a dispersion in each layer of $D_T$. Clearly, the output waveform corresponding to each pattern deviates more significantly with respect to the ideal one than in the case of analysis of each of the patterns separately, Fig. S7 a and b, respectively. A more significant deterioration is still expected for a narrower gap or a higher dispersion. However, by properly choosing the amount of dispersion and the length of the gap in between consecutive input patterns, as well as using a symmetrical dispersion strategy (see below), one can ensure that the classification of consecutive patterns or objects is performed with the desired accuracy, as shown through the example in Fig. S8. In this latest example, the gap is set to 16×.
the pulse repetition period and a symmetrical dispersion strategy is utilized, in which the same amount of dispersion is used in consecutive layers but with an opposite sign. The pattern separation implies that an additional latency by the gap length (~3.2 ns in this example) should be considered for the analysis of each of the input patterns. The results in Fig. S8 show that the peak pulses of each output waveform are located in the designed temporal positions, confirming that the proposed serial ONN strategy can be designed for a successful analysis of different patterns as they arrive sequentially into the network.

6. Discussion about the influence of the IV on the accuracy of the digital classification

Intuitively, if the data used for training and test diversifies, i.e., if the individuals of the same kind of object differ from each other, it becomes harder for the neural network to perform an accurate classification of the incoming patterns. In conventional deep learning, this concept is called dissimilarity or similarity. Here, we evaluate this parameter using the individuality variance (IV), which can be quantified through the individuality variance rate (IVR):

$$\text{IVR} = 10 \log_{10} \frac{P_s}{P_{IV}}$$  \hspace{1cm} (10)

where $P_s$ is the nominal peak amplitude of each pulse (after modulation by the pattern under analysis), and $P_{IV}$ quantifies the peak amplitude variance of each pulse. We used a Gaussian White Noise (GWN) function to generate the pulse-to-pulse amplitude difference, in which $P_{IV}$ is the standard deviation of the GWN. Then, the generated amplitude difference is imposed on each pulse amplitude. We train the SONN model for the classification of the digital signals described in the main text, and we test the accuracy of the model at IVRs of 0 dB, 5 dB, 10 dB, 15 dB, 20 dB, 25 dB, and 30 dB. The results of this study are shown in Fig. S9, confirming that the accuracy increases when the variance of the patterns/objects decreases. At an IVR of about 25 dB, the accuracy already reaches 100%. The reason why we choose the classification of digital patterns to explore the influence of the IVR to the SONN performance is that the IV added to the analog patterns will change the patterns greatly due to the different

![Fig. S9](image-url) Accuracy evolution of the serial deep learning scheme considered in the main text for the classification of digital signals, with IVRs from 0 to 30 dB.
amplitudes between 0 and 1, which eliminates the features of the patterns. In contrast, the digital patterns are more capable of enduring the IV because the amplitude of the digital pattern is 1 or 0. Still, we explored the impact of the IV on the accuracy rate: the accuracy remains 79.2% or lower before the IVR reaches 30 dB.

7. Comparison between the modulation types

In order to explore the influence of the modulation format on the recognition performance, we trained and simulated the SONNs with modulation types different from the phase-modulation case reported in the main text (amplitude modulation and complex modulation). Results are shown here for the analog signals’ classification problem described in the main text (Fig. 2). Besides the modulation type, all other parameters remain identical to those defined for the problem at hand in the main text. The results of this extended analysis are shown in Fig. S10. These results show that the accuracy rates of both modulation types are 79.2% at best, which is a lower value than that obtained with using pure phase modulation. Considering that the rest of the specifications are identical, except for the modulation type, the decrease in the recognition performance can be attributed to the change in the modulation type. Besides, numerous additional simulated training has been conducted to verify the reliability of this conclusion. However, we have also observed that the use of amplitude modulation helps in that it provides a significant mitigation of the undesired pulses around the main target peak pulse in the outcome waveform. If a recognition output with a high extinction ratio is preferred, the amplitude modulation could be added to the SONN. Notice that although the extinction ratio is increased in this case, the absolute value of the amplitude of the main peak is considerably decreased, i.e., approximately by 10 dB, which explains the observed deterioration in the obtained accuracy parameter.

8. Comparison with other ONNs

Here, we compare the proposed SONN with current ONNs in the aspects of latency, processing speed and power consumption for the classification of serial data. The serial architecture will inevitably introduce time delay including inherent delay (caused by the technique itself) and system or device delay (caused by the optical component or device and can be reduced; for SONN, the device delay can be greatly mitigated by using LCFBG). Given that we only consider the delay caused by the technique itself, the current ONN techniques, such as MZI-based ONN, diffraction-based ONN, etc; would have no delay. The delay of SONN would be generated from the temporal gaps we used to prevent the patterns from
interfering with each other, i.e; ~3.2 ns and 4.0 ns for the analog and digital patterns under analysis at a data rate of 5 GHz, respectively. However, in the aspect of processing speed (the data rate of a serial data flow), the SONN outperforms the current ONNs. To the best of our knowledge, the highest speed of the serial-to-parallel converter is 2.5 Gb/s, which can only support 4-bit conversion[2]. By contrast, the data rates in our simulation and experiment are 5 Gb/s and 12.61 Gb/s and the bit number is over 8. Besides, the power consumption of the converter is 1 W,[2] which can be eliminated by using this serial optical neural network. The power consumption of SONN can be significantly reduced by using LCFBG. According to the datasheet of the LCFBG made by Teraxion[3], the loss can be reduced to 3 dB for compensating the dispersion of 3400 ps/nm.

Reference

1. Keysight. M8100 Series Arbitrary Waveform Generators. Keysight Available at: https://www.keysight.com/zz/en/products/arbitrary-waveform-generators/m8100-series-arbitrary-waveform-generators.html. (Accessed: 25th June 2019)

2. Microchip. 2.5Gbps GPON/BPON ONU SERDES. Microchip Available at: http://ww1.microchip.com/downloads/en/DeviceDoc/sy87725l.pdf

3. Teraxion. Dispersion compensation module. Teraxion Available at: https://teraxion.blob.core.windows.net/media/1311/teraxion-clearspectrum-dcm-l-spcsheet.pdf