Neuromorphic information processing using silicon photonics

Peter Bienstman[a], Joni Dambre[b], Andrew Katumba[a], Matthias Freiberger[b], Floris Laporte[a], Alessio Lugnan[a], Stijn Sackesyn[a], and Chonghuai Ma[a]

[a]Photonics Research Group, Department of Information Technology, Ghent University - imec, Belgium

[b]Data Science Lab, Department of Electronics and Information Systems, Ghent University - imec, Belgium

ABSTRACT

We present our latest results on silicon photonics neuromorphic information processing based a.o. on techniques like reservoir computing. First, we dicuss how passive reservoir computing can be used to perform non-linear signal equalisation in telecom links. Then, we introduce a training method that can deal with limited weight resolution for a hardware implementation of a photonic readout.

Keywords: photonic neuromorphic information processing, reservoir computing, silicon photonics

1. INTRODUCTION

The continuous increase in demand for systems that can process the massive amounts of data available today has strained the currently employed transistor-based von Neumann architectures. Simultaneously, the growing demand for high-throughput, high-fidelity telecommunications systems has generated significant implementation hurdles for the associated signal processing systems.

To address the compounding challenges for these computation and communication systems, a major design revolution is underway for the next generations of these systems in the IT research world. The frantic search for potential solutions has initiated a revisit to analog computation platforms but with the aim of combining them with the state-of-the-art in large-scale integration technology. These platforms exploit the inherent dynamics of certain physical systems for processing and/or computing. Of these, prominently under consideration are biologically inspired techniques, and particularly brain-inspired computing approaches that employ artificial structures that mimic the brain's neural computational semantics.

Reservoir computing (RC) is a brain-inspired computing approach that initially emerged as a way around the intricacies associated with correctly training recurrent neural networks.[1–3] Classical software RC involves setting up a large randomly initialized nonlinear dynamical system (*the reservoir*) – usually an artificial neural network – that is tuned into a specific dynamical regime to allow for the following three conditions: separability of the inputs, generation of similar outputs for similar inputs and some form of finite memory of the previous inputs. Under these circumstances, the states of the reservoir can be linearly combined, following task-imposed optimization criteria, to extract the desired outputs for the specified inputs.

Beyond the initial software implementations, RC has evolved into a way to enable computing with physical nonlinear dynamical systems. Examples of the concept applied to mechanical systems, memristive systems, atomic switch networks, boolean logic elements and photonic systems can be found in.[4–8] Photonic RC particularly presents a number of benefits compared to e.g. electronics, as it offers a large bandwidth and is inherently massively parallel.

To date, experimental demonstrations of photonic reservoirs routinely achieve state of the art performance on various information processing tasks. Implementations based on a single nonlinear node with a delayed feedback architecture have proven that photonic RC is competitive for analog information processing.[9–17] Moreover, integrated photonic reservoirs can push computation speeds even higher for digital information processing. The

Corresponding author: Peter.Bienstman@UGent.be

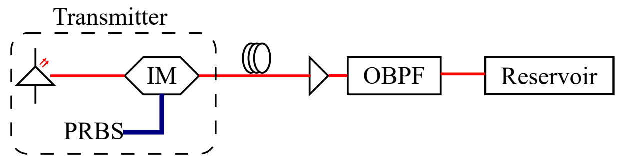

Figure 1: Schematic representation of the simulation setup to generate data for the signal equalization task. The input pseudo-random bit sequence (PRBS) signal is modulated onto a laser signal, transmitted over a fiber link, amplified and filtered, after which the field of the optical signal is saved to file to be used as input to the photonic reservoir simulation model. OBPF - Optical band pass filter.

performance of integrated photonic reservoirs has been studied numerically for networks of ring resonators,[18–22] networks of SOAs,[7] and experimentally with networks of delay lines and splitters in.[23] Integrated photonic reservoirs are particularly compelling, especially when implemented in the CMOS platform as they can take advantage of its associated benefits for technology reuse and mass production.

A recent development in the design of RC systems is the realization that for certain tasks that are not strongly nonlinear, it is possible to achieve state-of-the-art performance using a completely passive linear network, i.e., one without amplification or nonlinear elements. The required nonlinearity is introduced at the readout point, typically with a photodetector.[23] The work discussed in this paper is also based on this architecture. Aside from the integrated implementation introduced in,[23] the passive architecture has been adapted to the single node with delayed feedback architecture in form of a coherently driven passive cavity.[9]

The rest of this paper is organised as follows. First, we dicuss how passive reservoir computing can be used to perform non-linear signal equalisation in telecom links. Then, we introduce a training method that can deal with limited weight resolution for a hardware implementation of a photonic readout.

2. PASSIVE INTEGRATED PHOTONIC RESERVOIR COMPUTING FOR NON-LINEAR DISPERSION COMPENSATION

In this section, we investigate the performance of the PhRC equalizer on unrepeated fiber optic communications links in the metro regime for fiber lengths ranging from 100km to 250km for a 10 Gb/s NRZ OOK link. The setup is shown as in Figure 1. Unless stated otherwise, in this section we assumed a launch power of 5 mW (after the modulator). The amplifier, with a noise figure (NF) of 4.0 is set to entirely undo the link attenuation and the filter gets rid of out-of-band noise and is a 3$^{\text{rd}}$ order Bessel filter with bandwidth 4 times the data rate. Note that this data rate is chosen here to keep it compatible with the measurement capability in our characterization lab but the same procedure can be followed for higher speed links.

We investigate the performance of this reservoir design for equalization for links running at different data rates. The results, given in Figure 2, show that for link lengths less than 200 km in length, the PhRC equalizer can operate at data rates higher than the design data rate of 10 Gb/s. For example for the 100km link, we get error rates below the FEC limit for data rates close to 20 Gb/s. This result gives a measure of robustness of this particular PhRC equalizer. When designing equalizers for links with higher bit rates, the analysis would need to be repeated to find the best operating parameters (most notably the interconnection delay) for the equalizer.

We also compare the performance of the PhRC equalizer for a link with the launch power changed to 15 mW (from the 5 mW of the previous simulations) to an FIR FFE filter trained on the same amount of data. An adaptive FFE filter with 31 taps is used (The filter goes over the training data four times to allow for convergence). The results are shown in Figure 3. The PhRC equalizer outperforms the FFE equalizer with BERs over 5 orders of magnitude lower at, for example, 150 km and an order or two of magnitude lower at 200 km. The difference in performance originates in the fact that the PhRC equalizer is a nonlinear compensation device; it takes advantage of the nonlinear transformation in the reservoir to better model the distortion and outstrip the performance of the FFE filter. In Figure 3 we also plot the cases with and without the fiber nonlinearity. We observe that at the distances under consideration, the fiber nonlinearities are not yet deleterious. We do however observe that the reservoir is able to make use of its nonlinear nature to outperform the FFE equalizer for these links.

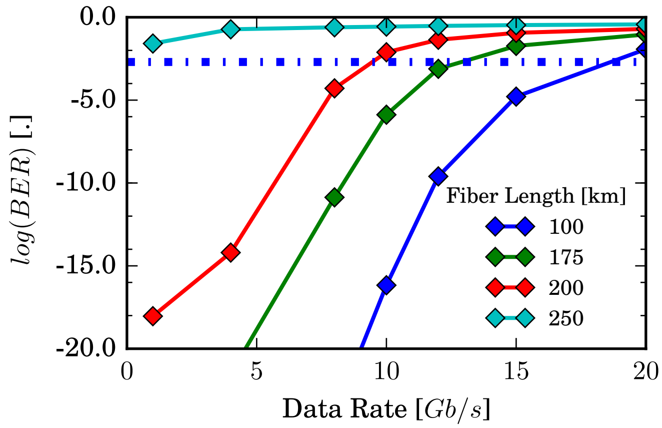

Figure 2: Error rate vs data rate for a reservoir with interconnection delay time equal to half the bit duration and latency 1 bit. A Hard Decision Forward Error Correction limit (HD-FEC limit) of 0.2×10^{-2} is also shown (dashed blue line). Error free operation is possible for all error rate values below this limit.

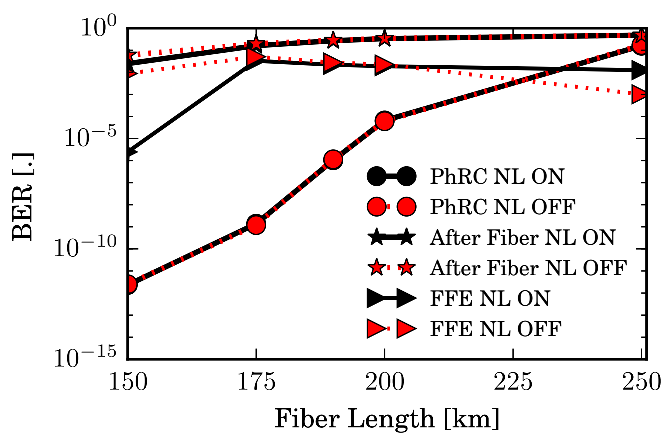

Figure 3: BER of the PhRC equalizer as compared to that of an FIR Feed Forward Equalizer (FFE) trained on the same amount of data for different fiber lengths. The launch power is set to 15 mW. NL ON- nonlinear propagation. NL OFF - nonlinear propagation is deactivated (Nonlinear Index of the fiber is set to 0).

3. ADDRESSING LIMITED WEIGHT RESOLUTION

When implementing a readout system in the optical domain, e.g. using optical elements to implement a weighted sum of the reservoir states, one problem is that, depending on the technology, the resolution of the optical weights could be much lower compared to applying weights in the electrical domain, which can easily reach 16 bit. In this paper, we will consider as an example an implementation of optical weights in Barium Titanate (BTO),[24] which has the critical advantage of being non-volatile, consuming nearly zero power while weighing the signal. The big drawback is that the resolution of the refraction index change limits the resolution to around 10 to 30 levels, coupled with some inevitable noise due to drifting of the elements.

To cope with this issue, we present an explorative quantization weight selection, inspired by methods that have been used in deep learning quantization. Typically, after a full-precision model has been trained, a subset of weights is identified to be either pruned[25] or kept fixed.[26] The other weights are then retrained in full precision and requantized. If necessary, this step can be repeated in an iterative fashion, retraining progressively smaller subsets of the weights in order to find the most optimal and stable solution.

A crucial part of these methods is selecting a subset of weights to be left fixed or to be pruned. Random selection of weights is not a good idea, because there is a high probability of eliminating 'good' weights that convey important information.[25][26] tackle this problem by choosing the weights with the smallest absolute value.

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

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

To get the best results, we conduct the procedure above in an iterative way. With each iteration, the ratio of fixed weights increases, starting from an initial value of 0.5. In each iteration step, we evaluate 20 different random weight partitions. We typically perform 4 iterations, each time increasing the ratio of fixed weights by a factor of two.

To illustrate this, we chose 4-bit delayed XOR (i.e. calculating the XOR of the current bit and 4 bits ago). Figure 4 presents the performance of our quantization retraining method on low-precision weights (8, 16, 32 levels) at different noise levels. From the blue curve, it can be seen that 4-bit delayed XOR is indeed a harder task to tackle since even with full precision, the BER increases significantly with a small amount of noise. Moreover, the task is also very sensitive to low precision weights, as shown from the naive quantization weights (orange curves), at the 8 levels resolution, the tasks is unsolvable even without any noise involved. In the meantime, we also observe that our explorative retrained weights are capable of providing very close performance to full resolution weights, except for 8 levels resolution, where due to the extra nonlinearity requirement of the task, retrained weights find themselves hard to follow. But still, naive direct quantized weights are severely outperformed by our explorative retraining method at the overall spectrum of the different noise levels.

4. CONCLUSIONS

We have described how passive photonic reservoir computing can be used to perform non-linear dispersion compensation in metro links. We have also shown how a special training scheme can be used to deal with the fact that the weight resolution in an optical readouts scheme can be limited.

Acknowledgments

This research was funded by the EU Horizon 2020 PHRESCO Grant (Grant No 688579), the EU Horizon 2020 Fun-COMP Grant (Grant No. 780848) and the Research Foundation Flanders (FWO) (Grant No G024715N).

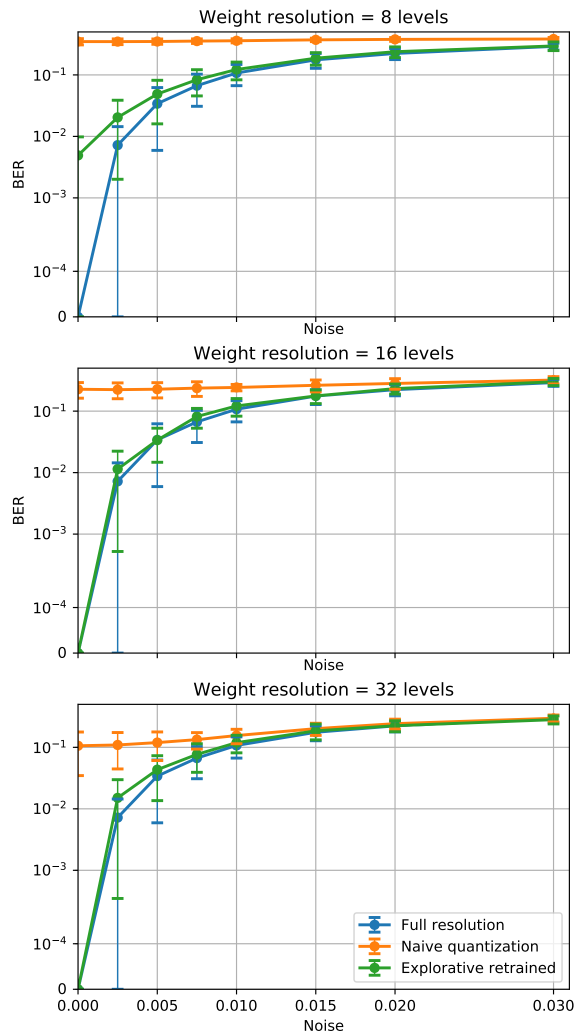

Figure 4: Performance comparison of three different weighting resolutions, 8 levels (top), 16 levels (middle), 32 levels (bottom) for the 4-bit delayed XOR task vs different noise levels. The blue curve represents the performance of full-resolution weights; the orange curve represents naive quantization weights; the green curve for explorative retrained quantization weights.

REFERENCES

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

[2] Jaeger, H. and Haas, H., "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication.," *Science (New York, N.Y.)* **304**, 78–80 (2004).

[3] Verstraeten, D., Schrauwen, B., D'Haene, M., and Stroobandt, D., "An experimental unification of reservoir computing methods," *Neural Networks* **20**, 391–403 (4 2007).

[4] Hauser, H., Ijspeert, A., Füchslin, R., Pfeifer, R., and Maass, W., "Towards a theoretical foundation for morphological computation with compliant bodies," *Biological Cybernetics* **105**, 355–370 (12 2011).

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

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

[7] Vandoorne, K., *Photonic reservoir computing with a network of coupled semiconductor optical amplifiers*, PhD thesis (2011).

[8] Paquot, Y., Duport, F., Smerieri, A., Dambre, J., Schrauwen, B., Haelterman, M., and Massar, S., "Optoelectronic Reservoir Computing," *Scientific Reports* **2**, 287 (2 2012).

[9] Vinckier, Q., Duport, F., Smerieri, A., Vandoorne, K., Bienstman, P., Haelterman, M., and Massar, S., "High-performance photonic reservoir computer based on a coherently driven passive cavity," *Optica* **2**(5), 438–446 (2015).

[10] Brunner, D., Soriano, M. C., Mirasso, C. R., and Fischer, I., "Parallel photonic information processing at gigabyte per second data rates using transient states.," *Nature communications* **4**, 1364 (1 2013).

[11] Appeltant, L., Soriano, M. C., Van der Sande, G., Danckaert, J., Massar, S., Dambre, J., Schrauwen, B., Mirasso, C. R., and Fischer, I., "Information processing using a single dynamical node as complex system.," *Nature communications* **2**, 468 (9 2011).

[12] Larger, L., Soriano, M. C., Brunner, D., Appeltant, L., Gutierrez, J. M., Pesquera, L., Mirasso, C. R., and Fischer, I., "Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing," *Optics Express* **20**, 3241 (1 2012).

[13] Duport, F., Schneider, B., Smerieri, A., Haelterman, M., and Massar, S., "All-optical reservoir computing," *Optics Express* **20**, 22783 (9 2012).

[14] Dejonckheere, A., Duport, F., Smerieri, A., Fang, L., Oudar, J.-L., Haelterman, M., and Massar, S., "All-optical reservoir computer based on saturation of absorption," *Optics Express* **22**, 10868 (5 2014).

[15] Soriano, M. C., Ortín, S., Brunner, D., Larger, L., Mirasso, C. R., Fischer, I., and Pesquera, L., "Optoelectronic reservoir computing: tackling noise-induced performance degradation," *Optics Express* **21**, 12 (1 2013).

[16] Nguimdo, R. M., Verschaffelt, G., Danckaert, J., and Van der Sande, G., "Fast photonic information processing using semiconductor lasers with delayed optical feedback: Role of phase dynamics," *Optics Express* **22**, 8672 (4 2014).

[17] Hicke, K., Escalona-Morán, M., Brunner, D., Soriano, M. C., Fischer, I., and Mirasso, C. R., "Information Processing Using Transient Dynamics of Semiconductor Lasers Subject to Delayed Feedback," *IEEE Journal of Selected Topics in Quantum Electronics* **19**, 1501610–1501610 (7 2013).

[18] Vandoorne, K., Dambre, J., Verstraeten, D., Schrauwen, B., and Bienstman, P., "Parallel reservoir computing using optical amplifiers.," *IEEE transactions on neural networks* **22**, 1469–81 (9 2011).

[19] Mesaritakis, C., Papataxiarhis, V., and Syvridis, D., "Micro ring resonators as building blocks for an all-optical high-speed reservoir-computing bit-pattern-recognition system," *JOSA B* (October) (2013).

[20] Fiers, M. A. A., Van Vaerenbergh, T., Wyffels, F., Verstraeten, D., Schrauwen, B., Dambre, J., and Bienstman, P., "Nanophotonic reservoir computing with photonic crystal cavities to generate periodic patterns," *IEEE Transactions on Neural Networks and Learning Systems* **25**(2), 344–355 (2014).

[21] Zhang, H., Feng, X., Li, B., Wang, Y., Cui, K., Liu, F., and Dou, W., "Integrated photonic reservoir computing based on hierarchical time-multiplexing structure," *Opt. Express* **22**, 31356–31370 (12 2014).

[22] Mesaritakis, C., Kapsalis, A., and Syvridis, D., "All-Optical Reservoir Computing system based on InGaAsP Ring Resonators for High-Speed Identification and Optical Routing in Optical Networks," **9370**, 1–7 (2015).

[23] Vandoorne, K., Mechet, P., Van Vaerenbergh, T., Fiers, M., Morthier, G., Verstraeten, D., Schrauwen, B., Dambre, J., and Bienstman, P., "Experimental demonstration of reservoir computing on a silicon photonics chip.," *Nature communications* **5**, 3541 (1 2014).

[24] Abel, S., Stferle, T., Marchiori, C., Caimi, D., Czornomaz, L., Stuckelberger, M., Sousa, M., Offrein, B. J., and Fompeyrine, J., "A hybrid barium titanatesilicon photonics platform for ultraefficient electro-optic tuning," *Journal of Lightwave Technology* **34**(8), 1688–1693 (2016).

[25] Han, S., Pool, J., Tran, J., and Dally, W. J., "Learning both Weights and Connections for Efficient Neural Networks," *papers.nips.cc* (2015).

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