Photonic neuromorphic information processing and reservoir computing

A. Lugnan, A. Katumba, F. Laporte, M. Freiberger, S. Sackesyn, C. Ma, E. Gooskens, J. Dambre, and P. Bienstman

View Online   Export Citation   CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Novel frontier of photonics for data processing−Photonic accelerator
APL Photonics 4, 090901 (2019); https://doi.org/10.1063/1.5108912

Photonic integration for UV to IR applications
APL Photonics 5, 020903 (2020); https://doi.org/10.1063/1.5131683

Tutorial: Photonic neural networks in delay systems
Journal of Applied Physics 124, 152004 (2018); https://doi.org/10.1063/1.5042342

Photonic neuromorphic information processing and reservoir computing

View Online　　Export Citation　　CrossMark

A. Lugnan,[1] ID A. Katumba,[1] F. Laporte,[1] M. Freiberger,[2] S. Sackesyn,[1] C. Ma,[1] E. Gooskens,[1]
J. Dambre,[2] and P. Bienstman[1,a] ID

AFFILIATIONS

[1] Department of Information Technology, Ghent University/imec, Technologiepark-Zwijnaarde 126, 9052 Gent-Zwijnaarde,
Belgium

[2] IDLab, Ghent University/imec, Technologiepark-Zwijnaarde 126, 9052 Gent-Zwijnaarde, Belgium

[a] Author to whom correspondence should be addressed: Peter.Bienstman@UGent.be

ABSTRACT

Photonic neuromorphic computing is attracting tremendous research interest now, catalyzed in no small part by the rise of deep learning in many applications. In this paper, we will review some of the exciting work that has been going in this area and then focus on one particular technology, namely, photonic reservoir computing.

I. INTRODUCTION

There are ever more signs that the scaling of transistors, as dictated by Moore's law, is starting to falter. This has prompted the research community to investigate alternative forms of computation, in which neuromorphic and bio-inspired architectures are prime contenders. Indeed, the human brain often still outperforms digital computers in flexibility and performance on various pattern recognition tasks. More importantly, it is able to achieve this at very modest power consumption levels, typically the power equivalent to a light bulb, whereas (super)computers require orders of magnitude more power for similar tasks.

Therefore, there is a clear need for novel techniques to perform information processing at levels that are well beyond the limit of today's conventional computing processing power, like in really high-throughput massively parallel classification problems. This is especially true in the context of the growing prevalence of big data, which leads to more and more applications that generate massive temporal data streams, like the data aggregated from large sensor networks.

One important class of these brain-inspired (or "neuromorphic") techniques are the so-called artificial neural networks (ANNs) that consist of a number of interconnected computational units, dubbed "artificial neurons." The layout and operation of the ANN are inspired by the structure and information processing mechanism of the human brain. The so-called spiking neural networks aim to include details about the timing of individual spikes in the communication between the neurons. Typically however, these spiking dynamics are abstracted into a single number, namely, the firing rate of the neuron. The most prominent class of neural networks today is the so-called "deep learning" feedforward architecture, which is characterized by a large number of hierarchical layers of neurons, where information only flows in the forward direction. The performance of this approach has been such that it has been dominating the fields of machine learning and artificial intelligence over the last couple of years.[1,2]

Spurred by the promise of energy-efficient high-performance neuromorphic computing and the success of deep learning, a lot of effort has been put into designing special-purpose hardware to either fully implement these principles, or to accelerate a time-sensitive subset of these algorithms.[3]

Photonics has been identified as a very interesting platform for such a hardware implementation, since light-based technologies boast tremendous bandwidths, multiplexing capabilities in terms of wavelength, and come with other practical advantages such as immunity to electromagnetic interference and the possibility of co-integration with micro-electronics. Moreover, in quite a few of the high-throughput applications that could benefit from a

neuromorphic approach, the input is in the optical domain. This is the case for, e.g., signals coming out of an optical fiber network, or image data generated by some biomedical applications. For reasons of speed, power efficiency, and latency, it makes a lot of sense to try and process these data directly in the optical domain, without first converting it to the electronic domain. Additionally, if one makes use of brain-inspired approaches, one can benefit from the speed and power efficiency that these bring as well.

The attractiveness of optics has been realized early on, e.g., in the pioneering work of Wagner and Psaltis[4] and Psaltis et al.,[5] who used holographic materials to implement neural networks. Although this work is rather old at this stage, it makes a lot of sense to try and revisit some of its concepts, especially in the light of the recent progress in the fields of optical materials and integrated photonics. In additon, the emerging field of programmable photonics[6,7] promises to provide vital technology to support the field of photonic neuromorphic computing.

II. PHOTONIC NEUROMORPHIC INFORMATION PROCESSING

In this section, we will review a number of recent approaches in the field of photonic neuromorphic information processing. This overview is neither intended to be exhaustive nor in-depth, but should serve to illustrate the variety within the field, as well as some of its recent successes. It is also worth noting that there are other interesting non-neuromorphic technologies that leverage the unique properties of photonics for information processing, like coherent Ising machines[8–10] and optical quantum computing.[11]

A. Optical spiking neurons

In the field of spiking neural networks, one strives to stay closer to the properties of biological neurons. More specifically, the fact that neurons communicate using a sequence of spikes (analog in their timing, but binary in their absence or presence) is explicitly taken into account.[12] These neurons are of the so-called leaky-integrate-and-fire (LIF) variety, and only show a response if the time-integrated input they receive exceeds a certain threshold. In that case, they send out a spike with a well-defined waveform, independent of the exact energy of the input above the threshold. In dynamical systems theory, this phenomenon is known as excitability.[13] Spiking neural networks are still actively researched and show promise in fields like temporal pattern detection,[14] where an interesting feature is their power efficiency because of the discrete nature of the spikes. They are also being explored in the context of deep learning.[15]

There have been numerous attempts to implement spiking neurons in photonics. We will now review some of these, as well as point to some relevant aspects when integrating single neurons in a larger system.

Typical photonic implementations of spiking neurons have mostly been in active photonic components like graphene excitable lasers,[16] distributed feedback (DFB) lasers[17] vertical-cavity surface-emitting lasers,[18–20] or micropillars.[21] Non-linear components like ring resonators have also been employed for this purpose.[22]

Another avenue of research consists in using phase change materials like GST to facilitate spike processing, either as a non-volatile weight in a synapse[23] or as a non-linear element in combination with other photonic circuitry to emulate an entire neuron.[24]

An aspect that has received somewhat less interest experimentally is the challenge of coupling these spiking elements together,[25,26] in order to create larger networks. For this, one needs to make sure the output spike has sufficient energy to trigger the next neuron. This is an area where more research is needed, if one wants to move further in the direction of applications.

Also important are learning mechanisms, i.e., the best way to adapt the synaptic weights between neurons in order to achieve a certain desired functionality. A mechanism to realize this has been identified in biological neurons as spike-timing dependent-plasticity (STDP),[27] where the time difference between the pre- and the post-synaptic spikes determines the strength of the weight change. To date, there have been several attempts to implement a similar functionality in photonics.[28–30]

B. Deep-learning accelerators

Given the importance of deep learning, people have been looking into ways to accelerate these architectures. Since a large part of the computation is spent doing matrix multiplication, this computational primitive is an important candidate for a hardware speedup. This field has a long history, see, e.g., Ref. 31, but has recently been revived in the context of integrated nanophotonics,[32] where the weights were set using thermal phase shifters in a network of Mach-Zehnder interferometers.

The scalability of systems like these is limited due to the footprint, but alternative free-space schemes have been put forward based on the multiplication that is happening when two coherent signals beat on a photodetector.[33] Their system combines time-division multiplexing and electrical integration, and allows us to calculate matrix multiplication.

Another proposed approach is to use tunable transmission through a GST-based modulator,[34] where the multiplicand is the input to the modulator, and the multiplier is the tuning signal of the modulator.

A conceptually similar approach is presented in Ref. 35, but here, the input is constant and the multiplicand and multiplier are the control inputs to two cascaded acousto-optical modulators. The setup is used to perform convolutions on an input image.

Experiments on a system combining both time and wavelength multiplexing to perform matrix operations are presented in Ref. 36. A dispersive medium is employed to temporally align the elements that need to be summed.

Lin et al.[37] demonstrated a deep learning network based on a series of diffraction gratings fabricated by 3D printing, where the structures are designed such that a useful function-like handwritten digit recognition is performed. This network can have a large number of parameters, but cannot be changed anymore after fabrication. A recurrent network version of this concept can be found in Ref. 38. The idea that the shaping of wavefront when propagating through a structure can perform a computation is somewhat related to the concept of computational metamaterials.[39]

Most of the techniques described above are essentially linear in nature. Efforts are also underway to implement the required

non-linear activation between layers in an optical or electro-optical fashion, e.g., based on modulators,[40] amplifiers,[41,42] or by having electro-optical interactions.[43] Zuo et al.[44] employ laser-cooled atoms to implement the non-linearity. In Ref. 45, a diffractive neural network is theoretically proposed that works in the Fourier domain and contains a photorefractive material to provide the non-linearity.

III. PHOTONIC RESERVOIR COMPUTING

A. Reservoir computing

Reservoir computing (RC)[46–48] was initially proposed as a methodology to ease the training of recurrent neural networks, which traditionally had been rather challenging. More recently, however, it has gained popularity as a neuromorphic computational paradigm to solve a variety of complex problems. Reservoir computing initially emerged as a software-only technique and merely presented another algorithmic way of processing temporal data on digital computers. However, it has evolved into much more over the past decade, as people have realized its suitability for hardware implementations, especially its robustness against fabrication tolerances.

The RC system consists of three basic parts: the input layer which couples the input signal into a non-linear dynamical system, the "reservoir" (i.e., the recurrent neural network, which is kept untrained), and finally the output layer that typically linearly combines the states of the reservoir to provide the time-dependent output signal. An illustration of this reservoir computing architecture is given in Fig. 1.

Another way of looking at reservoir computing is to consider the reservoir as a non-linear dynamical system that acts as pre-filter on the input data stream, transforming these data into a higher dimensional space. In this space, it will be easier to separate different classes using a hyperplane boundary, as provided by the linear readout. The use of linear boundaries has a positive influence on the capabilities of the system to generalize in a robust fashion to unseen input. (These advantages of linear classifiers are also used in, e.g., support vector machines.[50])

In discretized time, the reservoir state update equation is given in a general form by
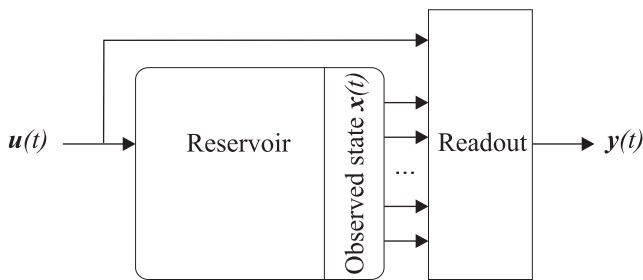

FIG. 1. Schematic representation of a reservoir computing system. The input signal u(t) is fed into the reservoir, and the resulting reservoir states x(t) together with the input are used to learn a linear readout that is then used to generate the output signal y(t) [Reprinted with permission from Katumba et al., IEEE J. Sel. Top. Quantum Electron. **24**, 1–10 (2018). Copyright 2018 IEEE.].

$$\vec{x}[k+1] = f\big(W_{res}\vec{x}[k] + \vec{w}_{in}(\vec{u}[k+1] + \vec{u}_{bias})\big), \qquad (1)$$

where f is a nonlinear function, $\vec{u}$ is the input to the reservoir, and $\vec{u}_{bias}$ is a fixed scalar bias applied to the inputs of the reservoir. For an N-node reservoir, W_{res} is an $N \times N$ matrix representing the interconnections between reservoir components. $\vec{w}_{in}$ is an N-dimensional column vector whose elements are nonzero for each active input node.

The output is given by a simple linear combination of the states

$$\vec{y}[k] = \vec{W}_{out} \cdot \vec{x}[k]. \qquad (2)$$

To use the reservoir to solve a particular task, a machine learning algorithm is used to train a set of weights (the readout) using a set of known labeled example data, such that a linear combination of the optical signals recorded at each node approximates a desired output as closely as possible. This algorithm typically takes the form of a least-square minimization, where the weighs are calculated through a Moore-Penrose pseudo-inverse. These weights are then used to generate the output signal for any unseen subsequently injected input signal sequences. RC systems are fast to train and find the global optimum without the need for an iterative method, as opposed to their traditional neural network counterparts. Reservoirs have shown state-of-the-art performance on a range of complex tasks on time-dependent data (such as speech recognition, non-linear channel equalization, robot control, time series prediction, financial forecasting, and handwriting recognition.[51,52]).

A key discovery was that the reservoir computing platform provides a natural framework for implementing learning systems on hardware platforms where the possibility to set all the internal parameters of the network is limited, and for which using a random, fixed network is, therefore, a great advantage. Examples of RC implementations in mechanical systems, memristive systems, atomic switch networksand Boolean logic elements systems can be found in Refs. 53–57.

Also in the field of photonics, several hardware implementations of the RC paradigm exist; see, e.g., the overviews in Refs. 49, 58, and 59. Broadly speaking, photonic reservoirs can be divided into systems with a single node coupled to a feedback loop (where the nodes are time-multiplexed and travel along the delay line) and spatial systems (where the nodes are explicitly localized at certain locations in a free-space setup or on a chip). We will now discuss some examples of these classes in more detail.

B. Systems with a single non-linear node with feedback

Historically, this was one of the earliest implementations of photonic reservoir computing, given the ease of implementation due to the low hardware complexity.[60–64] In these systems, there is only a single non-linear element, but it is coupled to a feedback loop with delay time τ (Fig. 2). The neurons (nodes) are virtual in nature and travel through the feedback loop. This means the nodes are time-multiplexed inside the system. This is done as follows: Every τ seconds, the input $x(t)$ is subject to a sample-and-hold operation. The resulting signal is then multiplied by a so-called masking signal, which typically has as period the round trip time τ as well. The mask period is subdivided into N different constant values, resulting
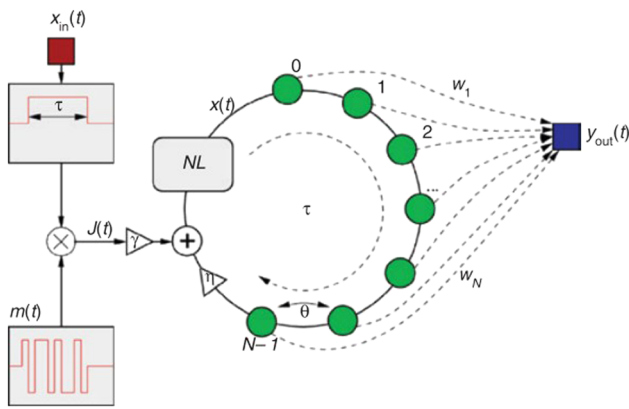
FIG. 2. In a delay-based reservoir, the input $x_{in}(t)$ is multiplied by masking signal $m(t)$. Virtual nodes (green) travel along the delay line and are processed in a non-linear element. The states of the nodes are linearly combined with weights w_i to generate the output signal $y_{out}(t)$. [Reprinted with permission from Van Der Sande *et al.*, Nanophotonics **6**, 561–576 (2017). Copyright 2017 De Gruyter Author(s), licensed under a Creative Commons Attribution 3.0 Unported License.].

in N virtual nodes inside the delay line with duration τ/N. The duration of the node should be such that the central non-linear element is always in a useful transient dynamic state. In addition, the node duration is typically short compared to the memory time scale of the non-linear element such that different (mostly neighboring) virtual nodes can interact.

To date, experimental demonstrations of these types of photonic reservoirs routinely achieve state-of-the-art performance on various information processing tasks, showing that photonic RC is competitive for analog information processing.[65–73]

Typically, these systems are rather bulky, since they rely on either a fiber loop or a free-space propagation path to implement the feedback loop. However, they have been efforts to integrate systems like these on-chip.[10,74] This has the advantage of increasing the stability and speed, and adds the option to more easily exploit coherence.

C. Free-space reservoir systems

The time-multiplexed systems described above have the fundamental drawback that the throughput is limited, simply because each node is processed sequentially. Other types of reservoirs employ spatial multiplexing, i.e., each node has an explicitly identifiable separate location and all the nodes can be processed in parallel.

In Ref. 75, an 8×8 array of VCSELs acts as the nodes of the reservoir, and they are diffractively coupled among them, with a spatial modulator setting the weights. Systems like these achieved excellent performance on 5-bit header recognition tasks. An different free-space implementation is described in Ref. 76, presenting a 9 $\times$ 9 2D array of nodes defined on a spatial light modulator.

A large-scale system was introduced in Ref. 77, consisting of a network of 2025 nodes. The non-linearity is performed in the electronic domain, which limits the update rate to 5 Hz. The read-out weights are implemented all-optically, using an array of digital micromirrors. The results in a set of binary weights, which are trained using reinforcement learning.

Another experimental demonstration of this concept is presented in Ref. 78, running at 640 Hz, used to predict the Mackey-Glass time series.[79] presents another free-space system, this time to perform image classification, capable of implementing networks of 16.384 nodes.

As an alternative to free-space systems, reservoirs have been proposed where the nodes consist of individual modes in a large-area waveguide like a plastic optical fiber.[80] A combination of scatterers and a cavity is responsible for creating rich dynamics.

D. Waveguide-based on-chip reservoir computing

Another option is to implement the different nodes on a photonic chip and connect them using waveguides. The performance of integrated photonic reservoirs has been studied numerically and/or experimentally for different types of nodes, starting with networks of optical amplifiers in Refs. 81 and 82. Later, networks of resonators were studied as well,[83–87] with applications ranging from speech recognition of isolated spoken digits, over binary tasks like header recognition, to image recognition. Integrated photonic reservoirs are particularly compelling, especially when implemented in a CMOS-compatible platform, as they can take advantage of its associated benefits for technology reuse and mass production.

A later development in the design of RC systems is the realization that for certain tasks that are not strongly non-linear, it is possible to achieve state-of-the-art performance using a completely passive linear network, i.e., one without amplification or non-linear elements. The required non-linearity is introduced at the readout point, typically with a photodetector.[88] Aside from the integrated implementation introduced in Ref. 88, the passive architecture has been adapted to the single-node-with-delayed-feedback architecture in the form of a coherently driven passive cavity.[65]

Apart from simplicity from a fabrication point-of-view, a further advantage of such a passive architecture is the reduced power consumption, since the computation itself does not require external energy.

The integrated photonic reservoirs typically studied in the past have been limited to planar architectures in a bid to minimize crossings that manifest as a source of signal cross-talk and extra losses. This constrains the design space from which reservoir configurations can be chosen. We introduced the original swirl reservoir architecture in Ref. 89 as a way to satisfy planarity constraints while allowing for a reasonable mixing of the input signals. The input to the integrated photonics reservoir chip could be to a single input node as is in Ref. 88 or to multiple inputs, which has some advantages over the former strategy, as is discussed extensively in Ref. 90.

Our work in Ref. 88 experimentally verified that a passive integrated photonic reservoir can yield error-free performance on the header recognition task for headers up to 3 bit in length with simulations indicating that it should be possible to go up to 8 bit headers (see Fig. 3).

The architecture used in Ref. 88, the so-called swirl architecture [Fig. 4(a)], has the disadvantage of containing nodes that are non-symmetrical. Therefore, these nodes suffer from modal radiation at each 2×1 combiner [for example, node 7 in Fig. 4(a)]. These radiation losses cannot be avoided in these single-mode combiners, and
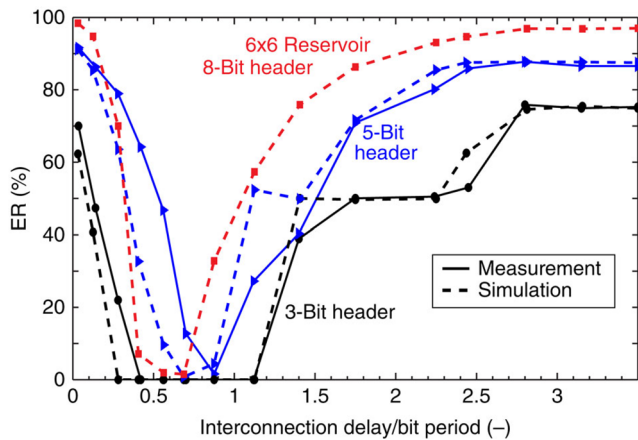
FIG. 3. Performance of a 6 × 6 swirl passive integrated photonics reservoir on the 3, 5, and 8 bit header recognition task. Good qualitative correspondence between theory and experiment is observed [Reprinted with permission from Katumba *et al.*, IEEE J. Sel. Top. Quantum Electron. **24**, 1–10 (2018). Copyright 2018 IEEE.].

FIG. 5. Light enters an on-chip photonic crystal cavity through a single waveguide, mixes inside the cavity and leaves the cavity from all connected waveguides. As such, this structure can provide all the required functionality for a passive reservoir. [Reprinted with permission from Laporte *et al.*, Opt. Express **26**, 7955 (2018). Copyright 2018 The Optical Society].

averaged over the different input phases, 50% of the power radiates away there.

Additionally, when focusing on the mixing behavior of the swirl architecture, there are some nodes that do not contribute significantly to the dynamics. For example, the nodes in the corners only play a very limited role in the reservoir dynamics, as they basically are only delay lines that do not redistribute the power to other nodes. Similarly, the other nodes at the edge of the reservoir, having fewer neighbors, contribute less to the dynamics than the nodes in the center.

Therefore, to address these two issues of losses and mixing, we presented a new architecture, the so-called the four-port architecture[91] [Fig. 4(b)]. In that architecture, the lossy 2 × 1 combiners are avoided and only 2 × 2 devices are used instead, and additional waveguides are added to increase the mixing.

Simulations of this four-port architecture for a 64-node network show that its losses can be an order of magnitude lower than those of the swirl architecture, as, apart from residual scattering and propagation loss, all the input power is redistributed to one of the
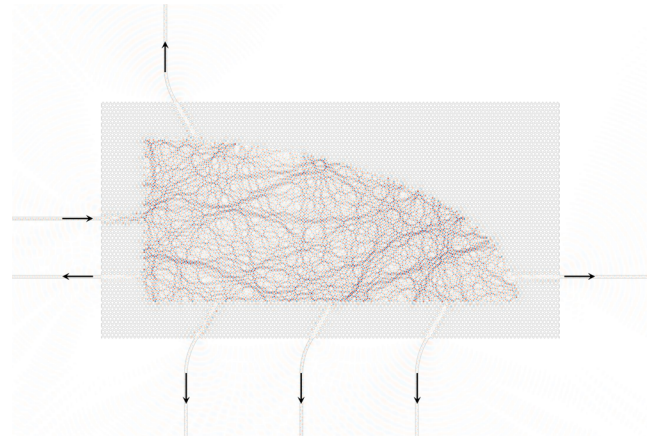
output channels in the four-port architecture. These simulations also show that the power is more evenly distributed through the reservoir than in the swirl-architecture and thus easier to measure in an experiment.

E. Cavity-based on-chip passive reservoir computing

An alternative architecture is based on a quasi-chaotic cavity, which can be orders of magnitude smaller than the waveguide-based approaches. (By quasi-chaotic we mean that the properties of the cavity are sensitive to the initial conditions set by fabrication intolerances, but that they are reproducible for a given device.) As can be seen in Fig. 5, by choosing the shape of the cavity carefully,[92–95] we can achieve complicated wave patterns and can therefore expect mixing between different delayed versions of the input signal, similar as in the waveguide-based approach. In addition, by tuning the Q-factor of the cavity, we can impact the memory of the reservoir.
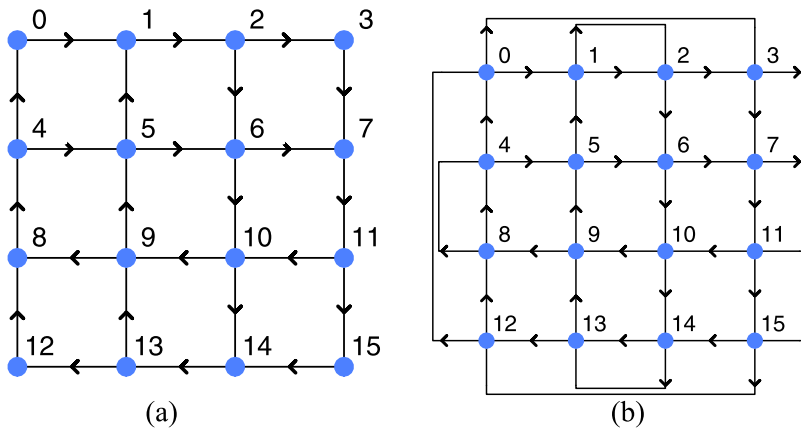

(a)

(b)

FIG. 4. Schematic view of two waveguide-based reservoir architectures. Compared to the swirl architecture, the four-port architecture does not suffer from the inherent 3 dB combiner loss.(a) Swirl architecture and (b) Four-port architecture [(a) Reprinted with permission from Katumba *et al.*, IEEE J. Sel. Top. Quantum Electron. **24**, 1–10 (2018). Copyright 2018 IEEE.].
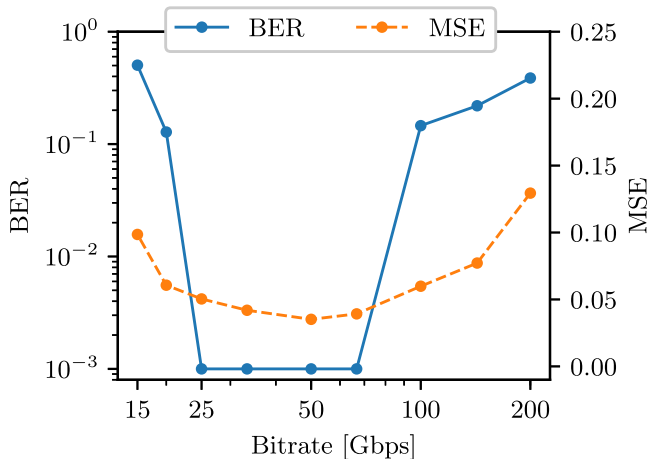
FIG. 6. BER and MSE for the XOR task for the 30 μm × 60 μm cavity for a wide range of frequencies. A clear frequency region of good performance is observed. (Reprinted with permission from Laporte *et al.*, Opt. Express **26**, 7955 (2018). Copyright 2018 The Optical Society).

This approach of using a cavity as reservoir results in extremely small on-chip footprints (<0.1 mm^2) while retaining similar performance on benchmarks.

We first illustrate this using simulations on the XOR task, where the reservoir needs to compute an XOR between two subsequent bits in the bit stream. This task is known in machine learning to be a non-linear task due to the fact that the output cannot be found by just performing a linear classification algorithm such as linear regression on the inputs. By performing the XOR task at different bit rates, a region of successful operation for the 30 μm × 60 μm cavity shown in Fig. 5 can be found, as can be seen in Fig. 6. In this figure, we can see by looking at the MSE (mean squared error) that the optimal bit
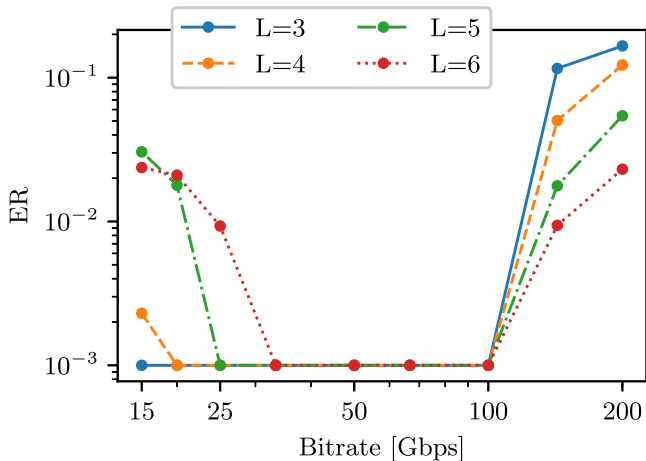

FIG. 7. Header recognition error rate for worst-performing header in the bit stream. Longer headers result in a narrower window of operation. [Reprinted with permission from Laporte *et al.*, Opt. Express **26**, 7955 (2018). Copyright 2018 The Optical Society].

rate is at 50 Gbps. However, looking at the BER (bit error rate), our simulations find a full band of suitable frequencies between 25 Gbps and 67 Gbps (Fig. 7).

However, for applications in telecom, recognizing headers in a bit stream is often more useful than performing an XOR. The exact same simple cavity design can perform this task as well. For each bit in the bit stream, a class label was given corresponding to the header of length L made by the current bit and the $L - 1$ previous bits. Again, by sweeping over the bit rate, we find a successful operation range of up to 100 Gbps for up to 6-bit headers.

IV. ARCHITECTURAL CONSIDERATIONS FOR PHOTONIC RESERVOIR COMPUTING

A. Improving performance through multi-reservoir architectures

The feasible size of a single integrated photonic reservoir currently remains relatively small, usually around 100 nodes or less. While a number of factors contribute to this limitation, a main reason is that the memory of the reservoir does not scale linearly with its size, e.g., due to losses in the system. Therefore, alternative solutions to improve the performance of integrated photonic reservoirs have to be pursued. One possible path of exploration is to increase the available computational power by combining multiple separate reservoirs into a single computing device.

Neural network literature[1] shows that performing subsequent non-linear transformations on the input data are highly beneficial in terms of performance on a wide variety of tasks. The space of possible architectures featuring multiple reservoirs is quite large, and we conducted simulations to evaluate a limited number of promising topologies in Ref. 96. Figure 8 shows two of these topologies which we found to perform best, which we term *ensembling* and *chaining*.

In ensembling,[97] several classifiers are trained for the same task and joined by taking a combination of the individual classifier predictions. While classifiers can be combined in many different ways, a simple approach is to average classifier predictions. A more sophisticated approach could be to train an additional set of weights to learn how to combine all obtained predictions. In general, ensembling aims to combine the strengths of all trained classifiers and mitigate their weaknesses. In order to build a simple ensemble of reservoirs, we connect the nodes of several reservoirs to a single readout [see Fig. 8(a)].

An important factor in order to improve performance of an ensemble over single reservoirs is that the models of an ensemble must differ from each other, i.e., their mistakes must be uncorrelated where possible. All passive photonic reservoirs differ by construction due to the silicon photonics manufacturing process. Since the effective index of the delay line waveguides cannot be entirely controlled, strong variations in phase affect the state signals of any fabricated reservoir. This makes these passive photonic reservoirs interesting candidates for reservoir ensembling in hardware.

In our second investigated connection scheme, chaining, reservoirs are combined by feeding the predicted output of a given reservoir into the readout stage of the next reservoir [see Fig. 8(b)]. This way, the prediction of earlier reservoirs is incrementally improved upon by later reservoirs. This is again a technique that depends on all used reservoirs differing from each other.
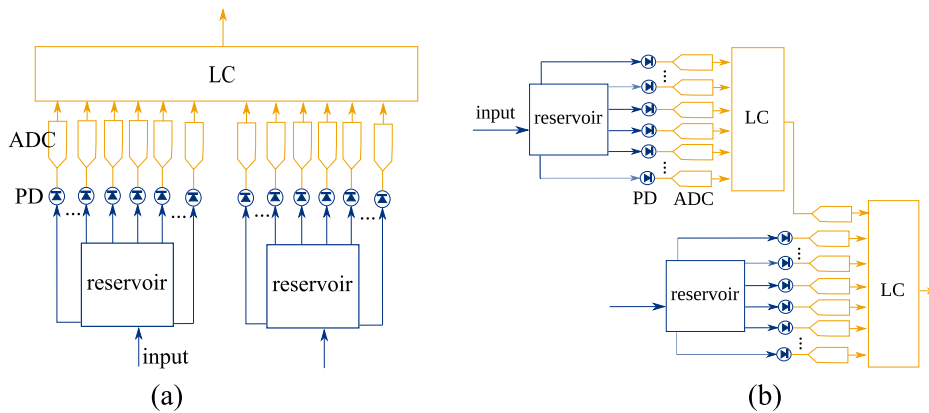
FIG. 8. Ensembling and chaining integrated photonic reservoirs with electrical readout PD: photodiode, ADC: analog-digital converter, LC: linear classifier, OM: optical modulator, and OC: optical combiner. The same input is sent to both reservoirs (a) ensemble and (b) chaining [Reprinted with permission from Freiberger et al., IEEE J. Sel. Top. Quantum Electron. **26**, 1–11 (2019). Copyright 2019 IEEE.].

Note that the chaining architecture remotely resembles the DeepESN architecture as proposed in Ref. 98, but has some significant differences. In Ref. 98, each reservoir module is driven with all the states of its predecessor. The readout is trained on the states of all reservoirs in the setup. While this architecture would likely exhibit a high memory capacity and its performance in software appears promising, it would have to be simplified for integrated photonics technology. For instance, a random combination of each reservoir's states could be projected into the subsequent reservoir after which the states of all reservoirs are readout. While our efforts so far have focused on more straight-forward architectures, bringing integrated photonic reservoirs closer to a DeepESN architecture appears to be an attractive direction for future research.

To evaluate the performance of our architectures, we simulated four 4×8 passive photonic reservoirs connected in both an ensembling and chaining setup and train each to solve the Santa Fe chaotic laser prediction task.[99] In this task, we predict the next sampling value for a time series recorded from a far IR laser driven in a chaotic regime. Figure 9 shows the results of our systems on this task. We compare our results with a 8×16 nodes baseline, i.e., a single reservoir that has the same overall number of nodes.

When taking a look at the results in Fig. 9, we can see that the ensemble of reservoirs slightly outperforms the chained reservoirs and the baseline. Both methods show improvement for all observed symbol rates with every reservoir added to the architecture. It is

notable that two ensembled reservoirs for some symbol rates match the larger baseline reservoir which makes use of twice the number of nodes. A possible explanation for this effect is that several small reservoirs introduce more richness and variation in the resulting combined reservoir states than would be possible for a single larger reservoir. Our chaining approach outperforms the baseline on a few bit rates, but is slightly outperformed on most bit rates.

We compare our simulations results with the delayed feedback approach in Ref. 68. Soriano et al.[68] report an normalized mean root squared error (NMSE) of 0.025 on the Santa Fe dataset for a 500-node system. They obtain lower error rates, but their system is much larger than the 128 nodes reservoir setups that have been used here.

B. Dealing with limited weight resolution

One of the significant advantages of photonic reservoir computing is its potential to achieve ultra-low power consumption. To realize this idea, one possibility is, instead of using traditional heater-based weighting elements, to use weighting elements that incorporate a non-volatile material. Whereas heaters require constant power to maintain the weights during operation, with non-volatile weights, no further power consumption is required after it has been set.[100,101]

However, depending on the technology, a non-volatile weighting element can come with a compromise on weighting accuracy. With Barium Titanate (BTO), for example, the resolution of the

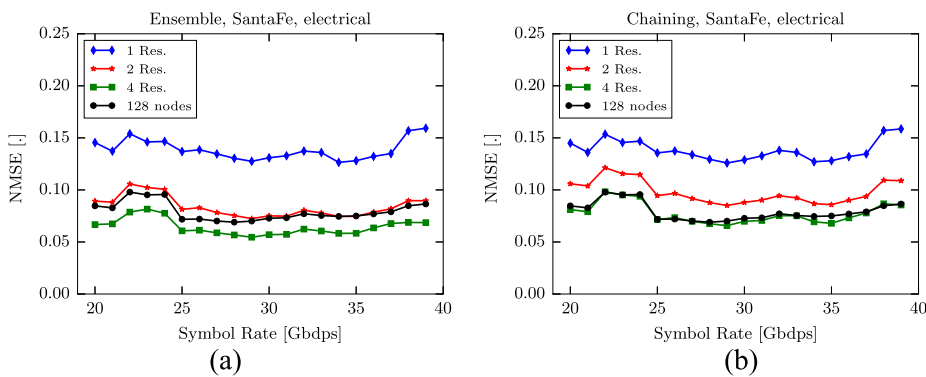

FIG. 9. Normalized mean root squared error (NMSE) of simulated reservoir on the Santa Fe time series prediction task as a function of bit rate for 1, 2, and 4 reservoirs combined using ensembling and chaining in the electrical domain. (a) Ensemble and (b) chaining [Reprinted with permission from Freiberger et al., IEEE J. Sel. Top. Quantum Electron. **26**, 1–11 (2019). Copyright 2019 IEEE.].

refractive index tuning is limited to around 10 to 30 levels.[101] On top of that, some inevitable drift causes further noise on the levels. For a system with such low weighting resolution and severe weighting noise, the performance could easily drop several orders of magnitude in terms of bit error rate (BER).

To tackle this problem, we proposed[102] a new training method inspired by methods that have been used in deep learning quantization. Typically, after a full-precision model has been trained and quantized, a subset of weights is identified to be either pruned[103] or kept fixed.[104] The other weights are then retrained in full precision and requantized. If necessary, this step can be repeated in an iterative fashion, retraining progressively smaller subsets of the weights in order to find the most optimal and stable solution.

A crucial part of these methods is selecting a subset of weights to be left fixed or to be pruned. Random selection of weights is not a good idea because there is a high probability of eliminating "good" weights that convey important information. Other authors[103,104] tackle this problem by choosing the weights with the smallest absolute value.

This is reasonable in deep learning models, since the millions of weights can provide enough tolerance when it comes to accidentally selecting the 'wrong' weights. However, in the readout systems for reservoir computing, we have much fewer weights and a much more limited resolution with severe noise. In this case, the absolute value will not provide enough information, as a combination of many small weights could be important in fine-tuning the performance of the network. This will lead to a risk of an accuracy loss when specific 'wrong' connections (that are more sensitive to perturbations) are chosen to be retrained.

Instead, we adapt a different (albeit more time-consuming) approach, where after quantization, we compare several different random partitions between weights that will be kept fixed and weights that will be retrained (in full precision) and requantized. By comparing the task performance for these different partitions, we are able to pick the best one.

Figure 10 illustrates the results based on our photonic reservoir simulation framework.[105] In this case, we chose the XOR task with 4-bit delay, i.e., calculating the XOR of the current bit with the bit from 4 periods ago. The noise profile is based on a Gaussian distribution, the value of the noise level represents the ratio between the standard deviation of the Gaussian distribution σ and the interval between two adjacent weighting levels Δw as

$$\text{Noise level} = \sigma/\Delta w.$$

The figure compares the performance of three sets of weights: our exploratively retrained weights, naively quantized weights, and the ideal full-precision weights. The naive weights are just a direct quantization of the full precision weights taking into account resolution and noise levels. The figure is clearly showing that the 4-bit delayed XOR task is not easy because even the system with full-precision weights performs significantly worse when even a little amount of noise is introduced. Naive quantization performs much worse: it can hardly provide convincing performance across the whole noise spectrum. However, the explorative retraining method provides up to 2 orders of magnitude better performance when the resolution is limited to only 8 levels. For a resolution of 16 levels, the retraining
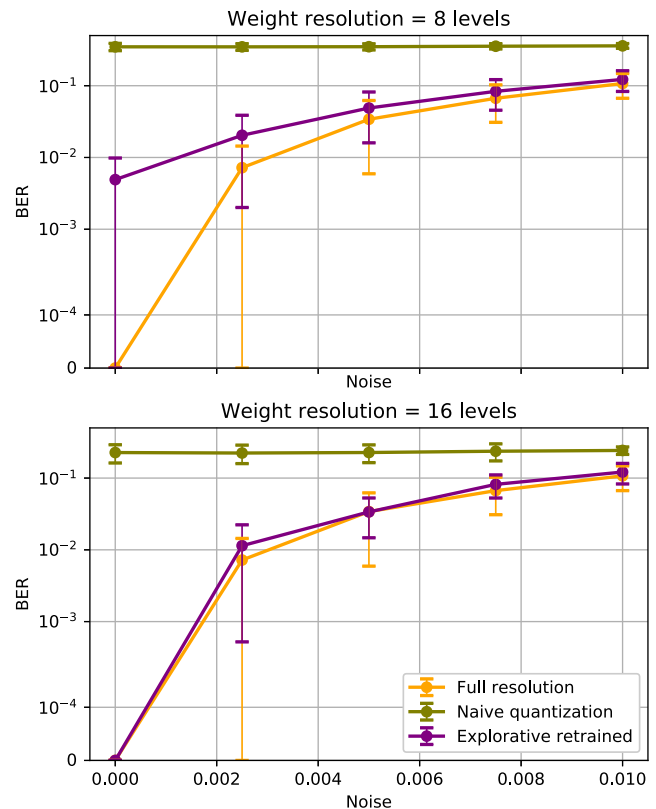

FIG. 10. Performance comparison of three different training methods for two weighting resolutions, 8 levels(top), and 16 levels(bottom) for XOR with 4-bit delay task vs different noise levels. The purple curve represents our explorative retraining method, the olive curve represents the naive quantization weights and the orange curve represents the full precision ideal weighting system. The error bars represent different random instantiations of the input weights and internal reservoir weights.

method can provide a steady performance that is very close to the full-precision system.

C. Multiwavelength reservoirs

One way to improve the scaling of photonic reservoirs is to exploit the wavelength dimension, as wavelength multiplexing is one of the key strengths of photonics that is already widely exploited in the fiber-optic telecommunications industry. It is therefore a natural solution when trying to decrease the footprint of reservoirs.

Indeed, wavelength multiplexing could enable the processing of (identical) tasks in parallel, such as equalizing several different wavelength channels. Of course, in order to still benefit from the improvement in footprint, we should be able to use the same set of weights for different wavelengths. Otherwise, we would need optical filters and a different optical readout for each wavelength.

One way to overcome this is to specifically train the reservoir for good performance on multiple wavelengths simultaneously. Figure 11 shows that, when training a reservoir on an XOR task at 1552.3 nm and 1552.4 nm, a single set of weights can be used to get
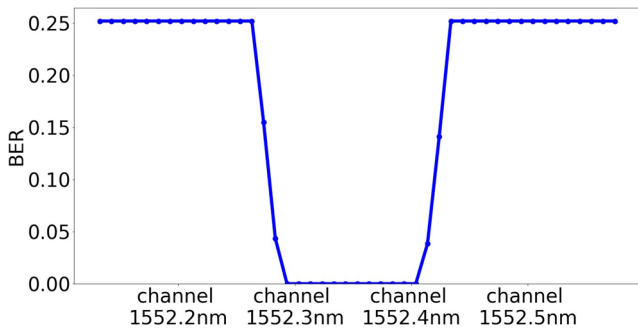
FIG. 11. Error rate for the XOR task. A single set of readout weights was trained for operation at 1552.3 nm and 1552.4 nm, resulting in a system that can work without change for these two adjacent WDM channels.

good performance over this wavelength range containing 2 WDM channels.

V. APPLICATIONS OF PHOTONIC RESERVOIR COMPUTING

In this section, we review two applications we have been focusing on recently, which share the property that the input is in the optical domain so that a photonic information processing scheme is well suited.

A. Non-linear dispersion compensation in telecom links

Reservoir computing can be used to perform equalization of telecom signals in the optical domain, without the need for extensive and power-hungry electronic DSP.[106–108] Here, we compare the performance of a PhRC equalizer to a traditional linear feedforward equalization (FFE) filter trained on the same amount of data. The link parameters can be found in Ref. 109. An adaptive FFE filter with
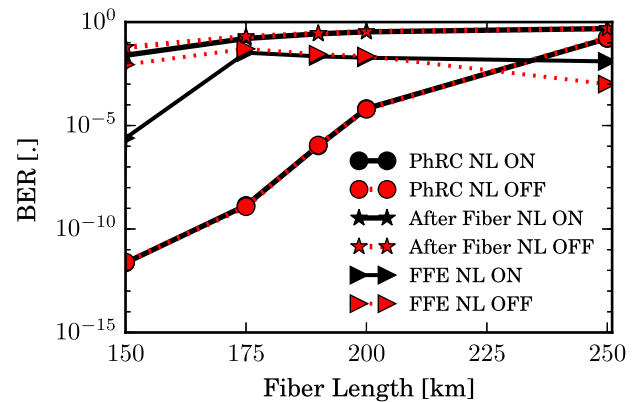

FIG. 12. BER of the PhRC equalizer as compared to that of an FIR Feed Forward Equalizer (FFE) trained on the same amount of data for different fiber lengths. The launch power is set to 15 mW. NL ON- nonlinear propagation. NL OFF—nonlinear propagation is deactivated. "After fiber": the uncorrected signal at the output of the fiber [Reprinted with permission from Katumba et al., J. Lightwave Technol. **37**, 2232–2239 (2019). Copyright 2019 IEEE.].

31 taps is used (the filter goes over the training data four times to allow for convergence). The results are shown in Fig. 12. The PhRC equalizer outperforms the FFE equalizer with BERs over 5 orders of magnitude lower for 150 km transmission length and an order or two of magnitude lower for 200 km. The difference in performance originates in the fact that the PhRC equalizer is a nonlinear compensation device; it takes advantage of the nonlinear transformation in the reservoir to better model the distortion and outstrip the performance of the FFE filter. In Fig. 12, we also plot the cases with and without the fiber nonlinearity. We observe that at the distances under consideration, the fiber nonlinearities are not yet deleterious. We do however observe that the reservoir is able to make use of its nonlinear nature to outperform the FFE equalizer for these links.
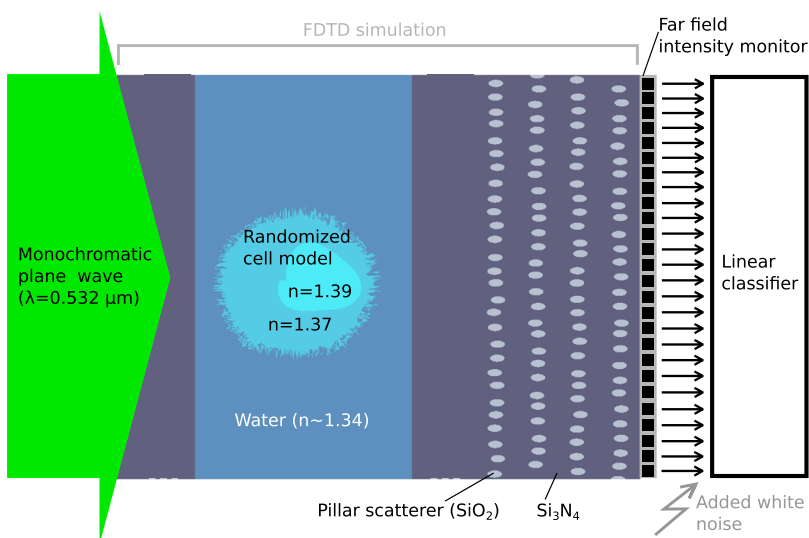
FIG. 13. Schematic of the classification process. A monochromatic plane wave impinges on a microfluidic channel containing a randomized cell model in water ($n_{H_2O} \sim 1.34$, $n_{cytoplasm}$ = 1.37, $n_{nucleus}$ = 1.39); the forward scattered light passes through a collection of silica scatterers ($n_{SiO_2} \sim 1.461$) embedded in silicon nitride ($n_{Si_3N_4} \sim 2.027$) and organized in layers; the radiation intensity is then collected by a far-field monitor, which is divided into bins (pixels); each pixel value is fed into a trained linear classifier (logistic regression) that consists of a weighted sum of the pixel values.

B. Biological cell sorting

In flow cytometry, a large number of suspended microparticles, e.g., biological cells, can be detected and analyzed one by one while flowing at high speed through a measuring device.[110] On top of this, cell sorting can be straightforwardly implemented by performing the cell classification on the fly and employing its outcome to trigger a subsequent particle sorter. Even though flow cytometry has been developing for decades, a lot of effort is still being spent to reduce instrumentation cost, size and complexity.[111] In addition, a high throughput is often needed to sort a statistically meaningful number of cells. These requirements can be met in label-free imaging cytometers,[112] where the light scattered by microparticles is imaged and automatically analyzed, e.g., through a machine learning algorithm. Still, the computation time needed by conventional and state-of-the-art classification algorithms, including deep neural networks,[113] significantly limits the throughput of cell sorting applications.[114–116] However, the same operating principle of RC, i.e., training a linear readout applied to a fixed high-dimensional nonlinear transformation of the input signals, can be applied in this case to classify cells by only computing a weighted sum of the pixel intensities acquired by an image sensor. This in principle allows extremely fast classifications and is particularly suitable for low-cost hardware implementations.

We recently presented a proof-of-concept of such approach employing the outcomes of thousands of 2D FDTD simulations of cells illuminated by a laser (Fig. 13) as training samples for a linear machine learning classifier.[117] A collection of dielectric pillar scatterers (silica in silicon nitride cladding) is interposed between the microfluidic channel containing the particles and an image sensor. This is done to enrich the non-linear (sinusoidal) mapping connecting the light phase modulation due to the cell presence and the corresponding interference pattern acquired by the image sensor. In holographic microscopy, the sample image is usually reconstructed from the acquired interference pattern by a computationally expensive algorithm. Instead, we proposed to directly apply a linear classifier on the acquired pattern so that the classification time is only given by the image acquisition and a weighted sum of pixel intensities. We showed that the use of pillar scatterers can halve the error in both the classifications of cell models with 2 different nucleus sizes and with 2 different nucleus shapes, using the same scatterer configuration.

We recently performed similar classifications as in Ref. 117, based on the nucleus size of cells, for 9 different pillar scatterer configurations, simulating the interference pattern with and without a far field projection (Fig. 14). To obtain better comparable classification performance estimations, we increased the number of simulated samples per scatterer configuration to 7200 and we optimized both the L2 regularization strength (strength^{-1} value chosen among 1/10, 1/25, 1/50, 1/100, 1/500) and the image sensor resolution (No. pixels is chosen among equidistant rounded values in the range [300, 700]). This was done through 2 nested 5-fold cross-validations: the inner one was used for hyperparameter optimization, the outer one for providing error bars to the performance estimations (see Ref. 117 for more background).

We conclude by noting that for all the explored scatterer configurations (top of Fig. 14, from B to I) the classification error using near field and far field diffraction patterns is reduced respectively
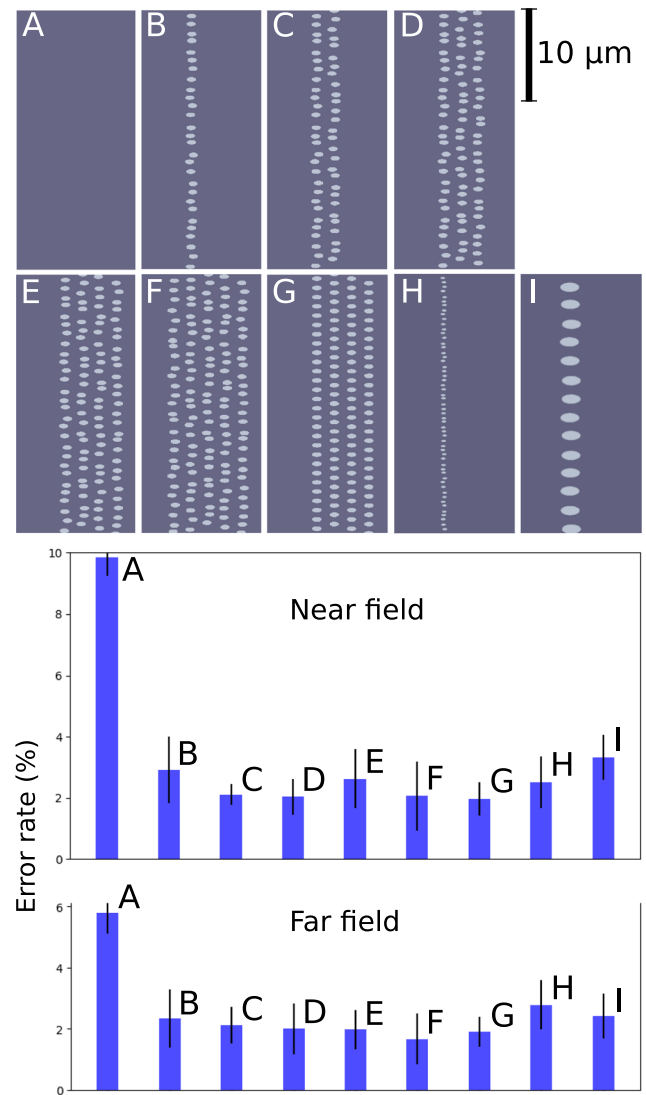
FIG. 14. Top: pillar scatterer configurations employed in the 2D FDTD simulations, as depicted in Fig. 13. Bottom: bar plots of the corresponding error in the classification of cells with 2 different nucleus size (as in Ref. 117), using near field and far field interference patterns respectively. The error bars (black thin bars) represent 2 standard deviations of the outer cross-validation results.

by a factor of ~4 and of ~2 with respect to the case without pillar scatterers (configuration A in Fig. 14). Therefore, the fast label-free cell classification performed by a simple linear classifier without other image processing or lenses, can be significantly improved by adding in the optical path a simple collection of diffractive elements. Finally, this improvement is shown to be surprisingly robust against imperfections of the diffracting layer, e.g., due to fabrication errors.

VI. CONCLUSIONS AND PERSPECTIVES

It is clear that the field of brain-inspired photonic computing has recently undergone an explosion in activity, where many

players investigate different interesting options and research avenues. An important factor contributing to this popularity is no doubt related to the rise in prominence of deep learning. This increased attention is an extremely positive evolution, given the complexity of the issues and the formidable technical challenges to overcome. Foremost among those is the issue of scalability. Compared to, e.g., software-based deep learning approaches, with tens of millions of free parameters, experimental realizations in photonics have been rather modest, and will probably remain so for a while, in view of issues related to power consumption, footprint and complexity of control.

Still, we should not limit ourselves to purely mimicking what is happening in deep learning, as other approaches and applications could perhaps provide a more natural fit with the inherent capabilities and strong points of photonics.

There are several aspects that need to be considered when deciding whether an application would be better served by an electronic or a photonic implementation.

In case the input is already in the optical domain, it makes a lot of sense to do the processing optically too. If we are dealing with electronic signals on the other hand, any cost related to electro-optical conversion will need to be weighted carefully against potential advantages.

A second aspect is the modulation speed of the input signal. Wave propagation and interference happen at time scales far shorter than anything that can be achieved with traditional electronics, and this can be a crucial asset. Obviously, the speed of the entire system is limited by the weakest link, which, depending on the implementation, could be either the modulation of the input source, the bandwidth of the detector at the end of the system, or the speed of any non-linear element inside the system. Therefore, the whole picture needs to be carefully considered.

Finally, another natural advantage of photonics over electronics is the extra degrees of freedom we can exploit, like wavelength and polarization. This multiplexing results in a virtual system that can be several times larger than the actual physical system that implements it.

In summary, these are exciting times to perform research in the field of analog information processing in photonics. Several recent research results show tremendous promise. However, the next steps and key challenges will be to translate these typically small-scale prototypes to a concrete, industrially relevant application, in such a way that it outperforms traditional electronic implementations.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436 (2015).

[2] W. G. Hatcher and W. Yu, "A survey of deep learning: Platforms, applications and emerging research trends," IEEE Access **6**, 24411–24432 (2018).

[3] C.-Y. Chen, B. Murmann, J.-S. Seo, and H.-J. Yoo, "Custom sub-systems and circuits for deep learning: Guest editorial overview," IEEE J. Emerging Sel. Top. Circuits Syst. **9**, 247–252 (2019).

[4] K. Wagner and D. Psaltis, "Multilayer optical learning networks," Appl. Opt. **26**, 5061–5076 (1987).

[5] D. Psaltis, D. Brady, X.-G. Gu, and S. Lin, "Holography in artificial neural networks," Nature **343**, 325–330 (1990).

[6] A. Ribeiro, A. Ruocco, L. Vanacker, and W. Bogaerts, "Demonstration of a 4 × 4-port universal linear circuit," Optica **3**, 1348–1357 (2016).

[7] D. Pérez, I. Gasulla, L. Crudgington, D. J. Thomson, A. Z. Khokhar, K. Li, W. Cao, G. Z. Mashanovich, and J. Capmany, "Multipurpose silicon photonics signal processor core," Nat. Commun. **8**, 636 (2017).

[8] P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, "A fully programmable 100-spin coherent Ising machine with all-to-all connections," Science **354**, 614–617 (2016).

[9] C. Tradonsky, I. Gershenzon, V. Pal, R. Chriki, A. A. Friesem, O. Raz, and N. Davidson, "Rapid laser solver for the phase retrieval problem," Sci. Adv. **5**, eaax4530 (2019).

[10] K. Harkhoe, G. Verschaffelt, A. Katumba, P. Bienstman, and G. Van der Sande, "Demonstrating delay-based reservoir computing using a compact photonic integrated chip," arXiv:1907.02804 (2019).

[11] J. L. O'Brien, A. Furusawa, and J. Vučković, "Photonic quantum technologies," Nat. Photonics **3**, 687–695 (2009).

[12] F. Ponulak and A. Kasinski, "Introduction to spiking neural networks: Information processing, learning and applications," Acta Neurobiol. Exp. **71**, 409–433 (2011).

[13] E. Izhikevich, *Dynamical Systems in Neuroscience* (MIT Press, 2007), p. 111.

[14] O. Bichler, D. Querlioz, S. J. Thorpe, J.-P. Bourgoin, and C. Gamrat, "Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity," Neural networks **32**, 339–348 (2012).

[15] M. Pfeiffer and T. Pfeil, "Deep learning with spiking neurons: Opportunities and challenges," Front. Neurosci. **12**, 774 (2018).

[16] P. Y. Ma, B. J. Shastri, T. F. de Lima, A. N. Tait, M. A. Nahmias, and P. R. Prucnal, "All-optical digital-to-spike conversion using a graphene excitable laser," Opt. Express **25**, 033504 (2017).

[17] H.-T Peng, G. Angelatos, T. Ferreira de Lima, M. A. Nahmias, A. N. Tait, S. Abbaslou, B. J. Shastri, and P. R. Prucnal, "Temporal information processing with an integrated laser neuron," IEEE J. Sel. Top. Quantum Electron. **26**, 1–9 (2019).

[18] T. Deng, J. Robertson, Z.-M. Wu, G.-Q. Xia, X.-D. Lin, X. Tang, Z.-J. Wang, and A. Hurtado, "Stable propagation of inhibited spiking dynamics in vertical-cavity surface-emitting lasers for neuromorphic photonic networks," IEEE Access **6**, 67951–67958 (2018).

[19] S. Xiang, A. Wen, and W. Pan, "Emulation of spiking response and spiking frequency property in VCSEL-based photonic neuron," IEEE Photonics J. **8**, 1–9 (2016).

[20] J. Robertson, T. Deng, J. Javaloyes, and A. Hurtado, "Controlled inhibition of spiking dynamics in VCSELs for neuromorphic photonics: Theory and experiments," Opt. Lett. **42**, 1560–1563 (2017).

[21] F. Selmi, R. Braive, G. Beaudoin, I. Sagnes, R. Kuszelewicz, T. Erneux, and S. Barbay, "Spike latency and response properties of an excitable micropillar laser," Phys. Rev. E **94**, 042219 (2016).

[22] T. Van Vaerenbergh, M. Fiers, J. Dambre, and P. Bienstman, "Simplified description of self-pulsation and excitability by thermal and free-carrier effects in semiconductor microcavities," Phys. Rev. A **86**, 063808 (2012).

[23] Z. Cheng, C. Ríos, W. H. P. Pernice, C. D. Wright, and H. Bhaskaran, "On-chip photonic synapse," Sci. Adv. **3**, e1700160 (2017).

[24] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," Nature **569**, 208 (2019).

[25] T. V. Vaerenbergh, M. Fiers, P. Mechet, T. Spuesens, R. Kumar, G. Morthier, J. Dambre, and P. Bienstman, "Cascadable excitability in microrings," Opt. Express **20**, 20292–20308 (2012).

[26] A. Tait, M. Nahmias, B. Shastri, and P. Prucnal, "Broadcast and weight: An integrated network for scalable photonic spike processing," J. Lightwave Technol. **32**, 4029 (2014).

[27] L. F. Abbott and S. B. Nelson, "Synaptic plasticity: Taming the beast," Nat. Neurosci. **3**(S11), 1178–1183 (2000).

[28] R. Toole, A. Tait, T. Ferreira de Lima, M. Nahmias, B. Shastri, P. Prucnal, and M. Fok, "Photonic implementation of spike timing dependent plasticity and learning algorithms of biological neural systems," J. Lightwave Technol. **34**, 470 (2015).

[29] Q. Ren, Y. Zhang, R. Wang, and J. Zhao, "Optical spike-timing-dependent plasticity with weight-dependent learning window and reward modulation," Opt. Express **23**, 025247 (2015).

[30] S. Xiang, J. Gong, Y. Zhang, X. Guo, Y. Han, A. Wen, and Y. Hao, "Numerical implementation of wavelength-dependent photonic spike timing dependent plasticity based on VCSOA," IEEE J. Quantum Electron. **54**, 1–7 (2018).

[31] R. Heinz, J. Artman, and S. Lee, "Matrix multiplication by optical methods," Appl. Opt. **9**, 2161 (1970).

[32] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," Nat. Photonics **11**, 441–446 (2017); arXiv:1610.02365.

[33] R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, and D. Englund, "Large-scale optical neural networks based on photoelectric multiplication," Phys. Rev. X **9**, 021032 (2019).

[34] C. Rios, N. Youngblood, Z. Cheng, M. Le Gallo, W. H. P. Pernice, C. D. Wright, A. Sebastian, and H. Bhaskaran, "In-memory computing on a photonic platform," Sci. Adv. **5**, eaau5759 (2019).

[35] S. Xu, J. Wang, R. Wang, J. Chen, and W. Zou, "High-accuracy optical convolution unit architecture for convolutional neural networks by cascaded acousto-optical modulator arrays," Opt. Express **27**, 19778–19787 (2019).

[36] Y. Huang, W. Zhang, F. Yang, J. Du, and Z. He, "Programmable matrix operation with reconfigurable time-wavelength plane manipulation and dispersed time delay," Opt. Express **27**, 20456–20467 (2019).

[37] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," Science **361**, 1004–1008 (2018).

[38] T. W. Hughes, I. A. D. Williamson, M. Minkov, and S. Fan, "Wave physics as an analog recurrent neural network," Sci. Adv. **5**, eaay6946 (2019); arXiv:1904.12831.

[39] A. Silva, F. Monticone, G. Castaldi, V. Galdi, A. Alù, and N. Engheta, "Performing mathematical operations with metamaterials," Science **343**, 160–163 (2014).

[40] J. K. George, A. Mehrabian, R. Amin, J. Meng, T. F. de Lima, A. N. Tait, B. J. Shastri, T. El-Ghazawi, P. R. Prucnal, and V. J. Sorger, "Neuromorphic photonics with electro-absorption modulators," Opt. Express **27**, 5181–5191 (2019).

[41] G. Mourgias-Alexandris, A. Tsakyridis, N. Passalis, A. Tefas, K. Vyrsokinos, and N. Pleros, "An all-optical neuron with sigmoid activation function," Opt. Express **27**, 9620 (2019).

[42] B. Shi, N. Calabretta, and R. Stabile, "Inp photonic circuit for deep neural networks," in OSA Advanced Photonics Congress (AP) 2019 (IPR, Networks, NOMA, SPPCom, PVLED) (Optical Society of America, 2019), p. IW2A.3.

[43] I. A. D. Williamson, T. W. Hughes, M. Minkov, B. Bartlett, S. Pai, and S. Fan, "Reprogrammable electro-optic nonlinear activation functions for optical neural networks," IEEE J. Sel. Top. Quantum Electron. **26**, 1–12 (2020).

[44] Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, and S. Du, "All optical neural network with nonlinear activation functions," Optica **6**, 1132–1137 (2019); arXiv:1904.10819.

[45] T. Yan, J. Wu, T. Zhou, H. Xie, F. Xu, J. Fan, L. Fang, X. Lin, and Q. Dai, "Fourier-space diffractive deep neural network," Phys. Rev. Lett. **123**, 023901 (2019).

[46] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," Neural Comput. **14**, 2531–2560 (2002).

[47] H. Jaeger, "The echo state approach to analysing and training recurrent neural networks," GMD Report No. 148, GMD - German National Research Institute for Computer Science, 2001.

[48] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, "An experimental unification of reservoir computing methods," Neural Networks **20**, 391–403 (2007).

[49] A. Katumba, M. Freiberger, F. Laporte, A. Lugnan, S. Sackesyn, C. Ma, J. Dambre, and P. Bienstman, "Neuromorphic computing based on silicon photonics and reservoir computing," IEEE J. Sel. Top. Quantum Electron. **24**, 1–10 (2018).

[50] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods (Cambridge University Press, Cambridge, 2000).

[51] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," Science **304**, 78–80 (2004).

[52] M. Lukoševičius, H. Jaeger, and B. Schrauwen, "Reservoir computing trends," KI - Künstliche Intelligenz **26**, 365–371 (2012).

[53] H. Hauser, A. Ijspeert, R. Füchslin, R. Pfeifer, and W. Maass, "Towards a theoretical foundation for morphological computation with compliant bodies," Biol. Cybern. **105**, 355–370 (2011).

[54] H. O. Sillin, R. Aguilera, H.-H. Shieh, A. V. Avizienis, A. Aono, A. Z. Stieg, and J. K. Gimzewski, "A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing," Nanotechnology **24**, 384004 (2013).

[55] M. S. Kulkarni and C. Teuscher, "Memristor-based reservoir computing," in Proceedings of the 2012 IEEE/ACM International Symposium on Nanoscale Architectures - NANOARCH '12 (ACM Press, New York, New York, USA, 2012), pp. 226–232.

[56] D. Marković, N. Leroux, M. Riou, F. Abreu Araujo, D. Querlioz, A. Fukushima, S. Yuasa, J. Trastoy, P. Bortolotti, and J. Grollier, "Reservoir computing with the frequency, phase, and amplitude of spin-torque nano-oscillators," Appl. Phys. Lett. **114**, 012409 (2019).

[57] J. C. Coulombe, M. C. A. York, and J. Sylvestre, "Computing with networks of nonlinear mechanical oscillators," PLoS One **12**, e0178663 (2017).

[58] G. Van Der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," Nanophotonics **6**, 561–576 (2017).

[59] Photonic Reservoir Computing - Optical Recurrent Neural Networks, edited by D. Brunner, M. C. Soriano, and G. Van der Sande (De Gruyter, 2019).

[60] D. Brunner, B. Penkovsky, B. A. Marquez, M. Jacquot, I. Fischer, and L. Larger, "Tutorial: Photonic neural networks in delay systems," J. Appl. Phys. **124**, 152004 (2018).

[61] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," Nat. Commun. **2**, 468 (2011).

[62] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," Sci. Rep. **2**, 287 (2012).

[63] 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 **20**, 3241 (2012).

[64] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," Opt. Express **20**, 22783 (2012).

[65] 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 **2**, 438–446 (2015); arXiv:1501.03024.

[66] 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. **4**, 1364 (2013).

[67] A. Dejonckheere, A. Smerieri, L. Fang, J.-l. Oudar, M. Haelterman, and S. Massar, "All-optical reservoir computer based on saturation of absorption," Opt. Express **22**, 10868 (2014).

[68] M. C. Soriano, S. Ortín, D. Brunner, L. Larger, C. R. Mirasso, I. Fischer, and L. Pesquera, "Optoelectronic reservoir computing: Tackling noise-induced performance degradation," Opt. Express **21**, 12 (2013).

[69] R. M. Nguimdo, G. Verschaffelt, J. Danckaert, and G. Van der Sande, "Fast photonic information processing using semiconductor lasers with delayed optical feedback: Role of phase dynamics," Opt. Express **22**, 8672 (2014).

[70]K. Hicke, M. Escalona-Morán, D. Brunner, M. C. Soriano, I. Fischer, and C. R. Mirasso, "Information processing using transient dynamics of semiconductor lasers subject to delayed feedback," IEEE J. Sel. Top. Quantum Electron. **19**, 1501610 (2013).

[71]R. M. Nguimdo and T. Erneux, "Enhanced performances of a photonic reservoir computer based on a single delayed quantum cascade laser," Opt. Lett. **44**, 49–52 (2019).

[72]Y. Hou, G. Xia, W. Yang, D. Wang, E. Jayaprasath, Z. Jiang, C. Hu, and Z. Wu, "Prediction performance of reservoir computing system based on a semiconductor laser subject to double optical feedback and optical injection," Opt. Express **26**, 10211–10219 (2018).

[73]J. Vatin, D. Rontani, and M. Sciamanna, "Enhanced performance of a reservoir computer using polarization dynamics in vcsels," Opt. Lett. **43**, 4497–4500 (2018).

[74]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 **26**, 29424–29439 (2018).

[75]D. Brunner and I. Fischer, "Reconfigurable semiconductor laser networks based on diffractive coupling," Opt. Lett. **40**, 3854–3857 (2015).

[76]J. Pauwels, G. Van der Sande, A. Bouwens, M. Haelterman, and S. Massar, "Towards high-performance spatially parallel optical reservoir computing," Proc. SPIE **10689**, 1068904 (2018).

[77]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 **5**, 756–760 (2018).

[78]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. **26**, 1–12 (2020).

[79]P. Antonik, N. Marsal, and D. Rontani, "Large-scale spatiotemporal photonic reservoir computer for image classification," IEEE J. Sel. Top. Quantum Electron. **26**, 1–12 (2019).

[80]C. Mesaritakis and D. Syvridis, "Reservoir computing based on transverse modes in a single optical waveguide," Opt. Lett. **44**, 1218 (2019).

[81]K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, and J. V. Campenhout, "Toward optical signal processing using photonic reservoir computing," Opt. Express **16**, 11182–11192 (2008).

[82]K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, and P. Bienstman, "Parallel reservoir computing using optical amplifiers," IEEE Trans. Neural Networks **22**, 1469–1481 (2011).

[83]C. Mesaritakis, V. Papataxiarhis, and D. Syvridis, "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 (2013).

[84]M. A. A. Fiers, T. Van Vaerenbergh, F. Wyffels, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Nanophotonic reservoir computing with photonic crystal cavities to generate periodic patterns," IEEE Trans. Neural Networks Learn. Syst. **25**, 344–355 (2014).

[85]H. Zhang, X. Feng, B. Li, Y. Wang, K. Cui, F. Liu, and W. Dou, "Integrated photonic reservoir computing based on hierarchical time-multiplexing structure," Opt. Express **22**, 31356–31370 (2014).

[86]C. Mesaritakis, A. Kapsalis, and D. Syvridis, "All-optical reservoir computing system based on InGaAsP ring resonators for high-speed identification and optical routing in optical networks," Proc SPIE **9370**, 937033 (2015).

[87]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. **24**, 1–8 (2018).

[88]K. Vandoorne, P. Mechet, T. Van 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. **5**, 3541 (2014).

[89]K. Vandoorne, T. Van Vaerenbergh, M. Fiers, P. Bienstman, D. Verstraeten, B. Schrauwen, and J. Dambre, "Photonic reservoir computing and information processing with coupled semiconductor optical amplifiers," in *2011 Fifth Rio De La Plata Workshop on Laser Dynamics and Nonlinear Photonics* (IEEE, 2011), pp. 1–3.

[90]A. Katumba, M. Freiberger, P. Bienstman, and J. Dambre, "A multiple-input strategy to efficient integrated photonic reservoir computing," Cognit. Comput. **9**, 307–314 (2017).

[91]S. Sackesyn, C. Ma, J. Dambre, and P. Bienstman, "An enhanced architecture for silicon photonic reservoir computing," in *Cognitive Computing 2018 - Merging Concepts with Hardware* (2018), pp. 1–2.

[92]F. Laporte, A. Katumba, J. Dambre, and P. Bienstman, "Numerical demonstration of neuromorphic computing with photonic crystal cavities," Opt. Express **26**, 7955 (2018).

[93]B. C. Grubel, B. T. Bosworth, M. R. Kossey, H. Sun, A. B. Cooper, M. A. Foster, and A. C. Foster, "Silicon photonic physical unclonable function," Opt. express **25**, 12710–12721 (2017).

[94]C. Liu, R. E. C. van der Wel, N. Rotenberg, L. Kuipers, T. F. Krauss, A. Di Falco, and A. Fratalocchi, "Triggering extreme events at the nanoscale in photonic seas," Nat. Phys. **11**, 358–363 (2015).

[95]H.-J. Stöckmann and J. Stein, "Quantum chaos in billiards studied by microwave absorption," Phys. Rev. Lett. **64**, 2215–2218 (1990).

[96]M. A. Freiberger, S. Sackesyn, C. Ma, A. Katumba, P. Bienstman, and J. Dambre, "Improving time series recognition and prediction with networks and ensembles of passive photonic reservoirs," IEEE J. Sel. Top. Quantum Electron. **26**, 1–11 (2019).

[97]J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning* (Springer series in statistics New York, NY, USA, 2001), Vol. 1.

[98]C. Gallicchio, A. Micheli, and L. Pedrelli, "Deep reservoir computing: A critical experimental analysis," Neurocomputing **268**, 87–99 (2017).

[99]C. Chatfield and A. S. Weigend, "Time series prediction: Forecasting the future and understanding the past: Neil A. Gershenfeld and Andreas S. Weigend, 1994, 'The future of time series', in: A.S. Weigend and N.A. Gershenfeld, eds., (Addison-Wesley, Reading, MA), 1-70.," Int. J. Forecasting **10**, 161–163 (1994).

[100]S. Abel, T. Stöferle, C. Marchiori, C. Rossel, M. D. Rossell, R. Erni, D. Caimi, M. Sousa, A. Chelnokov, B. J. Offrein, and J. Fompeyrine, "A strong electro-optically active lead-free ferroelectric integrated on silicon," Nat. Commun. **4**, 1671 (2013).

[101]S. Abel, T. Stöferle, C. Marchiori, D. Caimi, L. Czornomaz, M. Stuckelberger, M. Sousa, B. J. Offrein, and J. Fompeyrine, "A hybrid barium titanate–silicon photonics platform for ultraefficient electro-optic tuning," J. Lightwave Technol. **34**, 1688–1693 (2016).

[102]C. Ma, F. Laporte, J. Dambre, and P. Bienstman, "Addressing limited weight resolution in a fully optical neuromorphic reservoir computing readout," arXiv:1908.02728 (2019).

[103]S. Han, J. Pool, J. Tran, and W. Dally, "Learning both weights and connections for efficient neural network," Adv. Neural Inf. Process. Syst. **28**, 1135–1143 (2015).

[104]A. Zhou, A. Yao, Y. Guo, L. Xu, and Y. Chen, "Incremental network quantization: Towards lossless cnns with low-precision weights," arXiv:1702.03044 (2017).

[105]M. Fiers, T. V. Vaerenbergh, K. Caluwaerts, D. V. Ginste, B. Schrauwen, J. Dambre, and P. Bienstman, "Time-domain and frequency-domain modeling of nonlinear optical components at the circuit-level using a node-based approach," J. Opt. Soc. Am. B **29**, 896–900 (2012).

[106]A. Katumba, P. Bienstman, and J. Dambre, "Photonic reservoir computing approaches to nanoscale computation," in *Proceedings of the 2nd ACM International Conference on Nanoscale Computing and Communication, ACM NANOCOM 2015* (2015).

[107]A. Argyris, J. Bueno, and I. Fischer, "Photonic machine learning implementation for signal recovery in optical communications," Sci. Rep. **8**, 8487 (2018).

[108]M. Sorokina, S. Sergeyev, and S. Turitsyn, "Fiber echo state network analogue for high-bandwidth dual-quadrature signal processing," Opt. Express **27**, 2387 (2019).

[109]A. Katumba, X. Yin, J. Dambre, and P. Bienstman, "A neuromorphic silicon photonics nonlinear equalizer for optical communications with intensity modulation and direct detection," J. Lightwave Technol. **37**, 2232–2239 (2019).

[110]A. Adan, G. Alizada, Y. Kiraz, Y. Baran, and A. Nalbant, "Flow cytometry: Basic principles and applications," Crit. Rev. Biotechnol. **37**, 163–176 (2017).

[111] R. J. Yang, L. M. Fu, and H. H. Hou, "Review and perspectives on microfluidic flow cytometers," Sens. Actuators, B Chem. **266**, 26–45 (2018).

[112] Y. Han, Y. Gu, A. C. Zhang, and Y. H. Lo, "Review: Imaging technologies for flow cytometry," Lab Chip **16**, 4639–4647 (2016).

[113] Y. Li, A. Mahjoubfar, C. L. Chen, K. R. Niazi, L. Pei, and B. Jalali, "Deep cytometry: Deep learning with real-time inference in cell sorting and flow cytometry," Sci. Rep. **9**, 11088 (2019).

[114] L. Lagae, D. Vercruysse, A. Dusa, C. Liu, K. De Wijs, R. Stahl, G. Vanmeerbeeck, B. Majeed, Y. Li, and P. Peumans, "High throughput cell sorter based on lensfree imaging of cells," in *Technical Digest - International Electron Devices Meeting IEDM* (Institute of Electrical and Electronics Engineers Inc., 2015), Vol. 2016, pp. 13.3.1–13.3.4.

[115] Y. J. Heo, D. Lee, J. Kang, K. Lee, and W. K. Chung, "Real-time image processing for microscopy-based label-free imaging flow cytometry in a microfluidic chip," Sci. Rep. **7**, 11651 (2017).

[116] B. Cornelis, D. Blinder, B. Jansen, L. Lagae, and P. Schelkens, "Fast and robust Fourier domain-based classification for on-chip lens-free flow cytometry," Opt. Express **26**, 014329 (2018).

[117] A. Lugnan, J. Dambre, and P. Bienstman, "Integrated pillar scatterers for speeding up classification of cell holograms," Opt. Express **25**, 030526 (2017).