

Reservoir computing using a delayed feedback system: towards photonics

L. Appeltant¹, M.C. Soriano², G. Van der Sande¹, J. Danckaert¹, S. Massar³,
J. Dambre⁴, B. Schrauwen⁴, C.R. Mirasso², I. Fischer²

¹ Applied Physics Research Group (APHY), Vrije Universiteit Brussel,
Pleinlaan 2, B-1050 Brussel, Belgium

² Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (UIB-CSIC),
Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain

³ Laboratoire d' Information Quantique, CP 225, Université Libre de Bruxelles,
Boulevard du Triomphe, B-1050 Bruxelles, Belgium

⁴ Department of Electronics and Information Systems, Ghent University,
St. Pietersnieuwstraat 41, B-9000 Ghent, Belgium

Delayed feedback systems are known to exhibit a rich dynamical behavior, showing a wide variety of dynamical regimes. We use this richness to implement reservoir computing, a processing concept in machine learning. In this paper we demonstrate the proof of principle on an electronic system, however the approach is readily transferable to photonics, promising fast and computationally efficient all-optical processing. Using only one single node with delayed feedback instead of an entire network of nodes, we succeed in obtaining state-of-the-art results on benchmarks such as spoken digit recognition and system identification.

Reservoir Computing

Reservoir computing has recently been put forward as the generic name of a new research line in machine learning [1]. This field combines Echo State Networks, which were first introduced by Jaeger in ref. [2], and Liquid State Machines, published in ref. [3] by Maass. The general idea of reservoir computing is that by splitting the neural network into two separate layers, the training can be simplified drastically compared to traditional feedforward networks. The first layer is called the reservoir or liquid. This layer contains all the nonlinear nodes of the network and will transform the original input signal into a high number of reservoir states, every state corresponding to the value of a node. Different from the strategy used in traditional neural networks, the coupling weights between the nodes in the reservoir itself are not trained. They are chosen in a random way and remain fixed. The training only takes place in the second layer. This second layer does not provide any additional nonlinear transformation. It will only observe all the reservoir states in the first layer and subsequently calculate the coefficients of the linear combination that combines them into the desired target value.

The fact that a reservoir allows to transform a nonlinearly separable input into a linearly separable reservoir state is a direct consequence of the difference in dimensionality between the input layer and the reservoir. Since usually the number of nodes in the reservoir is high, this corresponds to a high-dimensional feature space. Inputs with the same

features will tend to cluster together and hence a linear hyperplane can be used to make a linear separation in the high-dimensional phase space [4]. As a complex nonlinear system with recurrent connections, the reservoir will store some of the previous inputs. In other words, they do not need to be remembered externally, the dynamics of the system will behave as a memory and a processing unit in one.

Delayed feedback systems as reservoirs

We propose a new approach to implement the reservoir or liquid, using delayed feedback systems [7]. Delay coupled systems have been intensively studied both in nature and technology. These systems range from the brain [8] to semiconductor lasers [9]. The delayed feedback can have a significant impact on the dynamical behavior of the system and can induce instabilities and synchronization between subsystems [10]. Delayed feedback systems are complex systems, which have a number of properties that are linked to well-functioning reservoirs. Because of the delayed states, they imply an infinite-dimensional phase space, with a fading memory and hence they allow the low-dimensional input to be mapped into a high-dimensional phase space.

In conventional neural networks, the input is injected in a large number of nodes in the reservoir. Every node has its proper input scaling, which is multiplied with the input before the nonlinear transformation takes place. In the case of a delayed feedback system, there is only one accessible nonlinear node and a delay line, see Figure 1(b). In order to benefit from all the delayed states in the delay line as nodes and to provide a high-dimensional mapping, we preprocess the input before injecting into the nonlinear node. Every sample undergoes a sample-and-hold operation such that every value is kept constant during one delay interval τ . Next, the input scaling factors, also referred to as the mask, of all the nodes are superimposed on the expanded input. The result can be seen in Figure 1(a). In this figure the mask consists of two discrete values, randomly assigned to all the nodes. The nodes are in fact delayed states of the physically present hardware node and we will refer to them as virtual nodes. The virtual nodes are located along the delay line and are a distance θ apart. Once the expanded input value, with the mask superimposed, has been completely injected into the nonlinear node, and consequently has filled the complete delay line, the values of all the virtual nodes are read out.

The hardware node exhibits some inertia in response to a rapidly changing input. When the virtual node distance θ is large compared to the internal time scale of the oscillator, the oscillator has the time to reach a steady-state that depends only on the input and the feedback signal, the state of the same node one delay time ago. It is completely independent of the states of the neighboring nodes. When decreasing θ down to values comparable to or smaller than the internal timescale the steady-state cannot be reached within the timespan of one node and hence the final value in that node will depend on the last value that was reached in the previous, neighboring node. This relation between adjacent nodes results in the construction of a virtual interconnectivity structure.

Results

We opted for a Mackey-Glass oscillator, first introduced by M.C. Mackey and L. Glass to illustrate the appearance of complex dynamics in physiological control systems by way

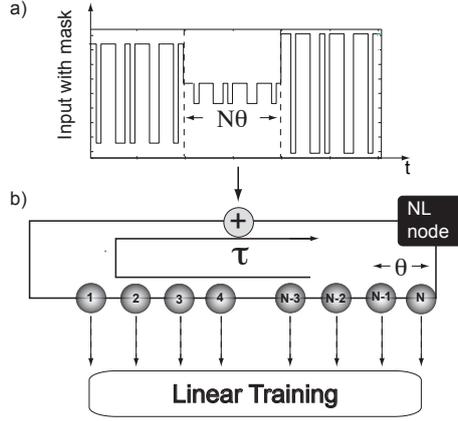


Figure 1: The system consists of one nonlinear node and a delay line that feeds back the output of that node after storing it for a time τ , as shown in scheme (b). The input generated by using the masking procedure is illustrated in (a). A linear training algorithm determines a set of coefficients that, when multiplied with the reservoir responses, leads to the desired target time trace (b).

of bifurcations in the dynamics [12], since can be easily implemented electronically [11]. After some renormalizations, the equation is given by

$$\dot{X}(t) = -X(t) + \frac{\eta \cdot [X(t - \tau) + \gamma \cdot I(t)]}{1 + [X(t - \tau) + \gamma \cdot I(t)]^p}, \quad (1)$$

with η the feedback strength, γ the input scaling, p the exponent and τ the delay time. The mixing of input and feedback signal happens just before the injection into the Mackey-Glass nonlinearity.

In order to test the performance of the system, several benchmark tasks were tested, but below only one is discussed in more detail.

Isolated Spoken digit Recognition

The TI46 corpus is commonly used in the field of reservoir computing as a benchmark for the performance of the system [5] and was introduced by Doddington and Schalk [6]. The input of the system is constructed by taking five female speakers and letting them utter ten different digits (0-9) ten different times. This results in 500 spoken words, all sampled at 12.5 kHz. The inputs are subsequently postprocessed using a cochlear ear model, leading to 86 different frequency bands for each sample. These bands are combined again by masking the input, as is described in ref. [7].

We use a delayed feedback Mackey-Glass system with $\tau = 80$, containing 400 virtual nodes with 0.2 separation each. The time was normalized to the internal timescale of the nonlinear node and hence the virtual nodes can be considered to be close to each other, which implies a connection with the neighboring nodes. Because for this benchmark non-linear mixing is more important than linear memory, the exponent p is taken to be 7. A

higher exponent corresponds to a more nonlinear nonlinearity shape. We applied cross-validation for the training-test set and overfitting was avoided by using ridge regression. The optimal parameter set was found to be $\gamma = 0.5$, $\eta = 0.8$ and $p = 7$. The corresponding word error rate was 0.14% in the simulation and 0.2 % in the experiment, corresponding to 1 misclassified sample out of 500. This result is comparable to state of the art performance of traditional reservoirs with hundreds of nodes.

Conclusion

We use delayed feedback systems for reservoir computing and our approach allows to drastically reduce the number of necessary nodes in a neural network. The result we obtain for the NIST TI-46 corpus benchmark is comparable to what is found in conventional reservoir computing systems. To employ the dynamics induced by delayed feedback, we create a set of virtual nodes, formed by the delayed states that reside in the delay line. This facilitates the hardware realization of a network significantly. The dynamics of the delayed feedback oscillator does not only serve as a way to map the input into a high-dimensional phase space, but it also allows to create a certain interconnectivity between the nodes that enables the mixing of the nonlinearly transformed signals.

References

- [1] D. Verstraeten *et al.*, "An experimental unification of reservoir computing methods," *Neural Networks*, vol. 20 (3), pp. 391-403, 2007.
- [2] H. Jaeger *et al.*, "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, pp. 78-80, 2004.
- [3] W. Maass *et al.*, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Comp.*, vol. 14 (11), pp. 2531-2560, 2002.
- [4] B. Schölkopf *et al.*, "Support Vector Machines, Regularization, Optimization, and Beyond," *Learning with Kernels*, MIT-Press (2001).
- [5] D. Verstraeten *et al.*, "Reservoir-based techniques for speech recognition," *Proceedings of IJCNN06, International Joint Conference on Neural Networks*, pp. 1050 - 1053, 2006.
- [6] G. Doddington, T. Schalk, "Speech recognition: turning theory to practice," *IEEE Spectrum*, vol. 18, pp. 26 - 32, 1981.
- [7] L. Appeltant *et al.*, "Information processing using a single dynamical node as complex system," *Nat. Commun.*, vol. 2:468 doi: 10.1038/ncomms1476, 2011.
- [8] P. R. Roelfsma *et al.*, "Visuomotor integration is associated with zero time-lag synchronization among cortical areas," *Nature*, vol. 385, pp. 157-161, 1997.
- [9] T. Heil *et al.*, "Chaos Synchronization and Spontaneous Symmetry-Breaking in Symmetrically Delay-Coupled Semiconductor Lasers," *Phys. Rev. Lett.*, vol. 86, pp. 5, 2001.
- [10] I. Fischer *et al.*, "Zero-lag long-range synchronization via dynamical relaying," *Phys. Rev. Lett.*, vol. 97, pp. 123902, 2006.
- [11] A. Namajunas *et al.*, "An electronic analog of the Mackey-Glass system," *Phys. Lett. A*, vol. 201, pp. 42-46, 1995.
- [12] M.C. Mackey *et al.*, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, pp. 287, 1977.