Reduction of analogue bandwidth requirement in photonic on-chip reservoir computing

B. Schneider,[1] P. Bienstman,[1] and J. Dambre[2]

[1] Ghent University, INTEC Photonics Research Group, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium
[2] Ghent University, INTEC ELIS Group, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium

Recurrent neural networks are brain-inspired dynamical systems that can compute cognitive tasks much more efficiently than traditional computing schemes. We study a special class of recurrent networks called Reservoir Computers (RC). This work demonstrates by means of simulations how an optical RC on chip can be made more tractable in terms of signal requirements. In fact, the short delay lines on a chip require a large analog bandwidth in order to achieve efficient computing. Using the example of the speech recognition task, we show that a fast digital masking operation together with additional feedback leads to a considerable decrease of the analog bandwidth requirements.

Introduction

Reservoir Computing is a concept in machine learning and is related to the training of recurrent (neural) networks. The latter are powerful tools inasmuch they combine the classification/regression abilities of feed-forward networks in high-dimensional spaces with the inherent memory and complex dynamics of recurrences. Unfortunately, it appears that these recurrences make it very hard to train all the internal network weights in an efficient manner. Here the concept of Reservoir Computing provides a clear-cut solution to this issue. Indeed, a linear combination of all the state variables associated with the network nodes allows for the application of a standard ridge regression algorithm in the read-out layer and makes it well-suited for possible experimental, online learning implementations. This was first proposed in the pioneering papers by H. Jaeger and W. Maass [1,2,3]. Much work has been devoted to the study of RC since then, leading to its successful application in financial forecasting and robotic navigation control [4,5]. Techniques like the linear and non linear memory function have been developed to understand what tasks a RC can compute efficiently [6,7].

Recently some groups successfully set up some opto-electronic or fibred delayed feedback hardware implementations of Reservoir Computers [8,9,10]. Striving towards photonic integration implies, on the one hand, an all-optical operating regime of the RC, with increased bandwidth for high-speed applications. On the other hand, one has to cope with the very short, on-chip delay lines, which need, in order to memorise and process information in the context of RC, a very high signal input rate. This is an important issue with respect to analogue signals like speech, because arbitrary waveform generators are limited in bandwidth and costly. That is why our simulations aim to reduce these requirements on the analogue signal bandwidth by performing a faster, digital masking procedure on the input signal and adding self-loops to the reservoir nodes.

Simulation of the Reservoir

For our simulations, we chose a network where a semiconductor optical amplifier is sitting on each node (cf. figure 1). The topology, usually a random digraph, is set to "swirl" which is better suited for on-chip routing. K. Vandoorne et al. [11] showed that this optical reservoir achieves good scores for a spoken digit speech recognition task. According to previous experiments [7,8,9] , a single non linear node with delayed feedback can be considered as a reservoir computer if it is subject to input masking. The masking procedure creates a periodic sequence of constant input time windows, called virtual neurons, each associated to a random weight factor. On a conceptual level we can think of virtual neurons as the subsequent, time-multiplexed input samples, keeping the node dynamics in a transient regime. Ridge regression then combines the virtual neurons into one or more output nodes.

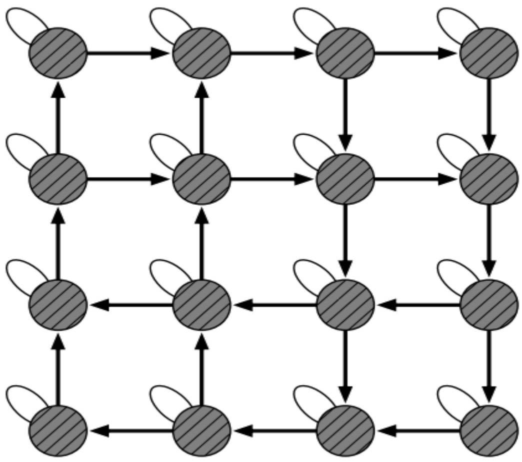

Figure 1: A 4x4 recurrent "swirl" network of non linear nodes (SOAs in gray, hatched). Interconnections (typically spiraled waveguides) are shown as arrows or arcs (self-loops).

Based on this observation, we added self-loops to each node (elliptical thin lines, cf. figure 1), on top of all the existing connections in the swirl network (black arrows). The length of the loop is equal for all the nodes and an integer multiple of the internodal distance, so as to store an integer number of virtual neurons in every loop. It is expected that this network consisting of single non linear nodes with delayed feedback outperforms both the single node reservoir and the multiple node swirl reservoir as mentioned in the paper of K. Vandoorne et alii [11].

Table 1: Listing of typical simulation parameters and their corresponding values.

Simulation Parameter	Value
Number of network nodes	81
Semiconductor injection current, delay	187 mA, 6.25 ps
Delay between nodes/ virtual node time window	187 ps
Interconnection weights	$0.25\cdot\exp(\pi/4)$
Number of virtual nodes	3-5 (self-loops with delay 3-5·187 ps)
Integration time step	6.25 ps
Input dimensions (speech channels)	77
Approximate speech signal duration (after speed-up)	100 ns

We list the most important simulation parameters in table 1. They correspond to the optimum values found for the speech recognition task in aforementioned reference [11]. For convenience this paper reports on the spoken digit task (without babble noise) [12], too. Keeping the semiconductor optical amplifier network related parameters fixed, we concentrate on a sweep of both phase and attenuation in the self-loops at different input peak power levels. In a second time, the effect of changing the period of the input mask or the internodal distance was investigated. It is worth noting that we simulated only a few virtual nodes in each self-loop due to the limited space of delay lines on chip. Moreover the same input masking was used for every node. Finally we note that the network was simulated with an in-house circuit simulator [13].

Results for the Spoken Digit Task

Figure 2 shows four different operational regimes with respect to peak input power. At low input powers, corresponding to a linear regime, the optimal word error rate, that is the average number of misclassified digits, is very localised around 90° phase shift in the self-loops. That proves that an optical network exhibiting phase control can boost the overall RC performance. Remarkably, the optimal peak power is around 100 mW; higher input peak powers showing only little change in performance. Consequently, there exists at least one optical amplifier within the network that operates in a strongly saturated regime. Interestingly, we exploit optical non linearity to obtain this result. The best error rate is 11% which is one order of magnitude higher than the result obtained by K. Vandoorne et alii (WER of 0.6%, with relabeling of unbalanced dataset) for the same network size (9x9 nodes).

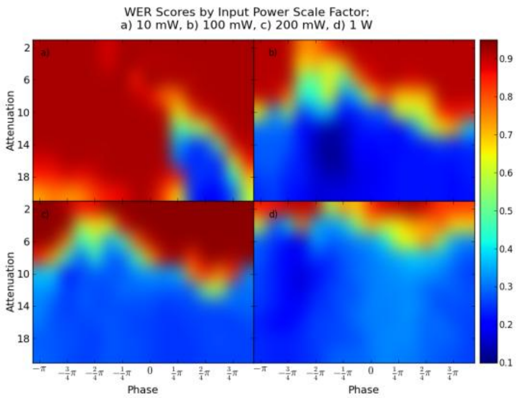

Figure 2: Word error rate for the spoken digit task in the (attenuation, phase) plane for different input peak powers. All the self-loops contain 5 virtual neurons.

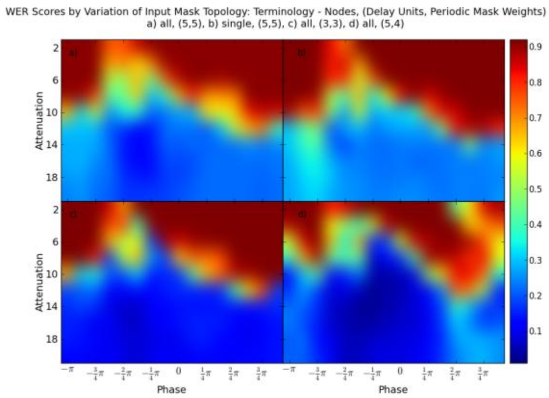

Figure 3: Word error rate for the spoken digit task in the (attenuation, phase) plane at 100 mW input peak power and different numbers of virtual nodes in the self-loops. In d) the masking period is detuned by one virtual neuron step.

This result is disappointing in view of the poor performance, although encouraging with respect to the analogue signal bandwidth which is decreased by a factor of five (when using five virtual nodes). Figure 3 confirms the trend that fewer virtual nodes, e.g. three (WER 7%), rather improve the performance of the reservoir, converging most likely to its best value when the self-loops are completely absent (WER 5.8% without relabeling). According to figure 3, the performance decreases if one applies the input signal to only one node instead of projecting it onto all of them. Five times longer internodal connections lower the error rate to 5% on the cost of increased waveguide losses and poor phase coherence. In addition, the same performance can be achieved with a five times lower analogue signal in a self-loop-free network possessing five times longer node interconnections.

Yet another interesting feature of figure 3 is the fact that detuning the masking period by a time corresponding to one “virtual node window”, the performance is improved to a word error rate of 2%. Because of its additional mixing of virtual nodes, this detuning indicates that the virtual node coupling is in general too weak. Therefore future simulations should focus on the case of shorter virtual neurons which mix stronger under the optical semiconductors' non linearity.

Conclusion

We showed by means of software simulation that a five times lower analogue signal bandwidth is necessary given that a periodic binary masking procedure is applied to the reservoir. The achieved word error rate, however, is increased by one order of magnitude compared to state-of-the-art studies. In this respect, this work is rather a proof of principle and further studies will address the question of decreased performance. First evidence is found that suggests the use of stronger correlated virtual nodes by reducing their time window to values typical of the intrinsic semiconductor timescale. The best word error rate documented in this work amounts to 2% using detuning of the masking period with respect to the number of virtual nodes.

References

[1] H. Jaeger, "The 'echo state' approach to analysing and training recurrent neural networks," in Technical Report GMD Report 148 (German National Research Centre for Information Technology), 2001.

[2] H. Jaeger and H. Haas, "Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication," in Science 2, vol. 304 (5667), pp. 78-80, 2004.

[3] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," in Neural Computation, 14 (11), pp. 2531-2560, 2002.

[4] K. Stanek, "Reservoir computing in financial forecasting with committee methods," M.Sc.Thesis in Engineering, Technical University of Denmark, Kongens Lyngby, Denmark, 2011.

[5] T. Waegeman et al., "Modular reservoir computing networks for imitation learning of multiple robot behaviors," in Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA), 2009.

[6] H. Jaeger, "Short term memory in echo state networks," GMD Report 152 (German National Research Institute for Computer Science), 2001.

[7] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, "Information processing capacity of dynamical systems," in Scientific Reports, Vol. 2, pp. 1-7, 2012.

[8] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," in Nature Communications, Vol. 2, pp. 6, 2011.

[9] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," in Optics Express, Vol. 20, Issue 20, pp. 22783-22795, 2012.

[10] R. Martinenghi, S. Rybalko, M. Jacquot, Y. K. Chembo, and L. Larger, "Photonic Nonlinear Transient Computing with Multiple-DelayWavelength Dynamics," in Physical Review Letters 108, 244101, 2012.

[11] K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, and P. Bienstman, "Parallel reservoir computing using optical amplifiers,", in IEEE Transactions on Neural Networks, Vol. 22(9), pp. 1469-1481, 2011.

[12] Texas Instruments-Developed 46-Word Speaker-Dependent Isolated Word Corpus (TI46), September 1991, NIST Speech Disc 7–1.1 (1 disc), 1991.

[13] M. Fiers et al., "CAPHE : a Circuit-level Time-domain and Frequency-domain Modeling Tool for Nonlinear Optical Components," in Proceedings of the 2011 Annual Symposium of the IEEE Photonics Benelux Chapter, Ed. Peter Bienstman et al., IEEE Photonics Society, pp. 277–280, 2011.