



# Wavelength dimension in waveguide-based photonic reservoir computing

EMMANUEL GOOSKENS,<sup>1,2,\*</sup>  FLORIS LAPORTE,<sup>1,2</sup>   
CHONGHUI MA,<sup>1,2</sup> STIJN SACKESYN,<sup>1,2</sup>  JONI DAMBRE,<sup>3</sup> AND  
PETER BIENSTMAN<sup>1,2</sup>

<sup>1</sup>Department of Information Technology, Photonics Research Group, Ghent University-Interuniversity Microelectronics Center, Sint-Pietersnieuwstraat 41, Gent 9000, Belgium

<sup>2</sup>Center for Nano-and Biophotonics (NB-Photonics), Ghent University, Sint-Pietersnieuwstraat 41, Gent 9000, Belgium

<sup>3</sup>Computer Systems Laboratory, Department of Electronics and Information Systems, Ghent University, Sint-Pietersnieuwstraat 41, Gent 9000, Belgium

\*[Emmanuel.Gooskens@UGent.be](mailto:Emmanuel.Gooskens@UGent.be)

**Abstract:** Existing work on coherent photonic reservoir computing (PRC) mostly concentrates on single-wavelength solutions. In this paper, we discuss the opportunities and challenges related to exploiting the wavelength dimension in integrated photonic reservoir computing systems. Different strategies are presented to be able to process several wavelengths in parallel using the same readout. Additionally, we present multiwavelength training techniques that allow to increase the stable operating wavelength range by at least a factor of two. It is shown that a single-readout photonic reservoir system can perform with  $\approx 0\%$  BER on several WDM channels in parallel for bit-level tasks and nonlinear signal equalization. This even when taking manufacturing deviations and laser wavelength drift into account.

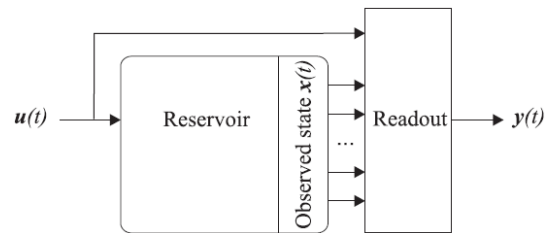
© 2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

## 1. Introduction

Reservoir computing (RC) employs a randomly initialised fixed recurrent neural network (RNN), called the reservoir, which is left untrained and to which a simple linear readout layer is added. Only this linear readout is trained, greatly facilitating the practical application of RNNs for hardware implementations [1–4] (Fig. 1). RNNs differ from feedforward neural networks by preserving in their internal states a nonlinear transformation of the input history. In other words, they have dynamical memory, which makes them ideally suited to process temporal information.

This reservoir does not need to be a traditional network of artificial neurons implemented in software, but it can be any dynamical system obeying a certain set of broad constraints. This means that a hardware implementation of RC is a logical choice. Indeed, letting a dynamical system evolve in hardware is typically more efficient than solving the equations that govern its behaviour on a general-purpose computer. Also, since the reservoir is initialised randomly anyway, deviations from the reservoir design during fabrication can be compensated for by tailoring the readout weights.

Photonics-based hardware implementations have an additional set of advantages. In particular, low power consumption and high data bandwidth make photonics-based hardware implementations attractive choices. In addition, exploiting wavelength division multiplexing, as is done in this paper, enables parallelism. There are many potential photonics-based hardware implementations. Among those investigated are systems consisting of a single non-linear node with feedback and free-space reservoir systems [6–22]. The former use of only a single node limits data bandwidth while the latter is not as compact, fast or cost-efficient as integrated systems. Here, we will focus on multi-node waveguide-based integrated RC systems developed on a silicon photonics platform. Such systems have been proven to perform well for various tasks such as bit-level tasks,

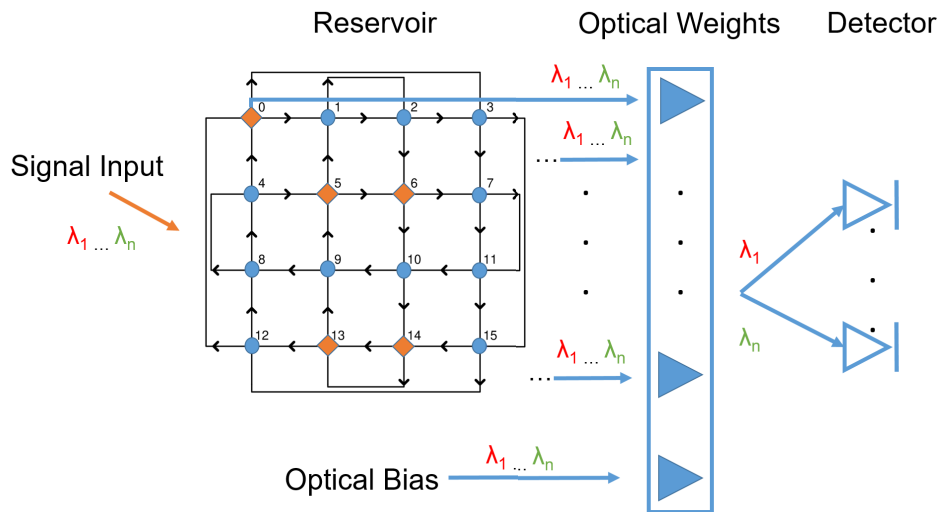


**Fig. 1.** Schematic representation of a reservoir computing system. The input signal  $u(t)$  is fed into the reservoir, and the resulting reservoir states  $x(t)$ , possibly together with the input or a bias signal, are used to learn a linear readout that is then used to generate the output signal  $y(t)$ . Reprinted with permission from Katumba et al. [5].

nonlinear dispersion compensation and isolated spoken digit recognition [5,23,24]. However, the footprint of waveguide-based photonic reservoirs, where waveguides form the interconnects and nodes consists of optical elements such as multimode interferometers [22], is typically on the order of one to a few tens of  $mm^2$ . This large footprint translates into added cost, which negatively impacts economic viability. One way to make more efficient use of a given chip area is to exploit wavelength-division multiplexing, as it allows to unlock the inherent parallelism in optical processing. This presents opportunities to increase processing power significantly, depending on the number of multiplexed channels used. Similar ideas have been explored in e.g. [25–27], although using different reservoir architectures and technologies. In the silicon photonics approach we discuss here, we employ a so-called optical readout, where the different nodes of the reservoir are weighted in the analog optical domain for both amplitude and phase [28], e.g. using heaters or reverse-biased pn-junctions in Mach-Zehnder configurations. We want to use a single set of optical weights for all wavelengths, as having a separate sets of optical weights for each wavelength would eliminate all the chip area savings. This can only be achieved in cases where a single task has to be executed for several wavelength channels in parallel. This is often the case in the telecommunication industry, an example being signal equalization.

The simulated reservoir is an integrated passive silicon photonics reservoir based on the designs outlined in [29–31] that strives to give an accurate reproduction of the reservoir dynamics and readout performance by including relevant manufacturing deviations as outlined in appendix A. (Fig. 2). The nodes consist of 3x3 multimode interferometers (MMIs) simulated using scatter matrix formalism. All nodes are connected to readout weights, consisting of both amplitude and phase weights, simulated by complex multiplication of complex weights with complex electrical field values. A 17th amplitude-phase weight set was connected to a continuous-wave optical signal serving as a trainable optical bias. The extra connections for the nodes at the edges of the reservoir help improve the reservoir dynamics and avoid the modal radiation losses at 2x1 combiners [30]. The multiple-input strategy increases the computational power of the reservoir through varied mixing between the multiple copies of the input signal with different phases. Additionally, multiple inputs lead to a more even power distribution throughout the reservoir benefiting node signal extraction, reservoir dynamics and reservoir memory. To limit the additional hardware needed with driving more nodes, only a subset of the 16 nodes was selected, based on a heuristic approach, as in [32]. The nonlinearity required for nonlinear tasks is supplied by the inherent nonlinearity of the photodetector.

The simulations are done in Photontorch [33], a set of photonic simulation tools for simulation and optimization of photonic circuits in time and frequency domain. The framework is built on top of the deep learning framework PyTorch, which enables the use of native PyTorch optimizers to optimize the (physical) parameters of the circuit.



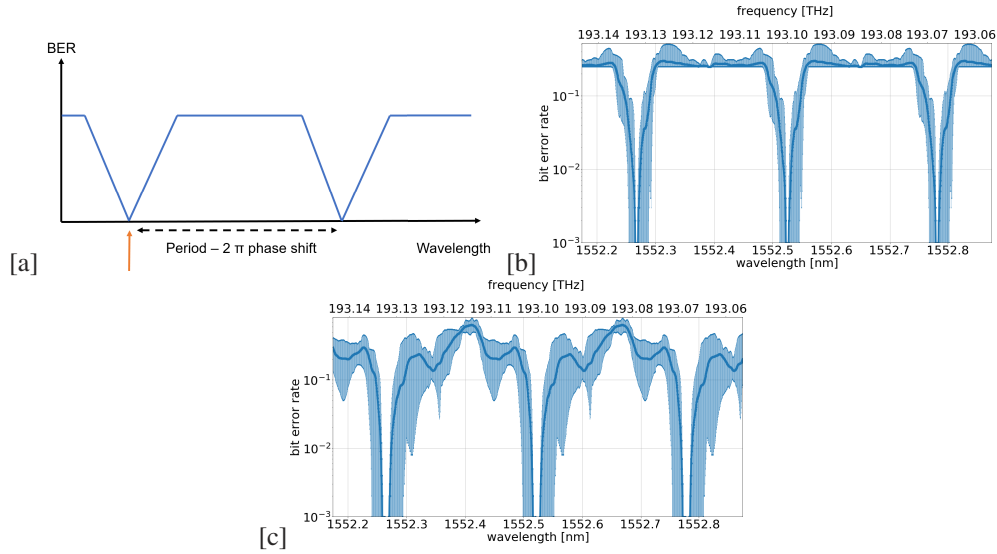
**Fig. 2.** Schematic of the simulated system. Signal input is injected in the orange diamond shaped nodes. All nodes are connected to the readout. Arrows indicate propagation direction of the signal throughout the reservoir.

The remainder of this paper is structured as follows. In section 2, it will be shown how good performance can be achieved at multiple wavelength channels by engineering the interconnection length parameter. In section 3, we propose a different approach to achieve good performance at multiple wavelength channels, based on adjusting the readout training. In section 4, we discuss how we can use similar techniques to mitigate laser wavelength drift. In section 5, this technique is then applied in combination with the method of section 2, leading to a demonstration of reliable operations at multiple wavelengths in parallel in a simulated photonic reservoir computing system.

## 2. Engineered interconnection lengths

The main reason why reservoir performance goes down when varying the wavelength/frequency is the resulting variation in phase shifts in the waveguide interconnections. This leads to altered signal mixing for which the readout was not trained, leading to an incorrect weighting and recombination of node outputs. However, when all interconnections are of identical length, there will be frequency changes for which the corresponding phase shift variation equals an integer multiple of  $2\pi$ . The frequency change inducing a  $2\pi$  phase shift is approximately constant, with variation being caused by dispersion. This gives rise to approximate frequency periodicity in reservoir performance (Fig. 3(a)). By engineering the waveguide interconnection length, one can ensure that the frequency spacing between DWDM or CWDM channels corresponds to this period.

An important boundary condition when engineering the interconnection length, is that the interconnection length should not surpass the distance that light can travel during a single bit period ( $d_{\max}$ ). This causes previously injected bits to be in transit in between nodes, hidden from the readout. These bits cannot contribute to the output until they reach the next nodes. Therefore, the system has gaps in its memory, which is detrimental for tasks requiring memory, such as the 2-bit delayed XOR task studied here. There is thus an upper limit imposed on the interconnection length, which depends on the input bitrate. This in turn imposes a lower limit on the frequency periodicity. For the bitrate  $B$  of 32 GHz, employed for the bit level tasks throughout this paper,



**Fig. 3.** Exploiting engineered interconnection length. (a) Training only occurred for one wavelength (1552.5244 nm), as indicated by the orange arrow. (b) Performance for the delayed 2-bit XOR task. (c) Performance for the nonlinear signal equalization task. Error bars indicate minimum and maximum achieved BER over 10 different reservoirs each with their own manufacturing deviations.

using Eqs. (1) and (2):

$$d_{\max} = \frac{c}{Bn_g}, \quad (1)$$

$$\frac{2\pi n_{\text{eff}}(\lambda)}{\lambda} d_{\max} = \frac{2\pi n_{\text{eff}}(\lambda + \Delta\lambda)}{\lambda + \Delta\lambda} d_{\max} + 2\pi, \quad (2)$$

the maximum interconnection length  $d_{\max} \approx 2$  mm and the minimum frequency periodicity  $\approx 32$  GHz corresponding to a wavelength shift  $\approx 0.257$  nm around 1552.5244 nm. Here,  $c$  stands for the speed of light and  $n_{\text{eff}}$  is the effective index of the waveguide. Throughout this paper, reservoir interconnections consist of strip waveguides with a width  $w = 0.45 \mu\text{m}$ , thickness  $h = 0.22 \mu\text{m}$ ,  $n_{\text{eff}}(1550\text{nm}) = 2.28$  and  $n_g(1550\text{nm}) = 4.56$ . First order dispersion of  $n_{\text{eff}}$  is considered (appendix Eq. (6)).

We test this method using a nonlinear bit-level task, namely the delayed 2-bit XOR task, and nonlinear signal equalization. The delayed 2-bit XOR task consists of performing the Boolean XOR operation using the current and previous bit of an idealized on-off keying (OOK) 32GHz optical signal. For the nonlinear signal equalization task we use as data a simulated OOK 10 GHz optical signal that has travelled through 2000km of dispersion- and loss-compensated optical fiber with Kerr nonlinearity and thus self phase modulation. Simulations for acquisition of this data were performed using VPIphotonics Design Suite. The readout weights are trained on a training bit stream of 1000 bits using the Adam optimization algorithm, as implemented by the PyTorch Python library based on [34,35] with as error metric the mean squared error (MSE) as defined by Eq. (3):

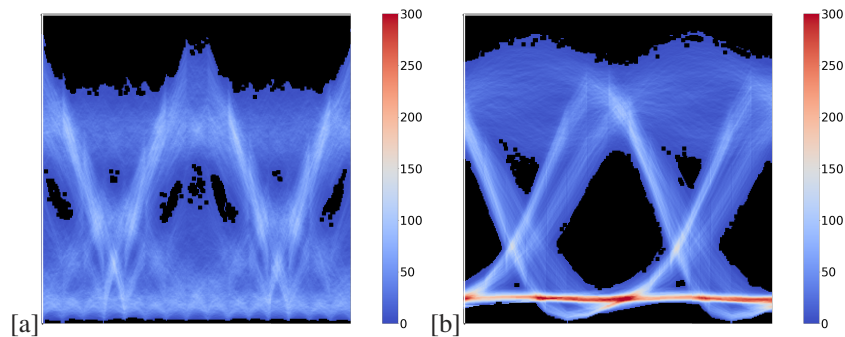
$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (3)$$

$y_i$  signifies the target value and  $\hat{y}_i$  signifies the predicted value. The threshold, based on which outputs are classified as 1 or 0, is then optimised for the lowest possible bit error rate (BER) on

the train data processed by this trained readout. The trained readout and optimised threshold are then applied to the states of the reservoir for a test bit stream of 100,000 bits to acquire the test BER. Since we use  $10^5$  bits in our simulations, good practice is to limit resolution of and crop the BER at  $10^{-3}$  i.e. 2 orders of magnitude higher than the lowest BER one can find in the simulation [36].

In practice, each manufactured reservoir has its own variations in interconnection lengths and interconnection effective indices, leading to variations in interconnection phases, due to manufacturing deviations. These manufacturing deviations are explained and quantified in appendix A. The main conclusions are that interconnection phases need to be considered random in simulations and that interconnection length variations are normally distributed with mean 0 and standard deviation 21.08 nm. To investigate if reliable performance can be achieved despite such manufacturing deviations, 10 different reservoirs, each with different manufacturing deviations, were trained and tested in the manner described above.

Results, using the minimum periodicity of 32 GHz, are shown in Figs. 3 and 4. At 1552.5244 nm, and the wavelengths separated from it by the period, a BER  $\approx 0\%$  is achieved for all reservoirs (Remember that the size of test set puts a lower limit on the BER we can reliably measure of  $10^{-3}$ ). For the nonlinear dispersion compensation task it is also clear that the eye diagrams are much improved. This shows that neither manufacturing deviations nor dispersion were an issue.



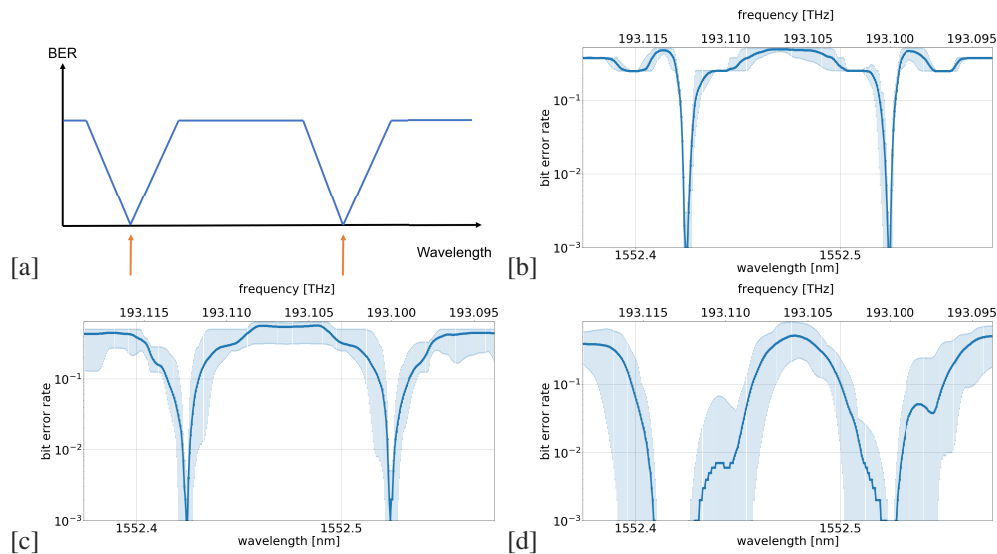
**Fig. 4.** Nonlinear signal equalization. (a) Eye diagram for original data stream after 2000km long dispersion- and loss-compensated optical fiber with Kerr nonlinearity. (b) Eye diagram after reservoir for wavelength 1552.5244 nm.

However, the wavelength range over which a low BER is reliably achieved for all reservoirs, is quite narrow. This would indicate a relatively high sensitivity to e.g. laser wavelength drift, this issue is addressed in sections 4. and 5.

### 3. Multiple-wavelength training

The method to achieve WDM described in section 2. works well but has a drawback. Indeed, there is a lower limit on the frequency spacing between channels, depending on the input bitrate. For bitrates of  $\approx 12.5$  GHz and higher, the smallest ITU-T DWDM spacing of 12.5 GHz [37] cannot be attained. Using Eqs. (2) and (6) with  $\lambda = 1552.5244\text{nm}$  and  $d_{max} \approx 2\text{mm}$  yields  $d_{wl} \approx 0.257\text{nm}$  which corresponds to 32 GHz frequency spacing. Therefore, in this section, we present another method, based on minimising the MSE for multiple wavelength channels simultaneously (Fig. 5(a)).

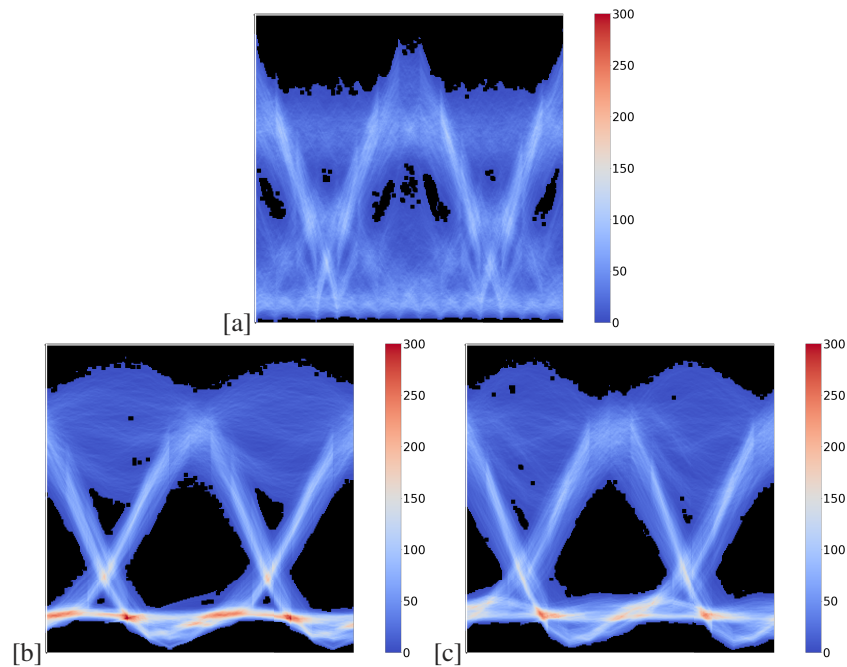
Taking into account more than one wavelength during training poses a greater challenge as a machine learning task. However, for the tasks we studied here, it turns out that this is not prohibitive and performance can be as good as that achieved for single wavelength training.



**Fig. 5.** Multiple-wavelength training for bit level tasks. (a) Training occurred at 2 wavelengths as indicated by the orange arrows. In this case, we used 1552.4239 nm and 1552.5244 nm, which correspond to the smallest wavelength channel spacing (12.5 GHz) on the ITU-T DWDM grid [37]. (b) Performance for the delayed 2-bit XOR task. Wavelength channels achieve 0.2% and 0.1% mean BER respectively. (c) Performance for a 3-bit sequence recognition task. Target sequences were [111],[110],[101] and [011]. Both wavelength channels achieve 0.2% mean BER. (d) Performance for the nonlinear signal equalization task. Both wavelengths achieve  $\approx 0\%$  BER. Error bars indicate minimum and maximum achieved BER for different reservoirs with their own simulated manufacturing deviations.

This is illustrated in Fig. 5, for the delayed 2-bit XOR task, for general bit sequence recognition tasks and for the nonlinear signal equalization task. For the recognition tasks, the output needs to be 1 when observing certain target sequences of an idealized OOK 32GHz optical signal, and 0 for the non-target sequences. The results show that the readouts achieve good performance at both trained wavelength channels for all 10 different reservoirs.

In Fig. 6 we visualize the eye diagrams of nonlinear signal equalization performed for 2 wavelengths in parallel through multiple wavelength training for one specific reservoir. It is clear that compared to the original eye diagram, the eye diagrams at both wavelengths are much improved.



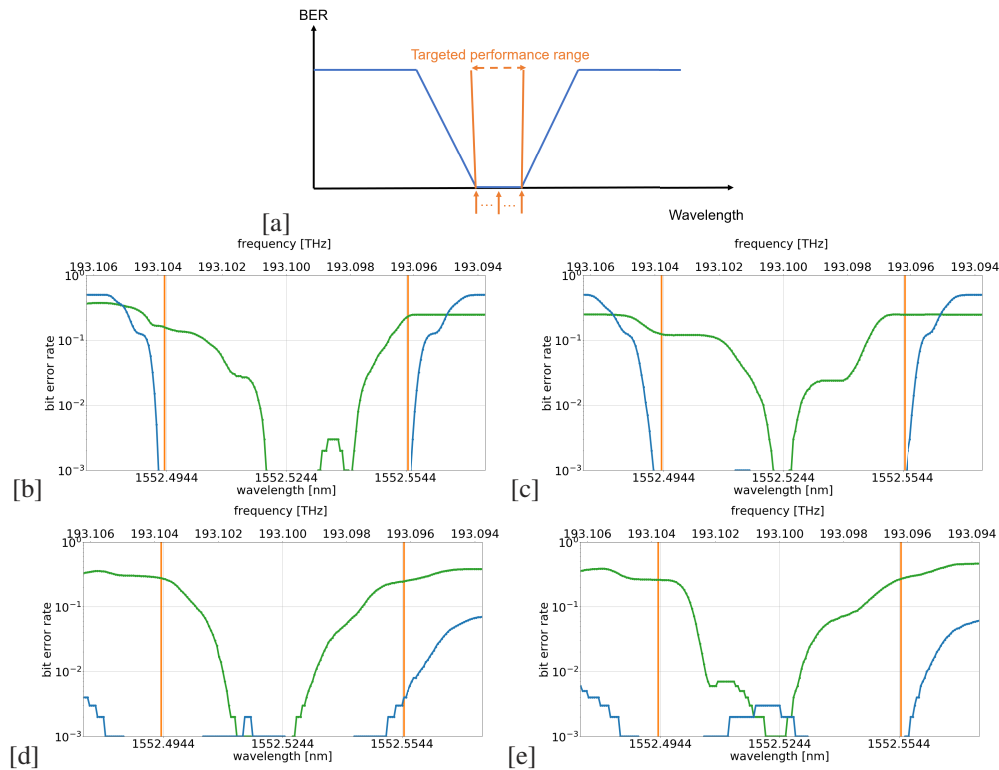
**Fig. 6.** Multiple-wavelength training for nonlinear signal equalization. (a) Eye diagram for original data stream after 2000km long dispersion- and loss-compensated optical fiber with Kerr nonlinearity. (b) Eye diagram after reservoir for wavelength 1552.4239 nm. (c) Eye diagram after reservoir for wavelength 1552.5244 nm.

#### 4. Mitigating laser wavelength drift

The input laser wavelength can drift over time, moving away from the optimal performance point, which can be a problem if the operating range is rather narrow. Therefore, in this section, we aim to enlarge the wavelength range over which good performance is achieved. Note that we suppose that this drift takes place on timescales much larger than the bit period. If not, the reservoir dynamics would be significantly altered compared to those used for training the readout, which would be detrimental to the performance.

As there is no closed-form analytical relation between this stable operating wavelength range and the readout weights, it is not possible to directly train the readout for this metric. Instead, the readout is trained to minimize the MSE for multiple wavelengths symmetrically situated around the targeted wavelength channel, similarly as in section 3. The number of wavelengths trained and the wavelength spacing between them need to be considered. For some reservoirs it suffices to train the target wavelength channel and the extremes of the targeted wavelength range. For other reservoirs, multiple closely spaced intermediary wavelengths are trained, so as to maintain consistent good performance over the targeted wavelength range. The difficulty thus lies in maximising the width of the wavelength range over which consistent good performance is achieved, without decreased performance in between trained wavelengths.

The targeted wavelength channel was chosen to be 1552.5244 nm. This is a wavelength channel on the ITU-T defined DWDM grid for all of the various possible frequency spacings (12.5 GHz, 25 GHz, 50 GHz, 100 GHz) [37]. Results are displayed in Fig. 7 for two specific reservoir initializations for a nonlinear bit level task, namely the 2-bit delayed XOR task, and the nonlinear signal equalization task.



**Fig. 7.** Mitigating laser drift (a) The orange arrows indicate the wavelengths that are trained. (b-c) Performance for the delayed 2-bit XOR task. (d-e) Performance for the nonlinear signal equalization task. For the bit level tasks it is common, and as is the case in (b) and (c), that only training 3 wavelengths is necessary. For the nonlinear signal equalization (d-e) a total of 5 training wavelengths, being the center and extremes of the targeted range and two intermediary wavelengths, yielded better results. Each figure corresponds to a particular reservoir initialization. Green indicates 1-wavelength training and blue 3-wavelength training. The orange boundaries indicate the targeted wavelength range over which to achieve good performance.

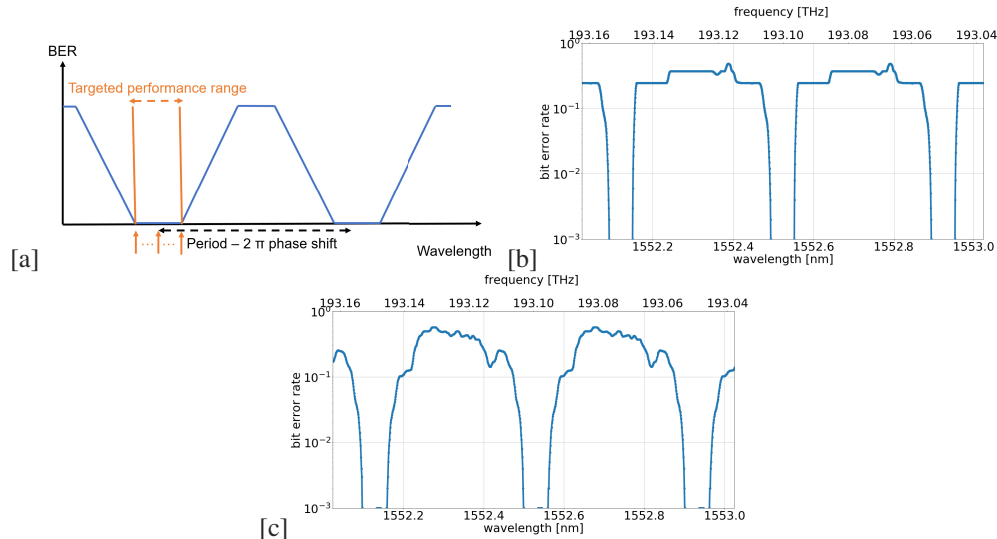
For the delayed 2-bit XOR task, averaged over 10 different reservoirs, good performance, defined as  $<1\%$  bit error rate (BER), is achieved over a wavelength range of 63.1 pm or a corresponding frequency range of 7.8 GHz in case of multiple wavelength training. This is a significant improvement compared to the 24.5 pm and 3 GHz ranges in case of 1-wavelength training.

For the nonlinear signal equalization task, averaged over 10 different reservoirs, good performance, again defined as  $<1\%$  bit error rate (BER), is achieved over a wavelength range of 95.4 pm or a corresponding frequency range of 11.9 GHz in case of multiple wavelength training. This is a significant improvement compared to the 41.0 pm and 5.1 GHz ranges in case of 1-wavelength training. For comparison, the commercially available Menara networks 5ZR0A00-TNBL 50GHz C-band tunable transceiver lists its wavelength stability after startup as  $\pm 25$  pm, thus a total wavelength variation of 50 pm [38].

Additionally, we remark that this increased stable operating range will also benefit robustness against other environmental effects (e.g. temperature).

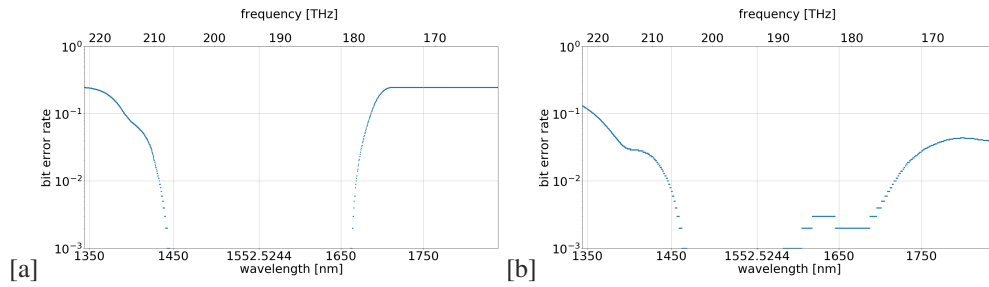
## 5. Robust WDM operation

Finally, in this section, we combine the tuning of the interconnection delay as in section 2. with the robustness technique of section 4. Figure 8 shows that such a combination is indeed successful. This results in a system where laser wavelength drift is taken into account during training and where processing power is greatly increased without the need for an increased system footprint. By broadening the operating range of the readout, any imperfection of the performance periodicity (e.g. due to dispersion) is also mitigated. This allows for multiplexing many wavelength channels even with large wavelength spacing.



**Fig. 8.** Robust WDM operation (a) The orange arrows indicate the wavelengths that are trained. For these particular reservoirs and tasks it was again sufficient to only train the target wavelength and the extremes of the targeted wavelength range over which to achieve good performance. Some other studied reservoirs needed inclusion of additional wavelengths in the targeted wavelength range during training to maintain good performance over the entire range. (b) Performance for the delayed 2-bit XOR task for one particular reservoir. (c) Performance for the nonlinear signal equalization task for one particular reservoir. Other reservoirs showed similar characteristics with some variation in the width of the good performance wavelength range. All achieve  $\approx 0$  BER at target wavelengths.

Dispersion ultimately limits the number of 50-GHz-spaced DWM channels for which good performance can be achieved. In practice, it turns out that we can easily accommodate tens to hundreds of 50-GHz-spaced channels, with 1552.5244 nm as the central channel. This is demonstrated in Fig. 9 where the BER for 50 GHz spaced channels were displayed. Note that in the wavelengths between these channels, for which BER is not shown, the BER will vary. It is clear that the BER remains near optimal value for several THz away from the central channel for which the interconnection length was designed.



**Fig. 9.** BER at target channels indicates robust WDM operation (BER outside of the target channels not plotted) (a) For the delayed 2-bit XOR task and reservoir of 8 [b] (b) For the nonlinear signal equalization task and reservoir of 8 [c]. In both of these cases and for all other reservoirs, dispersion only significantly impacts performance after  $>1$  THz wavelength spacing away from the designed wavelength channel. This corresponds to many tens to hundreds of DWDM channels.

## 6. Conclusion

We showed that including the wavelength dimension in both the design and training process of a photonic RC system opens up the possibility of greatly increasing system bandwidth through WDM. In particular, we demonstrated that a single-readout photonic RC system can perform with  $\approx 0$  BER at several wavelength channels for bit level tasks such as the delayed 2-bit XOR task and the nonlinear signal equalization task. This was done while taking into account manufacturing deviations, by reviewing 10 different reservoirs, and mitigating laser wavelength drift by increasing the stable operating wavelength range from 24.5pm/3GHz to 63.1pm/7.8GHz and from 41.0pm/5.1GHz to 95.4pm/11.9GHz for the bit level 2-bit delayed XOR task and nonlinear signal equalization task respectively. This clears the way toward commercial viability of photonic RC systems as the same chip footprint now has significantly increased processing power and reliability.

### A. Appendix: manufacturing deviations

#### A.1. Waveguide roughness

Due to the high refractive index contrast of the silicon-on-insulator (SOI) platform, the devices are very sensitive to geometric variations [39]. For this work, the intra-die geometric variation, specifically the variation related to the interconnection waveguides in a single reservoir, is of importance, as it impacts the phase information inside the reservoir. Phase information is a key aspect of photonic RC using coherent reservoirs, as interference effects play a key role in the reservoir dynamics.

Variations include thickness and width fluctuations caused by pattern density non-uniformity [40]. These waveguide width fluctuations can also be referred to as waveguide (sidewall) roughness. It is most often studied in the context of propagation losses, but here its effect on the phase is of interest. This waveguide roughness depends on the fabrication process and will be very different for devices fabricated with deep UV lithography or e-beam lithography, and vary between fabs. The thickness fluctuations on the other hand depend largely on the qualities of the source wafer. These width and thickness fluctuations impact the  $n_{\text{eff}}$  of the interconnection waveguides and thus the phase change light undergoes travelling through such an interconnection waveguide. From [40]  $\Delta n_{\text{eff, intradie}} \approx 0.015$  and  $\Delta n_{g, intradie} \approx 0.015$  for the IMEC Multi-Project Wafer (MPW) service, which uses 200 mm wafers and fabrication through 193 nm deep-UV lithography. Applying this to an interconnection waveguide with a length  $l = 1\text{mm}$  (a typical

length for the delay spirals inside the reservoir) at wavelength  $\lambda = 1550$  nm, using Eq. (4):

$$\Delta\phi = \frac{2\pi}{\lambda} l \Delta n_{\text{eff}}, \quad (4)$$

the resulting phase change due to the waveguide roughness  $\Delta\phi \approx 61 \gg 2\pi$ .

In other words, the exact phase change accumulated over interconnection waveguides cannot be known. This needs to be taken into account during simulations but does not preclude good system performance, as the readout can adapt to this during training. Indeed, not having to know or control the exact implementation of the reservoir is one of the key advantages of the reservoir computing paradigm. This does mean however that each reservoir needs its own individually trained readout and that it can not be readily assumed that different reservoirs will give rise to identical performance.

Of interest for our purposes is the additional wavelength dependence of this phase deviation, which, if sufficiently large, could cause the  $2\pi$  phase shifts exploited in section 2. to no longer be regularly spaced in the frequency dimension. From a first-order approximation of the group index as defined by Eq. (5) [41]:

$$n_g(\lambda_0) = n_{\text{eff}}(\lambda_0) - \lambda_0 \frac{n_{\text{eff}}(\lambda) - n_{\text{eff}}(\lambda_0)}{\lambda - \lambda_0}, \quad (5)$$

the effective index  $n_{\text{eff}}(\lambda)$  is calculated from the known effective index  $n_{\text{eff}}(\lambda_0)$  and group index  $n_g(\lambda_0)$  using Eq. (6):

$$n_{\text{eff}}(\lambda) = n_{\text{eff}}(\lambda_0) - \frac{\lambda - \lambda_0}{\lambda_0} (n_g(\lambda_0) - n_{\text{eff}}(\lambda_0)). \quad (6)$$

The manufacturing deviations can then be taken into account by adding extra deviations  $\Delta n_{\text{eff}}(\lambda_0)$  and  $\Delta n_g(\lambda_0)$  to the formula:

$$n_{\text{eff}}(\lambda) = n_{\text{eff}}(\lambda_0) + \Delta n_{\text{eff}}(\lambda_0) - \frac{\lambda - \lambda_0}{\lambda_0} (n_g(\lambda_0) + \Delta n_g(\lambda_0) - n_{\text{eff}}(\lambda_0) - \Delta n_{\text{eff}}(\lambda_0)). \quad (7)$$

Let us now consider the phase change undergone by light at wavelength  $\lambda$  passing through the waveguide of length  $l$  with refractive indices  $n_{\text{eff}}(\lambda_0) + \Delta n_{\text{eff}}(\lambda_0)$  and  $n_g(\lambda_0) + \Delta n_g(\lambda_0)$  as shown in Eqs. (8) and (9):

$$\phi(\lambda_2) - \phi(\lambda_1) = 2\pi l \left( \frac{n_{\text{eff}}(\lambda_2)}{\lambda_2} - \frac{n_{\text{eff}}(\lambda_1)}{\lambda_1} \right), \quad (8)$$

$$\begin{aligned} \phi(\lambda_2) - \phi(\lambda_1) = 2\pi l \left[ \left( \frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right) (n_{\text{eff}}(\lambda_0) + \Delta n_{\text{eff}}(\lambda_0)) \right. \\ \left. - \frac{1}{\lambda_0} \left( \frac{\lambda_2 - \lambda_0}{\lambda_2} - \frac{\lambda_1 - \lambda_0}{\lambda_1} \right) (n_g(\lambda_0) + \Delta n_g(\lambda_0) - n_{\text{eff}}(\lambda_0) - \Delta n_{\text{eff}}(\lambda_0)) \right], \end{aligned} \quad (9)$$

which give the total difference in phase change between wavelengths  $\lambda_1$  and  $\lambda_2$ .

From this one can deduce the difference in phase change for two wavelengths solely due to geometric variation by extracting the terms containing  $\Delta n_{\text{eff}}(\lambda_0)$  and  $\Delta n_g(\lambda_0)$  as shown in Eq. (10):

$$\begin{aligned} (\phi(\lambda_2) - \phi(\lambda_1))_{\Delta\text{geometric}} = 2\pi l \left[ \left( \frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right) \Delta n_{\text{eff}}(\lambda_0) \right. \\ \left. - \frac{1}{\lambda_0} \left( \frac{\lambda_2 - \lambda_0}{\lambda_2} - \frac{\lambda_1 - \lambda_0}{\lambda_1} \right) (\Delta n_g(\lambda_0) - \Delta n_{\text{eff}}(\lambda_0)) \right]. \end{aligned} \quad (10)$$

To estimate the magnitude of this expression, let us assume that the interconnection waveguide has a length of 1 mm. For the wavelengths, choose  $\lambda_0 = 1550$  nm and  $\lambda_1 = 1551.7208$  nm and  $\lambda_2 = 1552.5244$  nm, which corresponds to two neighbouring wavelength channels for a 100 GHz

channel spacing, the largest channel spacing for DWDM according to ITU-T specifications [37]. We assume  $\Delta n_{\text{eff}}(\lambda_0)$  and  $\Delta n_g(\lambda_0)$  to be independent from each other and set  $\Delta n_{\text{eff}}(\lambda_0) \approx -0.015$  and  $\Delta n_g(\lambda_0) \approx 0.015$ . These are the maximal possible intra-die variations according to [40] for the previously mentioned IMEC MPW service. Both  $n_{\text{eff}}$  and  $n_g$  thus have the maximum variation and in such a way that they contribute maximally to the extra wavelength dependence. This then leads to  $\phi_2 - \phi_1 \approx -\frac{\pi}{100}$ . This is a very modest extra phase shift due to the dispersion, which does not impact the performance of a readout in a significant way even when not improving its robustness as in section 4. This extra phase shift is linear with  $\lambda_2 - \lambda_1$ .

In conclusion, manufacturing deviations lead to a random but largely wavelength-independent phase change contribution. This means that it is possible to exploit phase periodicity to achieve good performance at regularly spaced wavelength channels, as long as interconnection length deviations and dispersion can be managed.

### A.2. Waveguide length deviations

Waveguide length deviations depend on the manufacturing process used. For this research, IMEC's Silicon Photonics platform is employed. According to [40] a fabricated 450 nm waveguide will have  $\pm 20$  nm variations. As length is also a 2D feature this same  $\pm 20$  nm variation is assumed for length. Interconnection waveguides often consist of spirals with multiple straight segments subject to possible length variation. Assuming a Gaussian distribution for length variation with mean 0 nm and standard deviation  $\frac{20}{3}$  nm and 10 straight segments per interconnection, the standard deviation for the total length variation in an interconnection =  $\sqrt{10 \cdot \left(\frac{20}{3}\right)^2} \approx 21.08$  nm. As the sum of independent normally distributed random variables is itself a normally distributed variable, its variance being the sum of the variances of the random variables [42].

**Funding.** Fonds Wetenschappelijk Onderzoek (3S044419).

**Acknowledgement.** Parts of this work were performed under the EU H2020 program under grant agreements 871658 (Nebula), 871330 (NEoteRIC) and 101017237 (PHOENICS).

We would like to acknowledge Nvidia for supplying our research group with 4 Geforce GTX 1080 GPUs on which the simulations performed for this paper were run.

**Disclosures.** The authors declare no conflicts of interest.

**Data availability.** Data and code underlying the results presented in this paper are available upon reasonable request.

### References

1. M. Wolfgang, T. Natschläger, and H. Markram, "Real-time computing without stable states: a new framework for neural computation based on perturbations," *Neural Comput.* **14**(11), 2531–2560 (2002).
2. D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, "An experimental unification of reservoir computing methods," *Neural Netw.* **20**(3), 391–403 (2007).
3. M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Comput. Sci. Rev.* **3**(3), 127–149 (2009).
4. G. V. der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," *Nanophotonics* **6**(3), 561–576 (2017).
5. 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**(6), 1–10 (2018).
6. 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**(15), 152004 (2018).
7. L. Appeltant, M. C. Soriano, G. V. 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**(1), 468 (2011).
8. Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," *Sci. Rep.* **2**(1), 287 (2012).
9. L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer, "Photonic information processing beyond turing: An optoelectronic implementation of reservoir computing," *Opt. Express* **20**(3), 3241–3249 (2012).

10. A. S. F. Duport, B. Schneider, M. Haelterman, and S. Massar, "All-optical reservoir computing," *Opt. Express* **20**(20), 22783–22795 (2012).
11. 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**(5), 438–446 (2015).
12. 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**(1), 1364 (2013).
13. A. Dejonckheere, A. Smerieri, L. Fang, J. I. Oudar, M. Haelterman, and S. Massar, "All-optical reservoir computer based on saturation of absorption," *Opt. Express* **22**(9), 10868–10881 (2014).
14. M. C. Soriano, S. Ortin, D. Brunner, L. Larger, C. R. Mirasso, I. Fischer, and L. Pesquera, "Optoelectronic reservoir computing: Tackling noise-induced performance degradation," *Opt. Express* **21**(1), 12–20 (2013).
15. R. M. Nguimdo, G. Verschaffelt, J. Danckaert, and G. V. der Sande, "Fast photonic information processing using semiconductor lasers with delayed optical feedback: Role of phase dynamics," *Opt. Express* **22**(7), 8672–8686 (2014).
16. 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**(4), 1501610 (2013).
17. R. M. Nguimdo and T. Erneux, "Enhanced performances of a photonic reservoir computer based on a single delayed quantum cascade laser," *Opt. Lett.* **44**(1), 49–52 (2019).
18. 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**(8), 10211–10219 (2018).
19. J. Vatin, D. Rontani, and M. Sciamanna, "Enhanced performance of a reservoir computer using polarization dynamics in vcsels," *Opt. Lett.* **43**(18), 4497–4500 (2018).
20. G. Mourgias-Alexandris, G. Dabos, N. Passalis, A. Totovic, A. Tefas, and N. Pleros, "All-optical wdm recurrent neural networks with gating," *IEEE J. Sel. Top. Quantum Electron.* **26**(5), 1–7 (2020).
21. A. Argyris, J. Bueno, and I. Fischer, "Photonic machine learning implementation for signal recovery in optical communications," *Sci. Rep.* **8**(1), 8487 (2018).
22. A. Lugnan, A. Katumba, F. Laporte, M. Freiberger, S. Sackesyn, C. Ma, E. Gooskens, J. Dambre, and P. Bienstman, "Photonic neuromorphic information processing and reservoir computing," *APL Photonics* **5**(2), 020901 (2020).
23. C. Mesaritakis, V. Papatziaris, 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**(11), 3048–3055 (2013).
24. 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).
25. R. M. Nguimdo, G. Verschaffelt, J. Danckaert, and G. V. der Sande, "Simultaneous computation of two independent tasks using reservoir computing based on a single photonic nonlinear node with optical feedback," *IEEE Trans. Neural Netw. Learning Syst.* **26**(12), 3301–3307 (2015).
26. F. Duport, A. Smerieri, A. Akrou, M. Haelterman, and S. Massar, "Virtualization of a photonic reservoir computer," *J. Lightwave Technol.* **34**(9), 2085–2091 (2016).
27. A. Akrou, A. Bouwens, F. Duport, Q. Vinckier, M. Haelterman, and S. Massar, "Parallel photonic reservoir computing using frequency multiplexing of neurons," <https://arxiv.org/abs/1612.08606>.
28. M. Freiberger, A. Katumba, P. Bienstman, and J. Dambre, "Training passive photonic reservoirs with integrated optical readout," *IEEE Trans. Neural Netw. Learning Syst.* **30**(7), 1943–1953 (2019).
29. K. Vandoorne, P. Mechet, T. V. Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nat. Commun.* **5**(1), 3541 (2014).
30. S. Sackesyn, C. Ma, J. Dambre, and P. Bienstman, "An enhanced architecture for silicon photonic reservoir computing," *Cognitive Computing 2018 - Merging Concepts with Hardware* (2018), pp.1–2.
31. S. Sackesyn, C. Ma, J. Dambre, and P. Bienstman, "Experimental realization of integrated photonic reservoir computing for nonlinear fiber distortion compensation," *Opt. Express* **29**(20), 30991–30997 (2021).
32. A. Katumba, M. Freiberger, P. Bienstman, and J. Dambre, "A multiple-input strategy to efficient integrated photonic reservoir computing," *Cogn. Comput.* **9**(3), 307–314 (2017).
33. F. Laporte, J. Dambre, and P. Bienstman, "Highly parallel simulation and optimization of photonic circuits in time and frequency domain based on the deep-learning framework pytorch," *Sci. Rep.* **9**(1), 5918 (2019).
34. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," (2017).
35. I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," <https://arxiv.org/abs/1711.05101>.
36. M. Jeruchim, "Techniques for estimating the bit error rate in the simulation of digital communication systems," *IEEE J. Select. Areas Commun.* **2**(1), 153–170 (1984).
37. ITU-T, "G694.1: SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS Transmission media and optical systems characteristics – Characteristics of optical systems Spectral grids for WDM applications: DWDM frequency grid," <https://www.itu.int/rec/T-REC-G.694.1-201202-I/en>.
38. Menara Networks, "Datasheet SZR0A00-TNBL," [http://menaranet.com/index.php?route=information/information&information\\_id=24](http://menaranet.com/index.php?route=information/information&information_id=24).

39. S. Pathak, D. V. Tourhout, and W. Bogaerts, "Design trade-offs for silicon-on-insulator-based awgs for (de)multiplexer applications," *Opt. Lett.* **38**(16), 2961–2964 (2013).
40. Y. Xing, J. Dong, S. Dwivedi, U. Khan, and W. Bogaerts, "Accurate extraction of fabricated geometry using optical measurement," *Photonics Res.* **6**(11), 1008–1020 (2018).
41. J. R. Rogers and M. D. Hopper, "Conversion of group refractive index to phase refractive index," *J. Opt. Soc. Am. A* **5**(10), 1595–1600 (1988).
42. D. S. Lemons, "Normal sum theorem," in *An Introduction to Stochastic Processes in Physics*, (The Johns Hopkins University, 2002), chap. 5.