IOP Publishing J. Phys. Photonics 3 (2021) 014002 https://doi.org/10.1088/2515-7647/abcb72

Journal of Physics: Photonics

PAPER

OPEN ACCESS

RECEIVED
16 September 2020

ACCEPTED FOR PUBLICATION
29 October 2020

PUBLISHED
15 December 2020

Multi-channel optical neuromorphic processor for frequency-multiplexed signals

Mariia Sorokina

Aston University, Aston Institute of Photonic Technologies, Birmingham B4 7ET, United Kingdom

E-mail: m.sorokina@aston.ac.uk

Keywords: neuromorphic computing, signal processing, non-linear optics

Abstract

Here we develop the first optical neuromorphic processor for frequency-multiplexed and multi-channel signals and demonstrate its application for orthogonal frequency-division multiplexing and wavelength-division multiplexing. The presented architecture supports multichannel operation through incorporation of the optical Fourier transform and dispersion-managed fiber echo state network analogue enabling processing of dual-quadrature high bandwidth signals. We demonstrate applicability of the design for prediction and equalization tasks. The proposed technology paves the way to multi-channel neuromorphic signal processing.

1. Introduction

The growing surge in optical neuromorphic computing [1] is driven by the emerging artificial intelligence (AI) and Internet of things (IoT) applications, which require fast signal processing and flexible design. Reservoir Computing, in particular echo state networks (ESN) [2], enable optical implementation via a non-linear delay dynamical systems, with underlying materials ranging from semiconductor laser [3] to silicon [4]. Fiber ESN analogue (FESNA) [5] and dispersion-multiplexed DM-FESNA [6] offer compatibility with current transmission networks. Moreover, it is the first design to enable dual-quadrature and high bandwidth neuromorphic signal processing. The existing neuromorphic systems are limited to a single-channel operation, while typical transmission systems are multi-channel, such as orthogonal frequency-division multiplexing (OFDM) [7, 8] or wavelength division multiplexing (WDM) [9]. In particular, OFDM and OFDM-based waveforms have broad applications [10, 11], which are common in 4G LTE and 5G. Furthermore, current trends, including 5G and IoT, envision even more applications for signal multiplexing. Moreover, the emerging technologies, such as massive connectivity and spectrum slicing, require flexibility in spectrum allocation, which poses a requirement for neuromorphic systems to operate frequency-multiplexed signals, while current neurmorphic systems are typically designed for time-multiplexing. Thus, there is a demand for neuromorphic technologies capable for processing frequency-multiplexed signals and support multi-channel operation.

Here we propose and examine the first neuromorphic technology, which enables processing of multi-channel and frequency/wavelength-multiplexed signals and demonstrate its application for OFDM and WDM signals. The design augments an established DM-FESNA technology with the fiber-based optical Fourier transform (OFT) [12–14], which performs time-frequency conversion optically. Both DM-FESNA and OFT are fiber-based technologies and easily compatible. Moreover DM-FESNA offers dual-quadrature high bandwidth signal processing, while OFT enables an adjustment of signal rate on demand. By varying dispersion and chirp one can adapt the scheme for processing signals with different bandwidth values, making it applicable for systems with fragmented spectrum. The possibility of processing various tasks for different channels is demonstrated here and is applicable for multi-user systems. Thus, the proposed OFT-FESNA is the first neuromorphic technology for multi-channel frequency- and wavelength-multiplexed dual-quadrature high-bandwidth signals.

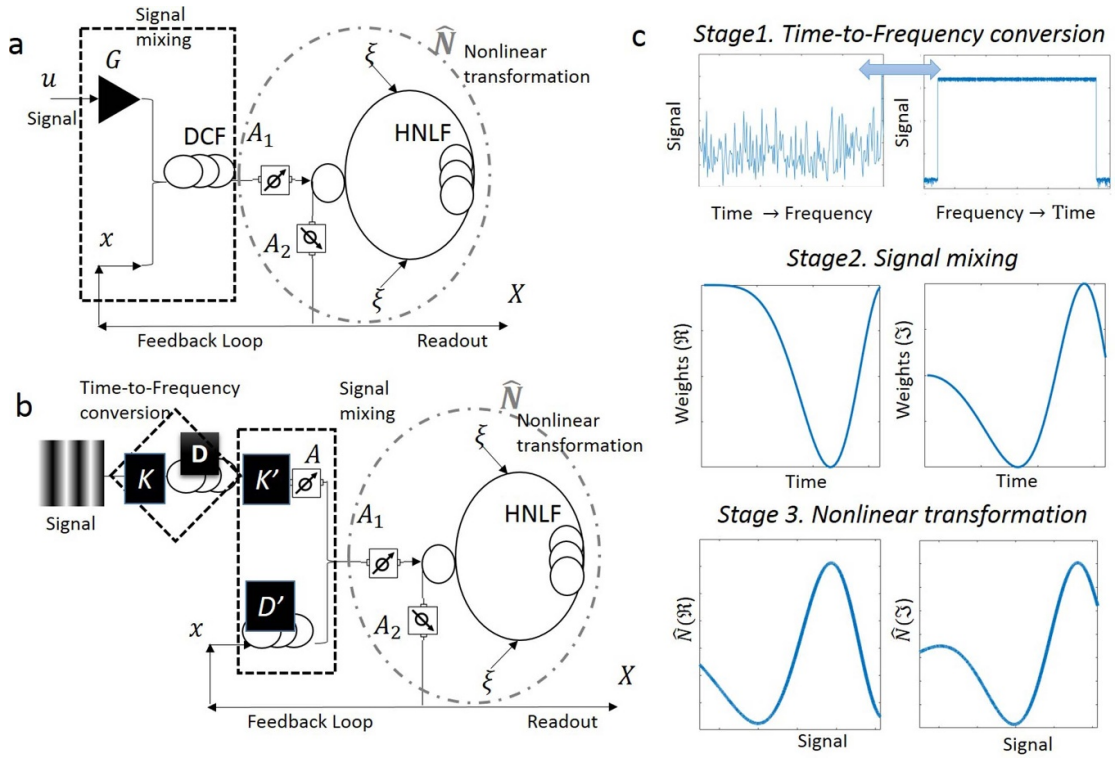

Figure 1. Design. (a) Fiber-optic ESN analogue (FESNA), which encompasses modulated pump, highly non-linear fiber (HNLF), attenuators, amplifier, and dispersion compensating fiber (DCF) and (b) the proposed multichannel scheme augmented by OFT functionality, which adds a chirp (K) and dispersion (D). The schematic representation is shown in panel (c).

2. Design

2.1. Problem formulation

The design incorporates the optical fiber-based neuromorphic technology DM-FESNA [6], see figure 1(a), which combined with the OFT enables multichannel neuromorphic signal processing, see figure 1(b). DM-FESNA realizes mixing of neural network nodes through fiber dispersion (large dashed square figures 1(a) and (b)) and performs non-linear processing via fiber non-linearity (dashed-dotted circle). The design is based on ESN, which relaxes training complexity (training requires only optimization of the output weights at the readout) and consists of a single non-linear element and a feedback delay loop.

Thus, the two major components on the network—signal mixing and non-linear transformation (represented in figures 1(a) and (b)) by a dashed square and dashed-dotted circle correspondingly) are both realized by incorporating fiber properties. The signal mixing mimics the synaptic interactions, where the neural nodes are represented by the samples $\mathbf{u}_n$ of the input signal $\mathbf{u}$ and are mixed between themselves and a gradually building feedback $\mathbf{x}_n$ due to applied chirp and dispersion. Afterwards, the signal undergoes a non-linear transformation denoted by a circle in figure 1(a) and (b). The obtained output is fed back to the system $\mathbf{x}_{n+1}$ and collected at the readout forming an augmented state matrix $\mathbf{X} = [\ldots \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots]$. The training is performed off-line via a linear regression or pseudo-inversion applied to the state matrix $\mathbf{X}$, thus, obtaining the optimum output weights $\mathbf{W}^{out}$.

We augment the scheme by incorporating OFT. As a result, a multichannel OFDM/WDM signal can be transformed into a serial temporal sequence with appropriate time/frequency-scaling and, thus, preserving the transmission rate, see figure 1(b). Conversion back to the original format is not necessary, as the transformed signal goes to the receiver, where the temporal sequence is collected and processed, while the feedback signal is coupled with upcoming OFT transformed input signal.

The scheme can be represented in three stages: (i) OFT to transform a signal into a serial temporal sequence with an appropriate time/frequency-scaling (see dashed rhombus of the design in figure 2(b), which realizes the time-to-frequency conversion illustrated in the top of figure 2(c)), then (ii) fiber dispersion for the signal mixing (dashed square in figure 2(b) and the corresponding dispersion-induced weights variation in the centre of figure 2(c)), and finally (iii) fiber non-linearity for the neural response (dashed-dotted circle in figure 2(b) and the resulting non-linear response for real/imaginary signal components at the bottom of figure 2(c)).

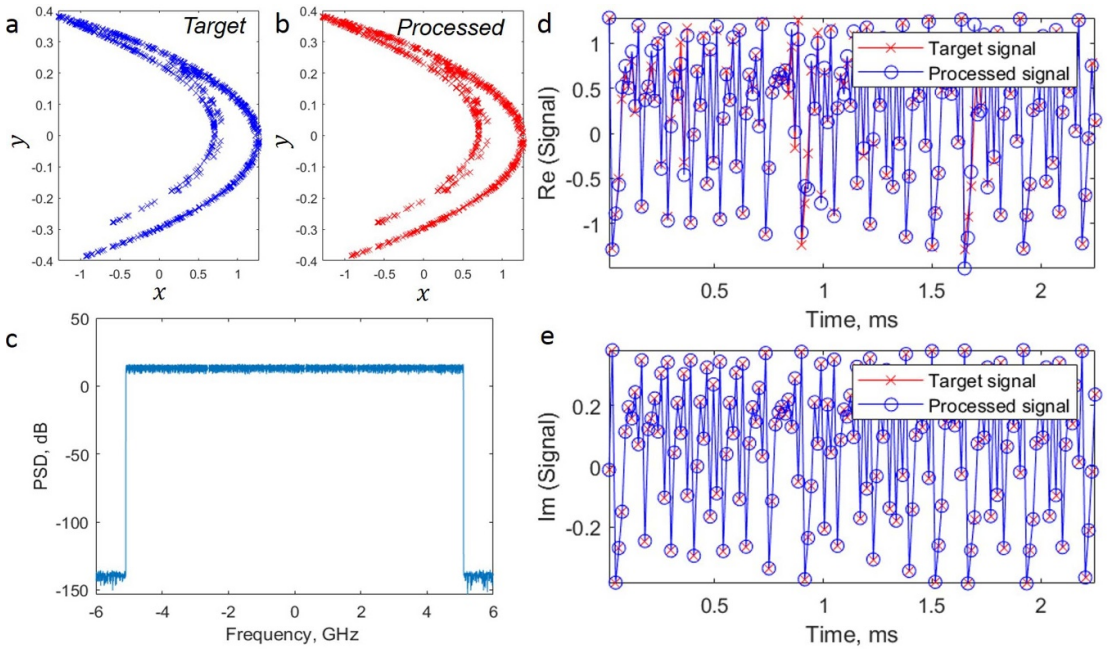

Figure 2. Single task operation: (a) target Hénon map (Set1) and (b) processed symbols modulated to (c) 10 GHz bandwidth OFDM signal. The time-domain target (red crosses) and processed (blue circles) signals are plotted in panels (d)–(e) depicting real and imaginary parts correspondingly.

2.2. Stage 1: Time-Frequency conversion

To achieve frequency-time conversion OFT is applied to temporal sequence converting the wavelength- or orthogonal frequency- division multiplexed (WDM/OFDM) signal into a serial temporal-multiplexed signal. OFT preserves signal coherency and is format transparent enabling numerous application, among which signal multiplexing and demultiplexing as well as multi-channel regeneration [14]. OFT is realized through a sequential application of a chirp K and dispersion D.

Assuming the signal $u(t)$ (having spectrum $U(\Omega)$) is undergoing the transformation, the OFT results in:

$$\widehat{\mathbf{OFT}}[u(t)] = \sqrt{\frac{i}{2\pi D}} \int_{-\infty}^{\infty} u(t') \exp\left(i\frac{K}{2}t'^2 - i\frac{(t-t')^2}{2D}\right) \tag{1}$$

Under condition $K = 1/D$, the OFT transforms the signal into its chirped and scaled spectrum:

$$\widehat{\mathbf{OFT}}[u(t)] = \sqrt{\frac{i}{2\pi D}} \exp\left(-i\frac{t^2}{2D}\right) U\left(\frac{t}{D}\right) \tag{2}$$

If required one can change the scaling size by varying the parameters K and D. Thus, the advantage of OFT is that the time-to-frequency conversion (and inverse operation) also enables variable scaling and the output in time-domain is proportional to the signal spectrum before OFT with the temporal dimension related to the original frequency as $\Delta t \to D\Delta\Omega$. Therefore, by choosing parameters (K and D) one can ensure the signal rate to remain the same or increase/decrease as required.

2.3. Stage 2: Signal mixing

Then the signal mixing is performed. Here we use a chirp and dispersion for the input and feedback signals correspondingly. The chirp induces variation of the signal samples:

$$\widehat{\mathbf{K}}'[u(t)] = u(t) \exp\left(i\frac{K'}{2}t^2\right). \tag{3}$$

Here and in the figure 2(b) we consider it as a separate function and denote it by K' for convenience purposes, one can consider the chirp K' as a part of the initial chirp $K = 1/D + K'$. For the feedback signal we utilize the dispersion D' of the DCF fiber, which induces signal mixing:

$$\hat{\mathbf{D}}'[U(t)] = \sqrt{\frac{i}{2\pi D'}} \int_{-\infty}^{\infty} u(t') \exp\left(-i\frac{(t-t')^2}{2D'}\right). \tag{4}$$

In many cases the residual chirp in equation (2) may be sufficient, thus $K' = 0$, while $D' = D = K^{-1}$. In general one can use additional chirp K' and dispersion D' to optimize weights.

The achieved masking results in the sine-modulated weights applied to real and imaginary components of the signals and induces mixing between the components. The weight matrices can be varied by changing the dispersion value resulting in higher oscillations—faster change of weight from symbol to symbol or lower oscillations—smoother mixing of symbols. To enforce the difference between the weight matrices one can change either the chirp/dispersion of K'/D' or by varying the attenuation A of the input signal. Note, the arbitrary chosen reservoir weights is an important advantage of reservoir computing and ESN, as only output weights need to be trained.

2.4. Stage 3: Nonlinear transformation

At the next stage the coupled input and feedback signals $s = u + x$ undergo the non-linear transformation, which is realized through non-linear phase shift induced by the Kerr-non-linearity. To apply the transformation to each quadrature (real and imaginary parts) independently, we split the signal in two copies, one of which acquires a $\pi/2$ phase shift, by a 3 dB coupler with the transformation matrix:

$$\hat{\mathbf{C}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \tag{5}$$

This results in the two copies of the signal in the upper and bottom branches:

$$v^u = s\sqrt{A_1/2}, \;\; v^b = is\sqrt{A_1/2}; \tag{6}$$

Then signals at each branch are coupled with the strong pump:

$$w^u = v^u + \xi; w^b = v^b + \xi; \tag{7}$$

and transmitted through the HNLF acquiring the phase shift $\widehat{\mathbf{HNLF}}[w] = w\exp(i\gamma L|w|^2)$. Under the condition of strong pump ($P_u \ll P_\xi$) and for parameters:
$\{\xi = 1/\sqrt{2}, \;\; \gamma L = 2\pi, \;\; A_1 = -18.5 \;\; dB, A_2 = -9 \;\; dB\}$, the signal-pump beating under exponent in the first order approximation leaves the product of $\xi\Re[v]$ and $-\xi\Im[v]$ (due to π phase shift asymmetry) in the upper and lower branches of NOLM correspondingly. In the simulations we assumed a standard off-shelf HNLF (with non-linear, dispersion, and attenuation coefficients of 10.5 1/W/km, −1.5ps/nm/km, and 0.8 dB km^{-1}, and a fiber length of $L_{HNLF} = 628$ m). Afterwards, both transformed quadratures are coupled back:

$$\hat{\mathbf{N}} = \sqrt{A_2/2}(\widehat{\mathbf{HNLF}}[v^u] + i\widehat{\mathbf{HNLF}}[v^b]) \simeq \sin(\Re u) - i\sin(\Im u)$$

These are performed with one HNLF in a mirror (non-linear optical loop mirror) configuration. NOLM enables efficient non-linear signal processing with many applications, including regeneration [15–17]. After transformation the signal is fed back and coupled with the new incoming signal and a copy is also collected at the receiver. The reservoir size is represented by the number of samples per symbol at the receiver: $\mathbf{X} = \{\ldots \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots\}$. The obtained samples are then multiplied with the output weights matrix, which is optimized according to each task during the training processes via a linear regression.

Thus, we can process frequency-multiplexed systems due to the OFT-FESNA.

3. Results and discussion

3.1. Signal prediction

Next we illustrate the application of the proposed design for neuromorphic processing, in particular, for prediction tasks. We choose the Hénon Map as one of the most studied dynamical systems with the inherent chaotic behaviour. In addition, the classical Hénon map is a discrete-time two-dimensional system. The Hénon map is an iterative process where each point $\{x_{n+1}, y_{n+1}\}$ is obtained via the previous $\{x_n, y_n\}$ according to the rule:

$$x_{n+1} = 1 - ax_n^2 + y_n,$$
$$y_{n+1} = bx_n.$$

Thus, the Hénon map can be represented on a complex plane as depicted in figure 3(a), where x and y denote real and imaginary parts of complex symbols. The iterative property of the Hénon map enables to use it for the prediction task and as a benchmark test for neuromorphic systems [18]. Indeed, here our aim is to predict future symbols $\{x_{n+1}, y_{n+1}\}$ based on the incoming $\{x_n, y_n\}$.

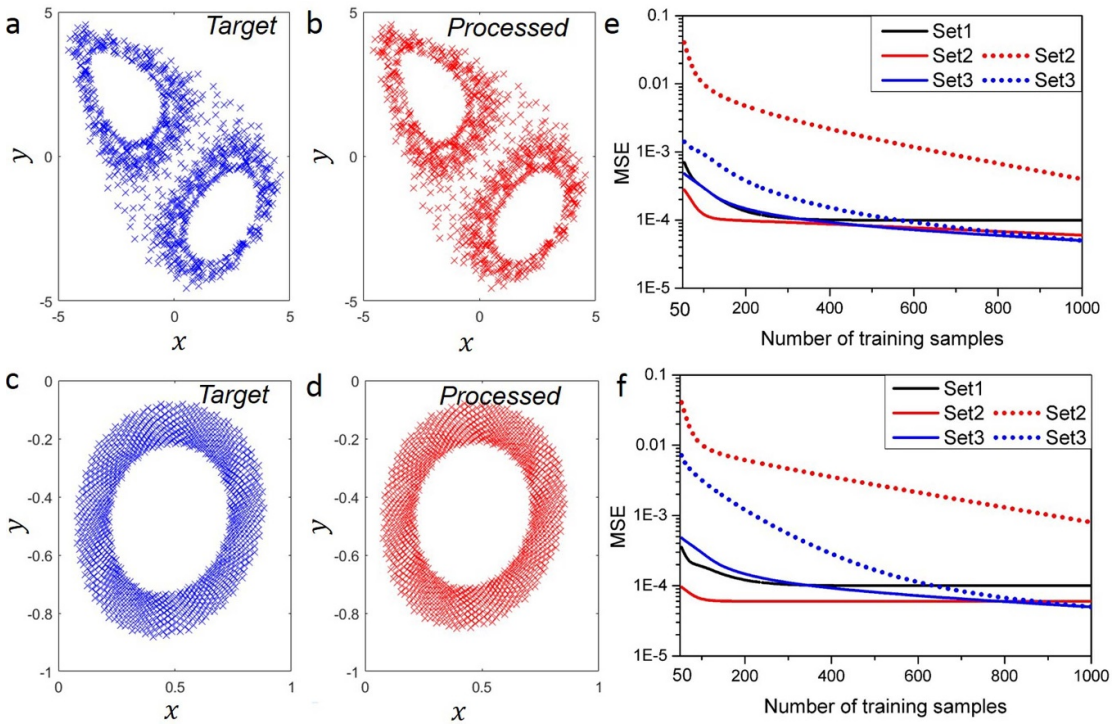

Figure 3. Multi-task operation: Hénon maps (a), (c) target (Set2,3) and (b), (d) processed symbols modulated OFDM signal. The mean squared errors for OFT-FESNA and benchmarking ESN (i.e. tanh-based with random weights) are depicted by solid and dotted curves in panels (e)–(f) for reservoir sizes 64 and 256 correspondingly.

We adapt this test for our purpose by modulating the complex discrete-time symbols into an OFDM signal. This is similar to modulating a signal with constellation symbols, where instead of a random selection of symbols, we choose to map the iterative Hénon symbols onto an OFDM carrier. As a result, we obtain a frequency-multiplexed signal, with a fixed bandwidth (see figure 3(c)). To illustrate applicability for various bandwidth parameters, we modelled two cases (i) future 5G-oriented case with the enhanced bandwidth (blocks of 2^{10} OFDM symbols mapped with 10 MHz carrier spacing, for $K = 1.55\ ns^{-2}$ resulting in 0.1 μs temporal spacing) plotted in figure 3 and (ii) standard OFDM parameters (the 2^{10} blocks of 128 symbols with 10 kHz carrier spacing for $K = 5.53\ ns^{-2}$ resulting in 45 ns temporal spacing) plotted in figures 4 and 5. Here the parameters were chosen to sustain the same rate of signal transmission. Increasing or decreasing parameters, in particular the relation between the chirp K and dispersion D, one can increase or decrease the speed of signal processing as mentioned in section 2.2. For this task we found the residual chirp in equation (2) to be sufficient, therefore $K' = 0$ and $D' = D = K^{-1}$. In both cases we simulated total number $N = 2^{20}$ symbols and used $T_g = 0.1\ T$ guard-band between the consequent intervals T (this is a typical guard-band in OFDM systems and in accordance with 3GPP standards). Using OFT we were able to perform frequency-time conversion optically with an appropriate scaling and perform efficient prediction task (for the signal, which was initially frequency-multiplexed), see figures 3(d) and (e).

Next we demonstrate the multi-channel operation by assigning to each channel a different prediction task by choosing symbols independently for each OFDM channel from three Hénon sets:

$$\text{Set1} : \{a = 1.4; b = 0.3\} - \text{figures 3(a) and (b)};$$
$$\text{Set2} : \{a = 0.2; b = 0.9991\} - \text{figures 4(a) and (b)};$$
$$\text{Set3} : \{a = 0.2; b = -0.9991 - \text{figures 4(c) and (d)}\}.$$

all three starting at the same point $\{x_1 = 0.1, y_1 = -0.3\}$. Each set represents a different chaotic map and requires independent training and prediction, thus emulating suitability of the design for simultaneous processing of different independent tasks. Moreover, each set exhibits different behavioural pattern and different complexity for processing. Thus, we demonstrate that one can utilize the proposed setup to solve different independent tasks and process multichannel signals solving various tasks for different channels. Although, for simplicity we used the same Hénon-based mapping, the variation of parameters resulted in different patterns of generated symbols leading to different dependence on the number of training symbols and reservoir size (i.e. sampling rate at the receiver: here 8 and 16 samples per symbol in figures 4(e) and (f))

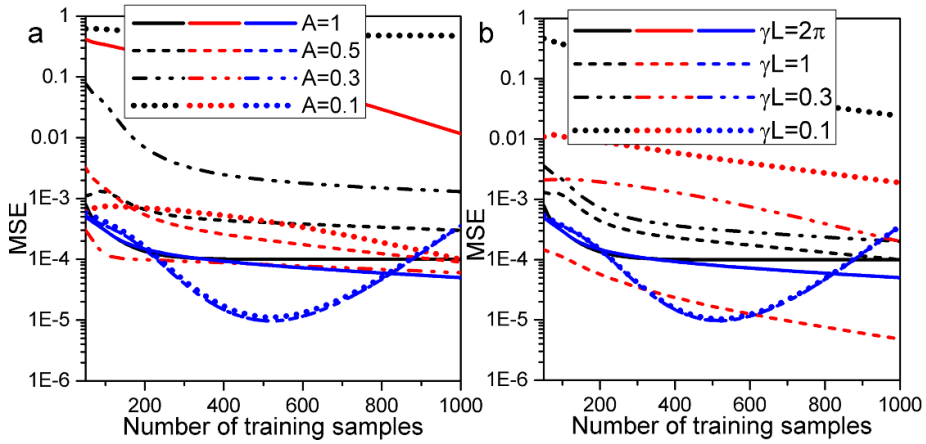

Figure 4. MSE dependence on optimization parameters: (a) attenuation coefficient for the feedback signal and (b) non-linear phase shift of HNLF plotted in black, red and blue for the Hénon sets 1–3 correspondingly.

correspondingly). Both parameters are easy to be changed at the receiver side. In addition for Set2 the value of attenuator A for the input signal had to be set at -10 dB (other sets did not require attenuator $A = 1$). This was done to create the difference between weight matrices $\mathbf{W}^{in}$ and $\mathbf{W}$ by changing a single parameter.

Finally, we compare OFT-FESNA performance with the standard ESN (for discrete symbols $\{x_n, y_n\}$) with the same training and reservoir sizes. The standard ESN design includes: (i) the weights chosen randomly from an interval $\{[-0.5 \ldots 0.5], i[-0.5 \ldots 0.5]\}$ and (ii) activation function is a step-function: $\tanh(x)+i\tanh(y)$. The results are plotted in figures 4(e) and (f) by dotted lines. One can see that OFT-FESNA slightly outperforms the considered ESN (until both reach the convergence regime dependent on the complexity of the task). Moreover, such architecture of ESN does not allow to perform prediction for the Set1 for the number of training samples below 1000. However, when in ESN the activation function is changed from tanh-based to sin-based, the ESN returns similar performance to that of OFT-FESNA (for all three sets). This highlights the importance of activation function optimization. Overall, OFT-FESNA performance is comparable to ESN and the sin-function basis for weights and activation function can be advantageous.

To demonstrate the flexibility of the design, the MSE for different values of attenuation coefficient A and HNLF parameter γL (here we changed L, while keeping $\gamma = 10.5$ 1/W/km fixed) is plotted in figures 4(a) and (b) with varied number of training samples (the reservoir size is 256). One can see in figure 4(a) that the optimum value of A is $\{1, 0.3, 1\}$ for each of three sets correspondingly. Moreover Set1 is the most sensitive to the variation of A. On the other hand, it is demonstrated that the value $\gamma L = 2\pi$ is not necessary and can be optimized accordingly. Although for 2π one can achieve a close approximation of tanh activation function. Nevertheless, 2π is clearly suboptimal for Sets 2,3, although it remains optimum for Set1. With values between 2π and 1 one can achieve reasonably good performance for all three sets. Thus, the activation function can be appropriately optimized by varying γL. This highlights flexibility of the design for performing various tasks.

Thus, the simple model of Hénon maps enables us to model complex systems exhibiting different behaviour for each set. Thus, each set presents a problem with different requirements for training and different optimum activation function. Consequently, the MSE dependence on the number of training samples also varies for different activation functions and different data sets (for example, when increasing the amount of training data, one can observe saturation for Sets1,2 for all considered activation functions and overfitting for Set3 for some activation functions). Moreover, by changing attenuator factor A and non-linear parameter γL one can vary normalization coefficient of the processed data and the slope of activation function, thus realizing different operation regimes. This is important for a variety of applications as performance of NNs depends strongly on normalization factor and activation function [19, 20].

3.2. Signal equalization

Next we demonstrate OFT-FESNA performance for equalization of optical communication signals. Here we consider 5 WDM channels with channel spacing 32 GHz with 64 quadrature amplitude modulation and root-raised-cosine 0.1 rolloff. The OFT with the chip parameter $K = 0.031$ ps^{-2} converts the signal spectrum into the time domain with temporal spacing 6 ps, thus preserving the same transmission rate. If required, one can perform the inverse time-to-frequency conversion to the original format by the second OFT (with

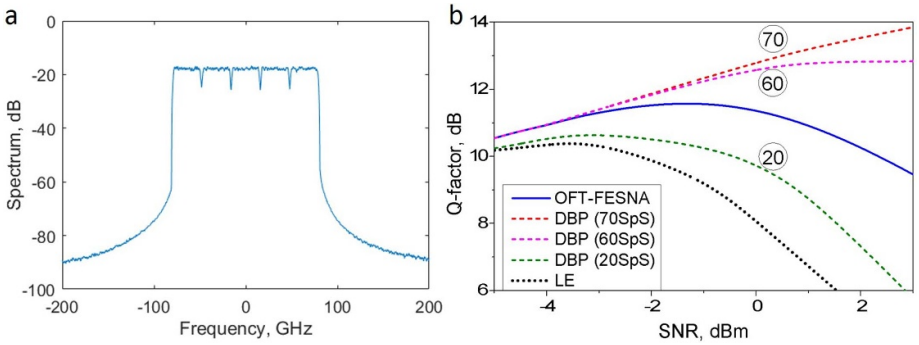

Figure 5. Q-factor as a function of input power for 64-QAM after transmission over 10 spans and processing with digital back propagation (dashed) with varied number of steps per span, linear equalization (dotted), and neuromorphic OFT-FESNA equalization (solid).

Table 1. Transmission system parameters.

Maximum phase change	0.01
Span length	80 km
Number of spans	10
Amplifier noise figure	5 dB
Attenuation coefficient	0.2 dB/km
Nonlinearity	1.2 1/W/km
Dispersion	17 ps/nm/km

the same parameters but inversed dispersion-chirp order) before the receiver. For this task the optimum parameters for signal mixing were found to be $D' = 0.016$ ns^{-2}, $K' = 0$, and $D = K^{-1}$.

The signal (with the total number of simulated symbols $N = 2^{25}$) was transmitted through a standard single-mode fiber communication system with parameters given in table 1. Using OFT-FESNA with the same parameters as for the previous task, while choosing $\gamma L = 2\pi$ and $A = 1$ (which were found to be optimum for this task), the optical frequency-time conversion and efficient equalization were performed and the resulted Q-factor as a function of input power is plotted in figure 5. For benchmarking the performance of neuromorphic OFT-FESNA equalizer is compared with the standard digital back propagation (DBP) equalization, which requires multiple steps per span (here 70, 60 and 20 are considered) and samples per symbol (here 16), as well as linear equalization (LE). While OFT-FESNA performance is lower than that of the ideal DBP due to non-ideal HNLF parameters (non-zero dispersion and losses), nevertheless the performance of optical OFT-FESNA is comparable to high complexity digital equalizer (DBP) and results in 1.2 dB gain over LE at peak performance.

4. Conclusions

We proposed a novel scheme for neuromorphic processing of multichannel frequency- and wavelength-division multiplexed signals. The concept was illustrated by the prediction of the Hénon map via processing OFDM-modulated symbols and signal equalization for WDM communication systems.
By varying fiber parameters one can optimize the design for various tasks. OFT-FESNA enables to realize neuromorphic processing optically for multi-channel signals sustaining the same transmission rate.

Acknowledgment

This project was supported by the Royal Academy of Engineering under the Research Fellowship scheme *RF*/201718/17154.

ORCID iD

Mariia Sorokina https://orcid.org/0000-0001-6082-0316

References

[1] Prucnal P R and Shastri B J 2016 *Neuromorphic Photonics* (Boca Raton, FL: CRC Press)
[2] Jaeger H The "echo state" approach to analysing and training recurrent neural networks, GMD Report 148, GMD - German National Research Institute for Computer Science (2001)
[3] Appeltant L, Soriano M C, Van der Sande G, Danckaert J, Massar S, Dambre J, Schrauwen B, Mirasso C R and Fischer I 2011 Information processing using a single dynamical node as complex system *Nat. Commun.* **2** 1–6
[4] Vandoorne K, Mechet P, Van Vaerenbergh T, Fiers M, Morthier G, Verstraeten D, Schrauwen B, Dambre J and Bienstman P 2014 Experimental demonstration of reservoir computing on a silicon photonics chip *Nat. Commun.* **5** 3541
[5] Sorokina M, Sergeyev S and Turitsyn S 2019 Fiber echo state network analogue for high-bandwidth dual-quadrature signal processing *Opt. Express* **27** 2387–95
[6] Sorokina M 2019 High bandwidth all-optical fiber-based neuromorphic signal processing *Conf. on Optical Communications* p 39
[7] Chang R W 1966 Synthesis of band-limited orthogonal signals for multi-channel data transmission *Bell Syst. Tech. J.* **45** 1775–96
[8] Weinstein S and Ebert P 1971 Data transmission by frequency-division multiplexing using the discrete Fourier transform *IEEE Trans. Commun. Technol.* **19** 628–34
[9] Le S T, Schuh K, Buchali F and Tan H N 2019 5 × 240 Gb/s WDM DD Transmission Over 80 km With Spectral Efficiency of 5.25 bit/s/Hz *IEEE Photonics Technol. Lett.* **31** 1830–3
[10] 3GPP TS 36.201 – v1.0.0, LTE Physical Layer – General Description (www.3gpp.org/ftp/Specs/archive/365Fseries/36.201/)
[11] 3GPP release 15 (www.3gpp.org/release-15)
[12] Kolner B H and Nazarathy M 1989 Temporal imaging with a time lens *Opt. Lett.* **14** 630–2
[13] Guan P, Røge K M, Lillieholm M, Galili M, Hu H, Morioka T and Oxenløwe L K 2017 Time lens based optical Fourier transformation for all-optical Signal processing of spectrally-efficient Data *J. Lightwave Technol.* **35** 799–806
[14] Guan P, Ros F D, Lillieholm M, Kjøller N-K, Hu H, Røge K M, Galili M, Morioka T and Oxenløwe L K 2018 Scalable WDM phase regeneration in a single phase-sensitive amplifier through optical time lenses *Nat. Commun.* **9** 1049
[15] Doran N J and Wood D 1988 Nonlinear-optical loop mirror *Opt. Lett.* **13** 56–8
[16] Sorokina M 2014 Design of multilevel amplitude regenerative system *Opt. Lett.* **39** 2499–502
[17] Sorokina M S and Turitsyn S K 2014 Regeneration limit of classical Shannon capacity *Nat. Commun.* **5** 3861
[18] Rodan A and Tino P 2011 Minimum complexity echo state network *IEEE T. Neural Netw.* **22** 131–44
[19] Ramachandran P, Zoph B, Le Q V Searching for activation functions (https://arxiv.org/abs/1710.05941)
[20] Misra D Mish: a self regularized non-monotonic activation function (https://arxiv.org/abs/1908.08681)