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] **and 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.

© 2013 Optical Society of America

OCIS codes: (190.3100) Instabilities and chaos; (200.3050) Information processing; (250.4745) Optical processing devices.

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," Nat. Photonics **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. Express **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 Phys. **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. Goedgebuer, and J. M. Merolla, "Chaotic oscillator in wavelength: a new setup for investigating differential difference equations describing nonlinear dynamics," IEEE J. Quantum Electron. **34**, 594–601 (1998).
19. L. Larger, J.-P. Goedgebuer, 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).
20. Y. Kouomou Chembo, L. Larger, H. Tavernier, R. Bendoula, E. Rubiola and P. Colet, "Dynamic instabilities of microwaves generated with optoelectronic oscillators," Opt. Lett. **32**, 2571–2573 (2007).
21. 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).
22. A. Rodan and P. Tiňo, "Minimum complexity echo state network," IEEE Trans. Neural Networks **22**, 131–144 (2011).
23. U. Huebner, N. B. Abraham, and C. O. Weiss, "Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH_3 laser", Phys. Rev. A **40**, 6354–6365 (1989).
24. 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).
25. L. Cao, "Support vector machines experts for time series forecasting", Neurocomputing **51**, 321–339 (2003).

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 sequentially, a computational paradigm known as reservoir computing has recently emerged [3–5]. 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 Fig. 1(a). 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].

In the context of RC, it has recently been shown that simple delay systems can replace complex networks without losing functionality [10]. Delay systems fulfill the required demands of high-dimensionality and they can be tuned to exhibit fading memory, two essential properties for RC. 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 Fig. 1(b). In this simple scheme, the delay line is then 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.

The suggested approach clearly simplifies the RC concept, opening new ways for high-speed photonics implementations and first implementations are already appearing [11–13]. A major
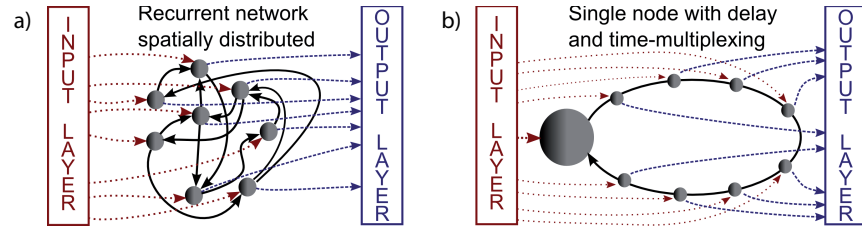
Fig. 1. (a) Sketch of a traditional RC setup. (b) RC based on a single nonlinear element subject to delayed feedback.

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 improve 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. Optoelectronic feedback scheme

Our hardware implementation of the RC concept is based on an optoelectronic oscillator with delayed feedback. The experimental setup, which is depicted in Fig. 2, consists of a well-known optoelectronic implementation of the (low-pass) Ikeda delay dynamics [17–19] (same setup as in [11]). The hardware implementation includes an optical and electronic delayed feedback loop, in which a $\sin^2$ nonlinear transformation is provided by a Mach-Zehnder modulator (MZM). The output of the MZM is delayed via a 4 km long optical fiber and, after optical detection, acts as a nonlinear driving force term at the input of a first order low pass filter with a cut-off frequency of $f_c = 1/2\pi T_R \sim 663$ KHz (DC preserving feedback). The overall gain of the nonlinear feedback plays the role of the bifurcation parameter β, which can be easily tuned by the light intensity level, seeding optically the MZM. For RC purposes, the nonlinearity gain is adjusted such that the steady state is stable when the oscillator is autonomous (no external information added). The electronic feedback signal serves as the radio-frequency drive of the MZM, closing the delayed feedback loop (delay time 20.82 μs). This optoelectronic oscillator exhibits the dynamical regimes typically observed for the Ikeda dynamics, including a period doubling route to chaos [19]. The signal, X(t) in Fig. 2, is detected via an 8-bit real-time digital

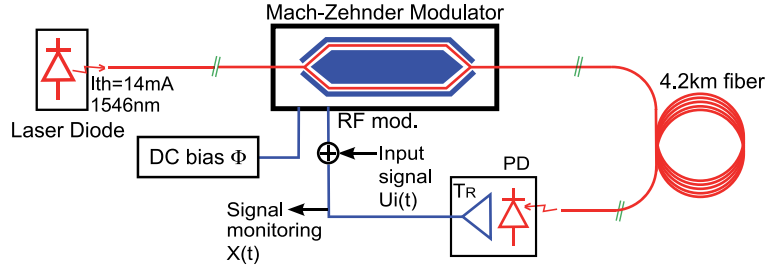oscilloscope for further analysis.

Fig. 2. Experimental implementation of the optoelectronic oscillator with delay.

The nonlinear optoelectronic oscillator can be described, in the presence of an external input signal $u_I(s)$, by the following dynamical equation [11, 18]:

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

where x is a dimensionless dynamical variable that corresponds to the voltage applied to the MZM by the low pass filter (X in Fig. 2) with the following renormalization $x = \pi \frac{X}{2V_\pi}$, where $V_\pi \sim 6.5$V is the voltage required to modulate the MZM over one π-range. In this manner, x is normalized such that it corresponds to the unit free argument of the $\sin^2$ nonlinearity. The time in normalized units is s ($s = t/T_R$), with $T_R = 240$ ns being the oscillator response time. Parameter β is the nonlinearity gain (total gain in the oscillation loop), and the delay time in normalized units is denoted as τ. The delay time in the experimental realization is $\tau' = 20.82$ μs, i.e. $\tau = \tau'/T_R = 86.75$ in normalized time units. Here, we consider a delay reservoir with 400 virtual nodes, i.e. the virtual node spacing is $\tau/400 \sim 0.2$ as suggested in [10]. 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, corresponding to the normalized input signal Ui(t) in Fig. 2. In order to keep a consistent time reference, $u_I(s)$ is also delayed in Eq. (1). The offset phase Φ can be tuned to adjust the rest state (stable steady state) around which the external input information is generating nonlinear transients, via modulation of the MZM voltage.

The proper control of the nonlinearity gain β is of particular importance for the dynamical characterization of the optoelectronic oscillator in the absence of external input, with β being typically employed as the bifurcation parameter in many experimental studies. In particular, this system starts to oscillate when $\beta > 1$ [18–20]. For $\beta < 1$, the system is stable (DC voltage) and the steady state of x depends on the offset phase Φ. The nonlinearity gain can be controlled in the experiments by varying the power emitted by the laser diode in Fig. 2. Assuming a linear dependence of the power with the injection current, we can define the nonlinearity gain as follows:

$$\beta = \frac{I - I_{th}}{I_{\beta_1} - I_{th}}, \tag{2}$$

where I is the laser injection current, $I_{th} = 14$ mA is the laser threshold current and $I_{\beta_1} = 88$ mA is the current at which the system starts to oscillate ($\beta = 1$).

3. Multi-valued masking scheme

A detailed numerical study allows for the identification of the best parameters of the optoelectronic system with delay for RC purposes in the presence of noise. 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(b). 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 [21]. 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 quantization noise in our numerical analysis to mimic the experimental conditions, where analog-to-digital conversion is so far unavoidable. The maximum SNR in the presence of quantization noise is given by the number of quantization bits Q in the ADC, and it can be calculated from $\text{SNR}_{max} \sim 6.02Q$ dB for a uniform distribution of the acquisition signal. We have checked numerically that our results on the importance of the input pre-processing also apply when a noise term is added to Eq. (1).

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. The function of the input pre-processing mask is to create transient responses in the system, which are then used for information processing. In standard RC, the weights connecting the input and the reservoir are usually drawn 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 [22]. 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]. Figure 3(a) shows an example of input samples $\{0.5, 0.25, 0.75\}$ expanded for a delay line with 10 virtual nodes and a two-valued pre-processing mask. Here, we find that the choice of the input mask has a major impact on the influence of the quantization error. In particular, input weights randomly drawn from a binary distribution cannot compensate for the negative effect of a quantized acquisition. As it will be shown below, we have found that a six-valued input mask with weights randomly drawn from a distribution with six discrete levels $\{1.5, 0.9, 0.3, -0.3, -0.9, -1.5\}$ yields a good compromise between performance and applicability in the case of quantized acquisition. This additional discrete levels, equally distributed, induce significantly different transient responses in the reservoir that can be optimally used for information processing. However, in the presence of quantization noise, it has been tested that the inclusion of more than six, equally distributed, levels does not longer improve the performance. Figure 3(b) shows an example of input samples $\{0.5, 0.25, 0.75\}$ expanded for a delay line with 10 virtual nodes and a six-valued pre-processing mask. For a proper comparison, we employ two- and six-valued pre-processing masks with zero mean and unity standard deviation.

For the illustration of the influence of the quantization noise on the performance of the delay reservoir, we choose a demanding time-series prediction task as a benchmark. The time-series that we use in this prediction task was recorded from a far-infrared laser operating in a (Lorenz-like) chaotic regime [23]. This time-series is one of the six data sets that were used in the "Time Series Prediction Competition" sponsored by the Santa Fe Institute, initiated by Neil Gershenfeld and Andreas Weigend in the early 90's [24]. We have chosen this particular time-series as it is a widely used benchmark test in the machine learning community [22,25]. In time series prediction, the goal is to predict the future value of the time series, $u(t+1)$, based on its previous values up to time t. To this end, we sequentially feed the RC system with the input stream $u(t), u(t-1), u(t-2)...u(t-n)$, already pre-processed with the input mask, and we try to predict the value one time step in the future, $u(t+1)$. In particular, the time-series that we
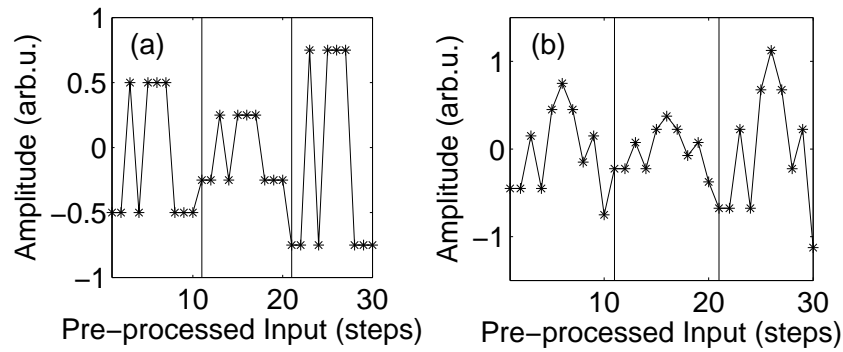
Fig. 3. Example of pre-processing for input samples {0.5, 0.25, 0.75} expanded for a delay line with 10 virtual nodes in the case of (a) a two-valued mask and (b) a six-valued mask. The vertical lines indicate the start of a new input sample.

use for the prediction task consists of 4000 samples of the Santa Fe laser dataset normalized to zero mean and unit variance. Out of the Santa Fe laser time-series, 75% of the samples (3000) are used for training and the remaining 25% (1000) are used for the prediction and the evaluation of the prediction error (testing). The training procedure, which is carried out off-line, consists of a standard multiple linear regression. The independent variables for the regression are the responses of the oscillator to the pre-processed inputs, sampled at the virtual node positions. The corresponding target values are the inputs samples one time step in the future. In the training procedure, an output weight is assigned to each virtual node, such that the weighted sum of all the virtual nodes values approximates the desired target value as closely as possible. The weights obtained in the training procedure are later used to test the prediction error with the remaining (untrained) input samples. Four different random partitions of the original data are evaluated for training and testing in order to obtain a statistically significant prediction error.
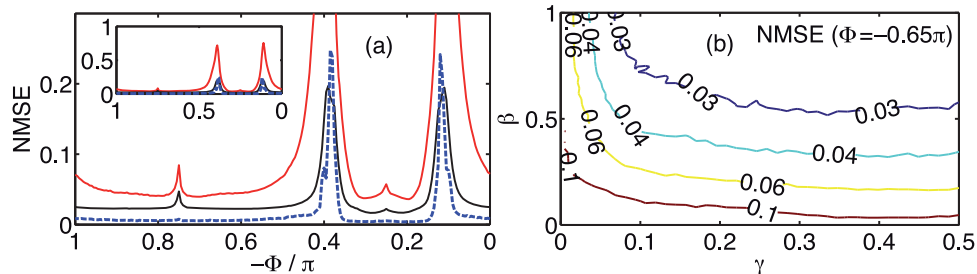

Fig. 4. (a) NMSE test prediction error in the Santa Fe laser time-series prediction task for $\beta = 0.8$ and $\gamma = 0.45$ as a function of the offset phase Φ. The red (blue) line corresponds to a binary mask in the presence (absence) of 10 bits quantization noise. The black line corresponds to a six-valued mask in the presence of 10 bits quantization noise. (b) Colour-coded NMSE prediction error in the $\beta - \gamma$ plane for $\Phi = -0.65\pi$ and a six-valued input mask.

Figure 4(a) shows the normalized mean square error (NMSE) for the time-series prediction task 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 10 bits quantization 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 checked that increasing the number of discrete values in the input mask does not improve the performance for the given SNR further. Figure 4(b) presents a summary of the numerical results 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 $\beta > 1$.

We have investigated the influence of the number of quantization bits on the performance degradation in more detail. Figure 5 shows the NMSE prediction error for ten random realizations of the two- and six-valued input pre-processing masks, respectively, as a function of the number of quantization bits. The prediction errors are consistently lower for the six-valued masks compared to the two-valued masks over the whole range of quantization bits. The difference between the NMSE obtained with the two-valued and six-valued masks decreases as the number of quantification bits increases. For infinite precision, the two types of masks yield the same NMSE for the Santa Fe time-series prediction task.
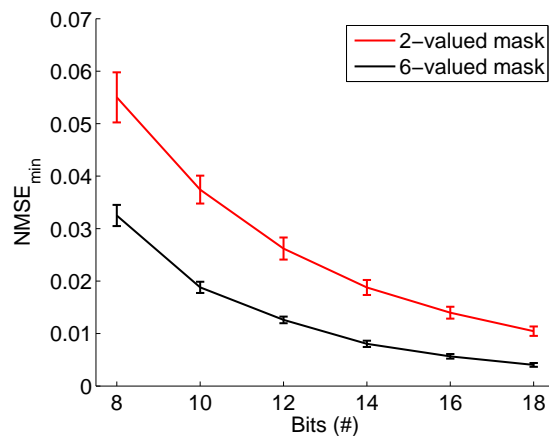

Fig. 5. Minimum NMSE test prediction error in the Santa Fe laser time-series prediction task for $\beta = 0.8$ and $\gamma = 0.45$ as a function of the number of quantization bits. The red (black) line corresponds to a binary (six-valued) mask. The error bars correspond to ten different random realizations of the masks.

Previous experimental results on the optoelectronic system reported a NMSE~12% for the Santa Fe laser time-series prediction 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 for 10 quantization bits. It is worth noting that it is possible to isolate the individual effects of each optimizing step via the numerical simulations (see Fig. 5). First of all, the numerical parameter scan allows for the optimization of the phase offset (Φ), the nonlinearity gain (β) and the input scaling (γ), yielding a prediction error ~6% with a two-valued mask and 8 bits quantization. Second, the use of a six-valued mask further reduces the error from ~6% to ~3% with 8 bits quantization. Finally, an additional increase in the SNR from 8 to 10 quantization bits, which can be obtained e.g. from oversampling and subsequent averaging, results in a decrease from ~3% to ~2% in the prediction error, as shown in Fig. 5.

4. Experimental evaluation

The chosen operating point of the reservoir in the absence of input is a fixed point, i.e. a constant DC voltage. 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 in the MHz range [21]. Therefore, our hardware implementation of a photonic realization of reservoir computing with an optoelectronic oscillator allows for information processing at Mbytes/s rates. All-optical photonic implementations and/or high-speed electronic components can eventually push the information processing speed towards the Gbytes/s range.

As described in Section 2, the optoelectronic set-up (see Fig. 2) 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 Fig. 6(a), we show the tuning of the nonlinearity (dotted line) and the operating point (solid line) when the MZM DC bias, i.e. offset phase of the nonlinearity, is varied for $\beta = 0.8$ and $\gamma = 0$.
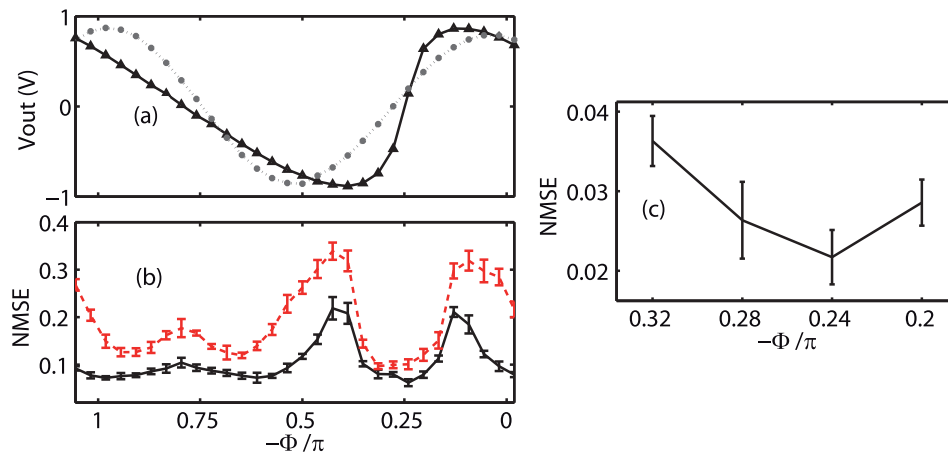

Fig. 6. (a) Experimentally recorded nonlinearity (dotted line) and operating point (solid line with triangles) as a function of the Mach-Zehnder offset phase for $\beta = 0.8$. (b) Test prediction error (NMSE) for the Santa Fe laser time-series prediction task with 400 virtual nodes for two-valued (dashed line) and six-valued (solid line) input masks ($\gamma = 0.45$). (c) NMSE of the predicted time-series for an improved detection with 5:1 oversampling and subsequent averaging obtained for the six-valued mask.

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) [23, 24] as input. In our experimental implementation, we feed the system with the pre-processed input samples. The pre-processing of the input samples, in which each input sample is multiplied by the input mask and expanded over one delay time, is carried out off-line. Figure 6(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. 4(a). 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 position of the operating points (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. In Fig. 6(b), the minimum prediction error achieved is 6% for the six-valued mask and 10% for the two-valued mask, respectively. These prediction errors are slightly higher than

the numerical results for 8 bits digitization of the ADC.

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. 6(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 the error obtained with a numerical noise-free implementation of standard reservoir computing with 50 nonlinear nodes [22]. 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.

5.   Conclusion

Our results show that the performance degradation by noise can be drastically reduced by improving 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 numerical and noise-free approaches [22, 24, 25]. In particular, the NMSE for the optoelectronic oscillator with feedback in the Santa Fe laser time-series prediction task is lowered to 2% for 10 bits quantization in the output layer, reducing significantly the 12% prediction error previously reported in this system [11]. The reasons for this improvement are twofold. First, the experiments are carried out with optimized conditions with respect to the properties of the nonlinearity, which can be estimated via numerical simulations. Second, we find that the use of a six-valued mask yields prediction errors significantly lower than a two-valued mask for a time-series prediction task in the presence of quantization. A two-valued (binary) mask is, for infinite precision, sufficient to induce the required diversity in the transient responses for a given input. However, we find that, for finite precision, the transients induced by the binary mask cannot be optimally used for information processing. We speculate that some of these transients become indistinguishable due to noise. In the presence of 10 bits quantization noise, a six-valued mask yields better results via inducing significantly different transient responses that can be optimally used for information processing. This approach illustrates and highlights that there is potential to further improve the performance of delay-based photonics RC.