

PAPER • OPEN ACCESS

Multimode Fabry-Perot laser as a reservoir computing and extreme learning machine photonic accelerator

To cite this article: Menelaos Skontranis *et al* 2023 *Neuromorph. Comput. Eng.* **3** 044003

View the [article online](#) for updates and enhancements.

You may also like

- [Reservoir computing and decision making using laser dynamics for photonic accelerator](#)
Atsushi Uchida, Kazutaka Kanno, Satoshi Sunada et al.
- [Exploring reservoir computing: implementation via double stochastic nanowire networks](#)
Jian-Feng Tang, , Lei Xia et al.
- [Hands-on reservoir computing: a tutorial for practical implementation](#)
Matteo Cucchi, Steven Abreu, Giuseppe Ciccone et al.



PAPER

OPEN ACCESS

RECEIVED

14 February 2023

REVISED

26 September 2023

ACCEPTED FOR PUBLICATION

11 October 2023

PUBLISHED

20 October 2023

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Multimode Fabry-Perot laser as a reservoir computing and extreme learning machine photonic accelerator

Menelaos Skontranis¹, George Sarantoglou¹, Kostas Sozos², Thomas Kamalakis³, Charis Mesaritakis^{1,*}  and Adonis Bogris²

¹ Department of Information and Communication Systems Engineering, University of the Aegean, Karlovassi-Samos 83100, Greece

² Department of Informatics and Computer Engineering, University of West Attica, Egaleo 12243, Greece

³ Department of Informatics & Telematics, Harokopio University of Athens, Athens 17778, Greece

* Author to whom any correspondence should be addressed.

E-mail: cmesar@aegean.gr

Keywords: neuromorphic computing, semiconductor laser, spectro-temporal multiplexing, time-delay reservoir computing, photonic accelerators

Abstract

In this work, we introduce Fabry–Perot lasers as neuromorphic nodes in the context of time-delayed reservoir computing and extreme learning machine (ELM) for the processing of temporal signals and the high-speed classification of images. By exploiting the multi-wavelength emission capabilities of the Fabry–Perot lasers, additional processing nodes can be introduced, thus raising the computational power without sacrificing processing speed. An experimental validation of this concept using a Fabry–Perot ELM is presented targeting a time dependent task such as channel equalization for a 50 km 28 Gbaud ‘PAM-4’ transmission, offering hard-decision forward error correction compatible performance. Additionally, the Fabry–Perot neuromorphic concept has been further strengthened by modifying the data entry technique by parallelly assigning different samples of the input signal to different modes so as to significantly reduce speed penalty. Numerical simulations revealed that this alternative data insertion technique can offer a reduction of the processing delay and physical footprint by 75% compared to the conventional approach assigning the same symbols to all Fabry–Perot modes. Moreover, by using a similar data processing scheme in ‘MNIST’ image classification task we were able to numerically achieve a processing speed of 255.1 Mimages s^{−1} and a classification accuracy up to 95.95%.

1. Introduction

During the last years, Artificial neural networks (ANNs) have drawn the spotlight of attention due to their ability to efficiently address a variety of highly complex and nonlinear tasks such as image classification, voice recognition, chaotic series prediction and channel equalization [1]. Their well desired performance stems from the fact that the processing of data is performed by the cooperation of multiple nonlinear nodes, known as neurons, in contrast to classical von Neuman processors which assign the same task to a single processing node, known as central processing unit. Consequently, the processing bottleneck of von Neuman processors is circumvented [2, 3] but their memory bottleneck is still an open issue [4].

The performance of ANNs is affected by their architecture, which is determined by two factors. The first one is the number of layers, which regulates the high level layout of the network and leads to the formation of ‘shallow’ or ‘deep’ networks. The second one is the connectivity scheme of the neurons, which is regulated by the adjustment of the synaptic weights, as it affects several attributes of the ANNs such as processing and dynamical memory. Up to now, two major categories of ANNs have emerged: recurrent neural networks (RNNs) which consist of neural assemblies that offer dynamic memory and are well fitted for time-dependent tasks, and feed forward neural networks which include a sequential assembly of layers with no memory; optimized to address problems such as image classification and voice recognition [1].

In RNNs, neurons of a specific layer are permitted to connect to one another. This triggers multiple time dependencies among the neurons which enhances the dynamical memory of RNNs and renders them perfect candidates for processing time-dependent data. However, these time dependencies in combination with the extensive number of training parameters significantly burden the training of RNNs. An RNN alternative that inherits its merits but is hardware friendly, is reservoir computing (RC). In particular RC systems allow the synapses at their hidden layer to remain random and untrained, whereas training is restricted only to the output layer [5].

In addition, up to now RC systems, despite their hardware friendliness, are mainly focused on time-dependent tasks, following the example of their RNN counterparts. On the contrary, the vast majority of real-world cases deal with the processing of static data such as image classification. In this framework, the most widely used network paradigm is the convolutional neural network (CNN). In this multi-layer feed-forward paradigm, each neuron detects a specific feature, whereas these features are combined by neurons in 'deeper' layers. Despite their success, the training of CNNs, similar to RNNs is a computational demanding task, due to the sheer number of trainable synapses that range from 10 000 to millions [1, 6, 7]. Following this lead, an interesting approach consists of expanding the time-unfolding property of RNNs and in particular RC systems to realize 'deep' architectures, whereas preserving the training simplicity [8, 9].

One of the most important aspects of the RC endeavour is the platform which will facilitate them. Up to now, various works have emerged based on spintronics [10, 11], mechanical systems [12, 13] and electronics (analog circuits [14], fully programmable gate array (FPGA)s [15] and memristive circuits [16]) with enhanced performance. From the aforementioned categories, electronics have shown great promise towards the development of the RC discipline primarily due to their mature technology which allows their large scale integration [5]. On the other hand, they are affected by the inherent impediments of electronics such as fan in/out bandwidth trade off and heat dissipation [2, 17].

However, photonics have risen as an alternative solution since they can counter most of the inherent problems of electronics, through their ultra-high bandwidth, low power consumption and parallel processing capabilities. In addition, photonic devices, such as lasers have been shown to exhibit neuro-computational properties and are isomorphic to biological neurons, thus tackling bio-inspired neural networks such as spiking neural networks [18]. Up to now, multiple photonic RC implementations have emerged [19–25] which can be characterized in three subcategories: the spatial RC, in which every node is represented by a separate component, the wavelength multiplexing RC in which every node is represented by a frequency of the comb [26, 27] and the time delayed RC (TDRC) in which the entire layout is replaced by a single node, equipped with a feedback loop [5]. Spatial RC schemes have achieved excellent performance in a wide variety of tasks [28, 29]. However, the complexity of the network remains an open issue especially when increased node count is needed [5]. On the other hand, in TDRC schemes node upscaling is straightforward, since the number of nodes is regulated by the length of the feedback loop. Moreover, if the feedback of TDRC scheme is eliminated then the network behaves as an in time unfolded extreme learning machine (ELMs), which make it suitable for memory less applications such as image classification. On the other hand, the hardware friendliness of the TDRC and time unfolded ELM comes at the cost of processing speed since an unavoidable speed penalty is inserted due to the time expansion of the original input which is a necessary step for the creation of the necessary time multiplexed virtual nodes [5].

In the majority of the photonic TDRCs or ELMs the nonlinear physical node is a single mode semiconductor laser under electrical or optical injection (DFBs, DBRs, VCSELs, Quantum Dots etc.), a choice that is based on the complex carrier dynamics of lasers [24, 25, 30–33]. However, their applicability to tasks that require real-time processing at elevated rates, is not straightforward, due to the inherent speed penalty triggered by time multiplexing [5, 19, 32]. Recently, a numerically simulated TDRC scheme based on a Fabry–Perot (FP) laser has been presented, which presents a significant improvement in terms of speed penalty, by exploiting multiple longitudinal lasing modes of Fabry–Perot lasers as quasi-parallel computation streams [19]. In particular, the multi-wavelength output of the FP laser allowed also for the inclusion of spectrally encoded nodes, thus decreasing the latency from 1 ns in the case of conventional TDRC schemes down to 240 ps.

In this work, we provide, for the first time to the best of our knowledge, experimental evidence regarding a FP based accelerator in a time delayed extreme learning machine (TDELM) configuration, targeting a 28 Gbaud PAM-4 equalization task after chromatic dispersion deteriorated transmission. Following this step, we provide numerical results concerning an alternative configuration of the original FP accelerator, where a new data insertion technique designated as spectro-temporal multiplexing (SPTM) is presented and numerically evaluated. The application of SPTM in the PAM-4 case offers an improvement in terms of accuracy and a doubling of the processing rate as the external loop is shrunked from 120 ps to 60 ps. In an image classification numerical simulation scenario our scheme offered an accuracy up to 95.95%, for the MNIST dataset, whereas it allowed a processing latency as low as 3.92 ns per image.

This work is organized as follows. In section 2, we present the theoretical frameworks of multimode FP based TDRC (FP-TDRC) and TDELM (FP-TDELM) layouts emphasizing on their capability to enable accelerated operation. In section 3, we experimentally validate the accelerated operation of the FP-TDELM scheme by addressing a PAM-4 channel equalization task for the first time to the best of our knowledge. In section 4, we numerically evaluate the performance of the FP-TDRC scheme by tackling the same task (PAM-4 equalization) with two different data insertion methods: the first one inserts the data in a sequential manner, mimicking the classical TDRC operation, whereas the second one applies the SPTM technique which enables parallel data insertion and minimizes the processing latency of the network. In section 5, the optimized SPTM concept is further numerically validated on a FP-TDELM scheme which targets the image classification of the MNIST dataset.

2. Theoretical framework of multi-mode FP-TDRC and FP-TDELM

Classical TDRCs and TDELMs are two types of neural networks which have drawn the spotlight of attention over the last two decades primarily due to their hardware friendliness and lightweight training. Even though both types of networks share a common principle of operation and an identical architecture, they have an important difference which is the existence or absence of feedback.

To be more specific, in TDRC implementations a single nonlinear node equipped with a feedback loop is used to replace the entire network, whereas virtual nodes are generated through time-multiplexing of incoming data (figure 1(a) feedback coefficient $k_f > 0$). In order to accomplish this, incoming data are subjected to two transformations: masking and time stretching [5, 19, 30]. Masking is realized by the vectorial multiplication of the input symbol (S) with a random value vector, known as mask, and its purpose is to project the low dimensional input to a higher dimensional space to facilitate the classification process by the TDRC. To be more specific, N samples of the incoming symbol are vertically concatenated in a S column vector (figure 1(a) S column vector) and are multiplied by a random row vector (figure 1(a) mask) which consists of K values drawn from a uniform distribution and range between 0 and 1. The result is a column vector (D) which contains $N \cdot K$ points. After this step, each masked point's duration is expanded to θ and is used as input. Therefore, each masked symbol D remains at the input for $T = N \cdot K \cdot \theta$, which is higher than the original duration of the symbol (T_s). This means that the preprocessing (masking and time-stretching) introduces an unavoidable speed penalty which is measured by the ratio $SP = T/T_s$. The masked and time-stretched input drives the single nonlinear node (figure 1(a) NL), whereas its output is sampled at a rate of $1/\theta$. Consequently, the output of the network consists of $N_v = T/\theta = N \cdot K$ points which is the number of the virtual nodes of the network. In TDRC schemes, the output state is strongly affected not only by the present input but also from previous outputs due to the existence of the feedback. The impact of the feedback is regulated by the feedback strength parameter (k_f) which is the ratio of the reinjected input power to the output power of the network whereas its time length is usually regulated to be linked to T (figure 1(a)). The existence of feedback allows the TDRC to implement the long-term fading memory mechanism which renders TDRCs suitable candidates for time-dependent tasks such as prediction of chaotic time-series and channel equalization of signals transmitted through a dispersive medium.

On the other hand, TDELM schemes (figure 1 $k_f = 0$) follow the same operational principles as their TDRC counterparts (time stretching, masking and sampling), except that in their case there is no feedback loop ($k_f = 0$). This means that their output is mainly affected by their current input, thus offering limited memory. Thus, TDELM require a digital storage mechanism at the output so as to tackle time-dependent problems through the correlation of preceding and succeeding symbols in the output layer.

In general, both TDRC and TDELM proliferate from the higher number of virtual nodes which in principle determines computational performance. However, the increase of the virtual node number has a twofold effect: either it causes the processing rate to decline (increasing T and keeping θ constant) or stressing the bandwidth capabilities of the scheme by keeping T constant and reducing θ . Up to now, TDRC and TDELM implementations have achieved satisfactory performance with minimum hardware requirements in a wide variety of tasks [5, 26, 34]. However, the augmented number of required nodes comes with a detrimental high speed penalty which reduces the applicability of these networks in real time systems.

To circumvent this trade-off, efforts have been focused on photonic TDRCs and TDELMs implementations so as to take advantage of their wide pallet of ultrafast dynamics and the high bandwidth operation. The majority of photonic TDRCs and TDELMs rely on a set up of two lasers which are connected in a injection signal generator (ISG)–injection locked laser (ILL) configuration (figures 1(b) and (c)). When the setup is based on single mode lasers ($M = 1$, M is the number of longitudinal modes), then the output of the ISG is injected to the ILL via a circulator (figures 1(b) and (c)) and its output is monitored by a single photodiode (figures 1(b) and (c) for $M = 1$). In this case, the number of nodes will be $N_v = T/\theta$ number on nodes, as it in classical TDRC and TDELM implementations. However, if $M > 1$ and each mode is monitored

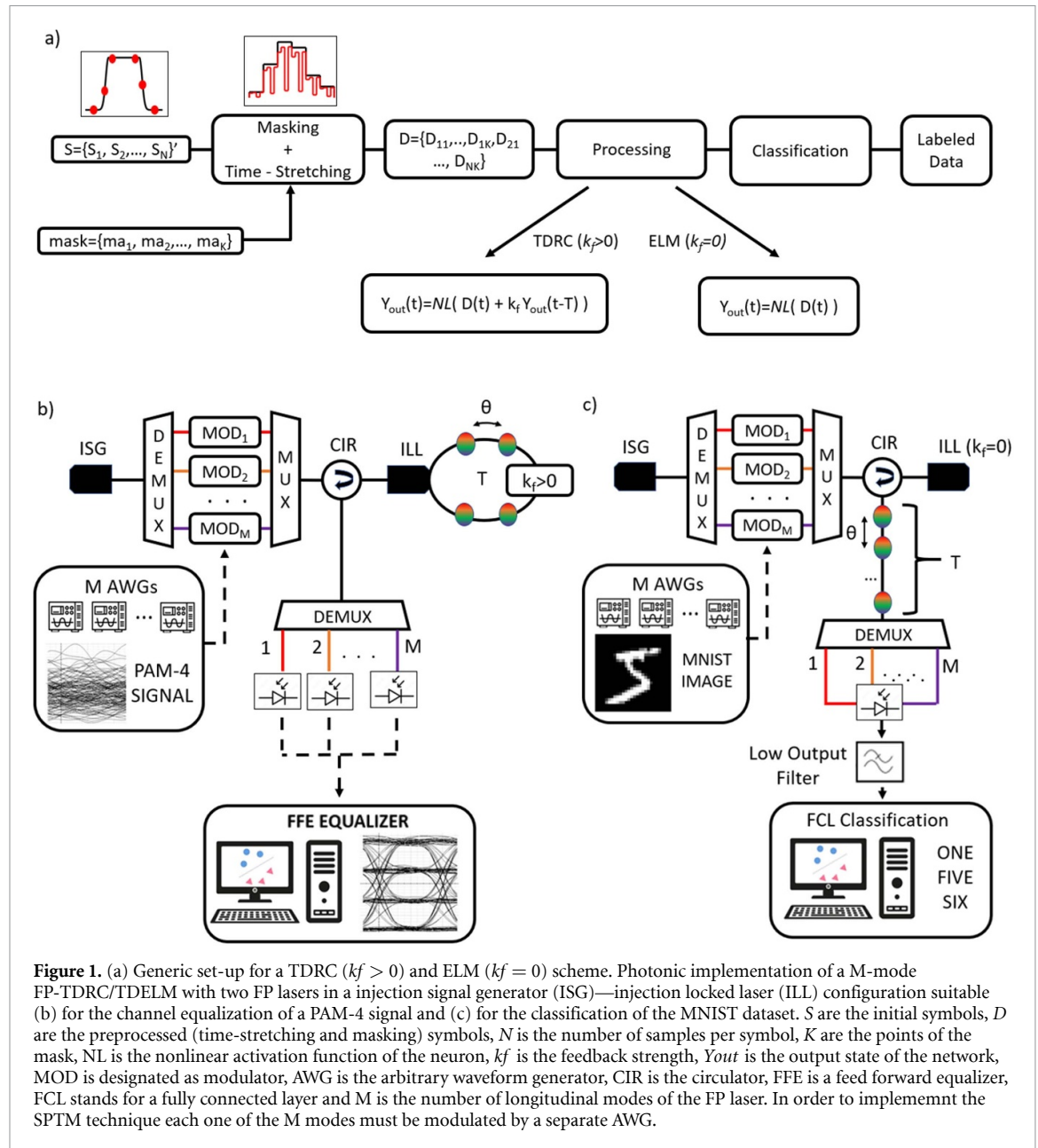
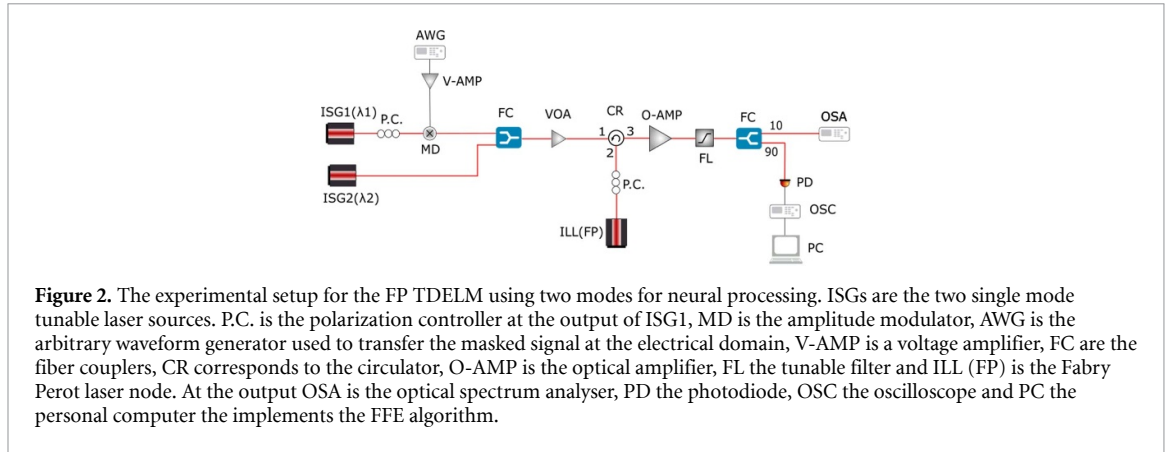


Figure 1. (a) Generic set-up for a TDRC ($k_f > 0$) and ELM ($k_f = 0$) scheme. Photonic implementation of a M-mode FP-TDRC/TDELM with two FP lasers in a injection signal generator (ISG)—injection locked laser (ILL) configuration suitable (b) for the channel equalization of a PAM-4 signal and (c) for the classification of the MNIST dataset. S are the initial symbols, D are the preprocessed (time-stretching and masking) symbols, N is the number of samples per symbol, K are the points of the mask, NL is the nonlinear activation function of the neuron, k_f is the feedback strength, Y_{out} is the output state of the network, MOD is designated as modulator, AWG is the arbitrary waveform generator, CIR is the circulator, FFE is a feed forward equalizer, FCL stands for a fully connected layer and M is the number of longitudinal modes of the FP laser. In order to implement the SPTM technique each one of the M modes must be modulated by a separate AWG.

independently (figures 1(b) and (c) for $M > 1$) then the multi-mode laser can support a $N'V = M \cdot T/\theta = M \times NV$ nodes for the same T resulting in a M -times higher number of neurons. The multimode node augmentation technique has been applied to photonic TDRCs by exploiting the multimode emission of a quantum-dot spin vertical cavity surface emitting laser ($M = 4$ two emission and two polarization modes) [32] and a FP lasers ($M = 8$ longitudinal modes) [19], allowing increased performance with significantly smaller T and SP. In these cases, the proof of concept has been shown only through numerical simulations. In the next section we will experimentally validate the multimode node augmentation technique in an TDELM set up for the first time to our knowledge.

3. Experimental validation of the Fabry Perot laser accelerator

The experimental setup for the FP laser-based TDELM is presented in figure 2. This scheme targets the equalization of a 28 Gbaud s^{-1} PAM 4 signal in the C-band after its transmission through an optical fibre of length equal to 40 km. We chose such a task so as to be compatible with the original work [19]. The signal consists of 40 000 randomly generated symbols and the transmission is implemented numerically, by using the split-Fourier method governed by the Manakov equations [35]. The signal at the output consists of 2 samples per symbol and afterwards it is resampled by performing interpolation so as to acquire $j = 4$ samples per symbol for masking reasons. Each input is projected to a higher dimensional plane by multiplying the



samples of each symbol with a weight matrix $W_{48 \times 4}$ consisting of randomly drawn fixed values originating from a uniform distribution in the range $[0, 1]$. This matrix provides the connectivity between the 4 input samples and 48 virtual nodes. The matrix $W_{48 \times 4}$ performs the masking procedure, that is encountered in TDRC and TDELM schemes. Since the overall number of virtual nodes is substantially high (48 virtual nodes in the single mode case and 96 virtual modes in the dual mode case), the performance of the system is essentially unaffected by the exact choice of the mask and therefore a single fixed matrix $W_{(48 \times 4)}$ is examined [36]. It is worth mentioning that this masking procedure has been recently employed for ELMs in [37] and utilizes linear combinations of masked samples, instead of simple masked samples as described in section 2. It essentially offers the same performance as the masking described in section 2. The virtual nodes values are serialized and the resulting time series is transferred from the digital to the analogue domain using an arbitrary waveform generator Tektronix AWG7082C arbitrary waveform generator (AWG) operating at a speed equal to 4 Gs s^{-1} . This corresponds to a temporal distance between the virtual nodes equal to $\theta = 250 \text{ ps}$.

This information is encoded via amplitude modulation at the optical mode that is generated by two tunable lasers (CoBrite IDPhotonics CDx). These lasers provide optical injection to a FP laser (Eblana Photonics), which acts as a non-linear element. The proposed setup consists of an unidirectional optical injection scheme, where the two lasers act as the ISGs and the FP-laser acts as the ILL. This ISG-ILL configuration allows for an increase in the modulation bandwidth of the ILL, thus allowing for faster processing rates [30]. In contrast to conventional schemes where two single mode lasers are used as ISG and ILL, in this case a multi-mode FP laser is used as the ILL and the two tunable IGSs inject light to different longitudinal modes of the FP-ILL. In detail, the first IGS is amplitude modulated via the information provided by the AWG, whereas the optical mode from the second IGS is directly injected to the FP-ILL. Information from the first mode will be non-linearly transferred to the second mode through the cross-gain mechanism inside the FP-ILL [38]. Since amplitude modulation is used, the best performance can be achieved by biasing the ILL slightly below its threshold ($I_{th10} \text{ mA}$) at 9.7 mA [25].

In order to implement the multi-mode processing scheme, each of the two longitudinal modes emanating from the FP-ILL needs to be read at the output and processed so as to create at the end the input of the feed forward equalizer (FFE) that forms the output layer and operates as a linear regression stage. Given the output of the FP-ILL that consists of M longitudinal mode ($M = 2$ in this case), an erbium doped fiber amplifier is used so as to amplify the output, followed by a tunable optical bandpass filter. The optical filter is used so as to isolate the output of a mode and drive it to the measuring devices. Therefore, in order to measure M longitudinal modes, the optical filter must be tuned M times, so as to isolate each time a different mode. Additionally, the filter is used to improve the optical signal to noise ratio of the measurements. Given a single optical mode, its output is driven to a 90/10 coupler to be distributed to a Tektronix 3-series MDO32 real-time oscilloscope operating at 5 Gs s^{-1} and to a Finisar Wave analyser 100 s optical spectrum analyser (OSA), respectively. The derived samples are used to feed the offline FFE procedure. In particular, since a TDELM possesses no intrinsic memory, the samples of the longitudinal mode are grouped in 40 000 labelled vectors in a manner that introduces the effect of memory. Each vector consists of 48 values corresponding the virtual nodes of the longitudinal mode, alongside with 26×48 values, that correspond to the virtual nodes of the 26 adjacent symbols (taps). Adjacent symbols possess information about the targeted symbol due to inter-symbol interference that takes place at the transmission stage. Additionally, since the FP-ILL filters out a part of the input due to its limited bandwidth, the input vector is also fed to the output vector, a technique that is also used in traditional RC schemes [39]. Thus, for each longitudinal mode 40 000 vectors labelled

vectors are formed, with each vector consisting of 28×48 values. Since this procedure is followed for M modes sequentially, the whole process results in 40 000 labelled vectors, with each vector consisting of $M \times 48 \times 28 = M \times 1344$ values. The 75% of the output vectors are used for training and the 25% for testing.

The presented experimental setup operates as a TDELM scheme, which compared with TDRC schemes holds some important benefits in terms of stability. Since the analogue system does not present an optical feedback loop, it is not affected by thermal variations and acoustic noise effects, that can randomly modify its phase and consequently the overall performance. Moreover, since the TDELM is based on simple back-to-back injection locking, it can be easily monolithic integrated, in contrast to conventional TDRC setups that require feedback loops with delay values in the order of nanosecond [30].

As mentioned above, the speed penalty is defined as the ratio between the duration of a symbol $T_s = 35.7$ ps and the total duration of its virtual nodes $T = N_v \times \theta = 12$ ns. The speed penalty is then equal to $SP = T/T_s = 336.13$, a value that is high owing to the low bandwidth of the experimental components. In order to showcase the computational power of the multi-mode scheme, the performance between using a single longitudinal mode ($M = 1$) and two longitudinal modes ($M = 2$) will be compared. Note, that by using two modes, the number of virtual nodes increases from 48 to $2 \times 48 = 96$, without inducing any increase in the speed penalty [19]. A similar increase in the case of a single mode ILL, would require the introduction of $N_v = 96$ virtual nodes at the pre-processing stage, leading to $T = N_v \times \theta = 24$ ns and a speed penalty $SP = 672.26$ which is twice the SP of the dual-mode case. Therefore, with higher bandwidth equipment, the FP-ILL is a promising candidate for real-time processing of ultrafast signals.

In order to evaluate the performance of the FP node, three different schemes are compared; (a) FFE, (b) single mode TDELM with FFE, (c) dual-mode TDELM with FFE. In the first case, the signal after its transmission is down-sampled at 2 points per symbol and it is driven directly to an FFE stage that performs the linear equalization. This corresponds to the direct drive of the input to the readout layer, which acts as the baseline to be compared with the two other analogue processing schemes. This method achieves a bit-error-rate (BER) equal to 0.042.

For the single mode TDELM with an FFE, the unmodulated ISG is switched off and the amplitude modulated ISG targets a single mode of the FP-ILL. This setup is equivalent to the conventional single mode case [30]. By shifting the wavelength of the ISG injection locking at different longitudinal modes of the FP-ILL can be achieved. Three different longitudinal modes are examined with optical frequencies equal to $\nu_1 = 192.7$ THz, $\nu_2 = 193.65$ THz, $\nu_3 = 194.435$ THz. The optical spectrum for these three different configurations is presented in figure 3(a). After achieving injection locking, the mode of the ISG is swept so as to test different frequency detuning (between ISG-ILL), ranging from -16 to 16 GHz. The absence of mode hopping effects indicates that the ILL is injection locked in this frequency range. The optical output for each frequency detuning is driven to the FFE stage. The process is repeated for all three optical frequencies ν_1, ν_2, ν_3 and the derived BER values are presented in figure 4. A substantial improvement in the performance is observed compared to the baseline with $BER = 0.042$. Moreover, it can be seen that the performance is wavelength transparent since it is unaffected by different choices of longitudinal modes.

Finally, when utilizing both ISGs, the amplitude modulated ISG is biased arbitrarily at 192.69 THz which corresponds to a -12 GHz frequency detuning from the corresponding longitudinal mode of the FP-ILL. The unmodulated ISG is also activated and it is biased at 194.435 THz, which corresponds to another longitudinal mode of the FP-ILL. The optical spectrum recorded for this configuration is presented in figure 3(b). By exploiting the tunable optical filter, the outputs of both longitudinal modes are separately recorded. During this process, the mode of the modulated ISG stays unaffected, whereas the mode of the unmodulated ISG is tuned so as to achieve different frequency detuning values ranging from -16 to 16 GHz, with respect to its target longitudinal mode. The outputs are driven to the FFE. The results are depicted in figure 4. A noticeable improvement in the overall performance is observed, achieving a BER as low as 3×10^{-3} . This value is below the hard-decision forward error correction (HD-FEC) limit, that is equal to 3.8×10^{-3} [40]. This boost in the performance can be attributed to the deployment of 96 virtual modes instead of 48 because of the existence of two intertwined processing lanes, offered by the two FP modes. Thus, more virtual nodes can be used without imposing additional burden at the processing speed.

4. Spectro-temporal data insertion strategy (SPTM)

Following the experimental validation of the FP TDRC-TDELM scheme, we aimed to push performance further, in terms of processing latency through a SPTM data insertion strategy. In detail, we preserve the classical masking procedure described in section 2 but with a basic exception. In the classical approach, the masked samples ($D_{11}, D_{12}, \dots, D_{NK}$) are sequentially inserted to the processor which means that the necessary time-delay of the external cavity is governed by the product of $N \times K \times \theta$. In our previous work [19], multiple copies of the same samples are broadcasted to the M modes of the FP laser, augmenting

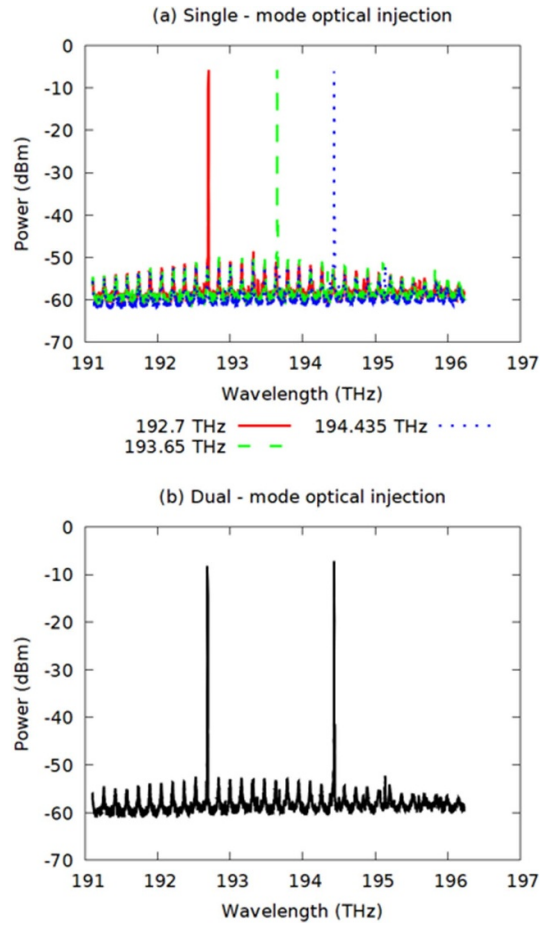


Figure 3. (a) The optical spectrum when the mode of the modulated ISG targets the longitudinal mode of the FP-ILL at $\nu_1 = 192.7$ THz (red continuous line), $\nu_2 = 193.65$ THz (green dashed line), $\nu_3 = 194.435$ THz (blue dotted line). (b) The optical spectrum when the two modes of the ISGs are optically injected at the optical frequencies $\nu_1 = 192.69$ THz (-12 GHz from the closest longitudinal mode) and $\nu_2 = 194.435$ THz (longitudinal mode of the FP-ILL).

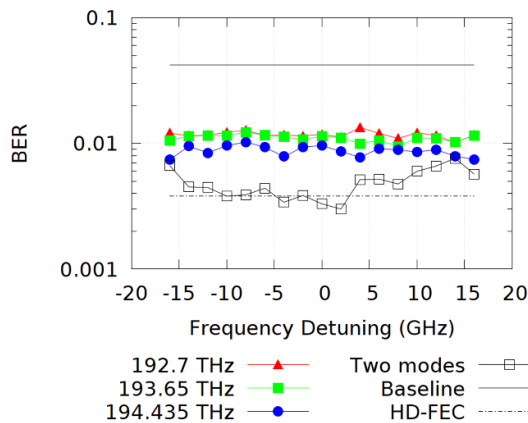


Figure 4. The performance of the evaluated systems in terms of the BER versus the frequency detuning between the optical modes of the FP-ILL and the tuneable laser source. The BER (a) for the baseline where a FFE is directly applied to the down-sampled received signal (continuous line), (b) for the TDELM that uses the optical injection around the optical modes of the FP-ILL located at $\nu_1 = 192.7$ THz (red triangle), $\nu_2 = 193.65$ THz (green square), $\nu_3 = 194.435$ THz (blue dot), (c) for the dual-mode case (white square). Alongside these results, the HD-FEC limit is depicted (dashed line).

neuron count for the same T and boosting performance in terms of accuracy and latency. However, the processing latency was still high as it was dictated by the product of $N \times K \times \theta$. To relax the processing latency constraint, the authors in [19] proposed the parallel insertion of symbols which may accelerate the performance of the network but has a detrimental impact on accuracy primarily due to intersymbol inference.

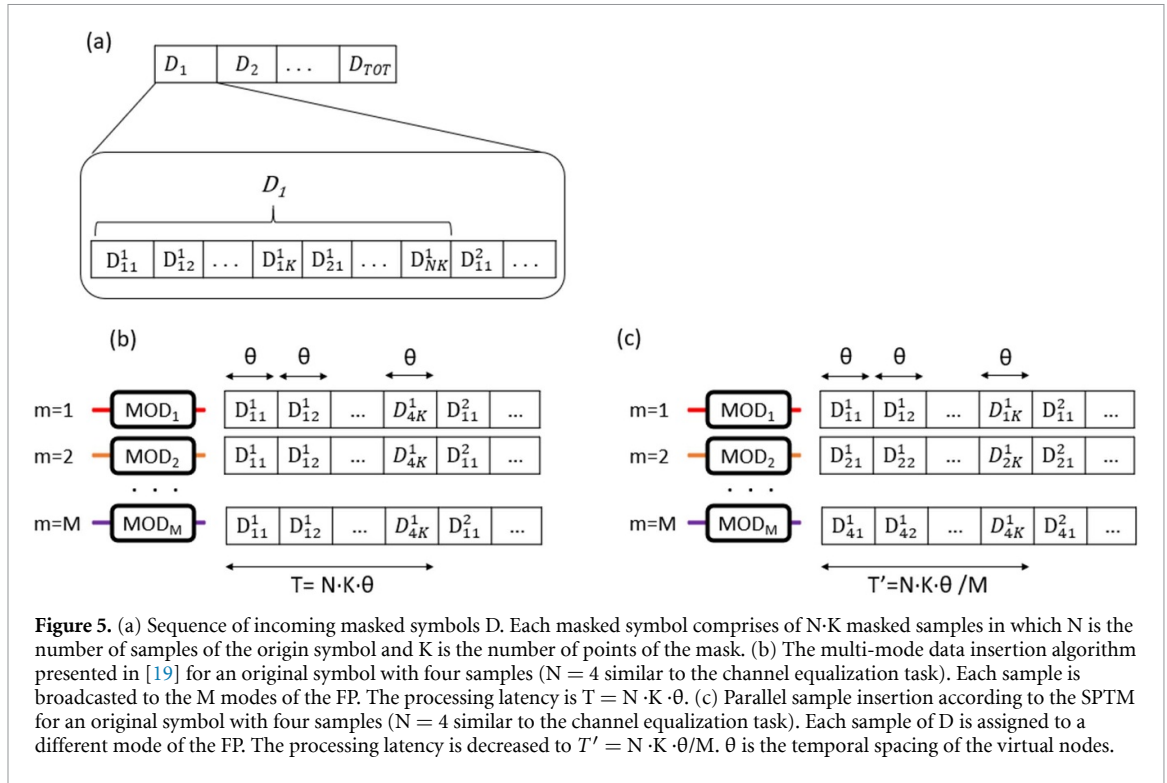


Figure 5. (a) Sequence of incoming masked symbols D . Each masked symbol comprises of $N \cdot K$ masked samples in which N is the number of samples of the origin symbol and K is the number of points of the mask. (b) The multi-mode data insertion algorithm presented in [19] for an original symbol with four samples ($N = 4$ similar to the channel equalization task). Each sample is broadcasted to the M modes of the FP. The processing latency is $T = N \cdot K \cdot \theta$. (c) Parallel sample insertion according to the SPTM for an original symbol with four samples ($N = 4$ similar to the channel equalization task). Each sample of D is assigned to a different mode of the FP. The processing latency is decreased to $T' = N \cdot K \cdot \theta / M$. θ is the temporal spacing of the virtual nodes.

To drastically address the latency issue of TDRCs, we propose the SPTM data insertion technique which allows the parallel insertion of data in the processor. This is accomplished by assigning different samples of a specific symbol to different modes of the FP, exploiting for instance a typical serial to parallel converter. When the SPTM technique is applied then an interesting differentiation takes place.

Firstly, different samples of the same symbol interact within the same feedback loop, thus exploiting intersample interaction and boosting the performance of the network through the nonlinear interplay of longitudinal modes. Secondly, although each sample remains at the input of the corresponding mode for the same time interval (θ), the processing latency is shrunk by M times due to parallel insertion of symbols. Therefore external cavity delay restrictions are relaxed. Based on the above, this new data insertion scheme will in principle affect the processing latency of the network by using a significant shorter external cavity, leading to reduced footprint and propagation losses. This concept is depicted in figure 5(c).

In order to validate these claims an FP TDRC was numerically simulated. The model is based on a set of $M + 1$ rate equations where M equations describe the output of the electrical field of each FP mode. Equations (2) and (3) account for the electrical carriers [19]:

$$\dot{E}_m = \frac{1 + ia}{2} \left[G_m(t) - \frac{1}{t_{ph}} \right] E_m(t) + \frac{k_f}{t_{rt}} E_m(t - T) e^{i\omega_m T} + \frac{k_{inj}}{t_{rt}} E_{inj}(t) e^{-i\Delta\omega_m t} + \sqrt{2\beta N(t)} \xi(t) \quad (1)$$

$$\dot{N} = \frac{I}{q} - \frac{N(t)}{t_n} - \sum_{1-M/2}^{M/2} G_m(t) |E_m(t)|^2 \quad (2)$$

$$G_m(t) = \frac{g[N(t) - N_0]}{1 + s \sum_{1-M/2}^{M/2} |E_m(t)|^2} \left[1 - \left(k \frac{\Delta f_L}{\Delta f_g} \right)^2 \right]. \quad (3)$$

Here, m is the index of the longitudinal mode, whereas E_m , ω_m and G_m are the envelope of the complex electrical field, oscillation frequency and gain of the m^{th} mode respectively. The total number of carriers inside the cavity are $N(t)$, $\xi(t)$ is the spontaneous emission process modelled as a complex gaussian process. The remaining parameters are the line enhancement factor ($\alpha = 3$), the photon lifetime ($t_{ph} = 2$ ps), the length of the external loop (T), the round trip cavity time (t_{rt}), the injected field from ISG FP to ILL FP (E_{inj}), the spontaneous emission rate ($\beta = 1.5 \times 10^{-10}$ ps⁻¹), the bias current (I) whose value is determined by the number of the modes, the electron charge ($q = 1.6 \times 10^{-19}$), the gain saturation coefficient ($s = 5 \times 10^{-7}$),

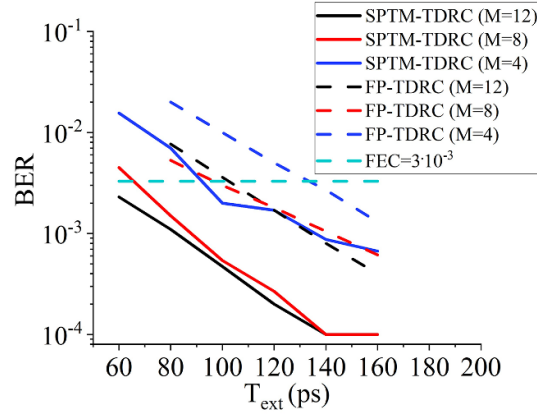


Figure 6. Evaluation of the performance of the SPTM-TDRC and a classical TDRC scheme based on a FP laser (FP-TDRC) for various lengths of the external loop and number of longitudinal modes.

the free spectral range ($\Delta f_L = 125$ GHz), the gain bandwidth ($\Delta f_g = 10$ THz) and the differential gain parameter ($g = 1.2 \times 10^{-8} \text{ ps}^{-1}$). In our model coherent mode mixing phenomena are neglected assuming high free-spectral-range [41, 42]. Based on this model a complete parameter scan was performed so as to investigate the optimum operational values for the optical injection strength ($k_{inj} = 0.75$) and for the optical feedback strength ($k_f = 0.01$). It is worth mentioning that in the case of TDELM k_f is set to zero.

Similar to the experimental investigation described above and [19], the architecture of the proposed FP-TDRC is based on two FP lasers in a ISG-ILL configuration, where the M modes of the ISG are modulated by the incoming data (figure 1(b)). The masked samples are imprinted on the phase of the M modes. The modulated outputs of the ISG are injected to the ILL whose output is divided in two different routes; the weaker output of the ILL is used as self-feedback through the external loop whereas the stronger output is directed to a set of M photodiodes (PD) which monitor the optical power of the M longitudinal modes. After M analog-to-digital converters the digitized outputs are fed to either a standard equalizer FFE equipped with 24 memory taps.

For benchmark reasons we chose as an evaluation task the retrieval of a 28 Gbaud PAM4 signal transmitted through a 50 km single mode optical fiber at 1550 nm. Transmission was again numerically simulated through the well-known split-step Fourier method governed by Manakov equations [35]. The transmitted optical signal was numerically detected by a 50 GHz photodiode and 100 Gsample s^{-1} analog-to-digital converter (ADC) which permits the representation of each received PAM-4 symbol by four samples.

Data wise, we simulated the transmission of 100 000 PAM-4 symbols, among which, 50 000 were used for training and 50 000 for testing. The standard metric in the evaluation process was the BER through error counting, whereas the performance target is the HD-FEC BER = 3×10^{-3} (figure 6 cyan dashed line).

In figure 6 the evaluation of the performance of the FP-TDRC is presented versus the length of the external loop T and versus the number of longitudinal modes ($M = 4, 8, 12$). Moreover, an evaluation of the performance of [19] (figure 6 FP-TDRC), for shorter external lengths is demonstrated for a direct comparison between simple FP-TDRC and SPTM assisted TDRC (SPTM-TDRC). The distribution of the samples is done with the following order. When $M = 4$, each sample is imprinted on one mode of the FP, whereas for $M = 8$ and $M = 12$ each sample is imprinted on two and three adjacent modes, respectively. Consequently, for $M = 8$ the first sample is carried on modes -3 and -2 , the second sample on modes -1 and 0 and so on whereas for $M = 12$ the first sample is imprinted on -5 , -4 and -3 , the second sample on -2 , -1 and 0 etc. The increase of the modes from four (figure 6 blue solid line) to eight (figure 6 red solid line) has a beneficial impact on the performance obviously due to the higher number of nodes whereas further increase results in a marginal improvement (figure 6 solid black line). The best performance of the SPTM-TDRC is achieved for low detuning values ($\Delta f \leq 2$ GHz), whereas in the vast majority of cases, SPTM-TDRC offers HD-FEC compatible performance. Specifically, figure 6 shows that HD-FEC performance is possible even when $T = 60$ ps for $M = 12$ ($\text{BER} = 10^{-3}$) which supports $N_V = 36$ nodes and inserts an SP = 1.68. Increasing T up to 140 ps improves performance reaching up to $\text{BER} = 10^{-4}$ for $N_V = 56$ nodes whereas conventional FP-TDRC [19] requires $T = 240$ ps for similar performance. The low performance of the SPTM-TDRC for $T = 60$ ps for $M = 4, 8$ is attributed to the low number of neurons ($N_V = 12$ and $N_V = 24$ respectively). The most significant result of our analysis is that the SPTM-TDRC achieves HD-FEC compatibility with $T = 60$ ps (SP = 1.68) and $N_V = 36$ nodes ($M = 12$) whereas

sequential data insertion techniques (FP-TDRC) require $T = 160$ ps ($SP = 4.48$) and $N_V = 32$ ($M = 4$) or $T = 120$ ps ($SP = 3.36$) and $N_V = 48$ nodes ($M = 8$). The use of shorter cavity lengths is of paramount importance since it simultaneously allows a reduction in the number of nodes and in the SP. This fact relaxes the major restriction of the TDRC, which is their processing latency, and paves the way towards their use in real time applications.

5. FP TDELM and SPTM for image classification

In this section we perform numerical simulations in order to study the extension of the original FP-TDRC/TDELM to four pairs of ISG-ILL TDELMs, where each pair is used to process a redistributed version of the original image. In detail, the operation of each ISG-ILL pair is similar to the PAM-4 case presented above but there are some key differences (figure 1(c)). Firstly, data are imprinted on the amplitude of the M modes, instead of the phase. Secondly, no external loop was used (TDELM configuration). Lastly, a single PD is used to monitor the output of the M modes of the ILL, thus signals are summed incoherently. The output of the PD is driven to a low pass filter (LPF) (figure 1(c)) and the filtered signal is sampled with a rate BW which is equal to the bandwidth of the LPF which determines the nodes count. The transfer function of the overall system comprising the FP response, the photodiode square law and LPF's spectral profile has the ability to operate as a convolutional filter and extract features mostly focusing on identifying correlations between adjacent temporal data (pixels). The correlation properties of the overall subsystem are a result of the interplay of different modes inside the FP and the integration (averaging) functionality offered by the LPF at the output.

Every pair of FP ILL is injected with pixels of the same image, but each ILL receives a rearranged sequence which facilitates the correlation mechanism. Moreover, each pixel is kept for $T_{pix} < 100$ ps in order to operate the FP in the transient regime. Due to the transient dynamics, the outputs of the laser are determined not only by the present inputs but also by the sequence of the previous ones. Although present inputs primarily affect the output of the modes, previous inputs have also an influence on the outputs where most recent pixels have greater impact due to the integration capabilities of the FP laser [43]. Consequently, apart from weighting related to the impulse response of each FP mode, a nonlinear correlation of adjacent pixels is also achieved by leveraging the transient response of the ILL. In addition, through the mode competition mechanism and the same carrier reservoir of the FP, the M pixels which are assigned to the M modes interact with one another and enhance the diversity of the incoming data. Finally, the processing of the outputs is performed on a fully connected layer (FCL).

Same as in the PAM-4 case, data insertion in this scenario is performed via the SPTM technique for four different redistributions of the original image. The redistributions of the pixels are equivalent to the scanning of the original image in four different orientations which are presented in figure 7. In particular, the 28×28 pixel images are transformed in four $M \times (784/M)$ pixel arrays where each array represents the input to the corresponding FP. Since a FP laser with $M = 4$ was simulated for the image classification task, the redistributed images are actually arrays comprising 4×196 pixels. The distribution of pixels is carried out in four different ways which are named as directions 1, 2, 3 and 4 (figure 7(c)). The different directions provide different time-series and contribute to revealing the correlation between adjacent pixels in different levels (vertical/horizontal). At first, the initial 28×28 pixel image is divided in 7×7 kernels and each kernel consists of $4 \times 4 = 16$ pixels. The 16 pixels of the kernel enter the FP in a Direction based order. In particular, for directions 1 and 2 (figure 7(a)) the 4×4 pixel kernel is divided in four 2×2 pixel sub-kernels whereas at directions 3 and 4 the sub-kernels consist of 4×1 pixels which is equivalent to a column of the kernel (figure 7(b)). Each sub-kernel is imprinted on a specific mode of the FP and they are processed in parallel by the FP TDELM which means that a total $4T_{pix}$ is needed for the processing of each kernel. After that, the next kernel that will enter the scheme is determined by the scanning direction. Consequently, directions 1 and 3 scan the image in a horizontal manner, whereas the directions 2 and 4 scan the image vertically. The difference between direction 1 and 2 is spotted on the orientation since direction 1 scans every line of kernels from left to right whereas direction 2 scans odd lines from left to right and even lines from right to left. The same principle is applied to direction 3 and 4 as directions 3 scans all the column from top to bottom whereas direction 4 scans the odd columns from top to bottom and even columns from bottom to top. Therefore, the four scanning directions create different sequences of pixels for the same image. These different sequences combined with the transient multi-modal dynamics of the FP laser and the integration properties of the LPF produce different depictions of the image so as to facilitate feature extraction through temporal correlations and hence classification. We must stress the fact that through in the pre-processing of the data no masking procedure takes place as in classical image classification tasks dimensionality reduction through feature extraction is the case, in contrast to RC where dimensionality expansion is required.

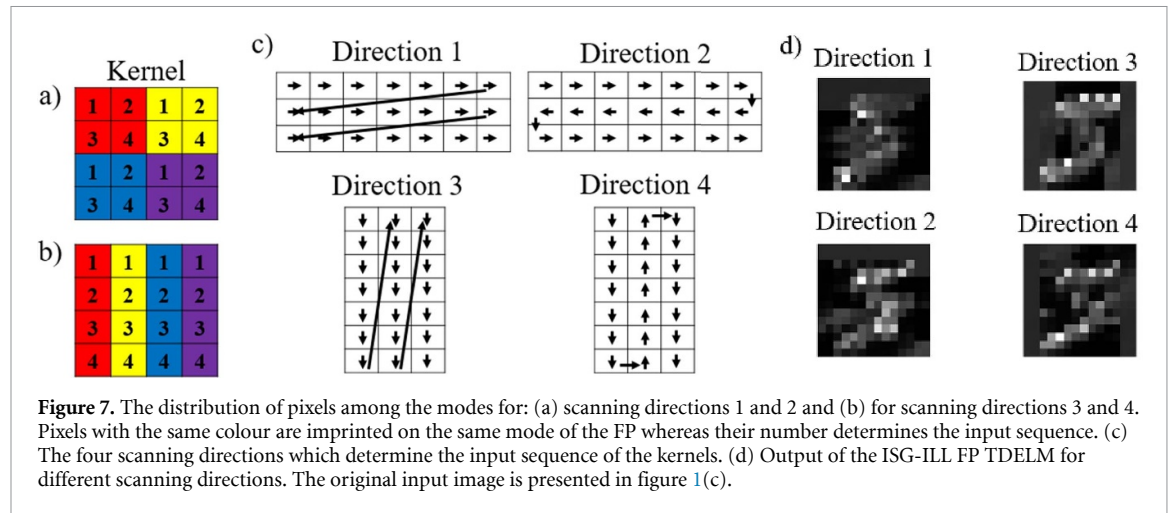


Figure 7. The distribution of pixels among the modes for: (a) scanning directions 1 and 2 and (b) for scanning directions 3 and 4. Pixels with the same colour are imprinted on the same mode of the FP whereas their number determines the input sequence. (c) The four scanning directions which determine the input sequence of the kernels. (d) Output of the ISG-ILL FP TDELm for different scanning directions. The original input image is presented in figure 1(c).

Table 1. Performance of the STM-FP TDELm in MNIST.

R \ BW				
	10 Gps	30 Gps	50 Gps	Inputs to FCL
R	94.90%	95.50%	95.95%	784
R/2	93.23%	94.95%	95%	392
R/4	93.05%	93.70%	94.32%	196

For the training of the proposed scheme, 49 000 images were used for training, 7000 for validation and 14 000 for testing. The FCL is trained by applying an adaptive stochastic gradient descent procedure with an initial learning rate equal to 7, which is decreased during training by a factor ranging from 0.65–0.72.

The evaluation of the SPTM-FP TDELm was performed for three different processing rates R ($R = 10, 30$ and 50 Gpixels per second) and for three different BW of the LPF ($BW = R, R/2$ and $BW = R/4$). The number of outputs was regulated by BW and was 784 for $BW = R$, 392 for $BW = R/2$ and 196 for $BW = R/4$.

The calculated performance of the SPTM-FP TDELm is presented in detail in table 1. From table 1 it is clear that as BW decreases so does the accuracy of the network for all R . This can be attributed to two factors: the first one is that smaller BW is equivalent to a higher compression since a wider area of the initial image is summed and averaged at the PD. Moreover, the value of BW determines the maximum possible sampling rate of the output. When $BW = R$ then every 4-pixel input is matched to a single value which means that the original image is compressed by a factor of 4. Consequently, smaller BW causes a higher compression of the image which results in an unavoidable accuracy loss.

The best performance for reduced BW ($R/2$ and $R/4$), and therefore for a smaller number of inputs, was 95% at 50 Gps which resulted in a processing time of $196 \cdot 20 = 3.92$ ns per image (255.1 MHz). For $BW = R$ the best performance was 95.95% at 50 Gps with an equivalent processing rate. In other state-of-the-art schemes accuracy similar performance was achieved but with more complicated architectures [44, 45], whereas higher performance also requires extra pre-processing of the data and very complicated structures with thousands of neurons [46]. In terms of processing rate, to the best of our knowledge the proposed FP-TDELm surpasses previous implementations relying RC inspired techniques due to the fast and parallel insertion of data achieving a processing time of 3.92 ns per image whereas other works require more than 17.1 ns for the same task [45, 46].

6. Conclusion

This work validates numerically and experimentally the use of FP lasers as photonic accelerators in various tasks and configurations. By exploiting multimode emission and an alternative STM data insertion strategy, overall footprint and processing latency are reduced by a factor equal to the number of utilized longitudinal modes. Therefore, increased classification accuracy and processing speed were achieved for two independent tasks: PAM-4 channel equalization and image classification using the MNIST dataset. The experimental and numerical findings pave the way for implementing this approach to other TDRC platforms such as spin VCSELs and quantum dot lasers. Finally, the reduction of the speed penalty associated with such neuromorphic hardware-friendly topologies is important, because it can unlock their usage in demanding real-world, high-speed tasks.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Funding

This work was supported from the EU H2020 NEoteRIC project under Grant Agreement (871330) and EU HORIZON PROMETHEUS project under Grand agreement (101070195).

ORCID iD

Charis Mesaritis  <https://orcid.org/0000-0002-0072-3116>

References

- [1] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44
- [2] Mahapatra N R and Venkatrao B V 1999 The processor-memory bottleneck: problems and solutions *Xrds* **5** 2–es
- [3] Abu-Mostafa Y S 1989 Information theory, complexity and neural networks *IEEE Commun. Mag.* **27** 25–28
- [4] Anon 2020 Beyond von Neumann *Nat. Nanotechnol.* **15** 507
- [5] Tanaka G, Yamane T, Héroux J B, Nakane R, Kanazawa N, Takeda S, Numata H, Nakano D and Hirose A 2019 Recent advances in physical reservoir computing: a review *Neural Netw.* **115** 100–23
- [6] Lecun Y, Bottou L, Bengio Y and Haffner P 1998 Gradient-based learning applied to document recognition *Proc. IEEE* **86** 2278–324
- [7] Radford N 2012 *Bayesian Learning for Neural Networks* vol 118 (Springer Science & Business media)
- [8] Jalalvand A, De Neve W, Van de Walle R and Martens J-P 2016 Towards using reservoir computing networks for noise-robust image recognition *2016 Int. Joint Conf. on Neural Networks (IJCNN) (Vancouver Canada)* (IEEE) pp 1666–72
- [9] Tong Z and Tanaka G 2018 Reservoir computing with untrained convolutional neural networks for image recognition *2018 24th Int. Conf. on Pattern Recognition (ICPR)* (IEEE) pp 1289–94
- [10] Torreon J et al 2017 Neuromorphic computing with nanoscale spintronic oscillators *Nature* **547** 428–31
- [11] Nakane R, Tanaka G and Hirose A 2018 Reservoir computing with spin waves excited in a garnet film *IEEE Access* **6** 4462–9
- [12] Eder M, Hisch F and Hauser H 2018 Morphological computation-based control of a modular, pneumatically driven, soft robotic arm *Adv. Robot.* **32** 375–85
- [13] Degraeve J, Caluwaerts K, Dambre J and Wyffels F 2015 Developing an embodied gait on a compliant quadrupedal robot *2015 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS) 2015 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)* (IEEE) pp 4486–91
- [14] Soriano M C, Ortin S, Keuninckx L, Appeltant L, Danckaert J, Pesquera L and van der Sande G 2015 Delay-based reservoir computing: noise effects in a combined analog and digital implementation *IEEE Trans. Neural Netw. Learn. Syst.* **26** 388–93
- [15] Haynes N D, Soriano M C, Rosin D P, Fischer I and Gauthier D J 2015 Reservoir computing with a single time-delay autonomous Boolean node *Phys. Rev. E* **91** 020801
- [16] Tran S J D and Teuscher C 2017 Memcapacitive reservoir computing *2017 IEEE/ACM Int. Symp. on Nanoscale Architectures (NANOARCH) 2017 IEEE/ACM Int. Symp. on Nanoscale Architectures (NANOARCH)* (IEEE) pp 115–6
- [17] Miller D A B 2000 Rationale and challenges for optical interconnects to electronic chips *Proc. IEEE* **88** 728–49
- [18] Prucnal P R, Shastri B J, Ferreira de Lima T, Nahmias M A and Tait A N 2016 Recent progress in semiconductor excitable lasers for photonic spike processing *Adv. Opt. Photon.* **8** 228
- [19] Bogris A, Mesaritis C, Deligiannidis S and Li P 2021 Fabry-perot lasers as enablers for parallel reservoir computing *IEEE J. Sel. Top. Quantum Electron.* **27** 1–7
- [20] Antonik P, Hermans M, Duport F, Haelterman M and Massar S 2016 *Towards Pattern Generation and Chaotic Series Prediction with Photonic Reservoir Computers SPIE LASE* ed B Jalali, S K Turitsyn, D R Solli and J M Dudley (SPIE) p 97320B
- [21] Nguimdo R M, Verschaffelt G, Danckaert J and der Sande G V 2015 Simultaneous computation of two independent tasks using reservoir computing based on a single photonic nonlinear node with optical feedback *IEEE Trans. Neural Netw. Learn. Syst.* **26** 3301–7
- [22] Héroux J B, Numata H, Kanazawa N and Nakano D 2018 Optoelectronic reservoir computing with VCSEL *2018 Int. Joint Conf. on Neural Networks (IJCNN) 2018 Int. Joint Conf. on Neural Networks (IJCNN)* (IEEE) pp 1–6
- [23] Paquot Y, Duport F, Smerieri A, Dambre J, Schrauwen B, Haelterman M and Massar S 2012 Optoelectronic reservoir computing *Sci. Rep.* **2** 287
- [24] Vatin J, Rontani D and Sciamanna M 2019 Experimental reservoir computing using VCSEL polarization dynamics *Opt. Express* **27** 18579
- [25] Brunner D, Soriano M C, Mirasso C R and Fischer I 2013 Parallel photonic information processing at gigabyte per second data rates using transient states *Nat. Commun.* **4** 1364
- [26] Lupo A, Butschek L and Massar S 2021 Photonic extreme learning machine based on frequency multiplexing *Opt. Express* **29** 28257
- [27] Butschek L, Akrouf A, Dimitriadou E, Lupo A, Haelterman M and Massar S 2022 Photonic reservoir computer based on frequency multiplexing *Opt. Lett.* **47** 782
- [28] Katumba A, Freiburger M, Bienstman P and Dambre J 2017 A multiple-input strategy to efficient integrated photonic reservoir computing *Cogn. Comput.* **9** 307–14
- [29] Mesaritis C, Papataxiarhis V and Syvridis D 2013 Micro ring resonators as building blocks for an all-optical high-speed reservoir-computing bit-pattern-recognition system *J. Opt. Soc. Am. B* **30** 3048
- [30] Argyris A, Bueno J and Fischer I 2018 Photonic machine learning implementation for signal recovery in optical communications *Sci. Rep.* **8** 1–13
- [31] Estébanez I, Schwind J, Fischer I and Argyris A 2020 Accelerating photonic computing by bandwidth enhancement of a time-delay reservoir *Nanophotonics* **9** 184

- [32] Skontranis M, Sarantoglou G, Bogris A and Mesaritakis C 2022 Time-delayed reservoir computing based on a dual-waveband quantum-dot spin polarized vertical cavity surface-emitting laser *Opt. Mater. Express* **12** 4047–60
- [33] Harkhoe K, Verschaffelt G, Katumba A, Bienstman P and Der Sande G V 2020 Demonstrating delay-based reservoir computing using a compact photonic integrated chip *Opt. Express* **28** 3086
- [34] Pierangeli D, Marcucci G and Conti C 2021 Photonic extreme learning machine by free-space optical propagation *Photon. Res.* **9** 1446
- [35] Marcuse D, Manyuk C and Wai P 1997 Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence *J. Lightwave Technol.* **15** 1735–46
- [36] Soriano M C, Brunner D, Escalona-Morán M, Mirasso C R and Fischer I 2015 Minimal approach to neuro-inspired information processing *Front. Comput. Neurosci.* **9** 68
- [37] Owen-Newns D, Robertson J, Hejda M and Hurtado A 2022 Photonic spiking neural networks with highly efficient training protocols for ultrafast neuromorphic computing systems *Intell. Comput.* **2** 0031
- [38] Nikas T, Bogris A and Syvridis D 2015 Two-mode injection-locked FP laser receiver: a regenerator for long-distance stable fiber delivery of radio-frequency standards *Opt. Lett.* **40** 886
- [39] Jaeger H 2001 The “echo state” approach to analysing and training recurrent neural networks-with an erratum note *Bonn, : German National Research Center for Information Technology GMD Technical Report* vol 148 p 13
- [40] Tzimpragos G, Kachris C, Djordjevic I B, Cvijetic M, Soudris D and Tomkos I 2014 A survey on FEC codes for 100 G and beyond optical networks *IEEE Commun. Surv. Tutor.* **18** 209–21
- [41] Horer J and Patzak E 1997 Large-signal analysis of all-optical wavelength conversion using two-mode injection-locking in semiconductor lasers *IEEE J. Quantum Electron.* **33** 596–608
- [42] Lenstra D and Yousefi M 2014 Rate-equation model for multi-mode semiconductor lasers with spatial hole burning *Opt. Express* **22** 8143
- [43] Ferrera M, Park Y, Razzari L, Little B E, Chu S T, Morandotti R, Moss D J and Azaña J 2010 On-chip CMOS-compatible all-optical integrator *Nat. Commun.* **1** 29
- [44] Feldmann J, Youngblood N, Karpov M, Gehring H, Li X, Stappers M, Le Gallo M, Fu X, Lukashchuk A and Raja A S 2021 Parallel convolutional processing using an integrated photonic tensor core *Nature* **589** 52–58
- [45] Nakajima M, Tanaka K and Hashimoto T 2021 Scalable reservoir computing on coherent linear photonic processor *Commun. Phys.* **4** 20
- [46] Huang Y, Zhou P, Yang Y, Chen T and Li N 2021 Time-delayed reservoir computing based on a two-element phased laser array for image identification *IEEE Photon. J.* **13** 1–9