Optoelectronic reservoir computing: tackling noise-induced performance degradation

M. C. Soriano[1], S. Ortín[2], D. Brunner[1], L. Larger[3], C. R. Mirasso[1], I. Fischer[1], L. Pesquera[2]

[1] *Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (UIB-CSIC), Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain*
[2] *Instituto de Física de Cantabria, IFCA (UNICAN-CSIC), Av. Los Castros s/n, E-39005 Santander, Spain*
[3] *UMR CNRS FEMTO-ST 6174/Optics Department, University of Franche-Comté, 16 Route de Gray, 25030 Besançon cedex, France*

miguel@ifisc.uib-csic.es

Abstract: We present improved strategies to perform photonic information processing using an optoelectronic oscillator with delayed feedback. In particular, we study, via numerical simulations and experiments, the influence of a finite signal-to-noise ratio on the computing performance. We illustrate that the performance degradation induced by noise can be compensated for via multi-level pre-processing masks.

OCIS codes: (190.3100, 200.3050, 250.4745)

References and links

1. J. L. O'Brien, "Optical quantum computing," Science **7**, 1567–1570 (2007).
2. H. J. Caulfield and S. Dolev, "Why future supercomputing requires optics," Nature Photon. **4**, 261 (2010).
3. W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," Neural Comput. **14**, 2531–2560 (2002).
4. H. Jaeger and H. Haas, "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication," Science **304**, 78–80 (2004).
5. D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, "An experimental unification of reservoir computing methods," Neural Networks **20**, 391–403 (2007).
6. M. Rabinovich, R. Huerta, and G. Laurent, "Transient dynamics of neural processing," Science **321**, 48–50 (2008).
7. W. Maass, and H. Markram, "On the computational power of recurrent circuits of spiking neurons," J. Comput. Syst. Sci. **69**, 593–616 (2004).
8. K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, and J. Campenhout, "Towards optical signal processing using photonic reservoir computing," Opt. Express **16**, 11182–11192 (2008).
9. K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, and P. Bienstman, "Parallel reservoir computing using optical amplifiers," IEEE Trans. Neural Networks **22**, 1469–1481 (2011).
10. 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," Nature Commun. **2**, 468 (2011).
11. L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer, "Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing," Opt. Exp. **20**, 3241–3249 (2012).
12. Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic Reservoir Computing," Sci. Rep. **2**, 287 (2012).
13. R. Martinenghi, S. Rybalko, M. Jacquot, Y. K. Chembo, and L. Larger, "Photonic Nonlinear Transient Computing with Multiple-Delay Wavelength Dynamics," Phys. Rev Lett. **108**, 244101 (2012).

14. D. Woods and T. J. Naughton, “Photonic neural networks,” Nature Physics **8**, 257–258 (2012).
15. J. P. Crutchfield, L. D. William, and S. Sudeshna, “Introduction to focus issue: Intrinsic and designed computation: Information processing in dynamical systems beyond the digital hegemony,” Chaos **20**, 037101 (2010).
16. J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, “Information Processing Capacity of Dynamical Systems,” Sci. Rep. **2**, 514 (2012).
17. K. Ikeda, “Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system,” Optics Commun. **30**, 257–261 (1979).
18. L. Larger, J.-P. Gordegbuer, and V. S. Udalsov, “Ikeda–based nonlinear delayed dynamics for application to secure optical transmission systems using chaos,” C.R. de Physique **5**, 669–681 (2004).
19. M.C. Soriano, L. Zunino, L. Larger, I. Fischer, and C.R. Mirasso, “Distinguishing fingerprints of hyperchaotic and stochastic dynamics in optical chaos from a delayed opto-electronic oscillator,” Opt. Lett. **36**, 2212–2214 (2011).
20. A. Rodan and P. Tino, “Minimum complexity echo state network,” IEEE Trans. Neural Networks **22**, 131–144 (2011).
21. A. S. Weigend and N. A. Gershenfeld, “Time series prediction: Forecasting the future and understanding the past,” http://www-psych.stanford.edu/ andreas/Time-Series/SantaFe.html (1993).

---

1. Introduction

Photonic-based information processing has been experiencing renewed interest over the last decade following the evolution of photonic technologies and quantum computing [1, 2]. A main issue in the real success of photonic-based information processing is that special-purpose, computationally efficient, optical devices should be presented in terms of their energy costs and applicability to general-purpose computation [2].

Unlike traditional computers, where the processing of information typically is handled in a sequentially, a computational paradigm known as reservoir computing has recently emerged [3, 4, 5]. Reservoir computing (RC) is inspired by the way our brain appears to process information [6]. In conventional RC, an untrained recurrent neural network (RNN) forms the reservoir, which is read out by a simple external classification layer as shown in Figure 1 (left). This recurrent network can perform information processing by using the reservoir’s transient responses induced by an input signal. It has been shown that RC serves universal computational properties; any potential operation can be realized, for certain tasks even outperforming other approaches [3, 7]. Interestingly, most implementations of this concept are being done via computer algorithms and numerical simulations with few attempts to exploit the full potential of this concept by implementing it in hardware [8, 9].

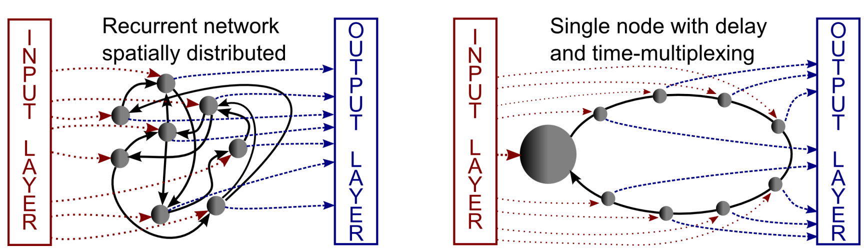

Fig. 1. (Left) Sketch of a traditional RC setup. (Right) RC based on a single nonlinear element subject to delayed feedback.

In the context of RC, it has recently been shown that simple delay systems can replace complex networks without losing functionality [10]. Delay-based architectures reduce the usually required large number of nodes to only few, or even a single nonlinear element with delayed coupling, as shown in Figure 1 (right). The delay line is divided in equidistant virtual nodes, which can be addressed via time multiplexing [10]. Even though the input layer is only connected to the nonlinear node, time multiplexing allows for implicitly accessing each virtual node with a different input weight. Delay systems fulfill the required demands of high-dimensionality and they can be tuned to exhibit fading memory, two essential properties for RC [10]. The suggested approach clearly simplifies the RC concept, opening new ways for high-speed photonics implementations and first implementations are already appearing [11, 12, 13]. A major advantage of these photonics implementations is the use of standard telecommunications hardware [14].

Hardware implementations of optoelectronic reservoir computing have proven to be remarkably successful in several computationally hard tasks such as spoken digit identification and pattern recognition [11, 12]. In particular, the here presented optoelectronic dynamical system is capable of identifying isolated spoken digits with excellent performance (0.2% word error rate) [11]. It has been found that, for RC purposes, it is preferable to bias the system in a stable regime (fixed point of the dynamics) without external input. The addition of an external input induces a complex transient response in the dynamical system, which is employed to perform information processing tasks. In delay-based reservoirs, the input signal is expanded over a time interval of length τ (delay time) and multiplied by an input mask (pre-processing) before it is injected into the nonlinear oscillator [10]. The pre-processing mask serves two purposes: defining the input connectivity weights and keeping the nonlinear node in the transient regime.

In contrast to digital electronic computing [15], our approach computes in an analog fashion by using the amount of light intensity to encode data. In analog optoelectronic computing, the finite signal-to-noise ratio (SNR) of the practical implementation is a major limiting factor [16]. In particular, we have found that the performance of certain computational tasks, e.g. time-series prediction, strongly degrades when the SNR is lowered. Here, we optimize the pre-processing technique and the parameters of our analog optoelectronic system to minimize its noise sensitivity. We elaborate on the robustness of different input masks in the presence of noise and find an optimal parameter range of operation for noise-sensitive tasks.

2. Modeling and numerical predictions

The hardware implementation of the optoelectronic oscillator with delayed feedback is based on an Ikeda-type nonlinearity [17, 18], which has been modified to account for the influence of an external input signal $u_I(s)$. This nonlinear oscillator can be described, in the presence of an external input, by the following dynamical equation :

$$\dot{x}(s) = -x(s) + \beta\,(\sin^2[x(s-\tau) + \gamma u_I(s) + \Phi] - 0.5), \tag{1}$$

where s is the time in normalized units ($s = t/T_R$), with T_R being the oscillator response time. Parameter β is the nonlinearity strength, the delay time is denoted as τ, and parameters Φ and γ are the offset phase of the nonlinearity and the input scaling, respectively. The external input signal $u_I(s)$ is introduced as a modulation to the nonlinearity.

A detailed numerical study allows for the identification of the best parameters of this optoelectronic system with delay for RC purposes. So far the influence of a finite signal-to-noise ratio in the hardware implementation has not been evaluated in detail, as it was not required for the proof-of-principle of the concept. Several noise sources can be present in the different layers of the information processing system shown in Fig. 1 (right). In particular, noise can appear in the reservoir itself, and/or the input and output layers. An analysis of the noise sources affecting the signal recorded from this optoelectronic system, including detection, showed that quantization noise in the acquisition procedure, i.e. output layer, was the strongest stochastic contribution to the signal [19]. The noise in the output layer originates from the quantization of the oscillator's output in the analog-to-digital conversion (ADC), which acts as an interface between the analog and digital worlds. In order to evaluate the influence of noise, we consider a 10 bits quantization in our numerical analysis to mimic the experimental conditions.

Primarily, we would like to highlight the influence of the limited SNR on the choice of the pre-processing mask between input and delay reservoir. In standard RC, the weights connecting the input and the reservoir are usually generated randomly from a uniform distribution over a given interval. In contrast, simple implementations of RC demonstrated that it is sufficient to consider a single absolute value for the input weights to the reservoir with aperiodic pattern of input signs, i.e. a binary mask [20]. According to this insight, a binary mask with input weights $\{1,-1\}$ was chosen on the first hardware implementation of RC with a single element with delay [10]. Here, we find that the quantization error has a major impact on the choice of the input mask. In particular, a binary distribution of input weights cannot compensate for the negative effect of a 10 bits acquisition. As it will be shown below, we have found that a six-valued input mask with weights $\{1.5, 0.9, 0.3, -0.3, -0.9, -1.5\}$, aperiodically distributed, yields a good compromise between performance and applicability. For a proper comparison, we employ two- and six-valued pre-processing masks with zero mean and unity standard deviation.

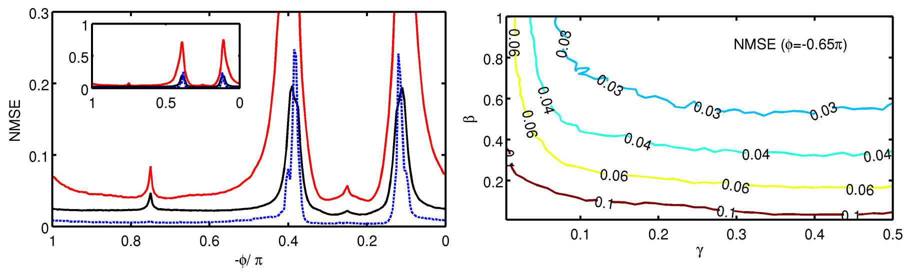

Fig. 2. (Left) NMSE test prediction error in the Santa-Fe time series prediction for $\beta = 0.8$ and $\gamma = 0.45$ as a function of the offset phase Φ. Red (blue) line corresponds to a binary mask in the presence (absence) of quantization noise. Black line corresponds to a six-valued mask in the presence of quantization noise. (Right) Colour-coded NMSE prediction error in the $\beta-\gamma$ plane for $\Phi = -0.65\pi$ and a six-valued input mask.

As an example, we compute the performance of the delay reservoir for a demanding time-series prediction task, using the experimental data of Lorenz-like laser chaos (Santa Fe test) as input [21]. We consider a delay reservoir with 400 virtual nodes, i.e. the virtual node spacing is $\tau/400 \sim 0.22T_R$, with $\tau = 21$ μs and $T_R = 240$ ns [11]. The input consists of 4000 samples of the Santa Fe dataset and the (off-line) training procedure is a standard linear regression with 4-fold cross-validation, i.e. 3000 samples are used for training and 1000 for testing.

Figure 2 (left) shows the normalized mean square error (NMSE) for the Santa-Fe time series prediction with $\beta = 0.8$ and $\gamma = 0.45$ as a function of the offset phase of the nonlinearity Φ when two-valued and six-valued input masks are implemented. In the absence of noise, the binary mask (blue dashed line) yields a minimum prediction error of about 1%. In the presence of noise, however, the prediction error for a six-valued mask (solid black line) is significantly lower than for the binary mask (solid red line), over the entire parameter range, with a minimum error of about 2%. We have confirmed that increasing the number of discrete values in the input mask no longer improves the performance for the given SNR. Figure 2 (right) presents a summary of the numerical predictions for the NMSE prediction error around an optimum region of operation ($\Phi = -0.65\pi$) in the $\beta-\gamma$ plane for a six-valued input mask. An extended region of low prediction error (NMSE$<$3%) can be identified at the upper right part of the $\beta-\gamma$ plane. The NMSE rapidly degrades for $\beta > 1$.

Previous experimental results on this system reported a NMSE$\sim$12% for this task [11]. The numerical simulations suggest that the NMSE can be reduced down to 2% with the six-valued mask and optimal β-γ values.

3. Experimental setup and validation

Our hardware implementation of the RC concept is based on an optoelectronic oscillator with delay, which is depicted in Figure 3. The experimental implementation is composed of a semiconductor laser diode, an integrated optics Mach-Zehnder modulator (MZM) performing a sine squared non-linear transformation, a fiber delay line, and an optoelectronic feedback for intensity detection, filtering and amplification. The feedback signal serves as the drive of the MZ modulator, closing the delayed feedback loop (delay time 21 μs). This optoelectronic oscillator exhibits the dynamical regimes typically observed for the Ikeda dynamics, including a period doubling route to chaos [18].

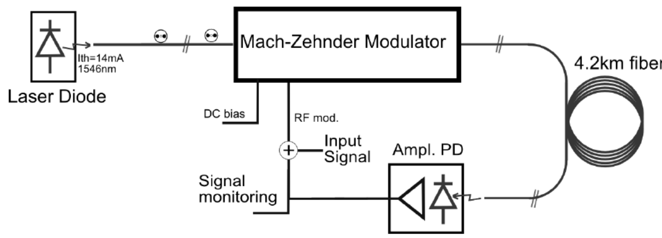

Fig. 3. Experimental implementation of the optoelectronic oscillator with delay.

The chosen operating point of the reservoir in the absence of input is a fixed point. The external input then induces a transient response on the reservoir. This transient response is expected to exploit the full bandwidth of the system, which is measured to be in the MHz range [19]. Therefore, this hardware implementation of a photonic realization of reservoir computing with an optoelectronic oscillator allows for information processing at Mbytes/s rates. However, all-optical photonic implementations and/or high-speed electronic components can eventually push the information processing speed towards the Gbytes/s range.

This optoelectronic set-up is a versatile system, in which the properties of the nonlinearity can be easily tuned. The operating point can be shifted along the nonlinearity by tuning the MZM bias. In Figure 4 (a), we show the tuning of the nonlinearity (dotted line) and the operating point (solid line) when the MZM voltage, i.e. offset phase of the nonlinearity, is varied for $\beta = 0.8$ and $\gamma = 0$. The operating points have qualitatively different properties at the positive and negative slopes of the nonlinearity.

We evaluate the performance of the practical implementation for a time-series prediction task using the experimental data of Lorenz-like laser chaos (Santa Fe test) [21] as input. Figure 4 (b) shows the performance of the system as a function of the operating point for two-valued and six-valued pre-processing masks. The dependence of the prediction error on the offset phase in the experiment agrees with the numerical findings reported in Fig. 2 (left). First, the six-valued mask (solid line) performs significantly better than the two-valued mask (dashed line). Second, there is a clear interdependence between the properties of the fixed points and the performance. The prediction errors are larger when the system operates around the extreme values of the fixed points curve. In addition, the inflection point of the fixed points curve should also be avoided.

The quantization error can be decreased by averaging and oversampling the signal if the acquisition is significantly faster than the oscillator's response. Alternatively, it can also be decreased by recording several times the oscillator's response to the same input, and subsequent averaging. As shown in Fig. 4 (c), an even lower prediction error (NMSE) of 2% can be found for an improved parameter scan around the best operating range when the signal is detected with five times oversampling and averaging. This error is comparable to a standard reservoir

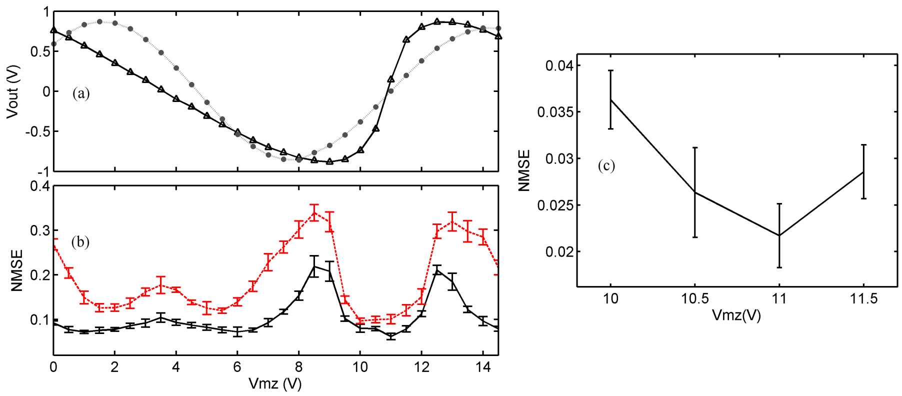

Fig. 4. (a) Experimentally recorded nonlinearity (dotted line) and operating point (solid line with triangles) as a function of the Mach-Zehnder bias voltage (offset phase). (b) Test prediction error (NMSE) for the Santa Fe time-series prediction task with 400 virtual nodes for two-valued (dashed line) and six-valued (solid line) input masks. (c) NMSE of the predicted time-series for an improved detection with 5:1 oversampling and six-valued mask.

with 50 nonlinear nodes [20]. Therefore, an excellent performance is obtained despite the finite SNR of the detection apparatus, which is estimated to be equivalent to a 10 bits dynamic range. The 10 bits dynamic range stems from the 8 bits digitization of the ADC, together with the 5:1 oversampling and smooth averaging.

4. Conclusion

Our results show that the performance degradation by noise can be drastically reduced optimizing the pre-processing technique. The prediction errors for the time-series prediction task reported in this manuscript are comparable or even better than current state-of-the-art approaches [20, 21]. In particular, the NMSE for the optoelectronic oscillator with feedback in the Santa-Fe time-series prediction task is lowered from 12% to 2%. The reasons for this improvement are twofold. First, we find an interdependence between the number of discrete amplitude levels chosen in the input pre-processing mask and the finite Signal-to-Noise ratio of the experimental implementation. The use of a six-valued mask yields the lowest prediction errors for a time-series prediction task in the presence of 10 bits quantization in the output layer. Second, the experiments are carried out with optimized conditions with respect to the properties of the nonlinearity, which can be estimated via numerical simulations. This approach illustrates and highlights that there might be even more potential to improve the performance of delay-based photonics RC using smart approaches like the one presented here.

Acknowledgements

The authors would like to thank members of the European Project PHOCUS for their fruitful suggestions. We acknowledge the financial support of the European Commission with projects No:275840 (NOVALIS) and FP7-ICT-2009-C No:240763 (PHOCUS). We also acknowledge for funding from the Spanish Ministry and Feder through projects FIS2007-60327 (FISICOS) and TEC2009-14101 (DeCodiCa).