

<https://doi.org/10.1038/s44172-024-00314-0>

# Robust neural networks using stochastic resonance neurons



Egor Manuylovich , Diego Argüello Ron, Morteza Kamalian-Kopae &amp; Sergei K. Turitsyn

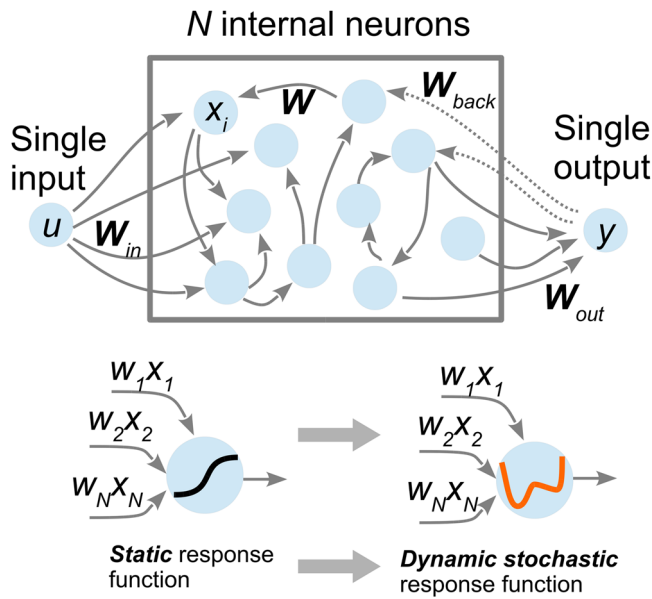
Various successful applications of deep artificial neural networks are effectively facilitated by the possibility to increase the number of layers and neurons in the network at the expense of the growing computational complexity. Increasing computational complexity to improve performance makes hardware implementation more difficult and directly affects both power consumption and the accumulation of signal processing latency, which are critical issues in many applications. Power consumption can be potentially reduced using analog neural networks, the performance of which, however, is limited by noise aggregation. Following the idea of physics-inspired machine learning, we propose here a type of neural network using stochastic resonances as a dynamic nonlinear node and demonstrate the possibility of considerably reducing the number of neurons required for a given prediction accuracy. We also observe that the performance of such neural networks is more robust against the impact of noise in the training data compared to conventional networks.

Artificial neural networks (ANNs) are capable of solving certain problems using only the accessible sets of training examples, without knowledge of the underlying systems responsible for the generation of this data (see e.g., refs. 1–3 and references therein). In scientific and engineering applications, ANNs can be employed as a nonlinear statistical tool that learns low-dimensional representations from complex data and uses this to model nontrivial relationships between inputs and outputs. The ability of ANNs to predict and approximate from given data is linked to the effective dimensionality—the number of independent free parameters in the model. The approximation capability of the ANNs is quantified by the universal approximation theorems (see for details, e.g., ref. 4). Note a recent proposal of the Kolmogorov-Arnold Networks<sup>5</sup> that exploits as the foundation for approximation the Kolmogorov-Arnold representation theorem<sup>6,7</sup>. While traditional deep learning ANNs based on multi-layer perceptron with a fixed activation functions on nodes (“neurons”), KANs have adjustable activation functions on edges (“weights”) and no linear weight matrices at all. However, each weight parameter is replaced by a learnable one-dimensional function parametrized as a spline that can also have high complexity. The complexity of the ANNs plays a critical role in the trade-off between their performance and accuracy of the predictions on the one side and power consumption and speed of operation on the other. Many modern applications of ANNs require over-parametrized models to ensure optimal performance, resulting in high computational complexity and corresponding increased power consumption. Power efficiency can be improved using physically implemented (not necessarily digital) neural networks, for instance, designed from layers of controllable physical elements<sup>8</sup>. Physics-inspired or physics-informed analog neural networks, capable of combining data processing with the knowledge of the underlying physical

systems embedded into their architecture<sup>8–12</sup>, is a fascinating area of research offering, in particular, a potential pathway to power-efficient ANNs. There is, however, an essential trade-off between improvements in energy efficiency and susceptibility to noise in analog networks. The fundamental challenge in non-digital systems is the accumulation of noise originating from analog components<sup>13–16</sup>. To overcome this challenge and unlock the full potential of analog ANNs, systems must be developed capable of absorbing, converting, and transforming noise. This paper proposes a new artificial neural network design using a stochastic resonance (SR)<sup>17</sup> as a network node—ANN-SR. The robustness of the ANNs operation in the presence of noise depends on the properties of the nonlinear activation function<sup>16</sup>. Instead of using a conventional static nonlinear element, we employ a dynamical system with bi-stable features (see Fig. 1) that can make a positive use of the noise—the stochastic resonance (SR). Note that various experimental implementations of SRs pave the way to new attractive, physically implementable designs of ANNs<sup>18,19</sup>. SR is a phenomenon that occurs when adding noise to a nonlinear system can improve the system’s performance. It is often observed in physical systems, for example, in electronic<sup>18,20,21</sup>, mechanical<sup>22,23</sup>, or even quantum<sup>24</sup> systems. SR has also been observed in a wide range of biological systems, including neurons<sup>25</sup>, cells<sup>26</sup>, and even hearing<sup>27</sup>. Note that the idea of using SR as a nonlinear activation function in ANNs has been studied<sup>28–30</sup>. However, these works did not address the key potential advantage of SR - its ability to make under certain condition a positive use of noise.

Here we demonstrate that SR being employed as a nonlinear activation function can notably improve the performance of ANNs when trained on noisy data. We also show that use of SR activation function can substantially reduce the computational complexity and the required number of neurons

Aston Institute of Photonic Technologies, Aston University, Birmingham, UK. e-mail: [e.manuylovich@aston.ac.uk](mailto:e.manuylovich@aston.ac.uk)



**Fig. 1** | Scheme of the considered echo state network and the proposed use of dynamic stochastic response functions instead of static ones.

for a given prediction accuracy. Our work uniquely demonstrates the robustness of SR neurons when trained with noisy datasets, showing improved prediction accuracy compared to traditional sigmoid functions. This application is particularly relevant for real-world scenarios where data is often imperfect. Additionally, we have performed a detailed computational complexity analysis, showing that while our SR-based approach is marginally more computationally expensive than classical approaches, it offers superior performance benefits. This trade-off is crucial for practical implementations where both accuracy and computational resources are considered. An error of  $10^{-2.76}$  or 0.0017 was achieved for predicting 100 points of the Mackey-Glass series by using only 40800 multiplications per point. For the classical sigmoid activation function, an error of  $10^{-1.45}$  or 0.036 corresponds to a similar number of multiplications. This means that we obtain 20 times more accurate results by using the same number of multiplications. To achieve the same accuracy using the sigmoid function, 203850 multiplications are needed. Thus, to achieve the same accuracy as using the classical approach, one needs to perform five times more multiplications and spend five times more energy. Moreover, ANN-SR performance is more robust against noise in the case of training on noisy data.

### Results

As an illustration of our general idea, we consider a particular recurrent neural network schematically shown in Fig. 1, known as echo state network (ESN)<sup>31</sup>. However, the idea of SR neurons can be easily extended to other architectures, such as e.g., deep neural networks. ESNs are fixed recurrent neural networks that constitute a “reservoir” with multiple (fixed) internal interconnections providing a complex nonlinear multidimensional response to an input signal. An output signal is obtained by training linear combinations only of these readout ESN responses. The ESN has an internal dynamic memory. The internal connections of the ESN are randomly set up. During the training procedure, the training sequence is fed into the ESN. After an initial transient period, the neurons start showing variations of the fed signal or “echoing” it. The readout weights are calculated after feeding the training sequence to the ESN.

To explain the mathematics behind our idea, we consider the classical ESN model<sup>31</sup>:

$$x_{n+1} = f(W_{in}u_{n+1} + Wx_n + W_{back}y_n) \tag{1}$$

where  $u \in \mathbb{R}$  is the current state of the ESN’s input neuron,  $x_n \in \mathbb{R}^{N \times 1}$  is the vector with dimension  $N$  corresponding to the network’s internal state, matrix  $W_{in} \in \mathbb{R}^{N \times 1}$  is the map of the input  $u$  to the vector of dimension  $N$ , the matrix  $W \in \mathbb{R}^{N \times N}$  is the map of the previous state of the recurrent layer, and  $f(x)$  is the nonlinear activation function. The recurrent layer consists of  $N$  neurons and the input information  $u$  contains a single element at each instance of time. The linear readout layer is defined as

$$y_n = W_{out}x_n \tag{2}$$

where vector  $W_{out} \in \mathbb{R}^{1 \times N}$  maps the internal state of the ESN to a single output.

In the classical approach, the nonlinear activation function  $f(x)$  is typically a sigmoid or hyperbolic tangent function<sup>31</sup>. In our model, we replace the analytical nonlinear function  $f(x)$  with a stochastic ordinary differential equation known as stochastic resonance (SR)<sup>32,33</sup>. The SR phenomenon has been studied in various physical systems, including climate modeling, electronic circuits, neural models, chemical reactions, and photonic systems<sup>17,32–35</sup>. This dynamic can be described using a bi-stable system with two inputs: a coherent signal and noise<sup>34,36</sup>. A standard example of the SR model is given by:

$$\frac{d\xi(t)}{dt} = -\frac{dU_0(\xi)}{d\xi} + s(t) + \sigma\mathcal{N}(t) \tag{3}$$

where  $s(t)$  represents the input signal,  $\mathcal{N}(t)$  is Gaussian noise with zero mean and unit variance, and  $\sigma$  is the noise amplitude. The potential  $U(\xi, t)$  is defined as:

$$U(\xi, t) = U_0(\xi) + \xi s(t) \tag{4}$$

Considering a symmetric bi-stable stationary potential well  $U_0$ :

$$U_0(\xi) = -\alpha\frac{\xi^2}{2} + \beta\frac{\xi^4}{4} \tag{5}$$

This model represents the limit of a heavily damped harmonic oscillator, depicting the movement of a particle in a time-dependent bi-stable potential  $U(x, t) = U_0(x) - xs(t)$ <sup>36,37</sup>. For  $\alpha > 0$ , the potential is bi-stable with two stationary points  $x_s = \pm\sqrt{\alpha/\beta}$  and a barrier  $\Delta U = \alpha^2/(4\beta)$ . In this work, we consider the SR function with the potential function with local minima of  $-1$  and  $1$ . Therefore, we assume, without loss of generality,  $\alpha = \beta$ . The time-dependent tilted potential  $U(\xi, t)$  is defined by the static potential function  $U_0$  combined with the time-dependent input signal  $s(t)$ , resulting in:

$$U(\xi, t) = -\alpha\left(\frac{\xi^2}{2} - \frac{\xi^4}{4}\right) + \xi s(t) \tag{6}$$

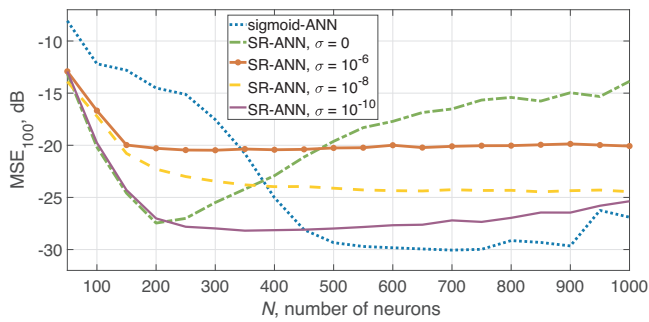
We demonstrate that the SR function, under certain conditions, outperforms the classical sigmoid as a nonlinear activation function in terms of prediction accuracy and computational complexity. The proposed SR neuron differs from a standard neuron by having its own internal state  $\xi(t)$ , which acts as memory. This state, along with incoming signals, forms the neuron’s output through integration over a time interval  $\Delta t$ .

To evaluate the performance of ANNs with SR activation function, we applied it to the prediction of two time series for two classical cases: (i) the Mackey-Glass (MG) equation, and (ii) the Rössler attractor.

For the Mackey-Glass series, governed by:

$$\frac{dq(t)}{dt} = \frac{aq(t - \tau)}{1 + q^{10}(t - \tau)} - bq(t) \tag{7}$$

with parameters  $a = 0.2$ ,  $b = 0.1$ , and  $\tau = 17$ , we also considered a noisy version of the MG series,  $q_\eta(t) = q(t) + \eta\mathcal{N}(t)$ , to assess the signal-to-noise



**Fig. 2** | Mean squared error (MSE) for artificial neural networks (ANNs) employing sigmoid (sigmoid-ANN) or stochastic resonance (SR-ANN) nonlinear activation functions and depending on the number of neurons.

ratio (SNR):

$$SNR_{MG} = \frac{\max(q) - \min(q)}{2\eta} \tag{8}$$

For the Rössler attractor, governed by:

$$\frac{dx}{dt} = -y - z \tag{9}$$

$$\frac{dy}{dt} = x + ay \tag{10}$$

$$\frac{dz}{dt} = b + z(x - c) \tag{11}$$

with parameters  $a = 0.2$ ,  $b = 0.2$ , and  $c = 5.7$ , we similarly considered a noisy version of the Rössler series to evaluate its SNR:

$$SNR_{Rössler} = \frac{\max(x) - \min(x)}{2\eta} \tag{12}$$

We used the  $x$ -component of the Rössler attractor to test the proposed SR-ANN performance.

Parts of both the MG series and the Rössler attractor series were used to train the ESN, and the next several tens of points of each series were compared to the output of the freely running ESN to estimate the prediction accuracy.

During the training and testing, the nonlinear response of each SR neuron is calculated as a result of the integration of Eq. (3) with a weighted sum of input neurons as the variable  $s$  in Eq. (3) and previous internal state of the neuron as an initial condition. As the training and test sequences are parts of the time series (MG or Rössler), they are time-dependent. Therefore, we integrate Eq. (3) over time with  $s$  as a parameter. The integration interval  $\Delta t$  is equal to the time step of the time series, and we set the initial value of the SR function as  $\xi(t = 0) = \mathcal{N}(0, 1)$ . On each time step, we take the current value of the function  $\xi_n$  and calculate the next value by integrating Eq. (3) with  $\xi_n$  as an initial value:

$$f(s_n) \equiv \xi_{n+1} = \xi_n + [\alpha(\xi_n - \xi_n^3) + \sigma\mathcal{N}(t)] \cdot \Delta t + s_n \cdot \Delta t \tag{13}$$

This is done whenever we need to calculate the nonlinear response function  $f(s)$  using a 2nd-order Runge-Kutta also known as the Euler method.

The output of the nonlinear response function is calculated as a trajectory of a system described by the SR equation with input values as an external force applied to the system. The internal state of the SR system is described by a vector  $\xi$  that evolves with time. For a given initial state  $\xi_n$  and input value  $s_n$ , the output of the SR nonlinear function can be calculated as the next internal state  $\xi_{n+1}$  separated from the previous state by time  $\Delta t$ .

Note that there are two sources of noise: (i) implementation noise coming from the active nature of the nonlinear node and (ii) the noisy data, both practically important and considered in this work.

In the current work, we use the proposed approach to design an ESN and estimate its accuracy depending on the number of neurons and noise amplitude in the training data. We compare the designed ESN-SR with the classical ESN with sigmoid activation function in terms of accuracy and computational complexity.

To investigate the computational complexity of the proposed approach, we calculate the number of multiplications required for a single step of ESN evolution. This number includes both the linear part of the ESN evolution and the multiplications needed for calculating the nonlinear response of the neurons.

According to equations (1) and (2), the linear part of ESN evolution includes 4 matrix multiplications of size  $N \times 1$ ,  $N \times N$ ,  $N \times 1$ , and  $1 \times N$  giving a total of

$$Q_{\text{linear}} = N^2 + 3N \tag{14}$$

multiplications per 1 step of the linear part of the ESN evolution. This is also the total number of multiplications for the classical approach for the nonlinear activation function, such as sigmoid or tanh, obtained using a lookup table with no computational complexity.

On the other hand, for the SR nonlinear function, the values of  $\alpha(\xi - \xi^3) + \sigma\mathcal{N}(t)$  or  $[\alpha(\xi_n - \xi_n^3) + \sigma\mathcal{N}(t)] \cdot \Delta t$  can also be calculated using look-up tables and bear no computational burden. However,  $s_n \cdot \Delta t$  needs to be calculated and add up  $\xi_n$ ,  $[\alpha(\xi_n - \xi_n^3) + \sigma\mathcal{N}(t)] \cdot \Delta t$ , and  $s_n \cdot \Delta t$ . So an additional

$$Q_{\text{nonlinear}}^{SR} = N \tag{15}$$

multiplications are needed to calculate the SR nonlinear response function.

To sum up, the total numbers of multiplications for 1 step of calculating the evolution of the ESN are:

$$\begin{aligned} Q_{\text{total}}^{\text{classical}} &= N^2 + 3N \\ Q_{\text{total}}^{SR} &= N^2 + 4N \end{aligned} \tag{16}$$

Note that the number of additional multiplications grows linearly with the size of the ESN, and the total number of multiplications grows quadratically and  $N \gg 1$ , so the quadratic term dominates in the complexity.

The prediction accuracy is measured by the mean squared error between freely running ESN and the corresponding 100 samples of the time series (MG or Rössler). The parameters of the ESN under test are as below:

- $N = 50 \dots 1000$
- Connectivity of  $\mathbf{W}$  is 0.01
- Hyperparameter of the SR function  $\alpha = 0.01$
- SR noise amplitude  $\sigma = 0, 10^{-10}, 10^{-8}, 10^{-6}$

The result of this test is shown in Fig. 2, where each point represents a mean value averaged over 1000 samples.

The “classical” ESN with sigmoid activation function is trained under similar conditions. The linear regression problem for determining the readout weights was performed using singular value matrix decomposition.

The transfer function of an SR neuron depends on the number of ESN evolution steps. The initial values (step 1) are normally distributed with mean 0 and variance 1. The transfer function for steps 1, 10, 1000, and 3100 are shown in Fig. 3. Each neuron automatically converges into its transfer function during the training procedure, allowing self-adjusting activation functions in different parts of ESN using the same node design. For the lower number of neurons, the performance of ESN-SR is better than the classical approach with a sigmoid activation function. In particular, the SR method reaches its maximum accuracy at  $N = 200$  neurons with an averaged error of  $10^{-2.76}$ . The number of multiplications per 1 step is  $Q_{\text{total}}^{SR}(N = 200) = 40800$ . For the classical sigmoid an error of  $10^{-1.45}$  or

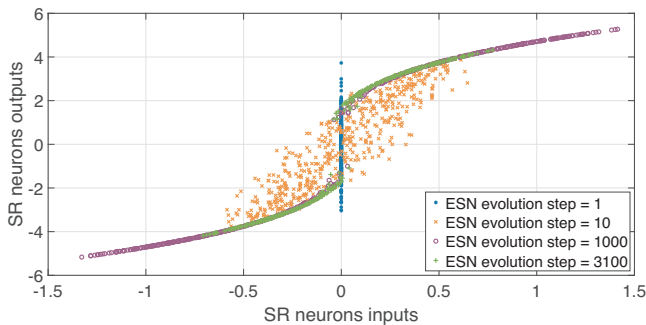
0.036 is achieved for a similar number of nodes  $Q_{total}^{classical}(N = 200) = 40600$ . So we obtain 20 times more accurate results by using the ESN of the same computational complexity. To achieve the same accuracy using the sigmoid activation function, one needs to take  $N = 450$  neurons leading to  $Q_{total}^{classical}(N = 450) = 203850$  multiplications per 1 step. So, to achieve the same accuracy using the classical sigmoid function, one needs to perform 5 times more multiplications and 2.5 times more nodes.

It should be mentioned that the maximum accuracy achieved by using the sigmoid function is slightly higher, but at the expense of much more computational cost: the maximum accuracy of  $10^{-3}$  is achieved at  $N = 700$  neurons that require  $Q_{total}^{sigmoid}(N = 700) = 492100$  multiplications per 1 step. So the best result is only 1.75 times better compared to the best SR result, while it is achieved by using 12 times more multiplications.

We investigated how the noise amplitude  $\sigma$  affects the accuracy and stability of the ESN, see Figs. 2 and 4. One can see that the low level of added noise slightly improves the maximum accuracy and increases stability and accuracy at a higher number of neurons by preventing overfitting. But even slightly higher noise amplitude reduces the accuracy. We also investigated the capability of predicting the continuation of the time series (MG or Rössler) when training on noisy sequences. The training sequence was corrupted by white Gaussian noise and fed into the ESN with the same training procedure.

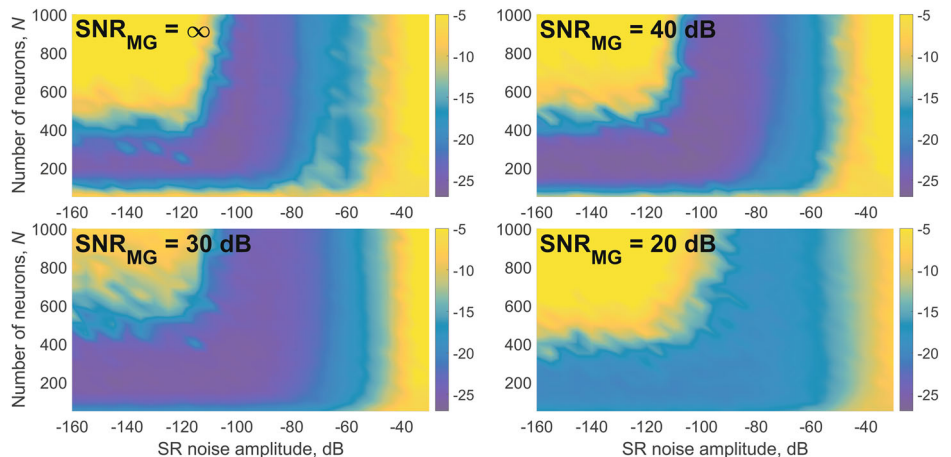
The dependence of the MSE (color-coded) on the number of neurons, internal noise in the nonlinear activation function, and SNR of the training sequence is shown in Fig. 4 showing the optimal internal SR noise level of  $10^{-10}$  for  $SNR_{MG} = \infty$ , level of  $10^{-9.5}$  for  $SNR_{MG} = 40$  dB, level of  $10^{-9}$  for  $SNR_{MG} = 30$  dB and level of  $10^{-8}$  for  $SNR_{MG} = 20$  dB.

We used the SR function with  $\sigma = 10^{-10}$  and compared it with the classical sigmoid function. Noise amplitudes in the training sequence  $\sigma$  corresponding to SNR of 20, 30, and 40 dB were chosen. Figure 5 shows how the MSE of the



**Fig. 3** | Stochastic resonance transfer function for different moments of the echo state network (ESN) evolution.

**Fig. 4** | Prediction accuracy  $MSE_{100}$ , dB (color-coded) depending on stochastic resonance noise level, number of neurons, and noise in the training sequence.



first 100 predicted values depends on the number of neurons for different nonlinear activation functions and various noise levels in the training sequence.

The proposed method shows superior performance compared to the classical approach in the case of the lower number of neurons and the same performance for the higher number of neurons. In particular, in the case of  $SNR = 20$  dB, the prediction accuracy is as good as 0.01 when the number of neurons is as low as 100 in the case of SR. The use of the sigmoid function provides 24 times less accurate results at this number of neurons. And this accuracy is never achieved with the classical sigmoid function, even at a higher number of neurons at this noise level in the training sequence. Figure 6 shows how the best prediction accuracy across various numbers of neurons depends on the SNR in the training sequence where shaded regions depict one standard deviation interval calculated on 1000 runs.

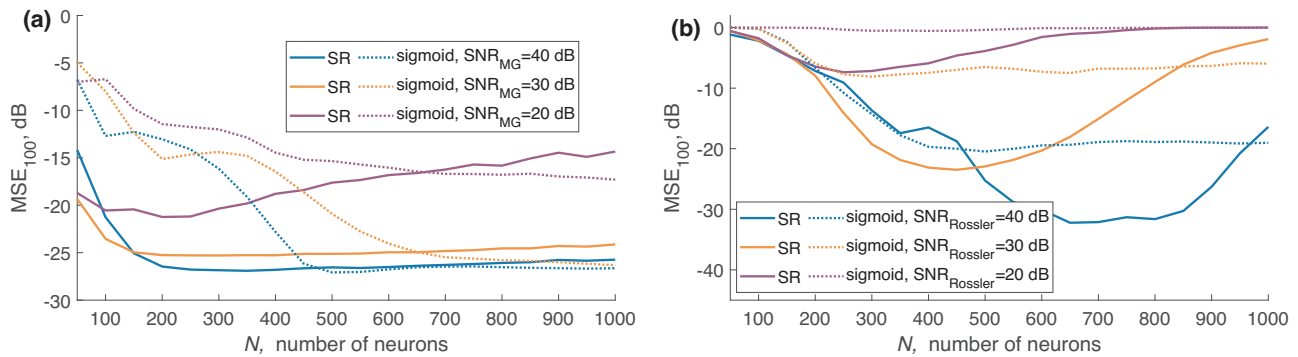
As can be seen in Fig. 6a, while there is no statistically significant difference between the performance of the ESN-SR and sigmoid systems on MG for  $SNR > 25$  dB, the latter outperforms the former in other regions, especially when the number of nodes is the same (violet curve).

In the context of Rössler attractor series, as depicted in Fig. 6b, the performance of the SR-ANN method remains consistently superior. Specifically, for SNR values below 25 dB, the SR-based approach exhibits notably better accuracy compared to the sigmoid-based approaches, irrespective of the number of neurons used. These results highlight the robustness of the SR-ANN method. This is particularly important for training on experimental data, which is often subject to noise, suggesting that SR-ANN might offer more reliable performance in practical scenarios where noise is prevalent.

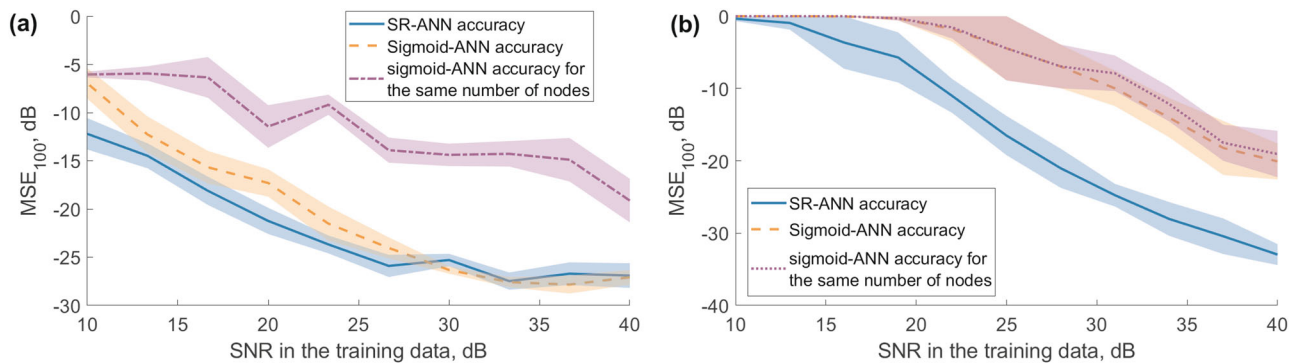
## Discussion

On a general level, we would like to note that our approach is different from the designing network through combining simple elements defined in the corresponding mathematical theory. Instead, we build computing system using the existing building block based on (relatively simple) physical system. In this particular example, we used stochastic resonance, however, this rather general approach can be applied to various physical systems and subsystems. We have shown that replacing the standard nonlinear function with SR considerably increased the accuracy of the ESN with fewer neurons when trained on a Mackey-Glass series. The proposed approach requires two times fewer neurons than the classical one to achieve an error of 0.0015: the computational complexity, in this case, has decreased by a factor of 4. Moreover, since noise is ubiquitous in any real implementation of an ESN, a more realistic investigation requires considering the impact of noise introduced to the system through the nonlinear activation node. Our simulations show that ESN-SR outperforms the one with conventional nodes by a margin that is dependent on the noise power. With a low noise level, the accuracy of the ESN is slightly improved, and overfitting is notably reduced for a larger number of neurons. When training on noisy data, the proposed





**Fig. 5 | Accuracy of the reservoir computer on noisy time series.** Accuracy of the reservoir computer when training on (a) noisy Mackey-Glass (MG) and (b) noisy Rössler sequences with various signal-to-noise ratios (SNR) in the training sequences. Solid lines represent performance using stochastic resonance (SR), while dashed lines correspond to the standard sigmoid function.



**Fig. 6 | Comparison of accuracy across ANNs with different nonlinear functions.** Best accuracy across artificial neural networks (ANNs) of varying complexity, comparing sigmoid-ANN and stochastic resonance ANN (SR-ANN) nonlinear functions on (a) Mackey-Glass series and (b) Rössler attractor series. Shaded areas depict one standard deviation interval.

approach is considerably superior to the classical one in both accuracy and computational complexity. When SNR = 40 dBm and with the same computational complexity, ESN-SR achieves a 25x accuracy improvement compared to the ESN-sigmoid. The classical approach requires more than four times more calculations to deliver the same accuracy. This difference becomes more prominent for noisier data: for SNR = 20 dB with the same computational complexity, the best accuracy of the proposed method achieved is 9.5 times better than the one with sigmoid nodes.

This indicates the capability of SR nodes to capture the underlying relations between samples of the input and manifesting memory properties in various ANNs.

SR-based neurons perform better in certain scenarios due to the unique properties of stochastic resonance that are leveraged in our network design. SR is a phenomenon where adding noise to a nonlinear system enhances its performance. This counterintuitive effect allows SR-based neurons to make positive use of noise, thus improving overall network performance. In traditional neural networks, performance of static nonlinear activation functions is often negatively affected by noise in the data. In contrast, SR-based neurons employ a dynamical system with bi-stable features that make positive use of noise, converting it into a beneficial element rather than a detrimental one. This is particularly advantageous in real-world applications where data is often imperfect and noisy.

We believe that the proposed idea of using a model (or, indeed, physical systems) governed by stochastic ordinary differential equations can be applied to a range of ANNs and can be generalized to different tasks. In particular, the proposed concept is compatible with high-bandwidth optical analog ANNs and reservoirs, offering potential solutions for high-speed parallel signal processing and reduction in power consumption in physical implementations.

**Methods**

We used the classical training procedure to train the designed ESN as described in ref. 31. This procedure is as follows:

1. Feed the first 3000 elements of the time series (MG or Rössler) into the output neuron of the ESN. The feeding is the following procedure:

$$\begin{aligned}
 y_{n+1} &= q_{n+1} \\
 x_{n+1} &= f(\mathbf{W}x_n + \mathbf{W}_{back}y_n)
 \end{aligned}
 \tag{17}$$

During this, the ESN must start “echoing” the fed signal: after some transitional period, the values of internal weights  $x$  start to show some variations of the fed signal.

2. Use the last 2000 internal states to train the output weights so at each step, the product of the output weights matrix and the internal state vector gives the corresponding output value:

$$\mathbf{W}_{out} = \arg \min_{\mathbf{W}_{out}} \sum |X\mathbf{W}_{out} - y|^2
 \tag{18}$$

Here  $X$  is a matrix composed of the last 2000 internal states  $x$ .  $\mathbf{W}_{out}$  can be determined by finding the minimum norm least-squares solution to the linear equation.

Then the training sequence is disconnected from the ESN, and the ESN runs freely. The next several output values the freely running ESN produces are compared to the true values produced by the MG equation.

## Data availability

The data that supports the findings of this study is available from the corresponding author upon reasonable request.

## Code availability

Code for implementing the SR neurons, ESN, and its training and validation algorithms are available from the corresponding author upon reasonable request.

Received: 22 May 2024; Accepted: 30 October 2024;

Published online: 13 November 2024

## References

- Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning* (MIT Press, 2016).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Brownlee, J. *Better deep learning: train faster, reduce overfitting, and make better predictions* (Machine Learning Mastery, 2018).
- Hornik, K., Stinchcombe, M. & White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366 (1989).
- Liu, Z. et al. Kan: Kolmogorov-arnold networks (2024). 2404.19756.
- Kolmogorov, A. On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. *Dokl. SSSR* **108**, 179–182 (1956).
- Arnold, V. On the representation of functions of several variables as a superposition of functions of a smaller number of variables, In the book “Collected Works: Representations of Functions, Celestial Mechanics and KAM Theory, 1957–1965”, 25–46 (Springer Berlin Heidelberg, 2009).
- Karniadakis, G. E. et al. Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
- Wright, L. G. et al. Deep physical neural networks trained with backpropagation. *Nature* **601**, 549–555 (2022).
- Li, M., Irsoy, O., Cardie, C. & Xing, H. G. Physics-inspired neural networks for efficient device compact modeling. *IEEE J. Exploratory Solid-State Comput. Devices Circ.* **2**, 44–49 (2016).
- Kutz, J. N. & Brunton, S. L. Parsimony as the ultimate regularizer for physics-informed machine learning. *Nonlinear Dyn.* **107**, 1801–1817 (2022).
- Pagnier, L. & Chertkov, M. Physics-informed graphical neural network for parameter & state estimations in power systems. *arXiv preprint arXiv:2102.06349* (2021).
- Draghici, S. Neural networks in analog hardware—design and implementation issues. *Int. J. Neural Syst.* **10**, 19–42 (2000).
- Semenova, N. et al. Fundamental aspects of noise in analog-hardware neural networks. *Chaos: An Interdiscip. J. Nonlinear Sci.* **29** (2019).
- Zhou, C., Kadambi, P., Mattina, M. & Whatmough, P. N. Noisy machines: Understanding noisy neural networks and enhancing robustness to analog hardware errors using distillation. *arXiv preprint arXiv:2001.04974* (2020).
- Semenova, N., Larger, L. & Brunner, D. Understanding and mitigating noise in trained deep neural networks. *Neural Netw.* **146**, 151–160 (2022).
- Gammaitoni, L., Hänggi, P., Jung, P. & Marchesoni, F. Stochastic resonance. *Rev. Mod. Phys.* **70**, 223 (1998).
- Dodda, A. et al. Stochastic resonance in MoS<sub>2</sub> photodetector. *Nat. Commun.* **11**, 4406 (2020).
- Singh, K. P., Ropars, G., Brunel, M., Bretenaker, F. & Le Floch, A. Stochastic resonances in an optical two-order parameter vectorial system. *Phys. Rev. Lett.* **87**, 213901 (2001).
- He, D., Chen, X., Pei, L., Jiang, L. & Yu, W. Improvement of noise uncertainty and signal-to-noise ratio wall in spectrum sensing based on optimal stochastic resonance. *Sensors* **19**, 841 (2019).
- Abbaspour, H., Trebaol, S., Morier-Genoud, F., Portella-Oberli, M. & Deveaud, B. Stochastic resonance in collective exciton-polariton excitations inside a GaAs microcavity. *Phys. Rev. Lett.* **113**, 057401 (2014).
- Monifi, F. et al. Optomechanically induced stochastic resonance and chaos transfer between optical fields. *Nat. Photonics* **10**, 399–405 (2016).
- Venstra, W. J., Westra, H. J. & Van Der Zant, H. S. Stochastic switching of cantilever motion. *Nat. Commun.* **4**, 2624 (2013).
- Wagner, T. et al. Quantum stochastic resonance in an AC-driven single-electron quantum dot. *Nat. Phys.* **15**, 330–334 (2019).
- Faisal, A. A., Selen, L. P. & Wolpert, D. M. Noise in the nervous system. *Nat. Rev. Neurosci.* **9**, 292–303 (2008).
- Bene, L., Bagdány, M. & Damjanovich, L. T-cell receptor is a threshold detector: sub-and supra-threshold stochastic resonance in TCR-MHC clusters on the cell surface. *Entropy* **24**, 389 (2022).
- Allen, K. M., Salles, A., Park, S., Elhilali, M. & Moss, C. F. Effect of background clutter on neural discrimination in the bat auditory midbrain. *J. Neurophysiol.* **126**, 1772–1782 (2021).
- Liao, Z., Wang, Z., Yamahara, H. & Tabata, H. Echo state network activation function based on bistable stochastic resonance. *Chaos Solitons Fractals* **153**, 111503 (2021).
- Liao, Z., Wang, Z., Yamahara, H. & Tabata, H. Low-power-consumption physical reservoir computing model based on overdamped bistable stochastic resonance system. *Neurocomputing* **468**, 137–147 (2022).
- Shi, Z., Liao, Z. & Tabata, H. Enhancing performance of convolutional neural network-based epileptic electroencephalogram diagnosis by asymmetric stochastic resonance. *IEEE J. Biomed. Health Inform.* **27**, 4228–4239 (2023).
- Jaeger, H. & Haas, H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80 (2004).
- Harikrishnan, N. B. & Nagaraj, N. When noise meets chaos: stochastic resonance in neurochaos learning. *Neural Netw.* **143**, 425–435 (2021).
- Anishchenko, V. S., Neiman, A. B., Moss, F. & Shimansky-Geier, L. Stochastic resonance: noise-enhanced order. *Phys. Uspekhi* **42**, 7 (1999).
- Harmer, G. P., Davis, B. R. & Abbott, D. A review of stochastic resonance: circuits and measurement. *IEEE Trans. Instrum. Meas.* **51**, 299–309 (2002).
- Soriano, M. C., García-Ojalvo, J., Mirasso, C. R. & Fischer, I. Complex photonics: dynamics and applications of delay-coupled semiconductors lasers. *Rev. Mod. Phys.* **85**, 421 (2013).
- Balakrishnan, H. N., Kathpalia, A., Saha, S. & Nagaraj, N. Chaosnet: a chaos based artificial neural network architecture for classification. *Chaos: An Interdiscip. J. Nonlinear Sci.* **29** (2019).
- Mingesz, R., Gingl, Z. & Makra, P. Marked signal improvement by stochastic resonance for aperiodic signals in the double-well system. *Eur. Phys. J. B-Condens. Matter Complex Syst.* **50**, 339–344 (2006).

## Acknowledgements

This work was supported by the EU ITN project POST-DIGITAL and the EPSRC project TRANSNET.

## Author contributions

Aston Institute of Photonic Technologies, Aston University, Birmingham B4 7ET, UK EM (corresponding author) wrote the code and analyzed the data, E. M., D. A. R., M. K. K., and S. K. T., wrote the article, S. K. T., conceived the research.

## Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to Egor Manuylovich.

**Peer review information** *Communications Engineering* thanks Natalia Berloff and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: [Rosamund Daw and Saleem Denholme].

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024