



HAL
open science

Impact of optical coherence on the performance of large-scale spatiotemporal photonic reservoir computing systems

Romain Modeste Nguimdo, Piotr Antonik, Nicolas Marsal, Damien Rontani

► To cite this version:

Romain Modeste Nguimdo, Piotr Antonik, Nicolas Marsal, Damien Rontani. Impact of optical coherence on the performance of large-scale spatiotemporal photonic reservoir computing systems. *Optics Express*, 2020, 28 (19), pp.27989. 10.1364/OE.400546 . hal-02963023

HAL Id: hal-02963023

<https://hal.science/hal-02963023v1>

Submitted on 1 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Impact of optical coherence on the performance of large-scale spatiotemporal photonic reservoir computer systems

Romain Modeste Nguimdo*, Piotr Antonik, Nicolas Marsal, and Damien Rontani

Abstract—Large-scale spatiotemporal photonic reservoir computer (RC) systems offer remarkable solutions for massively parallel processing of a wide variety of hard real-world tasks. In such systems, neural networks are created by either optical or electronic coupling. Here, we investigate the impact of the optical coherence on the performance of large-scale spatiotemporal photonic RCs by comparing a coherent (optical coupling between the reservoir nodes) and incoherent (digital coupling between the reservoir nodes) RC systems. Although the coherent configuration offers significant reduction on the computational load compared to the incoherent architecture, for image and video classification benchmark tasks, it is found that the incoherent RC configuration outperforms the coherent configuration. Moreover, the incoherent configuration is found to exhibit a larger memory capacity than the coherent scheme. Our results pave the way towards the optimization of implementation of large-scale RC systems.

I. INTRODUCTION

A wide variety of hard tasks such as speech recognition [1], nonlinear channel equalization and time-series prediction [2], detection of epileptic seizures [3], robot control [4], and automatic classification of images and videos [5]–[7] can be solved using brain-inspired systems that do not require explicit instructions, but rely on patterns and inferences [8], [9]. Reservoir computing belongs to that class of systems for information processing [10]. Over the last decade, its photonic implementations have received considerable attention due to their excellent performance, energy efficiency, and fast speed [11]–[14]. A large variety of photonic architectures have been already proposed such as single laser diode with optical or optoelectronic feedbacks [15]–[21], free-space optics [6], [22]–[24], optical cavities [25], [26], and photonic integrated technology [27]–[30].

A RC architecture typically consists of three parts: an input layer, a reservoir, and an output layer. The input and the output layers are the stages where the pre-processing and the post-processing are performed, respectively, while the reservoir is a usually large recurrent dynamical network (also called echo state network). The simplicity of the RC approach relies on the fact that the recurrent internal weights of RC are fixed and can be chosen randomly meaning that only the readout layer has to be trained [2], [31]. This dramatically reduces the

training time and also simplifies experimental implementations of such systems. Nevertheless, the physical implementations of RC systems have remained very challenging over a decade due to the large number of nonlinear nodes to be coupled and controlled. To get around this drawback, time-delay configurations have been successfully investigated [16], [17], [19], [25], [32], [33]. In this approach, the reservoir size increases linearly with the length of the feedback loop, but at the expense of the processing speed. That is, the data cannot be fed to the systems faster than the time-delay to get all the virtual nodes subjected to the data [13].

To perform complex tasks requiring a large number of nodes, a large-scale spatiotemporal photonic RC with a reservoir up to 2,500 nodes was, first, demonstrated [23]. Then, Antonik *et al.*, recently achieved the state-of-the-art experimental performance on image and video classification using spatiotemporal photonic RC system capable of implementing neural networks up to 16,384 nodes [6], [7]. In previous large-scale spatiotemporal photonic RC systems, complex and recurrently connected networks have been created using either imaging spatially structured via a diffractive optical element [23], [24] or from a computer [6], [7].

A part from the dimensionality, which plays a crucial role in the system performance, nonlinearity is another key attribute of the RC systems. Various implementations with different nonlinearities have been demonstrated, all showing good performance [6], [17], [19], [25], [32], [33]. However, it is still not clear whether a particular nonlinearity outperforms the others when used on specific tasks. For large-scale spatiotemporal photonic RC systems, the needed high-dimensionality and nonlinearity can be typically created in the optical or electronic domain using polarizers, high-dimensional cameras and spatial light modulator (SLM) [6], [23], [24].

Few investigations have separately focused on incoherent [6], [7] or coherent cases [22]–[24], but no comparative study has been carried out to evaluate the benefits of each configuration. In such schemes, the inherent nonlinearity implemented in the optical domain depends on the coherence on the light beam, which determines the internal connectivity of the network. Hence, the following question arises: between coherent and incoherent light beams, which is the most suitable for implementing spatiotemporal photonic RC systems?

In this work, we report numerical results of two SLM-based, large-scale, spatiotemporal RC systems, which only differ in the nature of the light beam and the physical coupling of the reservoir nodes. The paper is structured as follows.

Romain M. Nguimdo (email: romain.nguimdo@centralesupelec.fr), Piotr Antonik (piotr.antonik@centralesupelec.fr), Nicolas Marsal (nicolas.marsal@centralesupelec.fr), and Damien Rontani (damien.rontani@centralesupelec.fr) are with LMOPS EA 4423 Laboratory, CentraleSupélec, and Université de Lorraine, F-57000 Metz, France
*Corresponding author.

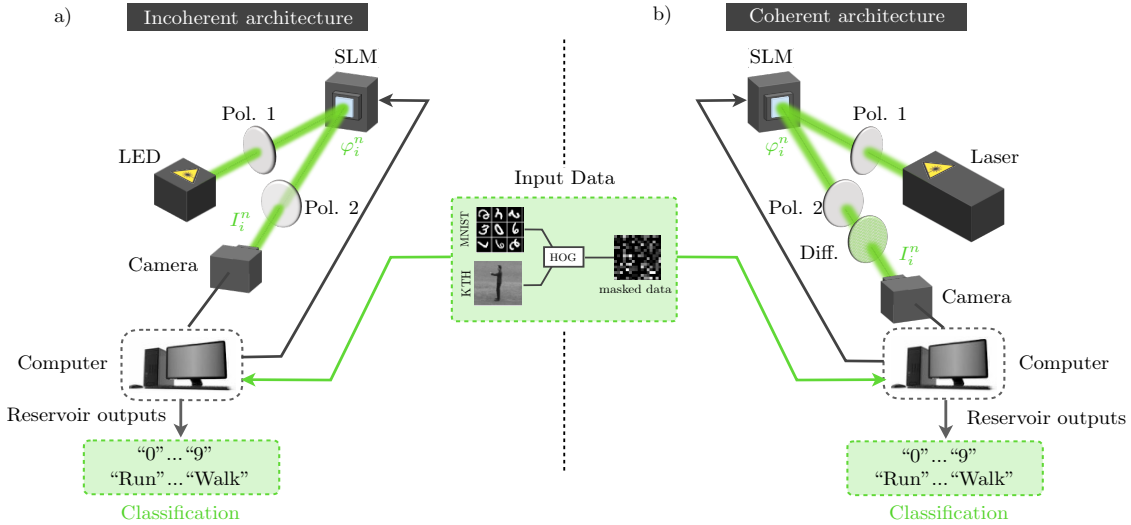


Fig. 1. Schemes of principle of the experimental setups. The data to be processed are the HOG features extracted from images or video streams. The features are multiplied by a random mask before being injected into the reservoir. The reservoir structure depends on the configuration. a) Incoherent configuration. The SLM is illuminated by an incoherent light from a LED. Then, a polarizer is used to transform the SLM phase-modulated signal into an intensity modulation before its detection by a camera. After the camera, the computer multiplies the signal by a random coupling matrix and superimposes it with the new masked data. The result is used to update the SLM state. b) Coherent configuration. The SLM is illuminated by a coherent light from a laser. After the second polarizer, a diffuser is inserted to ensure interconnections between the reservoir nodes via interferences before their detection on the camera. As such, the computer is used to inject the masked data into the reservoir and to update the phase states of the SLM.

Section I is devoted to the introduction. In Sec. II, we describe the operating principle of each configuration and give the corresponding numerical model. In Sec. III, we describe the benchmark tasks used to compare the performances of the two systems. Section IV highlights the main results and discusses the performance of each configuration. Section V concludes the paper while further modeling details are given in Appendix VI.

II. SYSTEM MODELING

The two architectures of large-scale, spatiotemporal, photonic RC systems investigated in this work are shown in Fig. 1. They are electro-optical systems composed of an optical path and an electrical path, both built by interconnecting stand-alone components. On one hand, a SLM is used to transform an electrical signal into optical signal through phase modulation, while a camera transforms the optical signal into an electrical signal. The number of nodes of the reservoir is essentially limited by the resolutions of the SLM and the camera. In our current setup, the lowest resolution is given by the SLM whose resolution equal to 512×512 could provide up to $N = 262,144$ nodes. However, we typically use only pixels located at the center of the SLM matrix due to physical limitations from (i) slight misalignment between the SLM and camera optical axe and (ii) inhomogeneous intensity distribution of the optical beam. Such a constraint limits our study to a maximum number of nodes N up to 16384 similar to [6].

A. Incoherent reservoirs

For the incoherent architecture, the SLM is illuminated by a continuous-intensity light beam emitted by a LED [Fig. 1(a)]. The reflected light beam then passes through a linear polarizer oriented at 45° with respect to the vertical axis so that it implements a nonlinear sin-square function by transforming the

phase-modulation into an intensity modulation. Afterwards, the obtained signal is detected by a camera and serves two purposes: the first is the readout for post-processing of the data on a computer, and the second is the implementation of the coupling between the reservoir's nodes by multiplying, digitally, this signal with an $N \times N$ random matrix. Subsequently, the data to be processed is added and the resulting signal is used to update the SLM phase for each pixel.

Experimentally, the camera and the SLM have a limited resolution of 8-bits. Hence, the dynamics of the system as shown in Fig. 1(a) can be described by the following coupled maps:

$$I_i^{n+1} = \left[I_0 \sin^2 \left[\sum_{k=1}^N w_{ik} I_k^n + \beta \sum_{l=1}^M b_{kl} U_l^n \right] \right]_{\lfloor \cdot \rfloor_8}, \quad (1)$$

where $\lfloor \cdot \rfloor_8$ refers to the 8-bit resolution and I_0 is the input intensity which illuminates the total pixels of the SLM. I_i^n is the light intensity at time step n for i^{th} -pixel at the output of the camera. $w_{ik} \in \mathbb{R}^{N \times N}$ are the coefficients of the random interconnection matrix, while $b \in \mathbb{R}^{N \times M}$ is the matrix associated to the input information U , M being the dimension of the input vector; β is the scaling parameter for the input data. The modeling details to obtain Eq. (1) are given in Appendix VI-1.

B. Coherent reservoirs

The architecture for the coherent case is shown in Fig. 1(b). It has structural similarities with the incoherent architecture with a few noticeable differences. First, the SLM is illuminated by a continuous-wave coherent light beam emitted by a laser (Fig. 1(b)). Then, a diffuser is inserted after the second polarizer (Pol. 2) so that the coherent light beam from this polarizer is scattered creating a complex interference pattern. Hence, the diffuser output is a randomly weighted sum of

the different contributions from all SLM pixels. In other words, the variability and the coupling between the RC nodes are ensured optically so that an input image imprinted as a phase-modulation on the SLM coherent light beam propagates through the optical diffuser to form a speckle pattern on the camera. Here, the computer is only used to inject the data and to update the SLM phases. This offers significant reduction on the computational load compared to the incoherent architecture.

Under this description, the RC system shown in Fig.1(b) can be modeled by:

$$I_i^{n+1} = \left[I_0 \left| \sum_{k=1}^N w_{ik} \sin \left[I_k^n + \beta \sum_{l=1}^M b_{kl} U_l^n \right] \right|_8^2 \right]_8. \quad (2)$$

Here, $w \in \mathbb{C}^{N \times N}$ is the complex-valued transmission matrix. The others parameters were defined in Sec. II-A. The detailed derivation of the Eq. (2) are given in Appendix VI-2.

Since we have assumed that the summation of the different internal weights contributing to the i^{th} pixel is achieved by a diffuser, the matrix elements of w are complex numbers fixed by this component with their phase values randomly distributed in the interval $[0, 2\pi]$. It has been demonstrated that the singular values of w follow the so-called ‘‘quarter-circle law’’ distribution [34]. According to this law, the module of w_{ik} is generated so that the singular values of a normally distributed square matrix lie on a quarter circle. More details on w are given in Appendix VI-3. Also, the elements of the matrix b can be freely fixed during the pre-processing on the computer. Hence, I_0 and β are the key parameters which can be tuned to observe different dynamical regimes.

For both configurations, we have normalized the matrix w so that $\sum_{i,j} |w_{ij}|^2 = 1$. Also, the training is done on the camera readout signal, which forms the following RC outputs:

$$y_{RC,p}^n = \sum_{k=1}^N w_{pk}^{out} [I_k^n]_8, \quad (3)$$

where $w^{out} \in \mathbb{R}^{P \times N}$ is a matrix whose elements are determined by the training. The number P of RC outputs depends on the type of benchmark tasks or metrics used.

III. BENCHMARK TASKS AND PERFORMANCE METRICS

Our comparative study is carried out on two different benchmark tasks, namely image classification of handwritten digits and the recognition of human actions in video streams. The image classification will be evaluated on the popular MNIST database [35] publicly available while the recognition of human actions in video streams will be performed on the popular KTH database also publicly available online [36]. In addition, we consider a memory capacity metric as a measure of system attribute.

A. MNIST task and pre-processing

The MNIST database contains 70,000 images of handwritten digits from 0 to 9. All images have been normalized to fit into a 28×28 pixels bounding box, antialiased, and

converted into gray-scale levels. We use 60,000 of these data to train the system and 10,000 others for testing. The output layer is composed of 10 binary outputs corresponding to the 10 different classes of digits. The pre-processing consists in extracting relevant spatial and shape informations from individual images. An efficient way to do this is to compute the histograms of oriented gradients (HOG) from the original images [7], [37]. For the HOG calculations, we used 9 bin orientations, 7×7 pixels per cell and 2×2 cells per block [7]. In addition, HOG values are normalized with their highest value so that the input scaling factor becomes the amplitude of the reservoir input signal. In such an approach, the RC input data are the HOG features convoluted with a random mask which ensures the variability over the reservoir nodes. The HOG features of the same image are simultaneously injected into the reservoir at the time step n and the next features are injected at time step $n+1$ meaning that the injection time interval between two consecutive images is the discrete time step. The order of the image injection does not matter. The linear regression training is applied to the different node responses so that the RC outputs ‘‘1’’ for the digit corresponding to the class the image is associated to, and ‘‘0’’ otherwise. The principle of ‘‘the winner-takes-all decision’’ strategy is further applied to classify each image based on the binary output with the highest value.

B. KTH task and pre-processing

The KTH database contains video recordings of six different motions (walking, jogging, running, boxing, hand waving, and hand clapping) performed by 25 subjects. Each motion is repeated four times. Similarly to [6], we limited the database to the ‘‘s1’’ scenario, *i.e.* videos shot over a uniform background without zoom or lightning variations. The ‘‘s1’’ subset contains a total of 599 video sequences, that we bring up to 600 by artificially duplicating a missing boxing sequence. Over a total of the 600 video sequences, a subset of 450 video sequences are used for training, while the remaining 150 sequences are used for testing. Both for training and testing, all videos are concatenated together and split into individual image frames. Then, the HOG algorithm is applied to each image to extract features, which will be used as the input for the RC. To reduce computational cost, we reduce the number of HOG features by further applying principal component analysis (PCA) based on the covariance method [38]. As such, we only keep the first 2,000 components (out of 9,576), whose eigenvalues account for 91.6% of the total variability in the data. From each video, one constructs a matrix where the rows are image frames, while the columns are the HOG features extracted from those images. At each discrete time step, we inject the HOG features of an image. The order of injecting the videos does not matter, but the images of each video should be injected as a sequence. In this work, the images will be injected so that the time interval between two consecutive images corresponds to the discrete time step of the model. In other words, the injection time interval between two consecutive images belonging to the same video is the same as that between two consecutive images from two different videos. After classifying the images from

the 600 videos through training and testing processes of our RC system, we subsequently classify the individual frame of these videos in their respective classes considering again “the winner-takes-all” approach. In practice, the classifier output is evaluated throughout the full video sequence (from the first frame to the last) and the final result corresponds to the class having the majority of frames within the sequence attributed to it.

C. Memory capacity metric

Memory capacity is a common way to evaluate the ability of machine learning systems for referring back to the past information. For time-dependent tasks, this property is crucial for the performance and efficiency of the information processing system. Considering a sequence of numbers $(u^k)_{k \geq 0}$ at discrete time steps drawn from a uniform distribution in the interval $[-0.5, 0.5]$, a typical way to address the linear memory capacity of a physical system consists in constructing the copies of given inputs u^k at time step k shifted by s time steps, *i.e.*, $\hat{y}^k = u^{k-s}$ with $s = 0, 1, 2, \dots$. To quantify this capacity, we define the memory function as the cross-correlation between u^{k-s} and \hat{y}^k [39]:

$$m^s = \frac{\langle [u^{k-s} - \langle u^{k-s} \rangle] [\hat{y}^k - \langle \hat{y}^k \rangle] \rangle}{\langle |u^{k-s} - \langle u^{k-s} \rangle|^2 \rangle^{1/2} \langle |\hat{y}^k - \langle \hat{y}^k \rangle|^2 \rangle^{1/2}}. \quad (4)$$

From this memory function, the memory capacity is calculated as:

$$MC = \sum_s m^s. \quad (5)$$

For all tasks, the training consists in using linear regression to determine the different weights, which should be assigned to the reservoir readout responses so that the output signal approaches the target as closely as possible. The simulations are carried out on Eq. (1) for the incoherent case and Eq. (2) for the coherent case.

IV. RESULTS AND DISCUSSIONS

A. Performances of the two system configurations

In order to identify the appropriate parameter ranges for the tasks processing, we first scan the system’s performance on the (β, I_0) -space parameters. The results for the coherent and incoherent configurations are shown in Fig. 2. For the incoherent configuration, classification errors are as low as 1.8% with 1,024 nodes even with 8-bit quantization considered in the model. Overall, classification errors less than 2% can be obtained for a broad range of parameter sets for this configuration. These errors are consistent with the state-of-the-art performance for similar reservoir’s sizes reported in the literature [7]. For the coherent configuration, however, worse classification errors are obtained for the MNIST task in the same parameter sets explored. The lowest classification error obtained is 2.8%. Overall, classification errors less than 3% can be obtained, but in a very narrow range of values for parameters (β, I_0) .

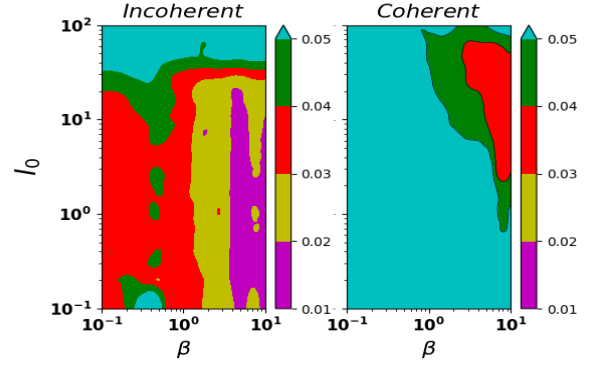


Fig. 2. Classification error for the MNIST task plotted in (β, I_0) -plane for the coherent case (left) and the incoherent case (right) with a reservoir of 1,024 nodes.

Guided by the results in Fig. 2, we now choose the optimal values of β and I_0 and investigate the influence of the reservoir size on the performance of the two configurations. These values are $I_0 = 5$, $\beta = 10$ for both for the coherent and the incoherent cases. They are kept fixed for all reservoir sizes as $\sum w_{ij}^2 = 1$ has been adopted. The results are shown in Fig. 3. We find that the increase of reservoir size has different effect on the system performance depending on the configuration. For the incoherent configuration, our results confirm a decrease of the classification errors with the increase of the reservoir size as previously reported in [6]. By way of illustration, the classification errors gradually decrease from $\sim 1.8\%$ for $N = 1.024$ nodes to $\sim 0.8\%$ for $N = 9.216$ nodes. For the coherent configuration, which has not been explored before, we observe that the classification error gradually decreases and reaches a minimum for a particular size before becoming worse with the increase of the reservoir size. For example, the classification error decreases from 3.7% for 1.024 nodes to 2.5% for 7.168 nodes, then, it increases again to 2.6% for 11.264 nodes. We have noted that, for some parameter sets, the increase of the errors with the increase of reservoir size after reaching the minimum is quite significant.

Next, we compare the performance of the two configurations on the classification of video streams. Figure 4 shows the classification errors over 150 videos (videos used for testing) for the two configurations in the (β, I_0) -plane. The analysis of the results shows that there is a large range of parameter sets for which misclassified videos are comprised between 20%

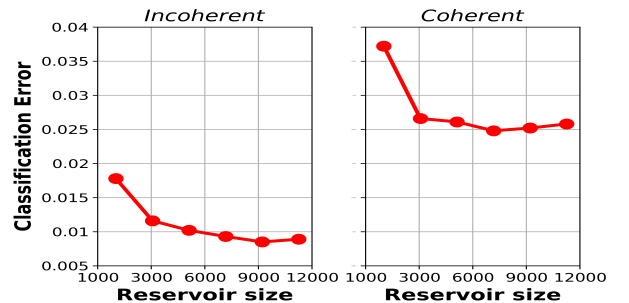


Fig. 3. Classification error for the MNIST task computed from Eqs. (2) and (1) as a function of the reservoir sizes. The parameters are $I_0 = 5$, $\beta = 10$ for both for the coherent and the incoherent cases.

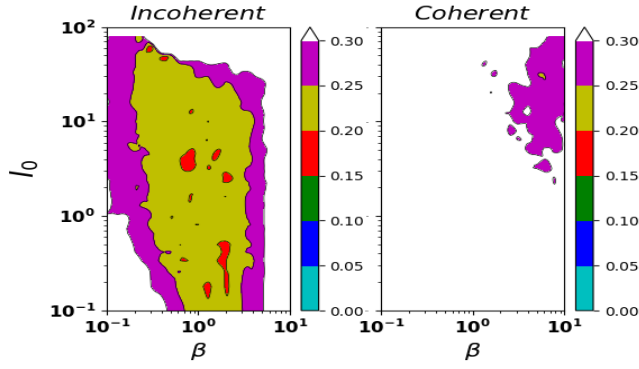


Fig. 4. Video classification error for the KTH task plotted in (β, I_0) -plane for the incoherent case (left) and the coherent case (right) with a reservoir of 1,024 nodes considering 8-bit resolution.

and 25%, and also few parameter sets with errors slightly below 20% for the incoherent configuration, while misclassified videos are in-between 25% and 30% in a narrow windows of parameter sets for the coherent configuration. Explicitly, up to 77.3% of videos can be successfully classified using our incoherent RC while a maximum of 73% of videos have been successfully classified using the coherent RC. The incoherent configuration therefore gives rise to a more flexibility in the choice of (β, I_0) while ensuring better performance.

In Fig. 5, we again choose the optimal parameter set for each configuration (i.e. coherent and incoherent) and plot the classification errors score on the videos as a function of the reservoir size. The parameters β and I_0 are kept fixed for all reservoir sizes. It is confirmed that the incoherent configuration outperforms the coherent configuration for all the reservoir sizes. Interestingly, the results for large reservoir sizes indicate a significant improvement of the performance for the two configurations. For example, we achieved an accuracy of 90% and 85% for the incoherent and coherent configurations, respectively, we a reservoir size of 11,264 nodes.

To deep into further comparison, we show in Fig. 6 the confusion matrices, which allow to have the explicit percentage of success of specific actions for each configuration. It appears that the two configurations achieve a good performance for the "boxing", "hand clapping" and "hand waving" actions which are typically less constraining to learn than the "jogging" and the "running" actions. For the two latter actions, the learning performance of the two systems are limited. Both from Fig. 6

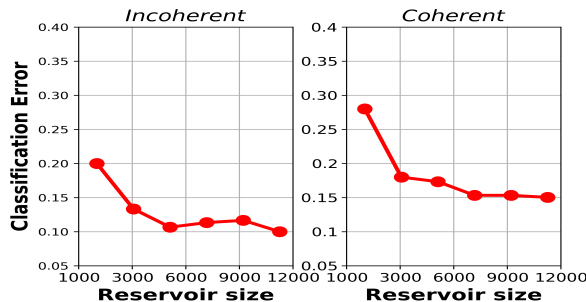


Fig. 5. Classification error of the KTH task as a function of the reservoir size for $\beta = 2$, $I_0 = 2.5$ for incoherent case and $\beta = 7$, $I_0 = 25$ for coherent case considering 8-bit resolution.

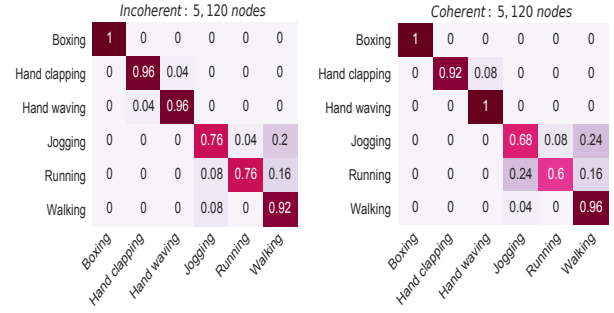


Fig. 6. Confusion matrices for $\beta = 2$, $I_0 = 2.5$ for incoherent case and $\beta = 7$, $I_0 = 25$ for coherent case considering a reservoir size of 5,120 nodes.

and our results with others reservoir sizes (not shown here), the incoherent architecture outperforms the coherent configuration on these two specific actions. In particular, with a relatively large reservoir size (e.g 5,120 nodes), it is found that the classification accuracy on the "jogging" and "running" actions can be increased to $\sim 65\%$ for the coherent configuration, while the accuracy exceeds $\sim 75\%$ for the incoherent configuration.

One should note that the results reported here are comparable but lower than those reported previously on the same experimental setup [6], [40]. The difference stems from the optimisation of the hyper-parameters β and I_0 . In this work, to avoid excessive computational times, we chose to keep the same hyper-parameters values for all the different reservoir sizes, whereas in [6], [40] they were optimised specifically for each size of the neural network. The similarity of our results to those reported previously shows the robustness of our system, since accuracies comparable to state-of-the-art can be obtained even with sub-optimal hyper-parameters.

B. Memory capacity

Another interesting property of RC systems is their memory capacity. Although low memory capacity may be sufficient for our above-mentioned tasks, some time-dependant machine learning tasks require a large memory capacity (e.g. NARMA10 [25], [32]). Therefore, for further insights regarding our two configurations, we show in Fig. 7 the results for

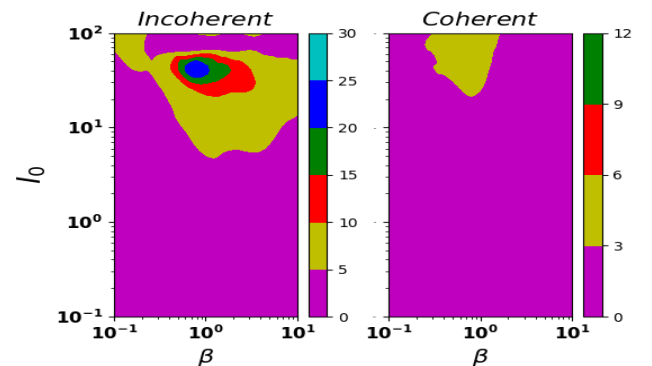


Fig. 7. Memory capacity plotted in the (β, α) -plane with a reservoir size $N = 1,024$. Note the difference of the scale for the colour bars for the two configurations.

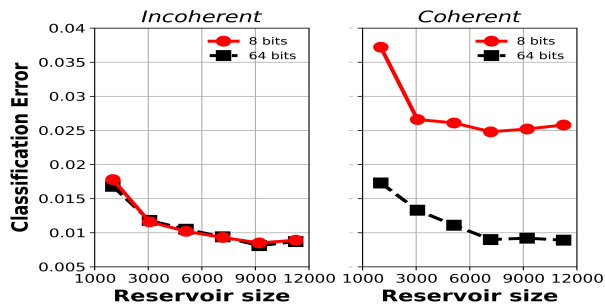


Fig. 8. Classification Error on MNIST task as a function of the reservoir sizes for different quantization levels. The parameters are the same as in Fig. 3.

the memory capacity of the system scanning in the (β, I_0) -plane with a reservoir size of $N = 1,024$ nodes. For the incoherent configuration, memory capacity up to 20 can be obtained for substantial parameter sets despite the presence of quantization noise. On the contrary, the coherent configuration yields a poor memory capacity. Precisely, the best memory capacity is less than 6 and only very few parameter sets allow a memory capacity larger than 3 for this configuration.

C. Effect of quantization noise

The decrease in performance for the coherent configuration on the MNIST task (see Fig. 3) is an unexpected result, as the RC performance typically improves with the increase of the dimensionality. In order to understand this result, we investigate the specific role of the bit quantization by comparing the results with and without quantization noise (*i.e.*, the bit resolution is that of the computer in our simulation model at 64-bit). Figure 8 shows the results for 64-bit resolution in comparison with those of 8-bit resolution already reported. For the incoherent configuration, the same classification errors are obtained both for 64-bit and 8-bit resolutions for the same reservoir sizes. The system is therefore robust to readout noise introduced by bit quantization for this parameter set. For the coherent configuration, we find classification errors as low as those obtained for the incoherent configuration at 64-bit resolution. This suggests that the large classification errors obtained with 8-bit resolution were caused by quantization noise. Thus, this configuration appears to be more sensitive to low bit resolution than the incoherent configuration: For example, the classification error increases from 1.8% for 64-bit resolution to 3.6% for 8-bit resolution for a RC size of 1,024 nodes. In addition, the results of the coherent configuration without bit quantization also indicate that the unexpected degradation of the system performance for large RC sizes is likely caused by the data overfitting due to the fact that data become more tricky to learn in the presence of strong noise. This suggests that more sophisticated learning techniques, such as ridge regularization methods [31] should be investigated to improve the system performance, when devices with low bit resolution are used to implement the coherent configuration. For future studies, the overall impact of noise (additive or multiplicative) on the performance could be investigated within the framework developed in Ref. [41].

Based on the various benchmark tasks used here, it appears that the incoherent configuration has a higher processing

capacities than the coherent configuration. For the sake of clarity, Table 1 summarizes the best performances of our two configurations on the different tasks for 8 and 64 bit resolutions for a reservoir with 1,024 nodes. It appears that the incoherent configuration outperforms the coherent configuration for all tasks. We also notice that the significant effect of low-bit resolution on the image classification is attenuated for the KTH task because each full video sequence is classified by simply choosing the class with the majority of frames within the sequence. For KTH and memory capacity, we have not investigated the influence of bit quantization for other reservoir sizes for 64 bits.

TABLE I
PERFORMANCES OF THE TWO RC SYSTEM CONFIGURATIONS ON THE DIFFERENT TASKS FOR 8 AND 64 BIT RESOLUTIONS CONSIDERING A RESERVOIR WITH 1,024 NODES.

Bit resolution	Incoherent case		Coherent case	
	8	64	8	64
MNIST: Lowest error	1.8%	1.8%	3.7%	1.8%
KTH: Highest accuracy	77.3%	80%	73%	73%
MC: Maximum value	20	23	4.07	6

V. CONCLUSION

In this work, we have compared the performances of an incoherent and a coherent free-space photonic RC systems on image and video classification, while preserving as much as possible the structure of each architectures by using similar components.

The results have shown a trade-off between the incoherent and the coherent architectures. On one hand, the coherent RC architecture allows physical coupling between the reservoir nodes and prevents the use of additional computational resources needed to perform the multiplication of the adjacency matrix with the RC's state vector, which considerably slows down the processing speed as the reservoir scales up. On the other hand, our results have shown that the incoherent configuration outperforms the coherent configuration in all benchmark tasks tested, but at the cost of processing speed. Furthermore, we have found that the incoherent configuration can exhibit a large memory capacity while the coherent configuration have shown comparatively low (five to tenfold lower) memory capacity for all the hyperparameter sets explored. Based on our simulations, we have found that the coherent configuration is more sensitive to quantization noise generated by low-bit resolution imposed by the physical component than the incoherent configuration. In particular, the incoherent configuration shows very-high robustness as no visible difference in performance between the 8 bit and 64 bit quantization were observed.

Our result has unveiled the impact of optical coherence on the processing capabilities of RC systems with similar structures on demanding computer-vision tasks; this will potentially pave the way towards informed design-choices and optimization of future large-scale, free-space, photonic RC architectures.

VI. APPENDIX

1) *Appendix 1:* For the incoherent configuration, a LED delivers an incident incoherent light beam propagating through a first linear polarizer (Pol. 1) oriented at 45° such that, according to the Jones formalism, the resulting electric field reads $\mathcal{E}_0 = E_0/\sqrt{2}e^{j\varphi_0}[1, 1]^T$, with E_0 and φ_0 the amplitude and the phase of \mathcal{E}_0 , respectively. On each pixel of the SLM, a specific time-discrete phase modulation is applied resulting in a local electric field \mathcal{E}_i^n for the i -th pixel at the n -th time step. Then, the SLM output passes through a second linear polarizer (Pol. 2) oriented at 45° with respect to the vertical axis. This polarizer transforms the electric field \mathcal{E}_i^n of each pixel into $E_0/\sqrt{2}\sin(\varphi_i^n/2)[1, 1]^T$, where φ_i^n is the phase of the i -th SLM's pixel at the n -th time step. We have assumed here homogeneity and constant electric field over a given SLM's pixel for a time-step duration. Afterwards, a camera is used to detect the intensity of the polarized light beam and its output gives the intensity of each pixel as [7]:

$$I_i^n = I_0 \sin^2(\varphi_i^n/2), \quad (6)$$

with $I_i \propto |\mathcal{E}_i|^2$ and $I_0 \propto |E_0|^2$. After the detection, the signal I_k^n is multiplied with a random matrix $W \in \mathbb{R}^{N \times N}$ whose elements are generated on computer before being added with the randomly-masked data. The resulting signal is used to update the programmable phase shifts of the SLM as follows

$$\varphi_i^{n+1} = \sum_{k=1}^N W_{ik} I_k^n + \beta \sum_{k=1}^M b_{ik} U_k^n. \quad (7)$$

Thus, by combining Eqs. (6) and (7) and rescaling the matrices and state variables, one finds

$$I_i^{n+1} = I_0 \sin^2 \left(\sum_{k=1}^N W_{ik} I_k^n + \beta \sum_{k=1}^M b_{ik} U_k^n \right), \quad (8)$$

which is the Eq. (1) given in Sec. II-A without considering the 8-bit quantization operation imposed on the SLM phase values and intensities detected by the camera.

2) *Appendix 2:* For the coherent configuration, the SLM is illuminated by a linearly polarized coherent light beam generated by a laser. After the second polarizer (Pol. 2) a diffuser is inserted so that the polarized, coherent light originating from each pixel $E_0/\sqrt{2}\sin(\varphi_i^n/2)[1, 1]^T$ is scattered. The electric field of the i -th pixel at the diffuser's output \mathcal{F}_i^n is then the coherent sum of the different contributions from other k -th pixels:

$$\mathcal{F}_i^n = E_0 \sum_{k=1}^N w_{ik} \sin(\varphi_k^n/2) \mathbf{e}_{i,k}, \quad (9)$$

where $\mathbf{e}_{i,k}$ is a normalized vector giving the polarization direction at the diffuser's output of the field amplitude contribution of the k -th to the i -th pixel. We assume that vectors $\mathbf{e}_{i,k}$ have independent, uncorrelated, random orientations, thus leading to a degree of polarization (DOP) of the scattered light with respect to its initial polarization approximately equal to zero (*i.e.* unpolarized light). Coefficients $w_{ik} = w_{ik}^d e^{j\phi_{ik}}$ form the complex-valued light transmission matrix with w_{ik}^d and ϕ_{ik} the real-valued transmission coefficient in amplitude and the

phase accumulation during the propagation of light from the k -th to the i -th pixel, respectively. The diffuser's output is then focused by an imaging lens to a camera. This camera reads the intensity of the signal of each pixel as $I_i^n \propto |\mathcal{F}_i^n|^2$. Similar to the incoherent configuration, the computer is used to include masked data and update the programmable phase-shift states of the SLM's i -th pixel according to the following update rule:

$$\varphi_i^{n+1} = I_i^n + \beta \sum_{k=1}^M b_{ik} U_k^n, \quad (10)$$

where β is the control parameters used to rescale the SLM phase-shift and the amplitude of the input data, respectively. $(b_{ik})_{i,k}$ is the input matrix. At time step $n+1$, the intensity detected by the i -th camera's pixel is such that $I_i^{n+1} \propto |\mathcal{F}_i^{n+1}|^2$, which becomes after substituting Eq. (9):

$$I_i^{n+1} \propto \left| E_0 \sum_{k=1}^N w_{ik} \sin(\varphi_k^{n+1}/2) \mathbf{e}_{i,k} \right|^2. \quad (11)$$

With a coherent light beam, there are cross terms reading $I_0 w_{ik} \bar{w}_{i,k'} \sin(\varphi_k^{n+1}/2) \sin(\varphi_{k'}^{n+1}/2) (\mathbf{e}_{i,k} \cdot \mathbf{e}_{i,k'})$, which come from pixel-to-pixel interference between the diffuser output signals detected by the i -th pixel of the camera. They play a crucial role both in the dynamics and the RC's processing capabilities. To simplify the notations, we have embedded the contribution of dot products $(\mathbf{e}_{i,k} \cdot \mathbf{e}_{i,k'}) = \cos(\angle(\mathbf{e}_{i,k}, \mathbf{e}_{i,k'}))$ directly in the transmission matrix w . Finally, using Eq. (10) and the same rescaling procedure used in Eq. (8), the model of the coherent RC reads,

$$I_i^{n+1} = I_0 \left| \sum_{k=1}^N w_{ik} \sin \left(I_k^n + \beta \sum_{l=1}^M b_{kl} U_l^n \right) \right|^2. \quad (12)$$

Equation (12) corresponds to Eq. (2) of Sec. II-B without considering the 8-bit quantization operation imposed on the SLM phase values and intensities detected by the camera.

3) *Appendix 3:* For our system, the optical power is conserved if we neglect losses which may occur during the propagation of the light between the input of the SLM and the output of the camera. To ensure this property, singular value decomposition (SVD) is applied on the coupling matrix w to find its singular values. Precisely, w is decomposed as:

$$w = U \Sigma V^T, \quad (13)$$

where U is the unitary change of basis matrix between transmission channels output modes and output free modes, Σ is a diagonal matrix whose elements are singular values λ_m of w , and V is a unitary change of basis matrix linking input free modes with transmission channel input modes of the system while the upper T refers to the matrix transpose [34]. Since the singular values of w are the square root of the energy transmission values of the transmission channels, we ensure the conservation of the total power transmission by normalizing the singular values as follows:

$$\lambda_m \longrightarrow \lambda_m / \sqrt{\sum_j \lambda_j^2}. \quad (14)$$

After this normalization, we reconstruct $w = U\Sigma'V^T$ where Σ' is a diagonal matrix whose elements are normalized singular values.

REFERENCES

- [1] F. Triefenbach, A. Jalal, B. Schrauwen, and J.-P. Martens, "Phoneme recognition with large hierarchical reservoirs," in *Advances in Neural Information Processing Systems*, vol. 23, 2010, pp. 2307–2315.
- [2] H. Jaeger and H. Haas, "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, pp. 78–80, 2004.
- [3] P. Buteneers, D. Verstraeten, T. W. P. Van Mierlo, D. Stroobandt, R. Raedt, H. Hallez, and B. Schrauwen, "Automatic detection of epileptic seizures on the intra-cranial electroencephalogram of rats using reservoir computing," *Artificial Intelligence Medicine*, vol. 53, pp. 215–223, 2011.
- [4] E. A. Antonelo, B. Schrauwen, and D. Stroobandt, "Event detection and localization for small mobile robots using reservoir computing," *Neural Networks*, vol. 21, pp. 862–871, 2008.
- [5] K. Lüdige and A. Röhm, "Computing with a camera," *Nat. Mach. Intell.*, vol. 1, pp. 551–552, 2019.
- [6] P. Antonik, N. Marsal, D. Brunner, and D. Rontani, "Human action recognition with a large-scale brain-inspired photonic computer," *Nat. Mach. Intell.*, vol. 1, pp. 530–537, 2019.
- [7] P. Antonik, N. Marsal, and D. Rontani, "Large-scale spatiotemporal photonic reservoir computer for image classification," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, p. 1, 2020.
- [8] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: a new framework for neural computation based on perturbations," *Neural Comput.*, vol. 14, p. 2531, 2002.
- [9] H. Jaeger, "Tech. Rep. GMD Report (German National Research Center for Information Technology)," Tech. Rep., 2001.
- [10] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, "An experimental unification of reservoir computing methods," *Neural Networks*, vol. 20, pp. 391–403, 2007.
- [11] G. V. der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," *Nanophotonics*, vol. 6, pp. 561–576, 2017.
- [12] G. Tanaka, T. Yamane, J.-B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: A review," *Neural Networks*, vol. 115, pp. 100–123, 2019.
- [13] Y. K. Chembo, "Machine learning based on reservoir computing with time-delayed optoelectronic and photonic systems," *Chaos*, vol. 30, p. 013111, 2020.
- [14] A. Lugnan, A. Katumba, F. Laporte, M. Freiberger, S. Sackesyn, C. Ma, E. Gooskens, J. Dambre, and P. Bienstman, "Photonic neuromorphic information processing and reservoir computing," *APL Photonics*, vol. 5, p. 020901, 2020.
- [15] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," *Sci. Rep.*, vol. 2, p. 287, 2012.
- [16] R. Martinenghi, S. Rybalko, M. Jacquot, Y. K. Chembo, and L. Larger, "Photonic nonlinear transient computing with multiple-delay wavelength dynamics," *Phys. Rev. Lett.*, vol. 108, p. 244101, 2012.
- [17] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," *Nat. Commun.*, vol. 4, p. 1364, 2013.
- [18] F. Duport, A. Smerieri, A. Akrouf, M. Haelterman, and S. Massar, "Fully analogue photonic reservoir computer," *Sci. Rep.*, vol. 6, p. 22381, 2016.
- [19] R. M. Nguimdo, E. Lacot, O. Jacquin, O. Hugon, G. V. der Sande, and H. G. de Chatellus, "Prediction performance of reservoir computing systems based on diode-pumped erbium doped microchip laser subject to optical feedback," *Opt. Lett.*, vol. 42, pp. 375–378, 2017.
- [20] L. Larger, A. Baylón-Fuentes, R. Martinenghi, V. S. Udaltsov, Y. K. Chembo, and M. Jacquot, "High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification," *Phys. Rev. X*, vol. 7, p. 1, 2017.
- [21] A. Argyris, J. Bueno, and I. Fischer, "Photonic machine learning implementation for signal recovery in optical communications," *Sci. Rep.*, vol. 8, p. 8487, 2018.
- [22] D. Brunner and I. Fischer, "Reconfigurable semiconductor laser networks based on diffractive coupling," *Opt. Lett.*, vol. 40, p. 3854, 2015.
- [23] J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, and D. Brunner, "Reinforcement learning in a large scale photonic recurrent neural network," *Optica*, vol. 5, p. 756, 2018.
- [24] J. Dong, M. Rafayelyan, F. Krzakala, and S. Gigan, "Optical reservoir computing using multiple light scattering for chaotic systems prediction," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, p. 1, 2020.
- [25] Q. Vinckier, F. Duport, A. Smerieri, K. Vandoorne, P. Bienstman, M. Haelterman, and S. Massar, "High-performance photonic reservoir computer based on a coherently driven passive cavity," *Optica*, vol. 2, pp. 438–446, 2015.
- [26] S. Sunada and A. Uchida, "Photonic reservoir computing based on nonlinear wave dynamics at microscale," *Sci. Rep.*, vol. 9, p. 19078, 2019.
- [27] K. Vandoorne, P. Mechet, T. V. Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nat. Commun.*, vol. 5, p. 4541, 2014.
- [28] K. Takano, C. Sugano, M. Inubushi, K. Yoshimura, S. Sunada, K. Kanno, and A. Uchida, "Compact reservoir computing with a photonic integrated circuit," *Opt. Express*, vol. 26, p. 29424, 2018.
- [29] F. D. L. Coarer, M. Sciamanna, A. Katumba, M. Freiberger, J. Dambre, P. Bienstman, and D. Rontani, "All-optical reservoir computing on a photonic chip using silicon-based ring resonators," *IEEE J. Sel. Top. Quantum Electron.*, vol. 24, pp. 1–8, 2018.
- [30] A. Katumba, X. Yin, J. Dambre, and P. Bienstman, "A neuromorphic silicon photonics nonlinear equalizer for optical communications with intensity modulation and direct detection," *IEEE J. Lightw. Technol.*, vol. 37, p. 2232, 2019.
- [31] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Comput. Sci. Rev.*, vol. 3, p. 127, 2009.
- [32] L. Appeltant, M. C. Soriano, J. D. G. Van der Sande, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," *Nat. Commun.*, vol. 2, p. 468, 2011.
- [33] L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer, "Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing," *Opt. Express*, vol. 20, p. 3241, 2012.
- [34] S. M. Popoff, G. Lerosey, M. Fink, A. C. Boccarda, and S. Gigan, "Controlling light through optical disordered media: transmission matrix approach," *New J. Phys.*, vol. 13, p. 123021, 2011.
- [35] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, 1998, p. 2278.
- [36] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local svm approach," in *Proceedings of the 17th International Conference on Pattern Recognition*, 2004.
- [37] H. E. Bahi, Z. Mahani, A. Zlati, and S. Saoud, "A robust system for printed and handwritten character recognition of images obtained by camera phone," Tech. Rep., 2015.
- [38] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.
- [39] R. M. Nguimdo, G. Verschaffelt, J. Danckaert, and G. V. der Sande, "Reducing the phase sensitivity of laser-based optical reservoir computing systems," *Opt. Express*, vol. 24, p. 1238, 2016.
- [40] P. Antonik, N. Marsal, D. Brunner, and D. Rontani, "Bayesian optimisation of large-scale photonic reservoir computers," *arXiv preprint arXiv:2004.02535*, 2020.
- [41] N. Semenova, X. Porte, L. Andreoli, M. Jacquot, L. Larger, and D. Brunner, "Fundamental aspects of noise in analog- hardware neural networks," *Chaos*, vol. 29, p. 103128, 2019.



Romain Modeste Nguimdo received the M.Sc. degree in theoretical physics from the University of Dschang, Dschang, Cameroon, and in mechanics from the University of Yaoundé I, Yaoundé, Cameroon in 2006, and the M.Sc. degree in theoretical physics and the Ph.D. degree from the Instituto de Física Interdisciplinar y Sistemas Complejos (CSIC-UIB), Palma de Mallorca, Spain, in 2008 and 2011, respectively. From 2012 to 2015, he was a Post-Doctoral fellow with Applied Physics Research Group, Vrije Universiteit Brussel (VUB), Brussels,

Belgium. From 2015 to 2018, he was awarded a 3 years F.N.R.S mandate with the affiliation at Université Libre de Bruxelles (U.L.B.), Brussels, Belgium. Since February 2019, he is a Post-Doctoral fellow at CentraleSupélec. Dr. R. M. Nguimdo has authored and co-authored 28 publications in international peer-reviewed journals. His research interests include nonlinear dynamics, optical chaos cryptography, and ultra-pure microwaves, random bit generation, and delay-based and large scale photonic reservoir computing.



Damien Rontani was born in Clamart, France. He received his Master of Science in Electrical Engineering (Supélec Curriculum) and a Master in Information Science, Energy and Systems from CentraleSupélec in Université Paris-Saclay (France) in 2004 and 2006, respectively. In parallel, he received his Master of Science in Electrical and Computer Engineering from the Georgia Institute of Technology (Georgia Tech, USA) in 2005. In 2011, he defended his PhD in Photonics at CentraleSupélec and his PhD in Electrical and Computer Engineering

at Georgia Tech. From 2011 to 2013, he was a Postdoctoral Fellow in the Physics Department at Duke University (USA). Since December 2013, he has been an assistant professor at CentraleSupélec, where he conducts leading-edge research in nonlinear dynamics of photonics systems, nonlinear optics, and machine learning at the physical layer for innovative applications in neuro-inspired information processing and cognitive computing. His research is conducted within the framework of the LMOPS EA-4423 Laboratory, a joint unit between CentraleSupélec and the Université de Lorraine (France). He has been the recipient of an IBM Faculty Award in Cognitive Computing in 2015 and a JSPS Fellowship for Overseas Researchers from the Japanese Society for the Promotion of Science in 2016. Dr. Rontani has authored and co-authored 29 publications in international peer-reviewed journals and has more than 60 contributions in international conferences.



Piotr Antonik was born in Minsk, Belarus, in 1989. He graduated in physics from the Université libre de Bruxelles (Brussels, Belgium) in 2013 and defended his PhD in 2017 under the direction of Prof. S. Massar. From 2017 to 2018 he was a Post-Doctoral researcher with the LMOPS EA-4423 and the Chair in Photonics at CentraleSupélec in Metz, France. In October 2018, he obtained a permanent position of Associate Professor at CentraleSupélec, Metz Campus, with the LMOPS EA-4423 and the Chair in Photonics. His research activities combine

machine learning, photonics, and FPGA programming, with applications in telecommunications, computer vision and biomedical imaging. To this day, Dr. Antonik authored and co-authored 1 book, 9 articles in peer-reviewed journals and more than 20 contributions to international conferences. His PhD thesis won the Springer Theses Award and was published in the Springer Theses collection.



Nicolas Marsal was born in Woippy, France. He received his Master of Physics from Université de Lorraine (France) in 2007. In 2010, he defended his PhD in Physics at CentraleSupélec. From 2011 to 2012, he was a Postdoctoral Fellow in the Nonlinear Optics department at Georgia Tech Lorraine (France). Since December 2012, he got a permanent position of assistant professor at CentraleSupélec. Its research activities are mainly focused on spatio-temporal dynamics of nonlinear photonic systems, propagation and interaction of non-conventional optical beams

and machine learning at the physical layer for innovative applications in neuro-inspired information processing and cognitive computing. His research is conducted within the framework of the LMOPS EA-4423 Laboratory, a joint unit between CentraleSupélec and the Université de Lorraine (France). Dr Marsal has authored and co-authored 26 publications in international peer-reviewed journals and presented more than 50 contributions in international conferences.