

Computing with Integrated Photonic Reservoirs



Joni Dambre, Andrew Katumba, Chonghuai Ma, Stijn Sackesyn, Floris Laporte, Matthias Freiberger, and Peter Bienstman

Abstract The idea of using photonic systems as reservoirs to perform general-purpose computing was first introduced in 2008. Since then, a wide range of systems using either discrete or integrated optical components has been explored. In this chapter, we summarise a decade of research into integrated coherent photonic reservoirs. In these systems, information is carried by the intensity and the phase of light waves. Computations emerge from the way the light propagates inside the system, and the ways in which light that travels along different paths is mixed and transformed. We discuss the computational capabilities of these reservoirs and the trade-offs that can be made to optimise them. We also discuss the technological constraints that are encountered in building such systems and the ways these reflect on their design and training. Finally, we give an overview of recent approaches to combining multiple such reservoirs into larger and computationally more powerful systems.

1 Introduction

Light is commonly used as an information carrier in telecommunication systems. Optical communication has almost entirely replaced electronics for all but very short distances, because electronic interconnect is inherently bandwidth limited and consumes far more power at high data rates. Optical telecommunication systems are now mostly held back by the fact that computation is still performed electronically. The limitations of electronics pop up at the boundaries between the optical and the electronic domain, where optical signals are converted into electronic ones and vice versa. This fact has revived the interest in pushing as much computation as possible into the optical domain. However, the devices and material systems used for integrated photonics do not match well with the typical behaviour of elementary digital building blocks such as logic gates and flipflops. For this reason, the general-purpose divide-and-conquer design methodology used for electronic digital computing systems is not portable to photonics. Also, in contrast to integrated electronics, device

J. Dambre (✉) · A. Katumba · C. Ma · S. Sackesyn · F. Laporte · M. Freiberger · P. Bienstman
Ghent University and imec, Technologiepark 126, 9052 Zwijnaarde, Ghent, Belgium
e-mail: Joni.Dambre@UGent.be

© Springer Nature Singapore Pte Ltd. 2021
K. Nakajima and I. Fischer (eds.), *Reservoir Computing*, Natural Computing Series,
https://doi.org/10.1007/978-981-13-1687-6_17

dimensions are hard coupled to the wavelength of the light that is used, which means that integrated photonics does not scale well. For this reason it is essential to find the sweet spot in the trade-off between computational power and area. Although dedicated analogue optical components are being designed for specific tasks, a design methodology for general-purpose analogue photonic components does not exist. On the positive side, the trade-off between bandwidth and power is generally better for photonic than for electronic devices, raising the hope that integrated photonic computing devices can push computation speeds even higher for digital information processing.

Around the beginning of the twenty-first-century, the principles of what is now known as *reservoir computing* showed up in several scientific publications. Echo state networks (ESN) (Jaeger 2001) and Liquid State Machines (LSM) (Maass et al. 2002) are the most commonly cited ones. What they have in common, is the fact that useful computation can be done by random dynamical systems, provided that the dynamics can be globally tuned to a useful regime: stable, but close to the edge of chaos.¹ Although the systems originally studied were simulated artificial neural networks, operating in discrete time, this fascinating idea soon stimulated other researchers to explore whether physical dynamical systems could also be used as reservoirs in this setup, giving birth to the field of *physical reservoir computing*. Photonic systems have been among the first physical reservoirs that were studied.

Two large families of photonic reservoirs exist. Time multiplexed reservoirs (addressed in other chapters of this book) use a single or very few physical non-linear processing units. They are time-multiplexed, using delayed feedback to provide dynamics and memory (Brunner et al. 2013; Larger et al. 2012, 2017; Paquot et al. 2012). They have been shown to be capable of solving relatively complex tasks, even image classification (Hermans et al. 2015), at very high bandwidths. They are in principle scalable towards complex tasks, but this requires longer delay lines, making their implementations rather bulky. In contrast, the parallel integrated photonic reservoirs which are the topic of this chapter are much closer to the echo state networks they are based on. They are integrated, and therefore small. They are also optimised for low power and very high bandwidths. The performance of integrated photonic reservoirs has been studied numerically for networks of ring resonators (Andre et al. 2014; Mesaritakis et al. 2013, 2015; Vandoorne et al. 2011; Zhang et al. 2014), networks of SOAs (Vandoorne 2011), and experimentally with networks of delay lines and splitters (Vandoorne et al. 2014). Integrated photonic reservoirs are particularly compelling when implemented in a CMOS-compatible platform as they can take advantage of its associated benefits for technology reuse and mass production.

This chapter describes our vision, the research route we followed and the lessons we learned. We start by giving an overview of the technological hurdles that have been (and are still being) taken to obtain low-power tunable computing modules. We then highlight recent first steps towards a scalable design methodology with such modules. We will not elaborate on state-of-the-art technologies for designing and

¹ Although *edge of chaos* is the common term in dynamical systems theory, the term *edge of stability* has always felt like a more appropriate term in this context.

fabricating the components we use, since that can be found in publications. Instead we focus on the interplay between system optimisation, which starts from known good practices in unconstrained (software) reservoir computing, and the technical possibilities and constraints of the chosen implementation substrate. In this way, we hope that our path and the lessons we learned can also inspire researchers who are exploring other physical substrates and systems.

2 From Ideas to First Prototypes

Early research on integrated photonic reservoir computing aimed to mimic Echo State networks with \tanh nonlinearities as closely as possible, while exploiting the potential benefits of integrated photonics: high bandwidth and low power. However, as software models, ESNs are entirely unconstrained with respect to properties like connectivity, connection weight values, neuron non-linearity or time scales. In contrast, in any physical implementation medium, limitations arise from physical and device constraints, or from fabrication issues. In addition, the often misleading near-infinite precision of software simulations is replaced by finite precision due to limited measurement accuracy and various noise sources. The very first conceptual description of a possible integrated photonic reservoir (Vandoorne and Bienstman 2007) proposed to use a regular 2D grid of Semiconductor Optical Amplifiers (SOAs) as reservoir nodes and had a “waterfall” interconnection topology (Fig. 1). Between that very first concept and the time of writing this chapter lies a decade of interdisciplinary research, combining systematic architecture and device optimisation, machine learning methods, targeted technological innovations and prototyping efforts. The result is systems that are now ready to bridge the final stretch between academic research and industrial take-up in the telecommunications sector.

2.1 Coherent Light and Planar Topologies

When using light as an information carrier, there are several options with respect to the way information is encoded into the signal. A first choice to make is to use either incoherent light or coherent light. In the first case, the information is carried only by the light intensity, represented by a positive real number. In the second case, the signal has both a magnitude (still positive) and a phase, and should therefore be modelled as a complex number. Like in our brains, the presence of *inhibition*, i.e., connections with negative weight between neurons, is crucial for computation in artificial neural networks. Early studies (unpublished) have confirmed that ESNs in which all signal values and all weights are positive perform very poorly. However, coherent light has both a phase and a magnitude. Equally, connections introduce both a phase shift and a change of magnitude (usually a loss). This implies that all operations in coherent optical reservoirs need to be modelled as complex-valued operations.

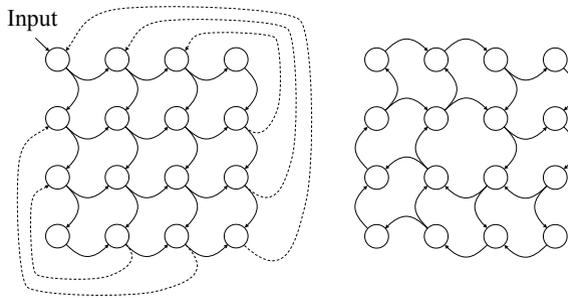


Fig. 1 Planar architectures for integrated photonic reservoirs: (left) the Waterfall topology from Vandoorne and Bienstman (2007) with the input signal inserted in the top-left node and optional feedback connection s(dashed lines) and (right) the Swirl topology that was first introduced in Vandoorne et al. (2008) and is still the basis for most recent work. No input injection is indicated here, since multiple variants have been investigated

As a consequence, even with only positive weights and positive light intensities, the addition of coherent light signals can be subtractive as well as additive w.r.t. to the magnitude of the signals. This is the main reason why coherent platforms were adopted early on for integrated photonic reservoirs.

Because of their integration on a 2D chip, integrated photonic reservoirs are also two-dimensional (planar) structures. Using photonic integrated waveguides for connecting neurons means that sharp bends and interconnect crossings cause prohibitive losses, excluding the stacking of layers of interconnect, as in electronic chips. For this reason, a purely planar connection graph is preferable. The most common approach is one in which each neuron has at most four neighbours. Initially, the input signal was injected into a single node only. After the initial concept that was based on the waterfall architecture (Vandoorne and Bienstman 2007) (Fig. 1, left), subsequent studies as well as first prototypes were based on the Swirl architecture (Vandoorne et al. 2008) (Fig. 1, right), which was designed to introduce optimal mixing of the internal state signals within the constraints of a planar structure. The connections are made by waveguides, which are passive components and therefore introduce no gain (only loss). They can be modelled as complex-valued weights with a magnitude smaller than 1 and a phase shift relative to their length. Each waveguide also has a propagation delay that is again relative to its length.

2.2 Readout and Training

In these reservoirs, a part of the light is split off at the output of each node and converted back into an electronic signal by a photodetector. This conversion of light intensity into an electrical signal non-linearly transforms each reservoir state (essentially quadratically) before linearly combining the states in the readout, trained with ridge regression.

In practice, the fabrication tolerances on the lengths of integrated waveguide connections are such that their loss and propagation delay are very close to the designed values, but the deviations of the phase shifts are so large that it is safest to consider each individual phase shift as random and different for each physical reservoir instance. In fact, due to the otherwise very regular topologies of the integrated photonic reservoirs, the variability of the phase shifts is their main source of richness. All simulation studies and design choices described in this chapter are based on performing simulations for a number of reservoirs with the same design parameters and random phases and reporting the average results.

However, there is also a downside to this uncontrollability: because of the differences between individual reservoirs, weights trained in simulation perform very poorly on the physical devices. Therefore, once they are fabricated, each individual reservoir will always need a calibration stage for training its readout. It is also not possible (or at least not within a reasonable time) to construct an accurate simulation model for individual physical reservoirs, since this would require a procedure to extract the actual values of all interconnection phases. This means that typical neural network training or optimisation approaches based on the application of gradient descent (backpropagation) cannot be used to optimise internal parameters of integrated photonic reservoirs.

2.3 Delays, Non-linearity and Power

The first integrated photonic reservoir architectures that were thoroughly studied and systematically optimised (simulation only) used a 4×4 array of Semiconductor Optical Amplifiers (SOAs) as reservoir nodes with a swirl connection structure. The input signal was initially inserted into a single node only (Vandoorne et al. 2008). The hyperparameters to be optimised for these reservoirs were the scaling and bias for the input signal, the bias current (determining the operating point of the SOAs) and the delay of the inter-node waveguides. The inter-node losses were fixed by the technology and consisted of the splitter losses, combiner losses and the waveguide losses, which were coupled to the waveguide length.

A first important conclusion from this initial work is the fact that tuning the inter-node connection delay is crucial for good performance. The importance of reservoir time scales was already pointed out in Verstraeten (2009) for simulated reservoirs (ESNs). The point of using a reservoir for computation is the fact that it naturally has memory, i.e., current state signals still depend to some extent on past inputs. This memory is weaker for inputs that lie further in the past. In order to solve a task on a given input signal (or signals), the reservoir has to be able to remember any past information from the input that is useful for the task and its bandwidth has to be high enough to respond to changes in the input signal. A physical reservoir's memory is determined by how fast signals (echo's) fade away (the losses) or are non-linearly mixed with other signals. Its bandwidth is affected by the bandwidths of all components, i.e., in photonic reservoirs: the signal generation and input modulation,

the SOAs, and the detector. Non-linear mixing occurs in the nodes (SOAs). The inter-node interconnections in integrated photonic reservoirs act as delay lines, i.e., sources of near-perfect memory. Although their loss is length dependent, it is dominated by splitter and combiner losses. For this reason, changing the inter-node delays directly impacts the memory without strongly affecting the non-linear mixing in the nodes. This means that for each task, inter-node interconnection lengths need to be tuned to match the relevant time scales in the input signal and the task.

A second conclusion was to drop the SOAs in future designs. It turns out that using SOAs nodes would increase the overall power budget above acceptable limits. They were originally introduced into the network to provide non-linearity, and because their static input-output characteristics more or less resemble the upper half of the hyperbolic tangent function that is commonly used in ESNs. It is generally understood that ESNs are in some sense universal approximators for fading memory functions, provided they can be made large enough (although some discussion exists about the precise conditions). Moving further away from ESNs by removing the non-linearity from the nodes and shifting it to the readout reduces the set of functions a reservoir can approximate. However, in that respect, the SOAs turned out to offer only limited advantages: at the optimal parameter settings for several benchmark tasks, they operated in an almost linear regime (Vandoorne 2011). In practice and for tasks that are not strongly non-linear, the non-linearity from the photodetectors alone is sufficient to achieve state-of-the-art performance (Vandoorne et al. 2014).

The two lessons above gave rise to the first actual implementation of an integrated photonic reservoir. Since the SOAs were omitted, it was a 4×4 purely passive Swirl network consisting of waveguides, splitters and combiners. For telecom tasks on bitstream signals, the lengths of the delay lines were optimised relative to the bit period. This made the design easily portable to operate at different bitrates, simply by changing the delay line lengths. Strangely enough, this also decreases the footprint of the reservoir for larger bitrates. For commonly used analogue benchmarks with much slower time scales like spoken digit recognition, the inputs were artificially speed up in order to achieve the optimal relation between signal and reservoir time scales. This first physical implementation contained only the reservoir itself. The readout was implemented in software after driving the reservoir with a long input sequence multiple times, each time measuring one reservoir state with a photodetector and storing it for further processing. The experimental results in Vandoorne et al. (2014) demonstrate that a 4×4 passive integrated photonic reservoir can yield error free performance on the header recognition task for headers up to 3-bit in length. Simulations for larger reservoirs indicated that it should be possible to go up to 8-bit headers (see Fig. 2). We additionally demonstrated that the passive integrated photonic reservoir can be used for bit-level manipulations on digital optical bit streams that could be useful for various tasks in telecommunication. Vandoorne et al. (2014) contains more information about the chip design and fabrication procedure.

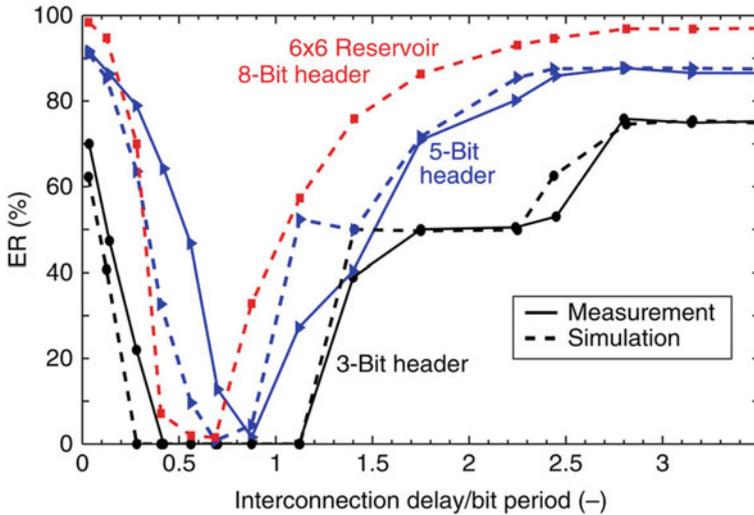


Fig. 2 Simulated and measured performance of a 4×4 swirl passive integrated photonics reservoir on the 3- and 5-bit header recognition task and simulated performance of a 6×6 reservoir on the 8-bit header recognition task (Vandoorne et al. 2014). The plots show the dependency of the error rate on the inter-node delays in the reservoir, relative to the bit period

2.4 Next Generation Reservoir Architectures

From the perspective of power consumption, the step to passive integrated reservoirs was crucial, since it removes all power consumption inside the reservoir. In addition, it allows the reservoir to be implemented in the CMOS-compatible silicon photonics technology, which drastically lowers the technological barrier for industrial take-up. However, the fact that there is no more gain inside the reservoir affects the reservoir performance: although signals still travel in the reservoir, their magnitudes now decrease very rapidly due to losses. When only a single reservoir node is driven by the input signal, the power in many nodes ends up below the detection limit (noise margin) of the photodetectors. Under these circumstances, even mildly scaling up the reservoir to address more difficult tasks makes no sense since this will not increase the number of nodes with detectable power levels. For this reason, follow-up simulation studies over the last few years have focused on designing more power-efficient reservoirs by reducing the losses or by shifting to totally different architectures. In what follows, we summarise recent progress on three lines of research.

Optimisation of the input distribution

Increasing the input power, even if this were desirable, does not help much because the losses accumulate in a multiplicative manner as light travels further from the input node. It turns out to be much more effective to split the available input power budget

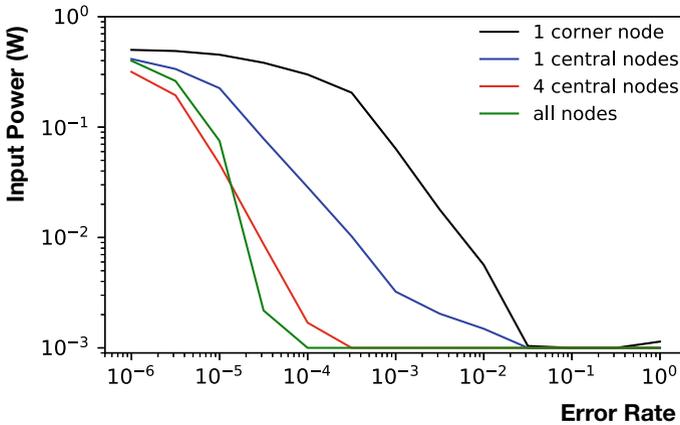


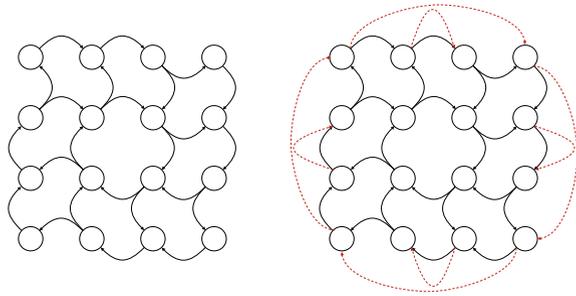
Fig. 3 Error rate for the one-bit-delayed XOR task versus total input power for different injection scenarios in a 4×4 passive Swirl reservoir. The minimum measurable error, given the number of bits used for testing, is 10^{-3} (Katumba et al. 2017). Note that the power budget only includes the power that is actually sent into the reservoir. It does not include the additional power that is consumed by the driving and receiving circuits as these have not been modelled in our simulations

across multiple input nodes (Katumba et al. 2017). Spreading the input power evenly across all nodes yields the best accuracy for a given power budget (Fig. 3 illustrates this).

However, the design and fabrication also become more difficult as the number of nodes that are driven by the input increases. For a 4×4 Swirl architecture, driving only the four central nodes is an excellent compromise between reservoir performance and technological complexity (Fig. 3). For the same accuracy, a reservoir with this input scheme needs about 2 orders of magnitude less power than when only a single node is driven to achieve almost the same performance as a reservoir for which the input power is distributed evenly across all nodes. Clearly, the outcome of this optimisation is closely linked to the size and connectivity of the reservoir. For the 16-node reservoir, the four central nodes form a single connected loop, from which power flows into the undriven boundary nodes. For larger architectures, the ideal trade-off may lie elsewhere. Ideally, for each future architecture, the input connectivity and reservoir connectivity must be jointly optimised.

Thus far, most benchmarks addressed with integrated photonic reservoirs have been relatively simple. However, scaling up to larger reservoirs in order to solve more difficult tasks only makes sense if the additional nodes actually contribute. Spreading the available input power across multiple nodes as in Katumba et al. (2017) makes this possible. However, it also makes the design more complex because the input signal needs to be coupled into more nodes. Also, driving nodes off the boundary of the planar network necessarily leads to crossings or the need for coupling the light into the nodes vertically (from the top). For this reason, it is desirable to keep the number of driven nodes as small as possible. If this can be complemented with

Fig. 4 Extension of swirl architecture (left) to more power-efficient four-port architecture (right): connections that were added are shown in red



technological solutions to reduce the losses, the fraction of nodes that need to be driven to achieve good performance can be further reduced. Two such approaches for loss reduction have been investigated.

Multi-mode reservoirs

The original passive reservoirs were designed to guide only a single mode. Note that the term *mode* in this context refers to the different transverse modes of the waveguide. Since part of the losses are due to light that leaks away to other modes, it could be beneficial to design a reservoir in which some of these other modes are also guided, such that the power that leaks into them is not lost. In this way, using multi-mode reservoirs can help to realise a better power balance.

The use of multi-mode rather than single-mode reservoirs was studied in simulation for passive reservoirs in which only a single node is driven (Katumba et al. 2018a). The successful application of this approach strongly hinges on the design of a novel multi-mode Y-junction with carefully tailored adiabaticity that lowers the losses at combination points in the photonic network constituting the reservoir. It turned out that, for a 36-node (6×6) reservoir, we can gain up to 30% in per node power, especially for nodes that are furthest from the input point. Although the power in the boundary nodes is still lower, this extra power boost could be the difference between being below or above the noise floor at a node. In future work, additional benefits of multi-mode reservoirs will be investigated. For example, separate modes can carry different degrees of freedom, which could be beneficial for mixing the input signals.

New architectures

In the swirl architecture (repeated in Fig. 4, left), each node has at most two inputs and two outputs. However, this maximum is not reached for all nodes. The four corner nodes only have a single input and output, which means that no signal mixing occurs in them. All other boundary nodes have three ports (either only one input or only one output). This means they need asymmetrical combiner or splitter devices which suffer from modal radiation losses. These losses are inherent for non-symmetrical reciprocal splitting devices, as on average there is a 50% modal mismatch between the two input channels. By changing the architecture in such a way that each node has exactly two inputs and two outputs, these losses can be avoided. From an architectural perspective,

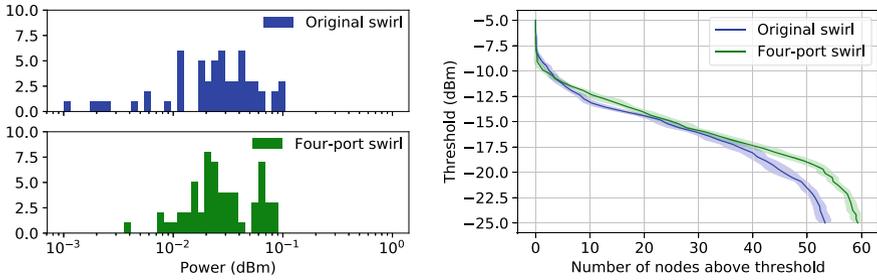


Fig. 5 Comparison between node power levels of 6×10 original swirl architecture and four-port architectures when driven with bitstream signals: (left) histograms of node powers for both architectures; (right) required detection thresholds for being able to detect a given number of nodes

the best way to do this would be to extend the architecture with so-called *wrap-around* connections at the edges. However, a wrap-around architecture is inherently not planar, which means that crossings would be needed in a planar implementation. A planar alternative was proposed in which wrap-around connections are added only for the corner nodes, and all other added boundary connections are non-crossing (Fig. 4, right). Note that the four corner nodes now effectively form a second four-node loop, similar to the inner four nodes. In Sackesyn et al. (2018), this new architecture is compared with the swirl architecture in simulations in terms of loss and in terms of performance on the equalisation of a non-linearly distorted BPSK signal. These simulations (Fig. 5) show that the four-port architecture indeed suffers less losses and thus has a better energy-efficiency as all input power at a node is redirected to one of the two output channels instead of radiating away.

3 Training Reservoirs with Integrated Optical Readouts

3.1 Motivation

To train the readout of a physical reservoir computing system, all states that used in that readout must be observable. In many cases, the reservoir is driven with the training input signal(s), the resulting reservoir states are recorded and one-shot learning through ridge regression is used to obtain the optimal readout weights. However, online learning (using recursive least squares or FORCE optimization Sussillo and Abbott 2009) is also possible. For the integrated photonic reservoirs discussed thus far, it was assumed that a photodetector would be needed for each reservoir state, followed by an AD convertor because training is usually done in the digital domain. Unfortunately, this solution would be challenging when scaling up to reservoirs with more nodes and tuned to operate at higher bandwidths, as high-speed photodetectors tend to be costly in terms of chip footprint.

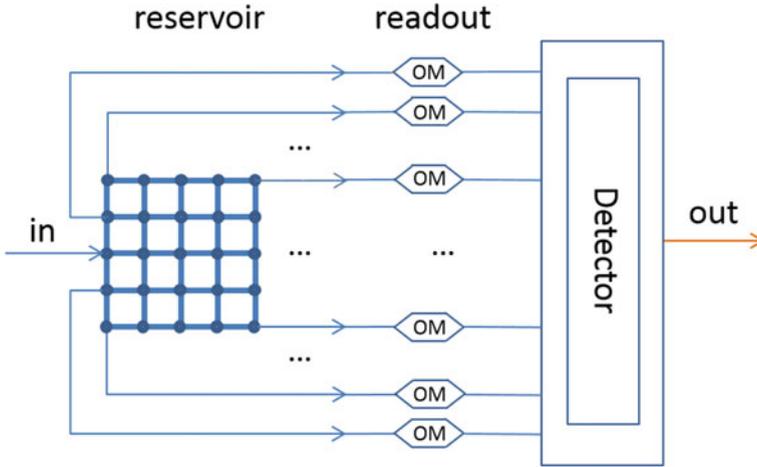


Fig. 6 Schematic of a fully optical readout. Each optical output signal is modulated by an Optical Modulator (OM) implementing the weights. The optical outputs are then sent to a photodiode where all signals are summed and then converted to a final electrical output signal

This chip area, as well as the power consumption of the photodetectors, could be avoided by implementing the readout optically. Since the aim is to reduce costs (chip area) and power consumption, the optical implementation of the readout should equally not involve power-hungry weighting components for each state signal. A straightforward tuneable optical weighting element can take the form of a reverse-biased pn-junction. An even better solution would be to use non-volatile optical weighting elements, such as the ones that are currently being developed by several groups (Abel et al. 2013; Ríos et al. 2015; Van Bilzen et al. 2015). Figure 6 illustrates the concept of a fully optical integrated readout.

From a system designer’s perspective, this means we end up with a readout with quite different properties from the original one (Fig. 6). In what follows, we summarise each of these differences and their implications and discuss how they have been taken into account in the designs for the next generation integrated reservoir prototypes.

3.2 Limited Observability

As mentioned before, the fabrication tolerances of integrated waveguides are such that the propagation phase of two nominally identical waveguides could be completely different. By averaging simulations across a number of randomly selected instances, the global architectural parameters can be tuned. However it is not possible to build a simulation model that exactly matches the behaviour of individual physical reservoirs, which means that training the readout weights cannot be done

in simulation. In fact, simulation studies have shown that, for realistic settings of the phase variability, weights trained in simulation are almost as bad as randomly initialised weights (Freiberger et al. 2017).

As long as the fabrication tolerances do not improve, this implies that device-specific training on the actual hardware will always be necessary. Thus far, no technical solutions have been found that indicate that the actual optimisation of the weight values could also be done optically, on the physical hardware, so opto-electronic conversion and digital training still remain necessary. For this reason, we maintain a single photodetector after the linear readout. However, in order to record all state signals for training, it no longer suffices to send the input sequences through the reservoir once. Remember that the photodetector's output is essentially the squared magnitude of its optical input. Obtaining the state signals implies that we need to know their magnitude, but also their phase, relative to a reference phase. It turns out that this is in fact possible with a single photodetector by driving the reservoir with the same input signals multiple times (just under three times per state signal), each time with different settings of the weights (Freiberger et al. 2018). Under ideal conditions, this non-linearity inversion procedure is exact. Each signal's magnitude is obtained by setting all weights to zero except for the observed signal. To obtain the relative phases, one signal's phase is chosen as a reference and two measurements are necessary to obtain the relative phase for each of the other signals. Once this procedure has been executed, training can be done offline. Once the weights have been trained, the reservoir can operate without the need for digital processing, although we expect that re-training or tuning may be necessary at regular intervals to compensate for any causes of drift or non-stationarity in the system.

3.3 *Non-linearities and Complex-Valued Regression*

With an integrated optical readout, the non-linearity of the single photodetector (now the only non-linearity in the system) is applied *after* the linear combination and the accuracy of the reservoir is evaluated after the photodetector. However, ridge regression is a linear technique. In order to use it in this new setting, the desired signal values after the non-linearity have to be converted back into desired complex-valued signals before the non-linearity. After that, the optical weighted sum at the readout is now a linear combination in the complex domain, which can be optimised using complex-valued ridge regression (Hoerl and Kennard 1970).

A straightforward way to obtain complex-valued target signals is to again invert the non-linearity by taking the square root of the original target signal as magnitude and setting the desired phase to an arbitrary value (e.g., zero). Unfortunately, this procedure yields rather disappointing results for tasks that really require non-linearity. A simple example is the highly non-linear XOR task on bitstream input signals, where the desired output is again a bitstream signal, representing the Boolean XOR of the current bit and the previous bit. This task could easily be solved by the reservoirs

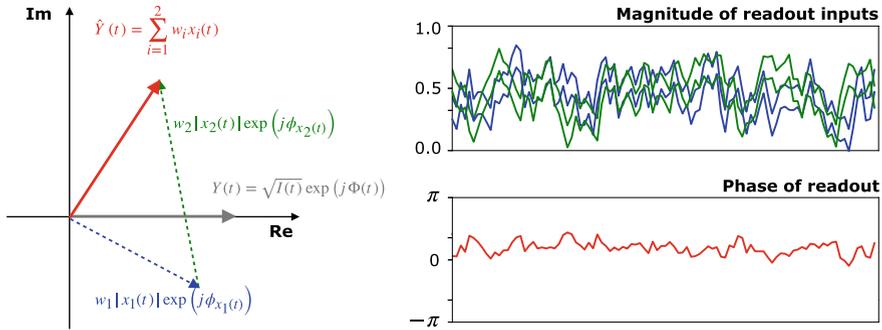


Fig. 7 Mathematics of a complex readout for two states: although each weighted input’s phase is constant in time (chosen at -30° and $+100^\circ$ in the figure), the phase of the linear combination varies in time

without optical readout, i.e., with a (non-linear) photodetector for each state signal, but turns out to be much harder with a single photodetector after the readout.

One part of the reason for this is the fact that, purely from a function approximation point of view, taking a weighted sum of non-linearly transformed signals offers more non-linear richness than taking the same non-linear transformation of a weighted sum. However, it also turns out that setting targets for the complex-valued signals *before* the photodetector is far from optimal. This can be explained as follows. First, fixing the magnitude of the signal also fixes the operating point of the output non-linearity. Since the output signal can easily be rescaled and shifted in the electrical domain, this is not necessary. Second, and more importantly, fixing the phase to a single value is overly restrictive. Let’s consider the simplest case of combining two states. The weights are constant in time and so is the phase for each of the state signals. The readout is therefore the sum of two complex-valued signals with magnitudes that vary in time and constant phases. Due to the properties of complex addition, the phase of this readout also varies in time, as illustrated in Fig. 7. This simple example shows why trying to enforce a constant phase to the outputs of the readout constrains the solution space unnecessarily. The phase in the complex domain is of no relevance to the signal after its transformation to the electrical domain, so it should be left unconstrained.

Since ridge regression is an optimisation technique for linear models, we cannot use it without setting targets in the complex domain. Instead, we use gradient descent to minimise the error between the electrical output signal and the complex-valued weights directly. We previously explained why using gradient descent is not possible for optimising the entire reservoir due to the uncertainty about inter-node connection phases. However, the phases in the paths between the readout weights and the photodetector are included in the signals obtained from non-linearity inversion procedure and a differentiable model for the photodetector is quite straightforward.

The replacement of ridge regression by gradient descent avoids enforcing any values for the complex-valued signal between the linear combination and the pho-

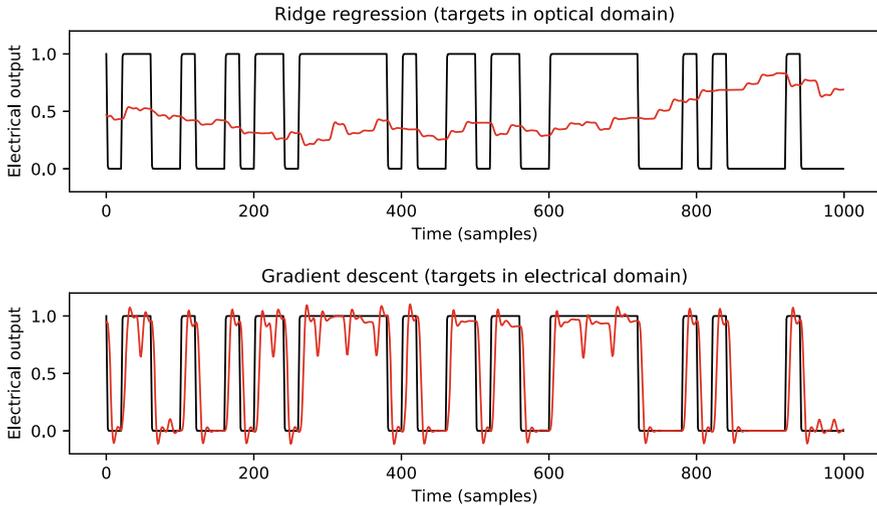


Fig. 8 Comparison of ridge regression with complex-valued targets and gradient descent for the XOR task on bitstream input (reservoir using four-port architecture of 4×4 nodes and input applied to all nodes)

todetector. The trained parameters are the magnitudes and phases of all weights (including the constant weight) as well as the bias and rescaling after the photodetector:

$$\hat{y}(t) = w_{\text{bias}} + w_{\text{scale}} \left(m_0 * e^{j\phi_0} + \sum_{i=1}^n m_i * e^{j\phi_i} * x_i \right)^2, \quad (1)$$

where n is the number of state signals used in the readout, w_{bias} and w_{scale} operate in the electrical domain (i.e., after the photodetector) and $m_0 \dots m_n$ and $\phi_0 \dots \phi_n$ are the tuneable magnitudes and phases of the optical weights. Figure 8 illustrates the striking difference between trained outputs obtained using ridge regression with complex-valued targets and using gradient descent directly on the real-valued targets at the output of the photodetector. This example indicates that, despite initial expectations, having a single non-linearity at the readout does not dramatically affect our reservoir performance. This is mostly due to the fact that we are working in the complex domain (with coherent light).

3.4 Limited Precision

Readout weights can be implemented using different approaches. From the perspective of minimising power, there is a large difference between volatile and non-volatile technologies. Volatile approaches, like reverse-biased PN-junctions, can typically be

tuned with fine resolution, but they need power to maintain their set value. In contrast, non-volatile weight elements only consume power when their value is being set, but the physics of the current candidate technologies do not allow for high resolution. A typical example with such a limitation is a weighing element based on barium titanate (BaTiO₃) (Abel et al. 2013), an integration of a transition metal oxide material. These elements are typically able to bring only 20 discrete weight levels. Clearly, this limited precision of the tuneable weights needs to be investigated since it can be expected to affect reservoir performance.

The impact of weight precision has been analysed in simulation. For a first study, we assumed the same number of levels for both the magnitudes and the phases of the weights. The readout was trained in the same way as with infinite precision and the weights were quantized only after training, using discretisation at 8, 16, 32 and 64 discrete levels (also referred to as 3-, 4-, 5- and 6-bit quantization, respectively). Figure 9 illustrates the results for a header recognition task, each time sweeping across different values for the inter-node delay (arbitrary units). We can conclude that, without taking any additional measures, the performance degenerates a lot for 3-bit quantisation. For 4 or more bits, the error rate at the optimal inter-node delay is not much higher than for the case with infinite precision but there remains some variation between the results for individual reservoirs. At 6 bits resolution, this variability disappears almost completely. These results are promising, given the fact that no measures have yet been taken to take weight quantisation into account during train-

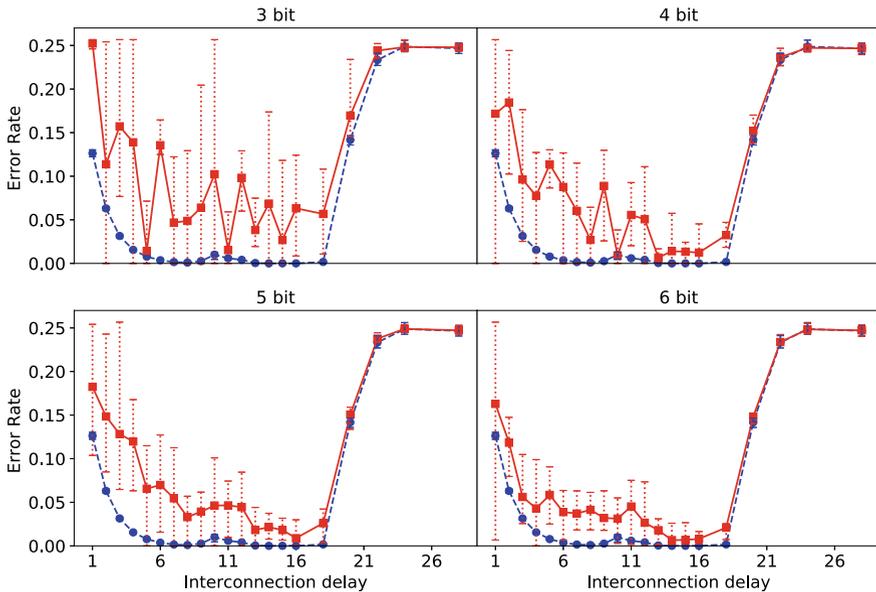


Fig. 9 Error rate when using quantised optical weights (red curves) in a header recognition task, as a function of interconnection delay. The weight resolution (in bits) is indicated at the top of each plot. For comparison, the blue curves show the case of infinite weight resolution

ing. Reduced weight precision in real-valued networks is currently being actively addressed in deep learning research. In view of introducing autonomous deep learning in embedded systems (without needing cloud connectivity) it is crucial to reduce the computational cost of inference in deep neural networks. One way to approach this is to use (extremely) reduced precision (Gupta et al. 2015; Hao et al. 2017). Results have even been proposed that go as far as using binary weights (Bengio and David 2015; Rastegari et al. 2016). By developing approaches for quantised training that are adapted to complex-valued readouts, we expect to be able to relax the precision requirements on integrated optical weights to 16 or even 8 quantisation levels (3 or 4 bits).

4 From Modules to Systems

4.1 *The Future of Single Reservoir Modules*

Based on simulation results, the architectural improvements mentioned in the previous sections should be sufficient to allow good reservoir performance for small to medium-sized reservoir modules. Second-generation prototypes to evaluate these claims have been designed and are being characterised at the time of writing this chapter. At the same time, more challenging applications in the telecom area are being addressed. Although a lot of optical processing is already being done in that domain, it is mostly linear. For this reason, we are focusing on non-linear tasks on optical telecom signals.

One problem that is being addressed is the restoration of optical signals that have degraded due to non-linear distortions during generation, transmission and reception phases (Djordjevic et al. 2010). Causes for this can be traced back to effects like dispersion, amplified spontaneous emission at amplification points, attenuation and reflections in fibre links, optical nonlinearities in fibres or timing jitter introduced at O/E and E/O points. Today, such restoration is typically done in the digital electronic domain using advanced DSP post-processing, but such an approach consumes a lot of power and chip real estate. Photonic reservoir computing could provide an alternative here, to undo (part of) these signal impairments already in the optical domain. Katumba et al. (2019) shows non-linear compensation in unrepeated metro links of up to 200km that outperform electrical FFE-based equalisers, and ultimately any linear compensation device. For a high-speed short-reach 40Gb/s link based on a Distributed Feedback Laser (DFB) and an Electroabsorptive Modulator (EAM), and considering an HD-FEC limit of 0.2102, the reach can be increased by almost 10km using a reservoir with only 16 nodes. These results show for the first time that integrated photonic reservoir modules can be competitive to the-state-of-the-art solutions in the telecommunications domain.

The next step is to address more complex tasks. Most often, tasks on optically encoded digital signals are performed on data encoded with a binary modulation format, but this can be extended to the much more complex case of computing directly on a PAM-4 signal, a two-bit per symbol amplitude modulation format. This is equivalent to performing Boolean computations in a 4-valued space. In Katumba et al. (2018b), initial results are given for operations on PAM-4 signals modulated at 10 GHz, which translates to a bitrate of 20 GHz. These indicate that small reservoirs like the ones discussed above cannot solve this. Even an 8×8 swirl reservoir has a symbol error rate of approximately 30%. However, the complexity of this task allows us to test how this can improve for increasing reservoir size. In this case, there was a close to linear improvement and for the largest simulated reservoir in that study (20×20 nodes), the symbol error rate had decreased to 5%. Although this is only a first study, it suggests that there is still quite a bit of room for improvement in addressing more complex tasks simply by scaling up the reservoir.

In parallel to the application-directed research, which aims to pave the road to industrial take-up, more fundamentally different reservoir architectures are also being explored. The reservoirs discussed thus far were still based on the earliest ones, which tried to follow the conventional node structure of neural networks quite closely. However, the field of physical reservoir computing has evolved a lot since the early days. In particular, as is clear from the variety of contributions in this book, the dynamical systems used as reservoirs do not have to follow the straitjacket of interconnected nodes. In particular, the inherently parallel nature of photonics allows for architectures in which the light propagates and mixes in free space. A possible design for an integrated free-space photonic reservoir consists of a photonic crystal cavity with a quarter stadium shape (Laporte et al. 2018), depicted in Fig. 10. In this design, the quarter stadium shape makes sure that an input signal gets mixed in a complicated manner (Liu et al. 2015; Sieber et al. 1993; Stöckmann and Stein 1990). The mixed light leaks out of the cavity along the connected waveguides which provide the state signals to the readout. As can be seen in the figure, the mixing of light within the cavity is very rich. In fact, it is richer than in the waveguide-based reservoirs, while at the same time using considerably less chip area. The time scales of the reservoir depend on the time-of-flight inside the cavity (the cavity dimensions) and how well the light is confined inside the cavity. The cavity in Fig. 10 was optimised for optical bitstream signals at a bitrate of around 50 Gbps. Its dimensions are $30 \mu\text{m} \times 60 \mu\text{m}$. In theory, cavities as small as $7 \mu\text{m} \times 7 \mu\text{m}$ are possible, which would allow bitrates up to 1 Tbps. Based on simulations, this photonic crystal design promises very low loss and excellent performance several benchmark telecom tasks, such as the highly non-linear XOR task and header recognition tasks, while still accepting bitrates in a wide region of operation (Laporte et al. 2018).

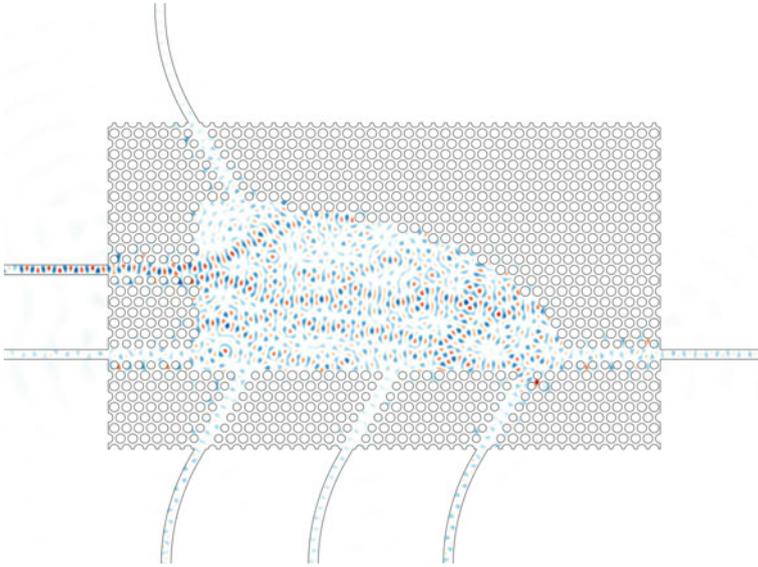


Fig. 10 Optical field profiles in a photonic crystal cavity. In the example, input is applied through a single channel. Thanks to the specific profile of the cavity's boundary, the mixing of the optical fields inside the cavity is very rich. The mixed light leaks out of the cavity along all the waveguides which are routed to a readout

4.2 Making Ensembles of Reservoir Modules

Although initial results indicate that there is still room for improvement with respect to task complexity by scaling up reservoirs, it is well known in the reservoir computing community that there are bounds to what you can achieve simply by making a reservoir larger and using more observed states in the readout. One of the main reasons for this is that, due to the interactions inside the reservoir, the states are usually highly correlated and it becomes more difficult to exploit the residual information that is added by additional states as the state space grows larger. For simulated reservoirs, such as infinite precision echo state networks, this would require an increase in the amount of training data. Physical reservoirs do not have infinite precision due to noise and measurement inaccuracies, so their performance tends to saturate even faster as a function of reservoir size. This limits the applicability of single reservoir modules.

In order to go beyond what can be achieved with a single reservoir module, more extensive training is required by building networks of interconnected reservoirs, in which training occurs only at the inter-module interfaces. Unfortunately, we are again (heavily) constrained by the fact that, due to process variability, the detailed behaviour of individual reservoirs cannot be modelled in simulation. This effectively rules

out backpropagation, the workhorse of deep learning, as a top-down optimisation approach. Instead, targets for each individual reservoir module must be explicitly defined based on the global task targets.

In machine learning, ensembling is an overarching name to achieve exactly that: increasing task performance beyond what can be achieved with a single model. In order to work well, the models in an ensemble must be different, i.e., the mistakes they make must be as uncorrelated as possible. Variation between models of the same type can be achieved by training them on different subsets of the training data or the input features. They can also be trained to correct each other's mistakes (*boosting*, Friedman et al. 2001).

In an initial simulation study of ensembling for photonic reservoirs, we consider ensembles in which each model is of the same type: an integrated 4×8 photonic four-port reservoir with interconnection delays tuned for operation at 30 Gbps. Since the interconnection phases of each reservoir are randomly chosen, each reservoir is already unique and different from the others. Since this work was performed in parallel with the research on optical readouts described in Sect. 3, it focused mostly on reservoirs with electrical readouts, i.e., with the non-linear transformation of a photodetector for each signal that is used in the readout.

We selected a number of approaches of combining four such reservoirs and compared them to the baseline case of a single 16×32 reservoir with the same architecture. The approaches we considered in this study are illustrated in Fig. 11. In the first type (single-stage ensemble), the observed states of all modules in the ensemble are concatenated into a single readout. In the second type, gradient boosting, only the first module is trained on the original task, while each consecutive module is trained to correct the mistakes of its predecessor's outputs. As a third approach, we have also evaluated the paradigm of stacking, which has already been applied in the context of reservoir computing (Keuninckx 2016; Nichele and Molund 2017). In this approach, each reservoir is trained on the original task. Only the first module is driven with the original input signal, while each consecutive module is driven with its predecessor's trained output (converted back into the optical domain). Finally, we introduce a new combination technique inspired by these approaches, which we refer to as chaining. Like in gradient boosting, each module is driven with the original input and like in stacking each readout is trained on the original task. However, from the second module onward, each readout also receives the trained output from the previous reservoir as an input. This offers a different way for the readouts to correct the mistakes of the previous reservoirs.

The architectures described above were evaluated on two benchmark tasks. The first task is the XOR task on bitstream data, but now with 3 bits delay. This means that the desired output is the Boolean XOR of the current binary symbol and that which lies three-bit periods in the past. This task has the same non-linearity requirements as the XOR task used before, but it requires more memory. The second task is the 1 sample ahead prediction Santa Fe task (Weigend and Gershenfeld 1993). Table 1 summarises the error rates obtained for these two tasks at 30 Gbps. Error rates printed in bold face indicate the best performing approach per task, error rates in italic the second best. This table shows that the single-stage ensembling systematically

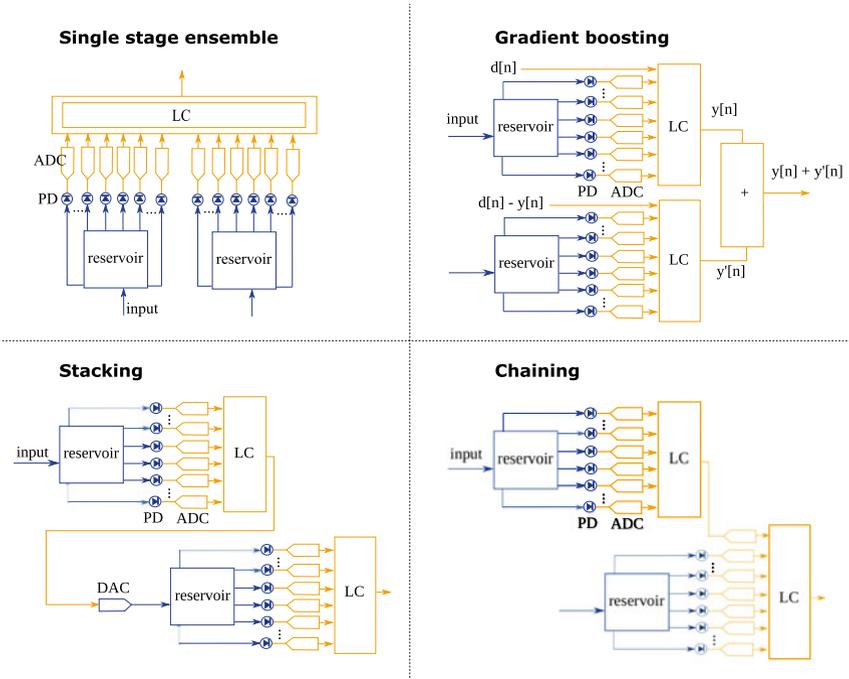


Fig. 11 Types of ensembles of reservoirs with electrical readout considered in the initial study. Components operating in the optical domain are shown in blue, components in the electrical domain in yellow

Table 1 Results for electrical training/coupling at 30 Gbps

Task	Baseline	Ensemble	Boosting	Stacking	Chaining
XOR 3-bit delay (BER)	0.006	0.001	0.041	0.222	0.038
Santa Fe (NMSE)	0.028	0.022	0.038	0.038	0.027

outperforms the baseline of a single large reservoir. This is good news, since we expect it will be technologically easier to combine multiple small reservoirs than to make a large one. The main reason for this is that optically routing power from all the nodes to the readout becomes more difficult when the dimensions of the architecture increase. Boosting and stacking are consistently worse than the baseline. This was also the case for experiments at other bandwidths (not shown). For our proposed chaining approach, conclusions are mixed. At 30Gbps, its performance is between the baseline and the ensemble for the Santa Fe task but worse than both for the XOR task. However, at other bandwidths, this gap is closed and chaining again matches the ensemble’s performance.

Clearly, this is only an initial study, in which many possibilities for optimisation are yet to be explored. In view of our recent progress on integrated readouts trained with gradient descent, the next study will focus directly on those architectures. It will also cover a more extensive range of combination approaches. For example, the recent approach followed in Gallicchio and Micheli (2017), Gallicchio et al. (2018) is similar to but not the same as the stacking performed above. In those works, each reservoir module is driven with all the states of its predecessor (with untrained weights). The readout is trained on the aggregated states of all reservoirs in the ensemble. This architecture performs better than the one studied here because, as the information flows from reservoir to reservoir, each subsequent reservoir has a memory that reaches further into the past. Translated to integrated photonics technology, it would have to be simplified, e.g., by projecting a random combination of each reservoir's states back into the next reservoir and training the readout on all states in the electrical domain.

5 Conclusion and Perspectives

In this chapter we outlined our research path on integrated photonic reservoir computing, from its first steps in 2007 until today, when second-generation prototypes have been designed are being characterised and at least one application is a promising candidate for industrial take-up.

The focus in this chapter was on the impact of technological limitations on system performance and on the architectural and operational solutions we developed to end up with ever better performing systems. In view of the present results, we are more confident now than a decade ago that this technology will eventually find its way into industrial applications.

The last part of this chapter reported on first steps towards designing multi-reservoir systems. It is clear that the ensembling approaches investigated in this study do not offer sufficient improvement by themselves. Their limitation lies in the fact that all targets for training individual modules are either equal to or directly derived from the original task targets. This is not sufficient for building really powerful multi-reservoir networks. Candidate starting points for exploring architectures as well as optimising them could be based on, e.g., reinforcement learning or large-scale genetic black box optimisation approaches. However, what is really needed is a new automatic and efficient divide-and-conquer design methodology for analogue computing. Only then can reservoir computing with integrated photonic systems, as well as other novel analog computing substrates, leverage general-purpose computation in the optical domain.

References

- S. Abel, T. Stferle, C. Marchiori, C. Rossel, M. Rossell, R. Erni, D. Caimi, M. Sousa, A. Chelnokov, B. Offrein, J. Fompeyrine, A strong electro-optically active lead-free ferroelectric integrated on silicon. *Nat. Commun.* **4**, 1671 (2013)
- D. Brunner, M.C. Soriano, C.R. Mirasso, I. Fischer, Parallel photonic information processing at gigabyte per second data rates using transient states. *Nat. Commun.* **4**, 1364 (2013)
- M. Courbariaux, Y. Bengio, J.-P. David, Binaryconnect: Training deep neural networks with binary weights during propagations, in *Advances in Neural Information Processing Systems* (2015), pp. 3123–3131
- I. Djordjevic, W. Ryan, B. Vasic, *Coding for Optical Channels* (Springer, US, 2010)
- M.A.A. Fiers, T. Van Vaerenbergh, F. Wyffels, D. Verstraeten, B. Schrauwen, J. Dambre, P. Bienstman, Nanophotonic reservoir computing with photonic crystal cavities to generate periodic patterns. *IEEE Trans. Neural Netw. Learn. Syst.* **25**(2), 344–355 (2014)
- M. Freiberger, A. Katumba, P. Bienstman, J. Dambre, On-chip passive photonic reservoir computing with integrated optical readout, in *2017 IEEE International Conference on Rebooting Computing (ICRC)* (2017)
- M. Freiberger, A. Katumba, P. Bienstman, J. Dambre, Training passive photonic reservoirs with integrated optical readout. *IEEE Trans. Neural Netw. Learn. Syst.* 1–11 (2018)
- J. Friedman, T. Hastie, R. Tibshirani, *The Elements of Statistical Learning*, Springer Series in Statistics, vol. 1. (Springer, New York, 2001)
- C. Gallicchio, A. Micheli, Deep echo state network (deepesn): a brief survey (2017), [arXiv:1712.04323](https://arxiv.org/abs/1712.04323)
- C. Gallicchio, A. Micheli, L. Pedrelli, Design of deep echo state networks. *Neural Netw.* **108**, 33–47 (2018)
- S. Gupta, A. Agrawal, K. Gopalakrishnan, P. Narayanan, Deep learning with limited numerical precision, in *Proceedings of the 32Nd International Conference on International Conference on Machine Learning, ICML'15*, vol. 37 (2015), pp. 1737–1746, JMLR.org
- M. Hermans, M.C. Soriano, J. Dambre, P. Bienstman, I. Fischer, Photonic delay systems as machine learning implementations. *J. Mach. Learn. Res.* **16**, 2081–2097 (2015)
- A. Hoerl, R. Kennard, Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12**(1), 55–67 (1970)
- H. Jaeger, The echo state approach to analysing and training recurrent neural networks—with an Erratum note 1. Technical report, GMD148, Bonn, Germany: German (2001), pp. 1–47
- A. Katumba, M. Freiberger, P. Bienstman, J. Dambre, A multiple-input strategy to efficient integrated photonic reservoir computing. *Cogn. Comput.* **4**, 1–8 (2017)
- A. Katumba, J. Heyvaert, B. Schneider, S. Uvin, J. Dambre, P. Bienstman, Low-loss photonic reservoir computing with multimode photonic integrated circuits. *Sci. Rep.* **8**(1) (2018a)
- A. Katumba, M. Freiberger, F. Laporte, A. Lugnan, S. Sackesyn, C. Ma, J. Dambre, P. Bienstman, Neuromorphic computing based on silicon photonics and reservoir computing. *IEEE J. Sel. Top. Quantum Electron.* **24**, 6 (2018b)
- A. Katumba, X. Yin, J. Dambre, P. Bienstman, A neuromorphic silicon photonics nonlinear equalizer for optical communications with intensity modulation and direct detection. *J. Light. Technol.* (2019)
- L. Keuninckx, Electronic systems as an experimental testbed to study nonlinear delay dynamics. PhD thesis, Vrije Universiteit Brussel (2016)
- F. Laporte, A. Katumba, J. Dambre, P. Bienstman, Numerical demonstration of neuromorphic computing with photonic crystal cavities. *Opt. Express* **26**(7), 7955–7964 (2018)
- L. Larger, M.C. Soriano, D. Brunner, L. Appeltant, J.M. Gutiérrez, L. Pesquera, C.R. Mirasso, I. Fischer, Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing. *Opt. Express* **20**(3), 3241–3249 (2012)

- L. Larger, A. Baylón-Fuentes, R. Martinenghi, V.S. Udaltsov, Y.K. Chembo, M. Jacquot, High-speed photonic reservoir computing using a time-delay-based architecture: million words per second classification. *Phys. Rev. X* **7**(1), 011015 (2017)
- H. Li, S. De, X. Zheng, C. Studer, H. Samet, T. Goldstein, Training quantized nets: A deeper understanding, in *NIPS* (2017)
- C. Liu, R.E.C. Van Der Wel, N. Rotenberg, L. Kuipers, T.F. Krauss, A. Di Falco, A. Fratalocchi, Triggering extreme events at the nanoscale in photonic seas. *Nat. Phys.* **11**(4), 358–363 (2015)
- W. Maass, T. Natschläger, H. Markram, Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **25**60, 2531–2560 (2002)
- C. Mesaritakis, V. Papataxiarhis, D. Syvridis, Micro ring resonators as building blocks for an all-optical high-speed reservoir-computing bit-pattern-recognition system, in *JOSA B*, October 2013 (2013)
- C. Mesaritakis, A. Kapsalis, D. Syvridis, All-optical reservoir computing system based on InGaAsP ring resonators for high-speed identification and optical routing in optical networks, vol. 9370, 2 (2015), p. 937033
- S. Nichele, A. Molund, Deep reservoir computing using cellular automata (2017), [arXiv:1703.02806](https://arxiv.org/abs/1703.02806)
- Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, S. Massar, Optoelectronic reservoir computing. *Sci. Rep.* **2**, 287, 2 (2012)
- M. Rastegari, V. Ordonez, J. Redmon, A. Farhadi, Xnor-net: Imagenet classification using binary convolutional neural networks, in *European Conference on Computer Vision* (Springer, 2016), pp. 525–542
- C. Ríos, M. Stegmaier, P. Hosseini, D. Wang, T. Scherer, C. Wright, H. Bhaskaran, W. Pernice, Integrated all-photonic non-volatile multi-level memory. *Nat. Photonics* **9**(11), 725–732 (2015)
- S. Sackesyn, C. Ma, J. Dambre, P. Bienstman, An enhanced architecture for silicon photonic reservoir computing, in *Cognitive Computing 2018 - Merging Concepts with Hardware* (2018), pp. 1–2
- M. Sieber, U. Smilansky, S.C. Creagh, R.G. Littlejohn, Non-generic spectral statistics in the quantized stadium billiard. *J. Phys. A: Math. Gen.* **26**(22), 6217 (1993)
- H.-J. Stöckmann, J. Stein, Quantum chaos in billiards studied by microwave absorption. *Phys. Rev. Lett.* **64**, 2215–2218 (1990)
- D. Sussillo, L.F. Abbott, Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63**(4), 544–557, 8 (2009)
- B. Van Bilzen, P. Homm, L. Dillemans, C. Su, M. Menghini, M. Sousa, C. Marchiori, L. Zhang, J. Seo, J. Locquet, Production of v_2 thin films through post-deposition annealing of $v_2 o_3$ and $v_2 o x$ films. *Thin Solid Films* **591**, 143–148 (2015)
- K. Vandoorne, Photonic reservoir computing with a network of coupled semiconductor optical amplifiers. PhD thesis, Ghent University (2011)
- K. Vandoorne, P. Bienstman, A photonic implementation of reservoir computing, in *2007 IEEE/LEOS Symposium Benelux Chapter Proceedings* (2007), pp. 195–198
- K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, J. Van Campenhout, Toward optical signal processing using photonic reservoir computing. *Opt. Express* **16**(15), 11182–11192 (2008)
- K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, P. Bienstman, Parallel reservoir computing using optical amplifiers. *IEEE Trans. Neural Netw.* **22**(9), 1469–1481, 9 (2011)
- K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, P. Bienstman, Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat. Communi.* **5**, 3541, 1 (2014)
- D. Verstraeten, Reservoir computing: computation with dynamical systems. PhD thesis, Ghent University (2009)
- A.S. Weigend, N.A. Gershenfeld, Results of the time series prediction competition at the santa fe institute, in *IEEE International Conference on Neural Networks* (IEEE, 1993), pp. 1786–1793
- H. Zhang, X. Feng, B. Li, Y. Wang, K. Cui, F. Liu, W. Dou, Integrated photonic reservoir computing based on hierarchical time-multiplexing structure. *Opt. Express* **22**(25), 31356–31370, 12 (2014)