



Reservoir computing system with double optoelectronic feedback loops

YAPING CHEN, LILIN YI,* JUNXIANG KE, ZHAO YANG, YUNPENG YANG, LUYAO HUANG, QUNBI ZHUGE, AND WEISHENG HU

State Key Lab of Advanced Communication Systems and Network, Shanghai Institute for Advanced Communication and Data Science, Shanghai Jiao Tong University, Shanghai, 200240, China

*lilinyi@sjtu.edu.cn

Abstract: Reservoir computing (RC) by supervised training, a bio-inspired paradigm, is gaining popularity for processing time-dependent data. Compared to conventional recurrent neural networks, RC is facilely implemented by available hardware and overcomes some obstacles in training period, such as slow convergence and local optimum. In this paper, we propose and characterize a novel reservoir computing system based on a semiconductor laser with double optoelectronic feedback loops. This system shows obvious improvement on prediction, speech recognition and nonlinear channel equalization compared to the traditional reservoir computing systems with single feedback loop. Then some influencing factors to optimize the performance of the new RC are numerically studied, and its great potential of addressing more complex and troubling problems in information processing is expected to be exploited.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

In today's internet era, the surge of informatization is sweeping around the world, in the sense that enormous amounts of heterogeneous data are subsequently accumulating explosively. Consequently, artificial neuron network (ANN) has been penetrating into every corner of social life unprecedentedly [1]. Generally speaking, recurrent neural network (RNN) has attracted great attentions as one of ANNs, and been widely implemented in varieties of complex tasks [2–5]. It must be noted that, however, RNN reserves some intrinsic drawbacks of local optimum, slow convergence, difficulties to learn long-range temporal dependencies due to vanishing or exploding gradient estimates, and etc.

Afterwards, a random recurrent network untrained in connection weights and processed by a simple classification/regression technique was first proposed by Jaeger as the Echo State Network (ESN) [6] and by Maass as the Liquid State Machine (LSM) [7]. In 2007, David Verstraeten et al. [8] proved that ESN and LSM are essentially identical and they named it as Reservoir Computing (RC). RC has many conspicuous advantages by the virtue of unique structure. It is unnecessary for RC to train in connection weights. Only you need a simple classification or regression technique for RC. And it is suitable for different kinds of tasks and apt to implementation hardware. RC as a new class of RNNs for processing information has accomplished a very large range of tasks, such as time series prediction, radar signal prediction, speech recognition, nonlinear channel equalization, handwritten numeral recognition and optical header recognition [9–14]. RC could be classified into electronic RC and photonic RC. Electronic RC proposed in [15] performs efficient information processing. Kristof Vandoorne et al. [16] first proposed photonic RC for optical signal processing in the context of large-scale pattern recognition problems. Subsequently, hardware implementations of photonic RC with error rates comparable to the state-of-the-art digital algorithms have been made as another breakthrough in optical information processing [17, 18]. Optoelectronic RC and all-optical RC have been developing rapidly in different structures such as optoelectronic systems based on Ikeda model [17, 19–21] and all-optical systems based on saturation of a semiconductor optical

amplifier as a nonlinear element [18], saturable absorption of a semiconductor mirror [22], InGaAsP micro-ring-resonators [23], a semiconductor ring laser of light in two directional modes with the same wavelength [24] and a semiconductor laser (SL) with double optical feedback and optical injection [25].

Despite there are intensive studies on RC as a nonlinear dynamical system with single nonlinear node and a delay line, the potential of RC has not yet been fully developed. In what follows we explore its potential from another perspective. For the first time, we propose a new RC structure with dual nonlinear nodes and double optoelectronic feedback loops. We study the performance of the new RC structure on three tasks that are widely considered in the reservoir computing community, namely the nonlinear autoregressive moving average (NARMA10) task, isolated spoken digits recognition and nonlinear channel equalization. The novel RC shows better performance than the previous ones with single nonlinear node and feedback loop. In addition, we study the influences of some typical parameters: feedback strength, length of feedback loop, and offset phase. Finally, we discuss the implications of our work for the future development of photonic reservoir computing.

2. Theory and system model

Traditional RC generally consists of three layers: an input layer, a reservoir and an output layer as depicted in Fig. 1(a). The input layer feeds the discrete input signals to the reservoir via fixed random weight connections called the input mask. The reservoir is composed of a certain quantity of nonlinear nodes randomly interconnected forming feedback loops. Once the input signals pre-processed by a mask signal come in, the reservoir produces internal variables called reservoir states. The input mask enriches the dynamics of the reservoir by breaking the symmetry that would occur if the same part of the input signals would be distributed to all the internal variables. Then the output layer linearly combines the reservoir states to produce the actual outputs. The desired output written into the output layer during training stage is often called teacher forcing signal. At last, training RC is just a simple linear regression task, and numerous batches or adaptive online algorithms are available. Photonic RC is also composed of three layers as shown in Fig. 1(b). Although it is simplified as a nonlinear node and feedback loop, its performances are comparable to the traditional one. And the principle is basically the same as the traditional RC.

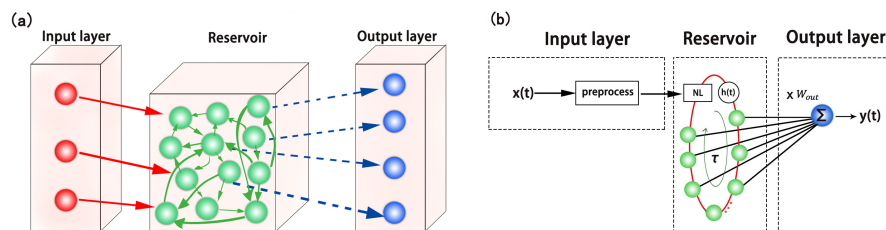


Fig. 1. (a) Schematic diagrams of traditional RC scheme. (b) Schematic diagrams of photonic RC scheme.

It is theoretically proved that the three-layer neural network with a hidden layer can remarkably approximate arbitrary continuous function infinitely [26,27]. Nevertheless, this kind of flat neural network with hundreds of nonlinear nodes in the hidden layer can hardly obtain satisfactory precision in tackling larger and more complex tasks in practice. Fortunately, a network with more hidden layers is of great value to approximate complex functions with much higher precision. And the more complex nonlinear nodes the RC has, the more conducive it is. The structure

with appropriate complexity takes data into the higher dimensions to process and the interaction between two feedback loops helps the system to deal with input signals. From the viewpoint of NN, in space, the single feedback-loop RC can be considered as a hidden layer. We introduce another nonlinear node and feedback-loop as a new hidden layer as shown in Fig. 2. So the novel RC we propose can be regarded as a NN with two hidden layers.

In this section, we analyze the theoretical model of the proposed RC with dual nonlinear nodes and double feedback loops. As illustrated in Fig. 2(a), the system framework includes a popular optoelectronic delay dynamic Ikeda model. We transform the structure diagram into the schematic diagram as described in Fig. 2(b). In both figures, the red lines represent the optical signal transmission, and the blue ones represent the electrical signal transmission. Figure 2(b) shows the reservoir states by some balls (green balls represent the state in the short feedback loop, red ones in the difference between two feedback loops) based on Fig. 2(a). In Fig. 2(b), another nonlinear node and delayed feedback loop are added and the structure serves as an optoelectronic oscillator with two optoelectronic delayed feedback loops, i.e. short feedback loop(τ_1) and long feedback loop(τ_2). $\Delta\tau$ is the difference between the two feedback loops i.e. $\Delta\tau = \tau_2 - \tau_1$. Two off-the-shelf voltages driven Mach-Zehnder Modulators (MZMs) are adopted to provide \cos^2 nonlinearity, placed at the output of a SL. The process of the input signal multiplied by the mask signal in the input layer is described in Fig. 2(c). The input signal $u(t)$ is sampled and each sample operates for a time interval of length T . Even if the input signal is time-discrete, similar operations should also be carried on. Before being injected into the SL, the piecewise input signal $u(t)$, $u(t) = u(n)$, $nT \leq t < (n+1)T$, is multiplied by an input mask signal $m(t)$ for preprocessing which is periodic with a period of T , i.e. $m(t+T) = m(t)$. The mask signal plays a significant role in defining the input connectivity weights and keeping the nonlinear nodes in the transient regime to obtain diverse transient responses to a data input. The period T is divided into N segments called virtual nodes, and each of duration θ : $\theta = T/N$, is the time span between two reservoir states $x(n)$ or two virtual nodes in the feedback loops. Proved in [17], when $T = N\tau_1/(N+k)$ where $1 \leq k < N$, we are in the unsynchronized regime where the reservoir states correspond to several dependent reservoirs with a fraction of neurons. In this regime, reservoir has rich dynamics. Therefore, we take $k = 1$, i.e. $\tau_1 = \theta + T$ in our work to get as rich dynamics as possible. Then $x(n-1)$ is adjacent to an internal variable of $x(n)$. Bias is added to the mask signal for different tasks to change the variability of the individual node's dynamics. In the reservoir, reservoir states are collected in from the short feedback loop. And several virtual nodes are kept for a while in the long feedback loop which can be adjusted to optimize the RC. In the output layer, the actual output is obtained by taking a linear combination of reservoir states $x_i(n)$: $y(n) = \sum_{i=0}^{N-1} W_{out} x_i(n)$, where the readout weights W_{out} are trained by a ridge regression algorithm in the training phase and fixed in the test phase. The ridge regularization parameter is used to make the RC more robust against overfitting. According to the above theories and the model in [13, 14], the nonlinear dynamics of optoelectronic feedback system with two optoelectronic feedback loops can be modeled by the following equation:

$$\tau \frac{dx(t)}{dt} + x(t) + \frac{1}{\sigma} \int_{t_0}^t x(\varepsilon) d\varepsilon = \beta \cos^2[\alpha p x(t - \tau_2) + \phi_2] * \cos^2[(1-p)x(t - \tau_1) + \phi_1 + s(t)] \quad (1)$$

where $x(t) = \pi V(t)/2V_\pi$ is the normalized bias voltage of MZM, $V(t)$ stands for the amplifier's output voltage, and V_π represents the half-wave voltage of the MZM. $\sigma = 1/(2\pi f_L)$ and $\tau = 1/(2\pi f_H)$ respectively represent the characteristic time scale of low-frequency cutoff and high-frequency cutoff, and f_L and f_H stand for cutoff frequencies of the low-frequency and high-frequency for the model, respectively. β denotes the feedback strength. α is the scaling factor to keep the power of variables of the long feedback loop. p stands for the distribution ratio

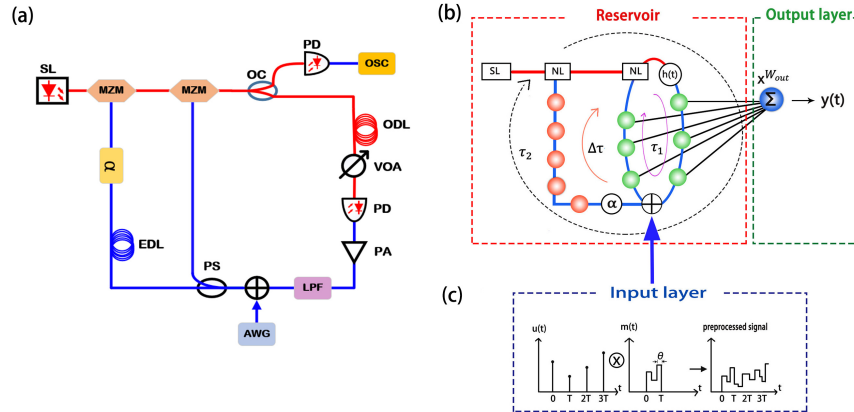


Fig. 2. The novel RC with double optoelectronic feedback loops. (a) Structure of the novel RC based on a SL with double feedback loops. SL, semiconductor laser; MZM, Mach-Zehnder modulator; OC, optical coupler; PD, photoelectric detector; OSC, oscilloscope; ODL, optical delay time; VOA, optical attenuator; PA, power amplifier; LPF, low pass filter; AWG, arbitrary waveform generator; PS, power splitter; EDL, electrical delay line. (b) Schematic of the novel RC based on a SL with two feedback loops. NL denotes the nonlinear transformation for the system. $h(t)$ stands for the system's impulse response. (c) The process of input signal multiplied by mask signal in the input layer.

of the coupler. ϕ_1 and ϕ_2 are the offset phase of two MZMs. τ_1 and τ_2 respectively stand for the short feedback time delay and the long feedback time delay. In addition, $s(t)$ stands for the input signal. In the numerical simulation, the parameters are selected as follows: $\tau = 19.89$ ps, $\sigma = 51.34$ ps, $V_\pi = 5$ V. We introduce $y = \int_{t_0}^t x(\varepsilon) d\varepsilon$ so that the system can be described by

$$\begin{cases} \frac{dx(t)}{dt} = -x/\tau - y/(\tau\sigma) + \beta \cos^2[\alpha p x(t - \tau_2 + \phi_2) \\ \quad * \cos^2[(1-p)x(t - \tau_1) + \phi_1 + s(t)]/\tau \\ \frac{dy}{dt} = x \end{cases} \quad (2)$$

We solve this system of ordinary differential equations by the fourth-order Runge-Kutta method, which is widely used in engineering as a high-precision one-step algorithm.

3. Results and discussion

With the above analysis about the proposed RC, we can now give a complete description of training RC for the following tasks. Three tasks are tested: NARMA10 task, Isolated spoken digits recognition, and Channel equalization. In this work, the interval of the virtual nodes is set to $\theta = 50$ ps. The number of virtual nodes N changes with different tasks. We measure τ_1 by the number of virtual nodes, i.e. $T = N\theta$ and $\tau_1 = \theta + T$. And τ_2 can be deduced by $\Delta\tau = \tau_2 - \tau_1$. The distribution ratio of the coupler p ranges from 0.1 to 0.4, $\beta \in [0.2, 1.4]$, ϕ_1 and $\phi_2 \in [-\pi/4, \pi/4]$ and $\alpha \in [2, 8]$. And the ridge parameter ranges from 10^{-6} to 10^{-4} .

3.1. NARMA10 task

The NARMA10 task is one of the most popular benchmark tasks in the RC community. In this task, we train our new reservoir computer to model the system behaving like a nonlinear autoregressive moving average equation of order 10 driven by white noise. The NARMA10 model is

widely used to simulate time series. The output $y^*(n)$ is expected to be as similar as possible to the response $y(n)$ of the NARMA10 model driven by the same white noise $u(n)$. The NARMA10 model is given by the following recursive formula:

$$y(n+1) = 0.3y(n) + 0.05y(n)\left(\sum_{i=0}^9 y(n-i)\right) + 1.5u(n-9)u(n) + 0.1 \quad (3)$$

where $u(n)$ is the random input drawn from a uniform distribution over the interval $[0, 0.5]$, and $y(n)$ is the output of the system. We use the normalized mean square error (NMSE) as a performance metric. In this task, the RC is trained over a sequence of 1000 time steps and tested over a subsequent sequence of 1000 time steps. And we repeat this procedure for 10 times to avoid contingency. NMSE is described by

$$NMSE = \frac{\sum_n [y^*(n) - y(n)]^2}{\sum_n \left\{ y^*(n) - \frac{\sum_n y^*(n)}{n} \right\}^2} \quad (4)$$

We set the number of the virtual nodes to $N = 50$. And the obtained NMSE is 0.103 ± 0.018 . For comparison as shown in Table 1, the $NMSE = 0.168 \pm 0.015$ was obtained in [17], and the $NMSE = 0.152 \pm 0.0138$ in [28] with the same number of virtual nodes. This result shows a great improvement to model NARMA10.

Table 1. Comparison to simulation results for NARMA10 task

	Simple Reservoir [28]	Optoelectronic RC [17]	Novel RC
NMSE	0.152±0.0138	0.168±0.01	0.103±0.018

3.2. Isolated spoken digits recognition task

Speech recognition remains a nontrivial task and we have made many efforts to go in quest of a better alternative to standard speech recognition methods which have hit a limit. Luckily, RC has proven its value for this recognition in [17–19, 29]. This task is the classification of isolated audio sequences, each one representing a digit (0-9) recorded ten times by five different female speakers. The dataset, a subset of the National Institute of Standards and Technology Texas Instrument-46 Corpus (NIST TI-46 Corpus) [30], has 500 sequences. Every input represents a spoken digit, preprocessed using the Lyon cochlear ear model [31]. We employ 10 linear classifiers, each one associated to one digit. The target function is set to 1 if the isolated spoken digit obtained just corresponds to the desired digit, and -1 otherwise. The results of the classifiers are averaged in time, and then the actual digit is obtained by applying a winner-takes-all method that the target function of the highest averaged classifier should be set to 1 and others be set to -1. The highest averaged classifier corresponds to the correct digit. The performance metric used to evaluate the digit recognition is the Word Error Rate (WER), i.e. the fraction of digits incorrectly classified. In addition, the corpus has only 500 sequences, so we divide them into five subsets and the estimation follows a standard cross-validation procedure to avoid the impact of the specific division of the available data in some operations such as training and testing. Five subsets are randomly chosen, four of which are used for training and one for testing. And this process is repeated 5 times so that each subset is evaluated once for the test stage. In this task, we set the virtual nodes to $N = 200$. We obtain the best result of all: $WER = 0\%$, which represents that all correct words are acquired. It is better than the previous ones. For comparison as summarized in Table 2, the traditional RC [29] gets $WER = 4.3\%$ composed of more than 1200 nodes and 3.8% using the Hopfield coding with a network of 320 neurons. LSTM with 121 units and 7791

weights [32] achieves an error rate of 2%. And the WERs are 0.14% at $N = 400$ in [15], 0.12% at $N = 308$ in [33] and 0.4% at $N = 200$ in the optoelectronic RC with single feedback loop [17]. So our result exceeds the previous achievements.

Table 2. Comparison to simulation results for isolated spoken digits recognition

	Traditional RC [29]	LSTM [32]	Electronic RC [15]	Optoelectronic RC [17]	Novel RC
WER	4.3%(N>1200)	2%(N=121)	0.14%(N=400)	0.4%(N=200)	0%(N=200)

3.3. Channel equalization task

This task was introduced in [17], the RC as a channel equalizer is capable of suppressing the inter-symbol interference (ISI) caused by multi-path fading communication channels. The wireless communication channel is modeled as a linear system followed by a memoryless nonlinear system with the second-order and third-order nonlinear distortions. The linear system is described as:

$$\begin{aligned} q(n) = & 0.08d(n+2) - 0.12d(n+1) + d(n) + 0.18d(n-1) \\ & - 0.1d(n-2) + 0.091d(n-3) - 0.05d(n-4) \\ & + 0.04d(n-5) + 0.03d(n-6) + 0.01d(n-7) \end{aligned} \quad (5)$$

where $d(n)$ is the input data to the channel and it is an independent, identically distributed random sequence with values from $\{-3, -1, +1, +3\}$. $q(n)$ is the output of the linear channel. Then, $q(n)$ goes through the nonlinear system, yielding

$$u(n) = q(n) + 0.036q(n)^2 - 0.011q(n)^3 + v(n) \quad (6)$$

where the additive noise $v(n)$ is a pseudo-random Gaussian with zero mean and variance adjusted to get the desired output signal-to-noise ratios (SNR) ranging from 12 to 32 dB. $u(n)$ is the ultimate output. The symbol error rate (SER), defined as the fraction of misclassified symbols, is used as a performance metric.

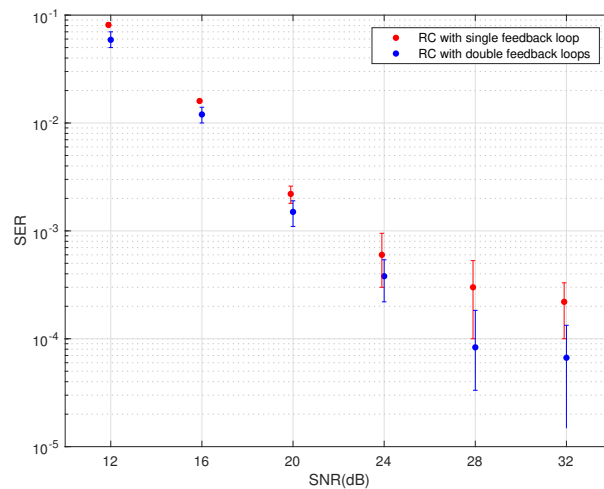


Fig. 3. Simulation results for the nonlinear channel equalization task.

In this task, we use 50 virtual nodes in the short feedback loop for comparison to previous results. The datasets include 10 different subsets and each subset contains 9000 samples, one third of them as a training set and the other as a testing set. And every subsets are reused for 10 times. This task is, given the output $u(n)$ of the channel, to reconstruct the input $d(n)$. The RC with double feedback loops has shown better performance of equalization than the RC with single feedback loop [17] as illustrated in Fig. 3. The horizontal axis means signal-to-noise ratio (SNR) of the nonlinear channel. And the vertical axis means symbol error rate (SER), the fraction of misclassified symbols. The red circles show the discrete simulation results in [17] for the similar Optoelectronic RC with single feedback loop. The blue circles represent the simulation results for the novel RC with double feedback loops. Besides, the error bars represent the deviation of the SER for the same trials. We can see that the SER drops at each SNR in comparison to [17]. It is noteworthy that the error symbols very close to 0 are obtained at 28 dB SNR and the SER appears zero in some subsets at 32 dB SNR, which means that all 60,000 symbols are nearly correct.

3.4. Influence factors

The feedback delay time and the feedback strength play crucial roles in the dynamics of optoelectronic feedback systems, as is the case for performances of the optoelectronic RC. For every task, the research methods of influence parameters on the performances of RC are similar. Therefore, we choose one of the three tasks (the first task) to analyze the influence parameters on the performance of RC as a reference for optimizing the system. Certainly, the analysis results can also be seen as a reference for other tasks. Here, we fix some factors: the intervals of nodes $\theta = 50$ ps, $N = 50$, $p = 0.2$ and $\alpha = 4$.

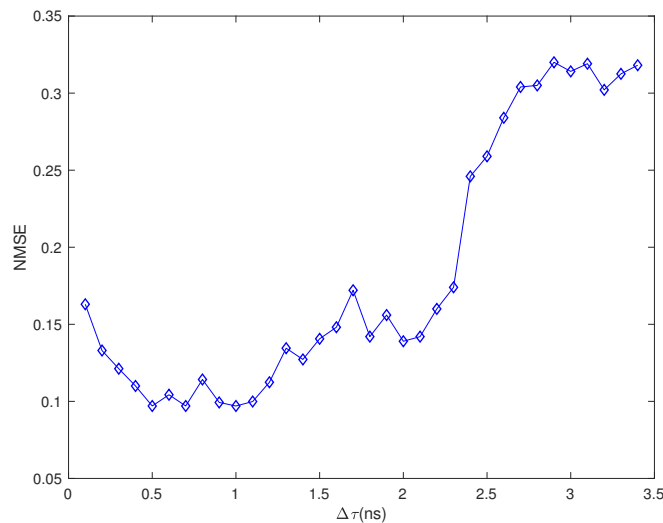


Fig. 4. NMSE dependence on $\Delta\tau$ from 0.1 ns to 3.4 ns for the first task.

Figure 4 shows how the difference $\Delta\tau$ between short and long feedback loop affects the performance of RC evaluated by NMSE. It shows a degradation trend with the increase of $\Delta\tau$. And then it is found that the NMSE fluctuates in pace with increasing $\Delta\tau$, and the minimum of NMSE appears at $\Delta\tau = 0.5$ ns, 0.7 ns and 1 ns corresponding to the difference between the number of virtual nodes of the feedback loops: 10, 14 and 20. Subsequently, the NMSE soars up over 20% after $\Delta\tau = 2.3$ ns and then changes slowly in last phase. It is speculated that $x(i)$

acted on some variables of $x(i - 1)$ can help the system reduce the NMSE. It indicates that strong independence between different virtual node states may have a positive influence on modeling NARMA. But the more variables added have not obvious effect and even subtly adverse impact. The result becomes worse when adding more previous transient states. So $\Delta\tau$ should be chosen appropriately to improve the performance of the new RC. From the viewpoint of application, we generally select $\Delta\tau = 0.5$ ns.

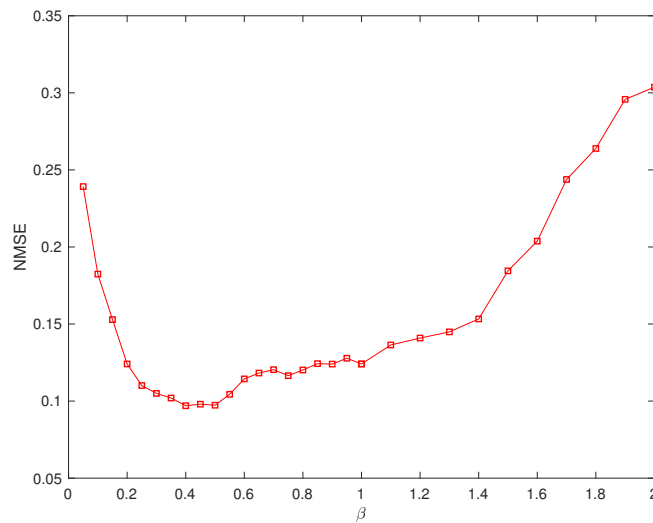


Fig. 5. NMSE as a function of the feedback strength β in the novel RC for the first task.

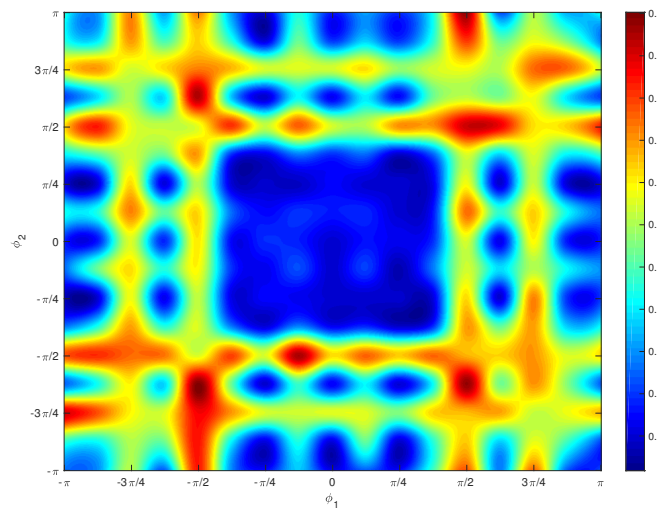


Fig. 6. Color distribution map of NMSE obtained by ϕ_1 and ϕ_2 ranging from $-\pi$ to π .

We then study the influence of the feedback strength on the performance of the new RC measured by NMSE. The system parameters are $\theta = 50$ ps, $p = 0.2$, $\alpha = 4$, $\Delta\tau = 0.5$ ns, ϕ_1 and

$\phi_2 \in [-\pi/4, \pi/4]$. It can be easily understood through a quantitative analysis in Fig. 5 that the power is not enough to process input signal when the feedback strength β is too small, so the NMSE is relatively high. And the NMSE fluctuates gently with the increase of the feedback strength β in the range of 0.25 to 1.4. But when the feedback strength β exceeds 1.4, the NMSE increases because chaos states in RC start appearing and dynamics of RC come to be out of order and extremely sensitive. So a proper feedback strength β should be used in order to keep the RC in a good state.

Finally, we study the impact of the offset phases ϕ_1 and ϕ_2 together on the NMSE of the novel RC. We set $\theta = 50$ ps, $\alpha = 4$, $\Delta\tau = 0.5$ ns and $\beta = 0.4$. Figure 6 gives a detailed description of the changes of NMSE of the novel RC. And it is conspicuously observed that the NMSE is relatively higher when ϕ_1 or ϕ_2 is close to $\pm\pi/2$ and $\pm 3\pi/4$. And if ϕ_1 and $\phi_2 \in [-\pi/4, \pi/4]$, the NMSE is relatively lower. It is also found that if $\phi_1 = \phi_2$, the NMSE is relatively lower in comparison to $\phi_1 \neq \phi_2$ by observing the detailed data.

4. Conclusions and outlook

In this paper, we introduce a new hidden layer and add more nonlinearities into RC forming a novel RC with double feedback loops. The performances of the optoelectronic reservoir computing system based on a semiconductor laser with double optoelectronic feedback loops for NARMA10 task, isolated spoken digits recognition and nonlinear channel equalization are studied and the obtained performances are compared with the previous results. The novel RC shows its advantages in prediction, spoken recognition and channel equalization. In addition, we numerically analyze some influence factors to optimize the performance of the new RC. Feedback delay time and feedback strength play crucial roles in the dynamics of optoelectronic feedback systems, as is the case for performances of the optoelectronic RC. And we also study the offset phases to optimize the parameters of the new RC.

Still, there are a number of challenges for the novel RC: how to realize hardware of the new RC; how many loops in the new RC are needed for a specific task (here we use two feedback loops for example). Despite the road to improve RC twists and turns, further study will be implemented to exploit its great potential to address more complex and troubling problems in information processing and other efforts will be made to push forward it a new-type computer.

Funding

National Natural Science Foundation of China (NSFC) (61575122).

References

1. C. L. P. Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Inf. Sci.* **275**, 314–347 (2014).
2. L. A. Feldkamp, D. V. Prokhorov, C. F. Eagen, and F. Yuan, *Enhanced Multi-Stream Kalman Filter Training for Recurrent Networks* (Springer US, 1998), pp. 29–54.
3. A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, (2013), pp. 6645–6649.
4. W. Mo, C. L. Gutterman, Y. Li, G. Zussman, and D. C. Kilper, "Deep neural network based dynamic resource reallocation of bbu pools in 5g c-ran roadm networks," in *Optical Fiber Communication Conference*, (Optical Society of America, 2018).
5. Y. Luo and Y. Huang, "Text steganography with high embedding rate: Using recurrent neural networks to generate chinese classic poetry," in *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, (ACM, 2017), pp. 99–104.
6. H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks—with an erratum note," *Ger. Natl. Res. Cent. Inf. Technol. GMD Tech. Rep.* **148**, 13 (2001).
7. W. Maass, T. Natschlädger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Comput.* **14**, 2531–2560 (2002).
8. D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, "An experimental unification of reservoir computing methods," *Neural Networks* **20**, 391–403 (2007).

9. F. Triefenbach, A. Jalalvand, B. Schrauwen, and J. P. Martens, "Phoneme recognition with large hierarchical reservoirs," in *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, vol. 2 (Curran Associates Inc., 2010), pp. 2307–2315.
10. L. Boccatto, A. Lopes, R. Attux, and F. J. Von Zuben, "An echo state network architecture based on volterra filtering and pca with application to the channel equalization problem," in *The 2011 International Joint Conference on Neural Networks*, (2011), pp. 580–587.
11. H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science* **304**, 78–80 (2004).
12. F. Duport, A. Akrou, A. Smerieri, M. Haelterman, and S. Massar, "Analog input layer for optical reservoir computers," *Eprint Arxiv* **146**, 460–4 (2014).
13. J. Qin, Q. Zhao, H. Yin, Y. Jin, and C. Liu, "Numerical simulation and experiment on optical packet header recognition utilizing reservoir computing based on optoelectronic feedback," *IEEE Photonics J.* **9**, 1–11 (2017).
14. Yu Jin, Q. Zhao, H. Yin, and Hehe Yue, "Handwritten numeral recognition utilizing reservoir computing subject to optoelectronic feedback," in *2015 11th International Conference on Natural Computation (ICNC)*, (2015), pp. 1165–1169.
15. 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," *Nat. Commun.* **2**, 468 (2011).
16. K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, and J. V. Campenhout, "Toward optical signal processing using photonic reservoir computing," *Opt. Express* **16**, 11182–11192 (2008).
17. Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," *Sci. Rep.* **2**, 287 (2012).
18. F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," *Opt. Express* **20**, 22783–22795 (2012).
19. 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).
20. R. Martinenghi, A. Baylón-Fuentes, F. Xiaole, M. Jacquot, Y. Chembo, and L. Larger, "Optoelectronic nonlinear transient computing with multiple delays," in *2013 Conference on Lasers Electro-Optics Europe International Quantum Electronics Conference CLEO EUROPE/IQEC*, (2013).
21. F. Duport, A. Smerieri, A. Akrou, M. Haelterman, and S. Massar, "Fully analogue photonic reservoir computer," *Sci. Rep.* **6**, 22381 (2016).
22. A. Dejonckheere, F. Duport, A. Smerieri, L. Fang, J.-L. Oudar, M. Haelterman, and S. Massar, "All-optical reservoir computer based on saturation of absorption," *Opt. Express* **22**, 10868–10881 (2014).
23. 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**, 1269–1277 (2015).
24. R. M. Nguimdo, G. Verschaffelt, J. Danckaert, and G. Van der Sande, "Simultaneous computation of two independent tasks using reservoir computing based on a single photonic nonlinear node with optical feedback," *IEEE Trans. Neural Networks Learn. Syst.* **26**, 3301–3307 (2015).
25. 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**, 10211–10219 (2018).
26. G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control. Signals Syst.* **2**, 303–314 (1989).
27. K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks* **2**, 359–366 (1989).
28. A. Rodan and P. Tiño, "Simple deterministically constructed recurrent neural networks," in *Intelligent Data Engineering and Automated Learning*, (Springer Berlin Heidelberg, 2010), pp. 267–274.
29. D. Verstraeten, B. Schrauwen, D. Stroobandt, and J. V. Campenhout, "Isolated word recognition with the liquid state machine: a case study," *Inf. Process. Lett.* **95**, 521–528 (2005).
30. T. Instruments-Developed, "46-word speaker-dependent isolated word corpus (ti46)," NIST Speech Disc pp. 7–1.1 (1991).
31. R. Lyon, "A computational model of filtering, detection, and compression in the cochlea," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 7 (1982), pp. 1282–1285.
32. A. Graves, D. Eck, N. Beringer, and J. Schmidhuber, "Biologically plausible speech recognition with lstm neural nets," in *Biologically Inspired Approaches to Advanced Information Technology*, (Springer Berlin Heidelberg, 2004), pp. 127–136.
33. D. Verstraeten, B. Schrauwen, and D. Stroobandt, "Reservoir-based techniques for speech recognition," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, (2006), pp. 1050–1053.