

**LA-UR-23-30482**

Accepted Manuscript

# **Ergodicity, lack thereof, and the performance of reservoir computing with memristive networks**

Caravelli, Francesco

Zhu, Ruomin

Baccetti, Valentina

Kuncic, Zdenka

Provided by the author(s) and the Los Alamos National Laboratory (2024-07-02).

**To be published in:** Nano Express

**DOI to publisher's version:** 10.1088/2632-959X/ad2999

**Permalink to record:**

<https://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-23-30482>



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.



PAPER • **OPEN ACCESS**

## Ergodicity, lack thereof, and the performance of reservoir computing with memristive networks

To cite this article: Valentina Baccetti *et al* 2024 *Nano Ex.* **5** 015021

View the [article online](#) for updates and enhancements.

You may also like

- [Ergodicity breaking and ageing of underdamped Brownian dynamics with quenched disorder](#)  
Wei Guo, Yong Li, Wen-Hua Song et al.
- [Conditioning problems for invariant sets of expanding piecewise affine mappings: application to loss of ergodicity in globally coupled maps](#)  
Bastien Fernandez and Fanni M Sélley
- [Ergodicity, hidden bias and the growth rate gain](#)  
Nash D Rochman, Dan M Popescu and Sean X Sun

**PRIME**  
PACIFIC RIM MEETING  
ON ELECTROCHEMICAL  
AND SOLID STATE SCIENCE

**HONOLULU, HI**  
October 6-11, 2024

*Joint International Meeting of*  
The Electrochemical Society of Japan (ECSJ)  
The Korean Electrochemical Society (KECS)  
The Electrochemical Society (ECS)

Early Registration Deadline:  
**September 3, 2024**

**MAKE YOUR PLANS NOW!**



## PAPER

## Ergodicity, lack thereof, and the performance of reservoir computing with memristive networks

## OPEN ACCESS

## RECEIVED

17 October 2023

## REVISED

21 January 2024

## ACCEPTED FOR PUBLICATION

15 February 2024

## PUBLISHED

13 March 2024

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Valentina Baccetti<sup>1</sup> , Ruomin Zhu<sup>2</sup> , Zdenka Kuncic<sup>2</sup> and Francesco Caravelli<sup>3</sup> <sup>1</sup> School of Science, RMIT University, Melbourne, Victoria 3000, Australia<sup>2</sup> School of Physics, University of Sydney, Sydney, NSW 2006, Australia<sup>3</sup> Theoretical Division (T4), Los Alamos National Laboratory, Los Alamos, NM 87545, United States of AmericaE-mail: [valentina.baccetti@rmit.edu.au](mailto:valentina.baccetti@rmit.edu.au)**Keywords:** ergodicity, nanoscale memristive networks, memory, reservoir computing**Abstract**

Networks composed of nanoscale memristive components, such as nanowire and nanoparticle networks, have recently received considerable attention because of their potential use as neuromorphic devices. In this study, we explore ergodicity in memristive networks, showing that the performance on machine learning tasks improves when these networks are tuned to operate at the edge between two global stability points. We find this lack of ergodicity is associated with the emergence of memory in the system. We measure the level of ergodicity using the Thirumalai-Mountain metric, and we show that in the absence of ergodicity, two different memristive network systems show improved performance when utilized as reservoir computers (RC). We highlight that it is also important to let the system synchronize to the input signal in order for the performance of the RC to exhibit improvements over the baseline.

**1. Introduction**

Memristive networks are electronic circuits that use memristive devices, a type of resistive switching memory element. These networks store and recall information by modifying the device's resistance. They find applications in computer memory, neuromorphic computing, and signal processing [1]. Memristive networks offer benefits such as high density, low power consumption, and potential for high-speed operation. Simultaneously, there is growing interest in alternative approaches to computation and optimization in response to the rapidly increasing demands on computing [2]. Various proposals have emerged to address this challenge, some of which involve the use of oscillators or frequency domain encoding [3–7], leveraging near- or in-memory computation [8–13], and exploring memcomputing [12, 13]. These innovative techniques aim to provide more efficient solutions for complex optimization problems [3–5, 12, 14–18]. However, understanding large assemblies of memristive devices in non-equilibrium statistical mechanics remains a challenge. Recent research has focused on exploring the geometric and statistical properties of nanowire networks, where electrical junctions exhibit memristive behavior due to the interplay between tunneling and filament formation phenomena [19, 20].

Conductive nano-filament formation at nanowire-nanowire junctions creates a memristive device, while quantum tunneling contributes to additional nonlinearity in switching behavior, making the dynamics more complex [21]. As most of the voltage drop occurs at the junctions, a basic model for the conductance evolution of these networks involves an assembly of memristive devices with voltage or current generators. The dynamic behavior of these systems is currently being studied, especially regarding bias-induced conductance transitions [21–24].

Various conductance transitions have been observed in memristive devices, often characterized by transient unstable dynamics of the memristive components and their internal memory parameter. One well-known transition, which has been defined as ‘rumbling’, has been analytically identified in the simplest model of memristive networks [23]. A similar transition has also been observed in nanowire networks [21]. This

transition involves the system effectively moving between different minima of an effective potential and is marked by bursts of transient positive Lyapunov exponents. It arises from the coexistence of multiple low-dimensional equilibrium points in the dynamics. The mean-field theory for such systems can be ensured by mapping them to a PEDS (projective embedding of dynamical systems) [25]. This is relevant to our study as we aim to identify non-ergodic behavior associated with these transitions.

In certain memristive systems, as the applied voltage increases, an effective mean potential develops with multiple minima. As one minimum becomes dominant, the system undergoes a rapid chaotic transition towards this emerging stable fixed point. This transition is distinct from the conventional Landau picture of symmetry breaking with bifurcation, as it arises from the competition between two minima.

While an understanding of these transitions is clearer in simplified memristive device models, it becomes less evident in more realistic systems. In this study, we aim to describe the dynamic behavior of these transitions in two systems using ergodicity measures. Ergodicity, a concept in statistical mechanics and thermodynamics, pertains to the long-term behavior of a system. It states that the time average of a system over an extended period is equivalent to the ensemble average of the system. In essence, it implies that the system reaches a steady state, and its long-term average behavior (time average) aligns with the average behavior across many instances of the same initial conditions (ensemble average).

Ergodicity plays a crucial role in various natural sciences, such as physics and chemistry, enabling an understanding of system evolution, equilibrium attainment, and the prediction of long-term behavior based on short-term observations. In the context of memory-dependent systems, the lack of ergodic behavior signifies (hard) ergodicity breaking, which occurs when symmetry breaking transpires in thermodynamic systems [26]. Typically, physical systems operate within a regime where ergodicity holds for a subset of possible phase space states.

Quantifying if and how a system's dynamics breaks ergodicity is therefore essential to gain insight into the dynamical state of the physical system, and is vital for comprehending the system's operational regime.

To be specific, we consider two models of memristive network: a resistive current-controlled, analytical toy model based on titanium dioxide devices  $\text{TiO}_2$ , [27]; a conductance-based, physically motivated model based on polymer-coated silver nanowires Ag-PVP [21, 24, 28]. Both models satisfy the pinched hysteresis property for the  $I - V$  curve, characteristic of memristive devices [29].

To test the hypothesis that behavior of these systems is rather different at these points, we use reservoir computing (RC) as our computational model [30]. RC is a popular approach to machine learning that utilizes a network of nodes, known as a reservoir, to generate temporal patterns in response to input signals, that has been recently shown to be universal [31]. This approach is particularly effective for processing time-varying data and has been applied to a range of tasks, such as speech recognition, natural language processing, and control systems. One of the key advantages of RC is that the reservoir does not require tuning during the training phase, reducing the complexity of the learning process. More details about RC can be found in appendix B.

Recently it has been reported that the 'edge of chaos', a phase transition between an ordered and a disordered state, may be important for the performance of RC [32], but it is important to stress that similar critical states have been reported both for biological neuronal networks in the brain [33] and in other artificial neural networks [34]. The edge of chaos hypothesis originates for the fact that several dynamical systems seem to achieve their best computational capacity near a phase transition, when they exhibit the most diverse behaviors [35, 36].

Goals of this study are to test the hypothesis that computation near a conductance transition point improves, and demonstrate that this transition point occurs when the system's dynamics is non-ergodic. This would mean that the lack of ergodicity can be used to predict such transitions, and more generally, transient instabilities.

This article is organized as follows: in section 2 we introduce the general definition of ergodicity and how it is related to memory; in section 3 we introduce the two memristive systems we are considering in this study; in section 4 we give the definition of the TM metric for the two memristive systems we are considering, and test the RC task results for both memristive systems in terms of ergodicity breaking. Conclusions follow.

## 2. Memory, ergodic convergence and the Thirumalai-Mountain metric

Ergodicity and the emergence of memory are closely intertwined concepts within the realms of statistical physics and complexity science. Ergodicity refers to a system's property of uniformly exploring all possible states in its phase space over an extended period. In essence, it characterizes a system's behavior as it traverses its accessible states in a time-averaged manner. On the other hand, memory denotes the influence of past states on the current state of a system.

In many systems, the emergence of memory is closely connected to the violation of ergodicity. When a system is ergodic, its behavior remains independent of its past history, exhibiting no memory effect. However,

when ergodicity is violated, the system may display persistent or long-term correlations, resulting in the emergence of memory (as it is the case for spin glasses, both in natural and artificial systems [37–39]). Ergodicity refers to a system's property where the long-term behavior can be deduced from a single, extended observation of the system, rather than relying on multiple independent realizations. When ergodicity is broken, a system develops memory [26, 37].

According to Boltzmann's hypothesis, the trajectories of any dynamical system in its phase space eventually evolve into regions where macroscopic properties reach thermodynamic equilibrium [40]. The ergodic hypothesis states that ensemble averages and time averages coincide as time progresses. This means that the ensemble-averaged value of an observable, denoted as  $\langle g \rangle$ , can be obtained by averaging its values over time during the observable's evolution. Mathematically for the observable  $g(t)$  we have:

$$\langle g \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T g(t) dt. \quad (1)$$

It is important to note that various definitions of ergodicity exist in the Physics and Mathematics literature [41–43], with significant distinctions. For example, in Markov chains, ergodic behavior requires a strongly connected transition graph, which is rarely the case. In practice, a system can exhibit ergodicity within a specific subset of its phase space. The concept of 'effective ergodicity' was introduced to describe scenarios where the system rapidly samples coarse-grained regions [42].

A measure of effective ergodic convergence is based on the observation that certain components of a system exhibit identical average characteristics at thermal equilibrium. These characteristics are determined by an observable defined on the system's phase space, denoted as  $\Gamma$ . To assess the effective ergodic behavior of the observable, it is necessary to estimate its average value using an ensemble approach, such as a thermal ensemble. This estimation is typically performed using the Thirumalai-Mountain (TM)  $g$ -fluctuating metric, denoted as  $\Omega_g(t)$ .

The TM  $g$ -fluctuating metric, introduced by Thirumalai and Mountain [42, 44, 45], quantifies the difference between the ensemble-averaged value of the observable,  $g(\Gamma)$ , and the sum of the instantaneous values of  $g(t)$  for each component of the system. At a given time  $t$ , the TM  $g$ -fluctuating metric is expressed as:

$$\Omega_g(t) = \frac{1}{N} \sum_{j=1}^N [\bar{g}_j(t) - \langle g(t) \rangle]^2, \quad (2)$$

Here,  $\bar{g}_j(t)$  represents the time-averaged value per component, and  $\langle g(t) \rangle$  denotes the instantaneous ensemble average, defined as:

$$\bar{g}_j(t) = \frac{1}{t} \sum_{i=0}^t g_j(t_i), \quad \langle g(t) \rangle = \frac{1}{N} \sum_{j=1}^N g_j(t). \quad (3)$$

This definition assumes that  $g(\Gamma)$  serves as a suitable physical order parameter, effectively characterizing the system's behavior. The definition of instantaneous ensemble average will depend on the considered system, as it will be shown in section 4.

The rate of ergodic convergence can be quantified using the derivative of the effective ergodic convergence, denoted as  $\Omega'_g$ . This quantity is given by:

$$\Omega'_g = \frac{\Omega_G(t)}{\Omega_G(0)} \rightarrow \frac{1}{tD_g}, \quad (4)$$

in the diffusive regime. When the system is instead in a sub-diffusive or super-diffusive regime, the power of the relaxation in time changes from 1 to a different exponent. Coarse graining of the phase space leads to the clustering of the system's accessible states, making the concept of effective ergodicity more applicable. Effective ergodicity is achieved when the system uniformly explores the coarse-grained regions within a finite time [44]. Here,  $D_G$  represents the diffusion coefficient associated with the property being studied, and  $\Omega'_g$  refers to the effective ergodic convergence. This definition aligns with other notions of ergodicity in cases where diffusion follows a power law [46]. The rate of ergodic convergence, determined using the TM metric, provides an estimate of the system's ergodicity. The behavior of the rate, as described by equation (4), indicates the system's attainment of effective ergodicity. For example, if the inverse of the rate scales linearly with time, the system reaches ergodicity in a diffusive manner:  $1/\Omega_g \rightarrow D_G t$ , where  $D_G = \Omega_G(0)$  represents the diffusion coefficient of the property  $G$ . It is generally expected that the rate scales with time as  $\Omega'_G(t) \sim t^{-p}$ . When the inverse of  $\Omega'_g$  exhibits a linear relationship with time, it indicates that all points in the phase space are equally likely, resembling the behavior of Brownian motion. When  $p > 1$ , the system reaches effective ergodicity faster than in the diffusive regime, while for  $p < 1$  the system may not reach effective ergodicity and therefore displays emergence of memory. This approach has been applied in various contexts, such as simple liquids [42], earthquake fault networks [47, 48], and the Ising model [49].

### 3. Models of memristive devices

As mentioned in the introduction, we consider two models of memristive networks, a resistive current-controlled toy model, that captures the essential features of  $TiO_2$  memristors, and a conductance-based system, that models the dynamics of Ag-PVP nanowire networks. Both models satisfy the pinched hysteresis property for the  $I - V$  curve, the signature characteristic of memristive devices [29]. We explain their functioning in more details below and in appendix A.

#### 3.1. Memristive network toy model

In [50], the dynamical equation for a circuit of memristors was derived under the assumption of a resistance current-controlled device. The resistance function used,  $R(x) = R_{on}(1 - x) + xR_{off}$ , approximates  $TiO_2$  memristors, where  $R_{on}$  and  $R_{off}$  represent the limiting resistances, and  $x \in [0, 1]$  is the *internal memory parameter*, the state variable describing the size of the oxygen-deficient conducting layer. At the lowest order, the evolution of the internal memory parameter can be described by a simple equation with hard boundaries:

$$\frac{dx}{dt} = \frac{R_{off}}{\beta} I - \alpha x = \frac{R_{off}}{\beta} \frac{S}{R(x)} - \alpha x \quad (5)$$

Here,  $\alpha$  and  $\beta$  are the decay constant and the effective activation voltage per unit of time, respectively, which determine the timescales of the system.  $S$  is the external potential applied to the memristor.

While the model presented above provides a simple description of a polar resistive device, various extensions have been explored in the literature. For instance, to account for diffusive effects near the boundaries, some models remove the hard boundaries and introduce a window function [51, 52]. Although these models better capture the detailed IS curves of physical devices, they still exhibit the fundamental pinched hysteresis behavior observed in the linear model.

In adimensional units ( $\tau = \alpha t$ ), the equation for  $x(\tau)$  in a single memristor device under an applied voltage  $S$  can be derived using Ohm's law ( $S = RI$ ), [23, 53]. The resulting equation is:

$$\frac{d}{d\tau} x = \frac{S}{\alpha\beta} \frac{1}{1 - \chi x} - x = -\partial_x V(x, s), \quad (6)$$

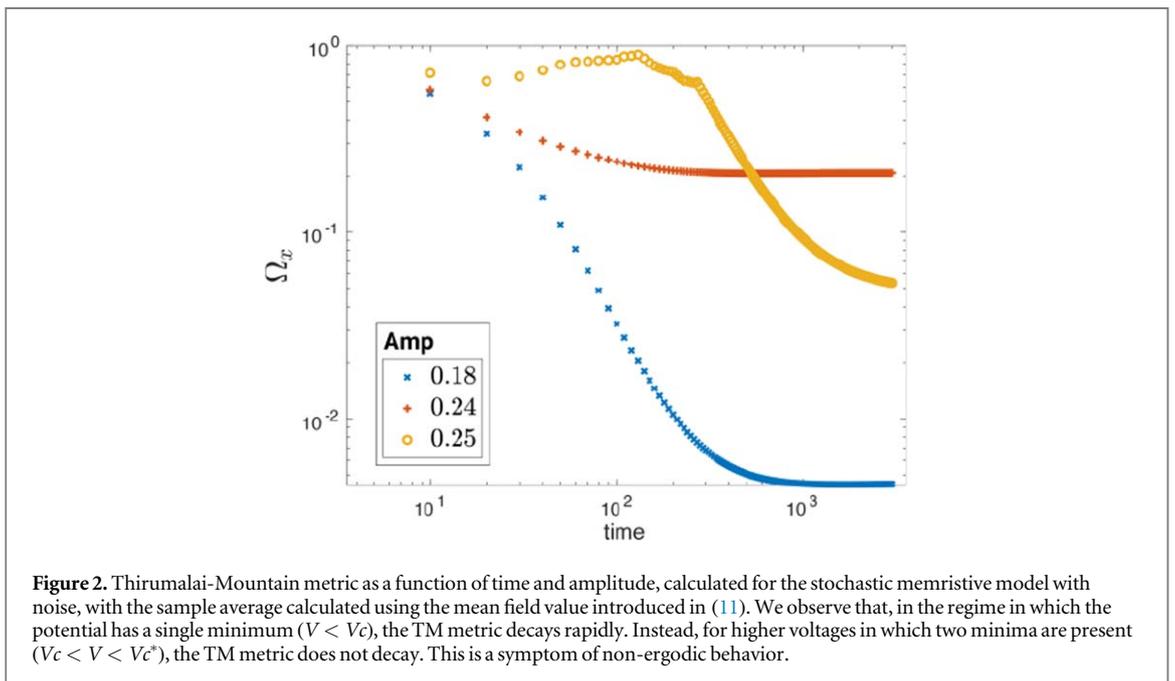
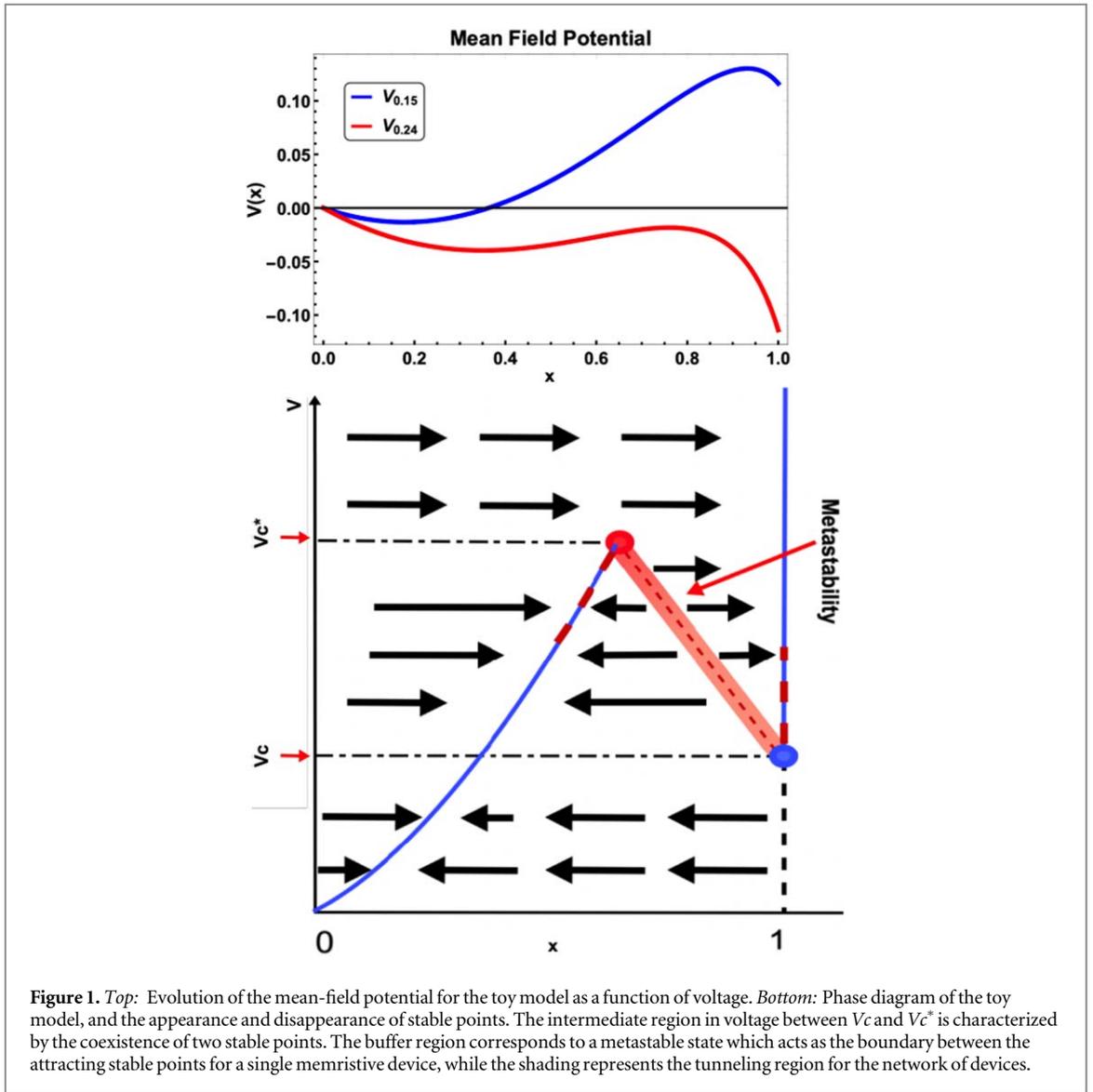
Here,  $\chi = \frac{R_{off} - R_{on}}{R_{off}}$  and  $s = \frac{S}{\alpha\beta}$ , with  $0 \leq \chi \leq 1$  in relevant physical cases. The dynamics of the system, represented by a single memristor device, are fully characterized by the gradient descent in the effective potential given by [23]:

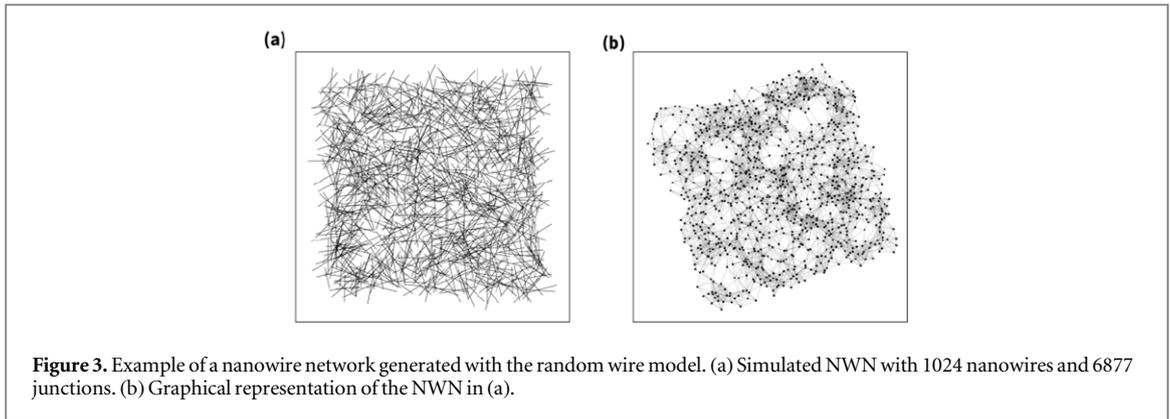
$$V(x, s) = \frac{1}{2}x^2 + \frac{s}{\chi} \log(1 - \chi x), \quad (7)$$

The potential exhibits two minima separated by a barrier, as depicted in figure 1 (top), for  $s = 0.15$  and  $s = 0.24$  with  $\chi = 0.9$ . The range of  $s$  for the existence of a barrier is limited, and when  $\chi$  approaches 1, the local minimum can move inside the domain  $[0, 1]$ , leading to the emergence of an unstable fixed point (i.e., the peak of the barrier). Consequently, two basins of attraction and locally stable minima are formed. A pictorial phase diagram illustrating this behavior is shown in figure 1 (bottom), highlighting the critical voltage points at which such behavior occurs. Below a critical voltage point  $V_c$ , only one fixed point is present. At the  $V_c$  value, a new fixed point emerges, but it becomes metastable in the presence of noise (dashed red line). At an intermediate point  $V_c < V_m < V_c^*$ , the two metastable points have equal energy, representing the switching point where the original stable fixed point becomes metastable. At higher values of  $V = V_c^*$ , the first fixed point disappears by merging with the barrier and becoming flat. Ultimately, only one fixed point remains at higher values.

For a network of memristors of this type, the mean-field potential resembles exactly the single memristor defined above. For details, see Supp. Mat. A.1. The key difference is that the mean-field potential is only an approximation when the system is high dimensional, e.g. now there is an effective metastable region, as shown in figure 1, instead of two stable minima, when one of the minima is lower than the other. Such effective 'tunneling' can be explained using the theory developed in [25], i.e. the fact that local mean field maxima become saddle points in high dimensions. Here we will use this fact to study effective ergodicity breaking in memristive networks.

In fact, these two regimes can be characterized via the Thirumalai-Mountain (TM) metric introduced before. In figure 2 we show the TM metric as a function of time for  $\bar{s} = 0.18$  (top),  $\bar{s} = 0.24$  (center), and  $\bar{s} = 0.25$  (numerically calculated for a circuit of  $N = 50$  randomly distributed memristors with parameters  $\alpha = \beta = 1$  and  $\chi = 0.9$ ). As we can see, in the former case the TM metric relaxes as a power law with an approximate intermediate exponent of  $p \in [-2, -1.5]$ , typical of a diffusive regime. For  $\bar{s} = 0.24$ , we see instead a typical behavior of weak ergodicity breaking, e.g. a transient non-monotonicity of the TM metric,





**Figure 3.** Example of a nanowire network generated with the random wire model. (a) Simulated NWN with 1024 nanowires and 6877 junctions. (b) Graphical representation of the NWN in (a).

associated with the trajectories effectively ‘tunneling’ through the mean field barrier. This behavior is indeed transient as for  $\bar{s} = 0.25$  the TM metric relaxes again as a power law.

Using the toy model, we can then pinpoint such non-monotonicity of the TM metric to a transient transition between two stable asymptotic states and the effective symmetry breaking of the potential. Thus, we can immediately identify the symmetry breaking of the potential as the culprit of such transient non-ergodic behavior.

### 3.2. Nanowire networks

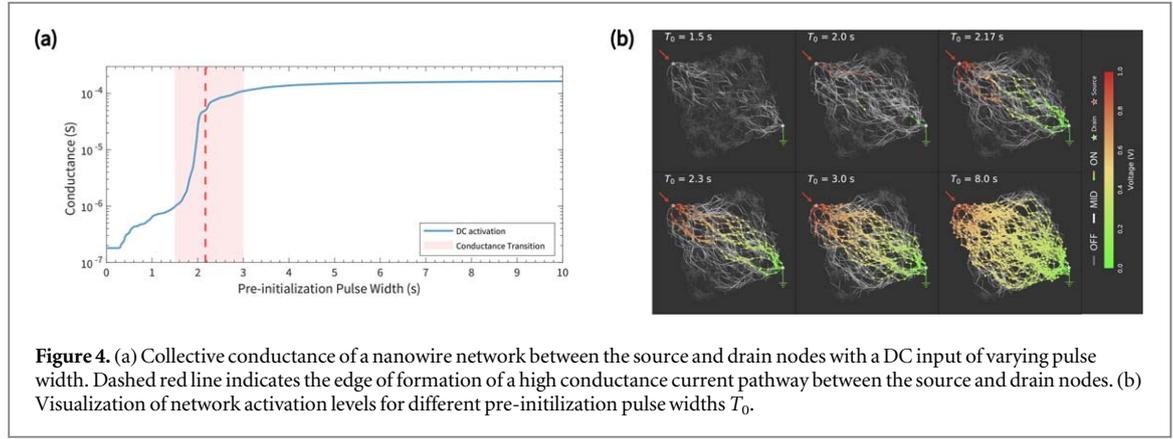
We now consider a more realistic model of memristive networks, which can be associated with self-assemblies of silver nanowires. Via established bottom-up self-assembly techniques, one can readily synthesize nanowire networks (NWNs) [22, 54]. These NWNs typically have a 2D spatial distribution of randomly oriented nanowires that are interconnected by cross-point Metal-Insulator-Metal (MIM) junctions. The NWN we consider have densities of 10 junctions/ $\mu\text{m}^2$  and 0.5 nanowires/ $\mu\text{m}^2$ . Device electrodes can be deposited onto the substrate using a mask, from which the conductance measurements can be readily performed. This bio-inspired structure is difficult to design and fabricate using top-down techniques.

Self-assembled NWNs exhibit topological properties, such as small-world propensity and modularity, that are similar to biological neural networks and distinct from random and grid-like networks [55]. Unlike fully connected bipartite networks in artificial neural networks, small-world networks have local connectivity and short path lengths, making them relatively sparse. Although small-worldness is necessary for important functional properties, such as synchronizability and information flow, it alone cannot explain the diverse range of dynamics across networks that exhibit this structural property.

A model for the simulation of realistic NWNs has been introduced previously in the literature [21, 24, 28]. Figure 3 (a) shows a visualization of a simulated nanowire network containing 1000 nanowires and 6877 junctions. Self-assembly is modeled by distributing nanowires on a  $3 \times 3 \mu\text{m}^2$  2D plane, with their centers uniformly sampled from  $[0, 3]$  and orientation uniformly sampled from  $[0, \pi]$ . The lengths of the nanowires are sampled from a gamma distribution (mean = 100 nm, standard deviation 10 nm), based on experimental measurements [22]. In theoretical studies, as illustrated in figure 3 (b), the NWN is transformed to the corresponding graphical representation, where the nodes represent nanowires and the edges are the cross-point junctions.

In this work, all simulation results for nanowire networks are generated using a network comprised of 1000 nanowires and 6877 junctions. All variables, except the adjacency matrix  $A$ , are time-dependent. A model for the conductance of a single junction, associated with the filament length is provided in the supplementary material, see appendix A.2. Each junction evolves as voltage bias is continuously applied to the network. The modified nodal analysis approach is applied to the graphical representation to solve Kirchhoff’s voltage and current conservation laws at each time step [56]. This is equivalent to the method used for the derivation of the exact network equation for the toy model, equation (A1). Although the NWN model is based on polymer-coated Ag nanowires, with memristive junction internal dynamics that differ from that of the toy model (based on metal-oxide memristors), the network dynamics are similar and one should think of the two models as equivalent from a physical perspective.

A DC input of varying duration is applied to the NWN to initialize the internal state of the network and prepare it for RC. We refer to this pre-initialization protocol as ‘priming the system’ [57]. Figure 4(a) shows the reservoir’s conductance (blue curve) as a function of the DC input length  $T_0$ . The shaded region represents the general conductance transition regime, identified from previous studies [21, 24], and the dashed line at  $T_0 = 2.17$  represents when the first conductance pathways form between the source and drain nodes. The



internal state of the network for different  $T_0$  is visualized in figure 4(b). In cases where the reservoir is under-activated ( $T_0 < 2$  s), the majority of memristive components remain inactive, resulting in insufficient dynamics from the network. When the reservoir is over-activated ( $T_0 > 8$  s), the internal dynamics of the system become saturated, limiting the system's capacity to process additional information. The conductance transition regime from figure 4(a) corresponds to an intermediate dynamical state of the reservoir, where conductance paths first span the network and the internal state of the system produces dynamical features that are more diverse than at other activation times.

#### 4. Ergodicity, reservoir computing and bistability

Reservoir computing using [58] was studied for the first time in [59], among other passive circuits. A brief recap of the procedures behind reservoir computing is provided in appendix B.

We have utilised the reservoir to perform a *chaotic time series prediction task*. This a benchmark task in machine learning as it evaluates the ability of a the system to predict future values of a chaotic time series, and it requires several memory steps and high consistency correlations to obtain good prediction results.

The time series originates from the Mackey-Glass (MG) delayed equation

$$\frac{dz(t)}{dt} = \frac{rz(t-\tau)}{1+z^n(t-\tau)}\gamma z(t). \quad (8)$$

with parameters  $r = 0.2$ ,  $\gamma = 0.1$ ,  $\tau = 17$ , and  $n = 10$  for which the system exhibits moderate chaotic dynamics [60].

The reservoir receives the MG time series in an input signal of the form

$$S(t) = aV_{MG}(t) + b, \quad (9)$$

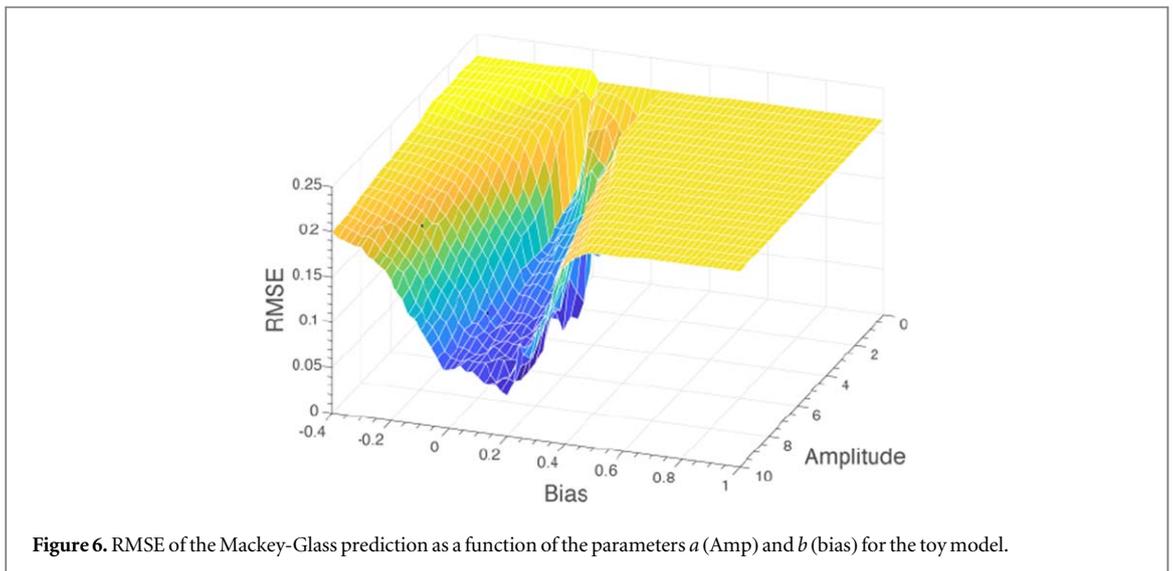
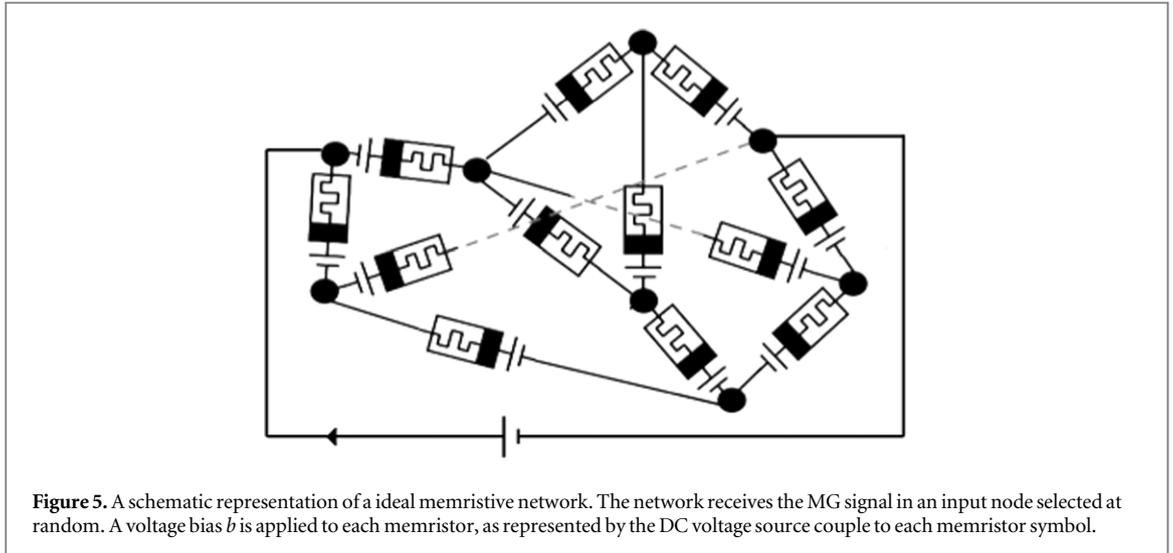
in one node chosen randomly, with amplitude  $a \in [0, 10]$  and  $b \in [-0.4, 1]$  a constant *bias* input.

##### 4.1. Toy model

Here, we have used a similar scheme to the reservoir, using a memristive circuit comprising  $N = 50$  idealized memristors. We have considered parameter values  $\alpha = \beta = 1$ , and  $\chi = 0.9$ , for which the system experiences a symmetry break, and where we expect its dynamics to become non-ergodic, see 3.1. This is the regime where the reservoir is at the *edge of stability*, as defined in [32], where in general, reservoirs can have optimal performance, although not always guaranteed. Weak ergodicity breaking in a dynamical system is associated with a strongly chaotic regime [46]. We will come back to this point later. We have also added a small noise value  $\sigma = 0.01$ . The equations (A2) were numerically integrated with a time step  $dt = 0.1$ . In figure 5 we present a schematic diagram of the network as it perform the prediction task. The MG time series is inputted in a node selected at random, while the bias  $b$  is applied to each memristor in the form of a DC voltage source.

For each value of  $a$  and  $b$ , we performed a MG prediction task using the internal memory states as the dynamical variables for the RC. The MG is delivered to the network as an input; the same MG signal with 5 steps ahead was employed as the target. The Root Mean Square error (RMSE), which we use as a measure of the performance of the RC, is shown in figure 6. We can see from the figure that varying  $a$  leads to a reduced value of the RMSE, right where the rumbling transition is present, approximately for  $b = 0.2$

The difference in the quality of the task can be observed, for the MG time series, in figure 7, for the toy model described in appendix A.1.

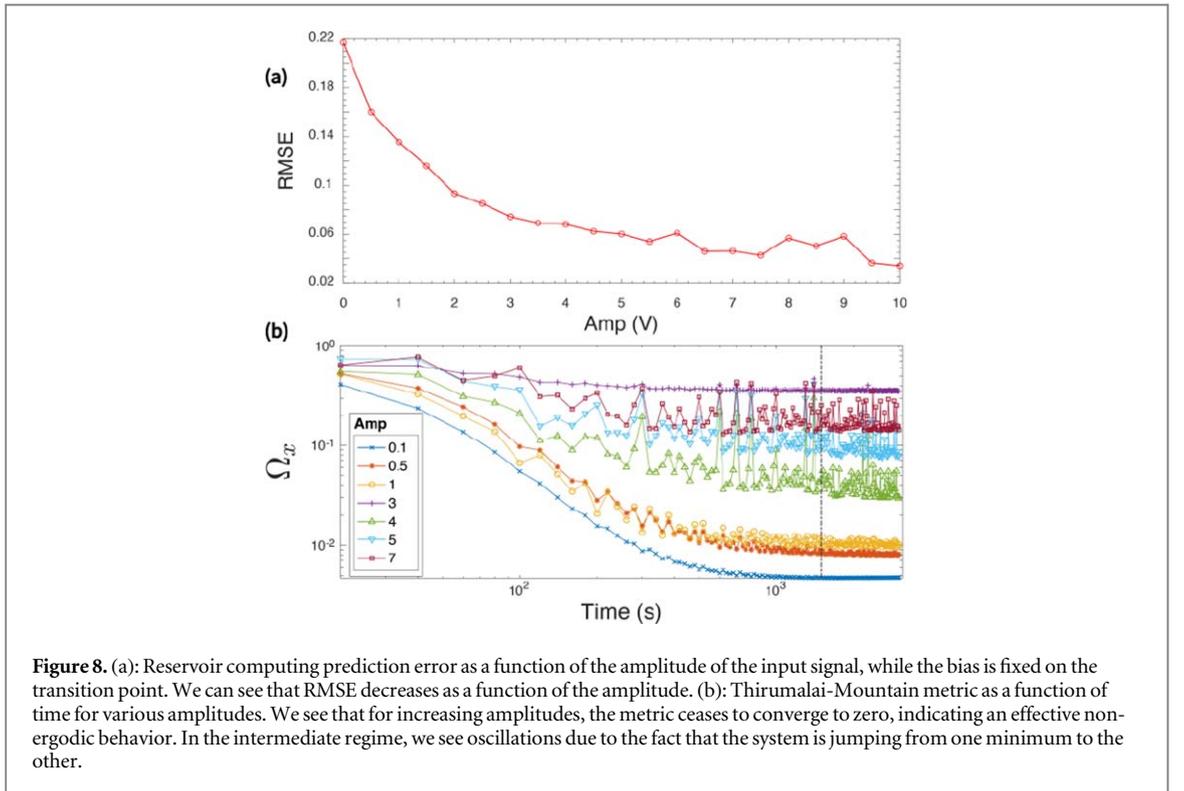
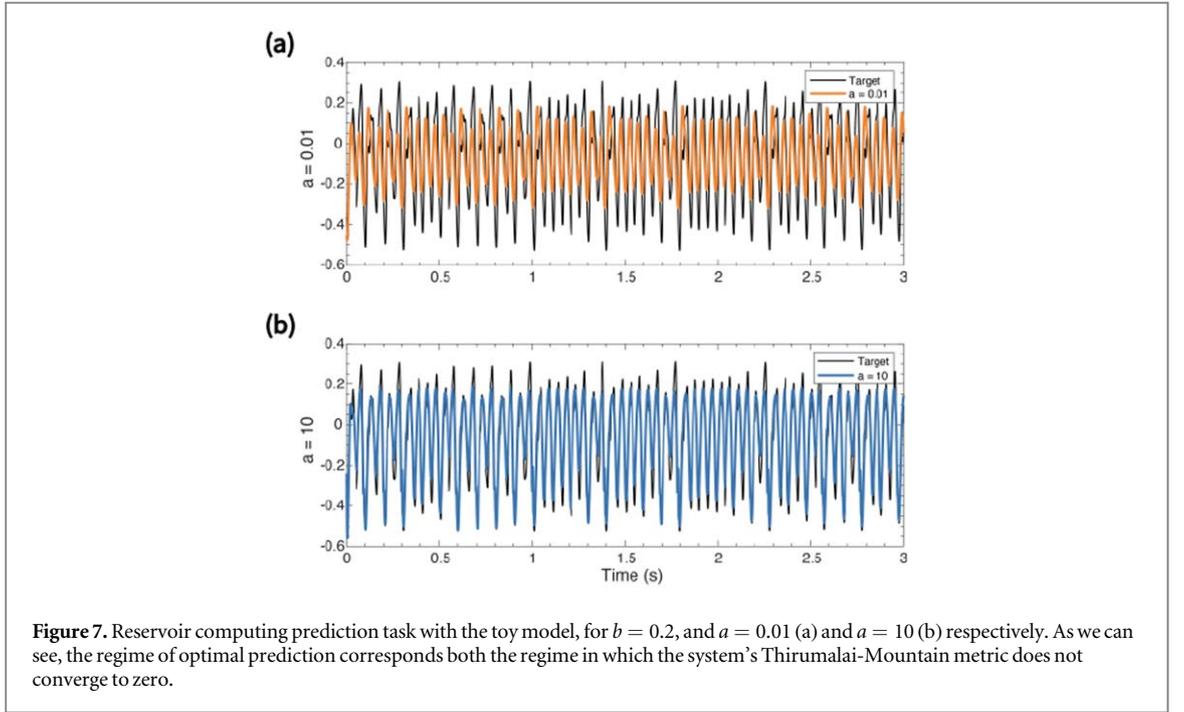


At a value of  $b = 0.2$ , a plot of the RMSE is shown in figure 8(a). As we can see, the value of the RMSE decreases as a function of  $a$ , meaning that the larger the magnitude of the input signal the better the performance, but this happens only when  $b$  is carefully chosen near the transition point. The reason why this occurs can be inferred by analyzing the response of the internal memory values, which are reported in the Appendix, as a function of the parameter  $a$  and for  $b = 0.2$ , below, near, and above the transition point. When  $b$  is such that the potential has a single minimum, the value of the memory oscillates in the vicinity of that minimum. When however  $b$  is such that the potential has two minima, e.g. in the symmetry-breaking regime, at sufficiently large values of  $a$  the memory values start to oscillate between the two minima. This implies that the time evolution of the system is much more dynamic in the symmetry-broken phase, and this effectively results in an improvement in the performance of the RC.

As we can see, increasing the value of the amplitude leads to the internal memory values fluctuating more prominently between the two stable states. This is also shown in figure 8 (b), where we plot the TM metric evaluated on the response of the system in the small  $a = 1$  and large  $a = 10$  amplitude regimes. In one first case, the system is still ergodic, while in the case  $a = 10$  the TM metric is not converging to a value zero.

Furthermore, we have calculated the  $TM_x$  metric near the transition bias  $b = 0.2$  for varying values of the amplitude  $a$  as

$$\Omega_x(t) = \frac{1}{N} \sum_j^N [\bar{x}_j(t) - \langle x(t) \rangle]^2. \quad (10)$$



As the ensemble average  $\langle x(t) \rangle$  we have consider the mean field value  $x_{cg}$  defined as [23]

$$x_{cg} = \frac{1}{N} \sum_{ij} \mathcal{P}_{ij} x_j, \quad (11)$$

where  $\mathcal{P}_{ij}$  is the *projection operator*, an operator that contains the information about the circuit topology. More information can be found in appendix A.1. We have chosen the mean field value  $x_{cg}$  as a natural order parameter for the *entire memristive network* that resembles, in definition, the general meaning of ensemble average.

This is actually one of the advantages of studying the analytical  $TiO_2$  toy model first, as in this case, we know the details of the order parameter for the whole memristive network.

## 4.2. Nanowire networks

The nanowire connectome dynamics has been extensively explored across various studies. Prior research has highlighted that nanowire Networks (NWNs) showcase brain-like dynamics, demonstrating their optimal information storage and processing capabilities at conductance transition points [21, 23, 24]. More recently, a dynamical mean-field theoretical technique for polymer-coated Ag nanowires has uncovered emergent dynamical features [61] such as transitions. In the context of the two-terminal setup, these transitions are commonly observed within a regime termed as the ‘edge of formation’ [21]. This ‘edge of formation’ is delineated by the activation of memristive components, but precedes an exponential surge in the formation of parallel paths. As depicted in figure 4, this regime establishes a select few high-conductance current paths between the two electrodes (a phenomenon known as the ‘winner-takes-all’) [62]. Within this state, the internal dynamics of the NWN intricately map the input signal to a diverse feature space.

In the case of the toy model, previous analytical studies provide a comprehensive understanding of both the system dynamics and the order parameters to be employed [23, 53], which allows us to apply a simplification for the TM metric. Nevertheless, the same technique cannot be utilized for the nanowire model since it is more realistic and cannot be characterized in the same way. For that reason, the collective conductance of the nanowire network is used as the order parameter (to calculate the ensemble average), which is determined based on the individual conductances of individual memristive junctions and the underlying circuitry shaped by the connectivity. In this case we use the collective network conductance  $G$  as it is experimentally accessible, and we employ *time multiplexing* to expand the number of features to learn from, see appendix B. In the meantime, the findings derived from the toy model will play an important role in interpreting the results of nanowires.

### 4.2.1. Thirumalai-Mountain metric

We now wish to link computation performance and ergodicity, also in the case of NWNs, we used the TM metric to understand the effective ergodicity of these dynamical systems; in particular, equation (10) is also used to calculate the TM metric for NWNs. However, the meaning associated with each quantity is closer to the definition of the metric used in the original work on supercooled liquid, [42]. The overall observable of interest is the two-probe conductance, defined earlier for nanowires, which is a scalar. Then, the time average is calculated using the global conductance, while the collective conductance is employed as an ensemble average across various realizations of the initial conditions:

$$\bar{g}(t) = \frac{1}{T} \int_0^T G^i(t) dt, \quad (12)$$

$$\langle g(t) \rangle = \frac{1}{N} \sum_i G^i(t), \quad (13)$$

where  $G^i(t)$  is the two-probe conductance of a particular realization at time  $t$  and  $G(t)$  is the collective and effective conductance of the network between the two points. For the time average quantity, note that we select a single element of the ensemble to evaluate the time average.

The ensemble of different realizations is generated by randomly perturbing the filament levels of junctions in the network:

$$\Lambda^i = (1 + \delta)\Lambda_0, \quad (14)$$

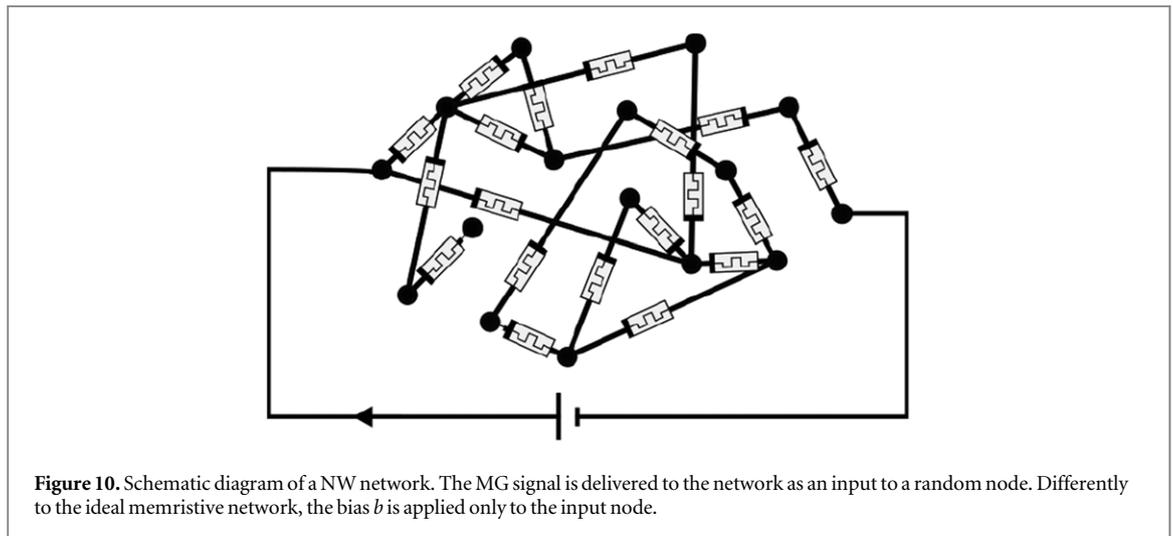
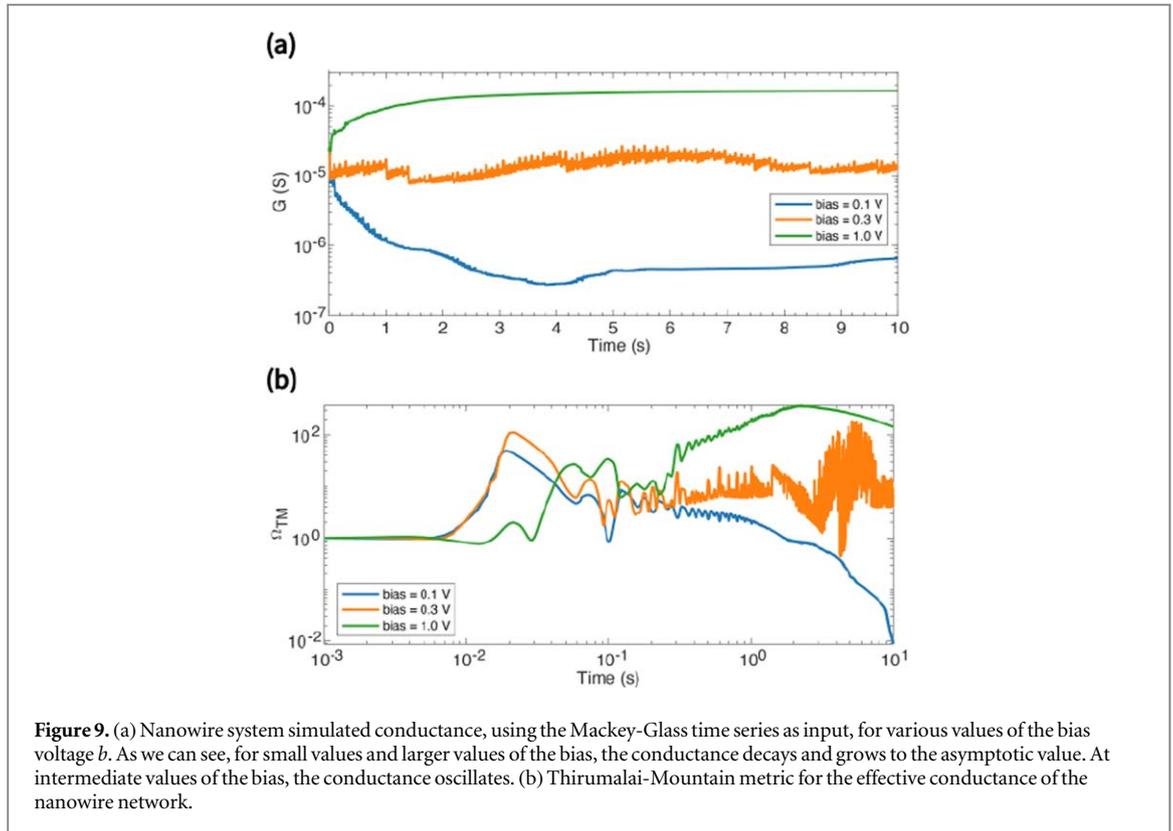
where  $\delta$  is randomly sampled from a flat distribution, while  $\Lambda^i$  is the parameter controlling the length of the filaments for all junctions, while  $\Lambda_0$  is the initial condition of the filament. The random variable  $\delta$  is sampled from a uniform distribution over  $(-0.1, 0.1)$ . Thus, effectively we are sampling over the initial conditions of the system.

The TM metric, as a function of the bias, is shown in figure 9 (b). As we can see, and consistently with the case of the toy model, for small and large values of the bias, the metric decays. At intermediate values of the bias, at the edge of the transition between low and high conductance states, where the conductance synchronizes with the input, the TM metric fails to decay. This is exactly the regime in which the conductance transition occurs and is a signature of a special state for the nanowire network, highly synchronized with the input.

### 4.2.2. Reservoir Computing

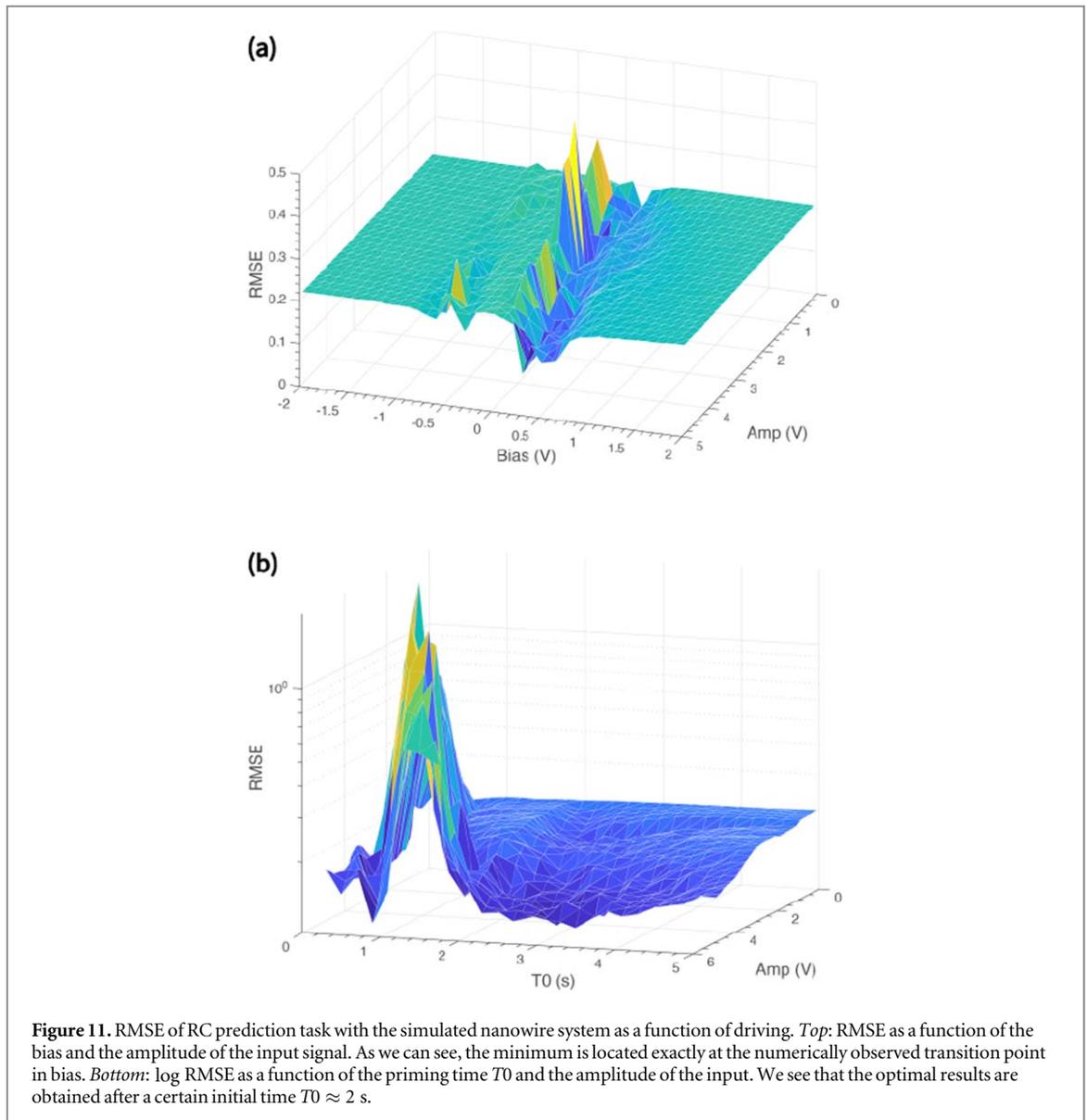
For the realistic nanowire network model above, the implementation of RC involved the two probe conductance, with designated input and readout nodes. Such construction has already been considered in the literature [57, 63].

The prediction task of the MG time series was performed using NWNs under the RC framework, with two nanowire nodes in the network selected as the source and the drain, see figure 10. The MG signal was linearly transformed as described in equation (9), and delivered to the network as input, while the same MG signal with 5 steps ahead was employed as the target. The task can be break down into three phases:



1. *Priming:* A 2 V DC input of varying length  $T_0$  was applied to drive the internal state of the network. The first 1000 data points of the MG time-series were delivered subsequently to wash out the influence from the initial state.
2. *Training:* The effective conductance of the network corresponding to  $t = 1000 - 4000$  ms was measured and multiplexed using the virtual node technique [64] to provide training features (see details in appendix B). The readout layer was trained using a linear regression with a ridge parameter  $r = 0.01$ .
3. *Testing:* The effective conductance during  $t = 4000 - 5000$  ms was measured and multiplexed in the same fashion as the training phase, while the trained readout layer was applied to make predictions. The performance of the reservoir can thus be evaluated accordingly.

We consider two regimes: one in which the system is initialized away from the voltage transition point between the high and low conductance state, and one in which the system sits at the boundary between the two. As we can see from figure 11 (top), the RMSE of the RC model has a behavior very similar to the one observed for

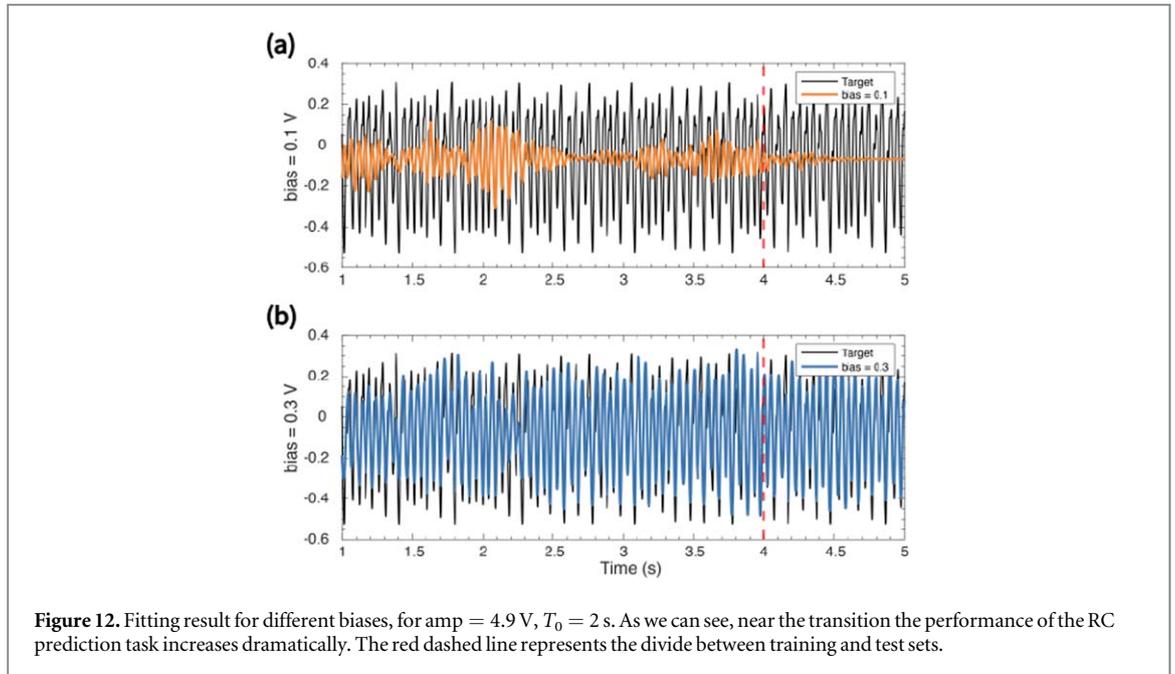


the toy model of figure 6. At values of the amplitude close to the transition point, in which the Thirumalai-Mountain fails to converge to zero, the physical RC performance peaks. Our intuition is that the system tuned at the point in which it is at the edge of a transition, is more prone to synchronization [32] with the input signal. This, combined with existing knowledge [57] on the average ‘priming’ time that it takes for the system to reach synchronization with the signal, shows that the combination of input time and choice of input voltage leads to optimal performance. We can see the difference between the tuned and non-tuned network in figure 12.

## 5. Discussion

Memory effects are a critical aspect of many physical and biological systems, and they have been shown to play a vital role in the behavior of complex systems. Meanwhile, ergodicity is a property of systems that describes the degree to which they explore their phase space. In recent years, in particular, physical systems with memory such as nanowires or nanoparticle connectomes, memristive, and other nanoscale devices, have become increasingly essential candidates as substrates for synthetic intelligent devices, e.g. brain-like physical material. In particular, there are strong indications, both in theoretical models and experiments, that these devices exhibit conductance transitions both of the first and second order. Additionally, memristive nanowire networks used as reservoirs show enhanced learning performance when in an edge of chaos state, between ordered and chaotic dynamics.

The present study explores the interplay between ergodicity breaking and memory in two models of memristive devices. The first model we studied was an analytical toy model introduced in the literature for ideal memristive, randomly connected networks, and has been instrumental in understanding their non-equilibrium



properties such as Lyapunov functions [65] or conductance transitions [23]. The second is a more realistic model for memristive networks composed of nanowires [21, 24, 57, 66], based on polymer-coated Ag nanowires [22, 67, 68]. Differently from the toy model, the nanowire network structure is heterogeneous, not random, with network properties such as small worldness and modularity, which appear to be important for efficient information trafficking [55, 69]. In both cases, we studied the Thirumalai-Mountain [42] metric to understand how the systems relax when driven by different inputs, and in particular the ergodic properties of these systems. Both in the case of the toy model, and in the case of the nanowire model, we found that the Thirumalai-Mountain metric signals lack of ergodicity near the voltage bias value where the conductance transitions are expected to be. This is a very important result as it suggests these complex physical systems are not sensitive to the specific details of the internal synapse-like dynamics or how closely they mimic biological systems. Instead, what matters most is that the functional connectivity is dynamically changing in a non-trivial way, as it also does in neuronal networks. This is interesting in view of the fact that it has been recently observed that edge of instability can be linked to a computational advantage [32, 36]. A similar result is observed in this paper. We did observe that, in particular, this is not necessarily true unless the dynamical system under study is synchronized to the input signal, as suggested in [32]. In fact, this improved performance occurs only after a transient period ('priming') which allows the nanowire network to synchronize to the input signal, as shown in figure 11. The results described above are consistent across two different models: the toy model in which conductance transitions can be understood quantitatively and analytically, and the more realistic nanowire model able to capture the experimentally observed conductance in Ag nanowires.

Overall, these results suggest the conductance transitions in memristive networks may be explained by a common underlying theory. While developing such a theory is beyond the scope of this paper, we can point to some clues. The emergent memory and learning performance found when memristive systems are non-ergodic hints at a symmetry breaking dynamical transition. Such transitions are often associated with disordered systems. In the memristive systems studied here, disorder is introduced in the initial conditions, network heterogeneity and intrinsic stochasticity. We aim to investigate this further in future work.

## 6. Conclusions

In conclusion, by connecting memory effects and ergodicity, we have identified a set of principles that can be used to create more effective computational models. Our results show for the first time that non-ergodic behavior can be linked to the effectiveness of reservoir computing, leading to new approaches for developing more advanced and efficient computational tools. Our findings will motivate further investigations into the connections between memory effects and ergodicity, and will lead to new and exciting developments in the field of physics informed machine learning. In future work we aim to apply the ideas developed in this study to working memory and meta-plasticity with memristive systems, [70, 71], in particular, in view of the recent results on the demonstration of cognitive function in nanowire networks shown in [68].

## Acknowledgments

The work of FC was carried out under the auspices of the NNSA of the U.S. DoE at LANL under Contract No. DE-AC52-06NA25396, and in particular support from LDRD via 20 230 338ER and 20 230 627ER. RZ is supported by a Postgraduate Research Excellence Award scholarship from the University of Sydney. VB acknowledges funding through the RMIT Vice-Chancellor's Research Fellowship. We would like to thank J.-P. Carbajal for comments on the paper.

## Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

## Appendix A. Memristive models

### A.1. $TiO_2$ toy model

For a circuit with multiple memristors having varying resistances  $R(x_i)$  and constant voltage generators  $S_i = \tilde{S}$  connected in series, the equation (6) for a single memristor extends to a system of coupled nonlinear ordinary differential equations. This system describes the network dynamics of the memory elements  $x_i(t)$  and can be expressed as [53, 65]:

$$\frac{d}{dt}\vec{x} = \frac{1}{\beta}(I - \chi\mathcal{P}X)^{-1}\mathcal{P}\vec{S} - \alpha\vec{x}, \quad (A1)$$

Here,  $\chi = \frac{R_{off} - R_{on}}{R_{off}} < 1$  and  $X_{ij}(t) = x_i(t)\delta_{ij}$ . The matrix  $\mathcal{P}$  represents the projection operator on the vector space of cycles in the circuit's graph  $\mathcal{G}$  [53], and it can be obtained from the directed incidence matrix  $B$  of  $\mathcal{G}$  as  $\mathcal{P} = I - B(B^t B)^{-1}B^t$  [27, 50, 56, 72]. The property of  $\mathcal{P}$  being a projection operator reflects Kirchhoff's circuit laws. In this paper, we consider  $B$  as a random matrix to abstract the dynamical system from any specific circuit topology.

It is interesting to see that there is a range of voltage in which the system experiences bistability. This is not a pitchfork bifurcation, but it is nonetheless the emergence of two (meta)stable points in the range  $[V_c, V_c^*]$  of figure 1 (bottom). We can see the relevance of this feature in the study of ergodicity as a function of voltage, as we explain below.

In order to study the ergodic versus non-ergodic behavior of such toy model, we extend the system to a stochastic one, introducing noise. The dynamical equations of a memristive device can incorporate noise through stochastic terms, introducing randomness into the system's behavior. This stochasticity is manifested in the memory component of the device. This is a diffusive memristive model of the form

$$d\vec{x} = \frac{1}{\beta}[I - \chi\mathcal{P}X]^{-1}\mathcal{P}\vec{S}dt - \alpha\vec{x}dt + \sigma(\vec{x}, t)d\vec{\eta}, \quad (A2)$$

the vector  $d\vec{\eta}$  contains  $n$  Wiener stochastic processes with  $\sigma$  the diffusion coefficient.

It is known that for the memristive toy model of equation (A1) and in the limit of dense networks, for values of  $\chi = 0.9$  and  $\bar{s} = \frac{1}{\alpha\beta} \frac{1}{N} \sum_{i=1}^N S_i \approx 0.24$  in which the system develops two mean-field minima, induces bulk dynamical transition characterized by a short transient of positive Lyapunov exponents [23]. For smaller or larger values of  $\bar{s}$  instead, the memory parameters are in a laminar regime, characterized by the presence of only a global minimum in the potential  $V(x, s)$ .

We discuss the properties of the effective potential. For  $\alpha = \beta = 1$  and  $\chi = 0.9$ , the range is  $1/10 < s < 5/18$ . The emergence of the two minima is non-perturbative requiring  $\chi$  to be close to 1. In the case of  $\chi = 0$ , the network consists of regular resistors, and the behavior of the single-element dynamical system is characterized by simple basins of attraction without any exotic features. To ensure the validity of the equations of motion for the single variables, cutoff functions, such as  $\frac{d}{d\tau}x = -W(x)f(x, y)$ , are applied, where  $W(x) = 1$  for  $0 < x < 1 \cup x = 1, f(x, y) > 0 \cup x = 0, f(x, y) < 0$ , and zero otherwise. An implicit analytical solution for this differential equation has been derived in [73].

### A.2. Junction Model

Nanowire-nanowire cross-points are modeled as electrically insulating, ionically conducting junctions [22, 28, 74, 75]. The junctions exhibit voltage-threshold memristive switching due to electro-chemical metallization and the influence of electron tunnelling transport [21]. However, the model neglects junction-

level stochastic fluctuations, which introduces further nonlinearities and lead to more complex dynamics [21]. Junction conductance,  $G = G(\lambda)$ , is determined by a state variable  $\lambda(t)$  which characterizes the length of the conducting filament responsible for memristive switching:

$$G(\lambda) = \frac{1}{R_t(\lambda) + R_{\text{on}} + \xi} + \frac{1}{R_{\text{off}}}, \quad (\text{A3})$$

where

$$\xi = \frac{R_{\text{on}}^2}{R_{\text{off}} - R_{\text{on}}}, \quad (\text{A4})$$

and the tunnelling conductance  $G_t(\lambda)$  is calculated using the low voltage Simmon's formula (equation (A5)) for a MIM junction [76]:

$$\begin{aligned} G_t(\lambda) &= [R_t(\lambda)]^{-1} \\ &= \frac{3(2m_*)^{1/2} e^{5/2} (\phi/e)^{1/2}}{A2h^2s^2} \\ &\quad \cdot \exp \left[ -\frac{4\pi(2m_*e)^{1/2}}{h} s \left( \frac{\phi}{e} \right)^{1/2} \right] \end{aligned} \quad (\text{A5})$$

$$s = \max \left[ (\Lambda_{\text{crit}} - |\Lambda|) \frac{s_{\text{max}}}{\Lambda_{\text{crit}}}, 0 \right], \quad (\text{A6})$$

with effective mass  $m_* = 0.99m_e$  and PVP layer (assumed homogeneous) thickness  $s_{\text{max}} = 5$  nm. The difference between Fermi levels of PVP and Ag determines the potential barrier  $\phi = 0.82$  eV.

$A = (0.41 \text{ nm})^2 = 0.17 \text{ nm}^2$  is the area of a face of the silver unit cell. The resistances of nanowires are considered to be negligible compared to the resistance of the junctions [77]. Consequently,  $G_t(\lambda)$  exhibits an additional nonlinear dependence on  $V$ , through the filament growth parameter  $s = s(\Lambda(V))$  which modulates junction switching due to filament formation.

All junctions start with a high resistance 'off' state initially. When the value of  $\lambda$  of a single junction reaches the predefined threshold  $\lambda_{\text{crit}}$ , the resistance of the junction switches to 'on' state. The ratio of these resistance states is  $R_{\text{off}}/R_{\text{on}} = 10^3$ , with  $R_{\text{on}} = 10^4$  Ohms.

The evolution of  $\lambda(t)$  is described by a polarity-dependent voltage-threshold model [21, 24, 28, 74]:

$$\frac{d\lambda}{dt} = \begin{cases} (|V(t)| - V_{\text{set}}) \text{sgn}[V(t)], & |V(t)| > V_{\text{set}} \\ 0, & V_{\text{reset}} < |V(t)| < V_{\text{set}} \\ b(|V(t)| - V_{\text{reset}}) \text{sgn}[\lambda(t)], & |V(t)| < V_{\text{reset}} \end{cases} \quad (\text{A7})$$

where  $V_{\text{set}}$  is the on-threshold and  $V_{\text{reset}}$  is the off-threshold, and  $b$  is a positive constant defining the relative rates of decay of the filament. All simulation results presented in this work are generated using the following parameters:  $V_{\text{set}} = 10^{-2}\text{V}$ ,  $V_{\text{reset}} = 10^{-3}\text{V}$  and  $b = 10$ . Experimental validation of this model is presented elsewhere [21, 22].

More specifically, the junction voltage distribution  $V$  across the network is obtained numerically by solving

$$\mathcal{L}^\dagger V = I, \quad (\text{A8})$$

in which  $I$  is current and  $\mathcal{L}^\dagger$  is the expanded graph Laplacian of the network, expressed in block matrix representation as

$$\mathcal{L}^\dagger = \begin{bmatrix} \mathcal{L} & C \\ C^\dagger & 0 \end{bmatrix}, \quad (\text{A9})$$

where  $\mathcal{L}$  is the graph Laplacian and where the elements of  $C$  are either 1 if the nanowire node is connected to an external electrode or 0 otherwise. The Laplacian  $\mathcal{L}$  is generated by

$$\mathcal{L} = D - W, \quad (\text{A10})$$

where  $W$  is the weighted adjacency matrix of the network, with the weights representing junction conductance distribution:

$$W_{ij} = A_{ij} G(i, j), \quad (\text{A11})$$

where  $G(i, j)$  is conductance on the edge connecting nodes  $i$  and  $j$  and  $D$  is the weighted degree matrix generated from  $W$ :

$$D = \text{diag}(d_i), \quad d_i = \sum_{k=1}^N W_{ik}. \quad (\text{A12})$$

After Kirchoff laws are solved, the parameters  $\lambda(t)$  are evolved accordingly.

## Appendix B. Reservoir computing and multiplexing

Reservoir computing is a popular approach to machine learning that utilizes network of nodes, known as a reservoir, to generate temporal patterns in response to input signals. The reservoir acts as a dynamical system that maps input signals to high-dimensional feature spaces, where a readout layer can then be trained to classify or predict the desired output. This approach is particularly effective for processing time-varying data and has been applied to a range of tasks, such as speech recognition, natural language processing, and control systems. One of the key advantages of reservoir computing is that the reservoir can be pre-designed and does not require tuning during the training phase, reducing the complexity of the learning process. Furthermore, the reservoir can be implemented using a variety of physical substrates, such as electronic circuits, optical systems, and biological neural networks, making it a versatile approach that can be adapted to a wide range of applications.

The training of a reservoir computer involves finding a set of optimal readout weights that map the reservoir states to the desired output. This can be formulated as a linear regression problem, where the readout weights  $\mathbf{W}_{out}$  are estimated by minimizing the mean squared error between the target output  $\mathbf{y}(t)$  and the estimated output  $\hat{\mathbf{y}}(t)$ :

$$\min_{\mathbf{W}_{out}} \frac{1}{T} \sum_{i=1}^T |\mathbf{y}(t_i) - \mathbf{W}_{out} \mathbf{x}(t_i)|^2$$

where  $\mathbf{x}(t_i)$  is the reservoir state at time step  $t_i$ . The solution to this optimization problem can be obtained using the pseudo-inverse of the reservoir state matrix  $\mathbf{X}$ :

$$\mathbf{W}_{out} = (\mathbf{X}^t \mathbf{X} + \alpha \mathbf{I})^{-1} \mathbf{X}^t \mathbf{Y}$$

with  $^t$  representing the vector or matrix transposition, where  $\mathbf{Y}$  is the target output matrix, and  $\alpha$  is a regularization parameter that prevents overfitting. The reservoir state matrix is defined as:

$$\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)]^t$$

and the target output matrix is defined as:

$$\mathbf{Y} = [\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(N)]^t,$$

where  $N$  is the total number of training samples.

Once the readout weights are estimated, they can be used to generate the output for new input signals by computing:

$$\hat{\mathbf{y}}(t_i) = \mathbf{W}_{out} \mathbf{x}(t_i)$$

where  $\hat{\mathbf{y}}(t_i)$  is the estimated output at time step  $t$ . In the following, we will assume that  $\mathbf{x}(t)$  are either the internal memory states or the conductance states.

In the case of a single conductance measured, in order to multiply the number of virtual nodes, we used the multiplexing technique introduced in [64]. The time series is time multiplexed at certain sampling frequencies. Then, the time series is parsed into  $N$  vectors of length  $K$ , such that  $NK$  is the length of the time series. Then, elements of such vectors at locations  $mN + l$  for integer  $l \in [1, K]$  are considered as different internal virtual nodes. The regression then follows the same scheme as the previous section.

Reservoir computing has been successfully applied in various physical systems, including opto-electronic and nanoscale systems such as silver nanowire networks (NWNs) to implement neuromorphic computing. NWNs are particularly interesting for this application due to their memristive behavior and complex network topology that resembles that of biological neural networks. By leveraging the dynamics of NWNs as a reservoir, the input signals can be mapped to a high-dimensional state space through a random projection, followed by a simple readout operation to obtain the desired output. This allows for the implementation of complex computational tasks such as pattern recognition, time-series prediction, and control in a highly efficient and parallel manner. In this way, reservoir computing offers a promising approach to harness the capabilities of NWNs for neuromorphic computing, and silver nanowire learning represents an exciting area of research at the intersection of nanotechnology and machine learning.

To be considered a feasible reservoir, a dynamical system must satisfy the condition that its state approaches a nontrivial function of the input trajectory in the long time limit. This can be expressed in terms of two requirements [59]:

*Fading Memory:* If the system were started from two different initial conditions while driven with the same input trajectory  $u$ , the system's trajectories should eventually converge to the same state. This implies that the system has a finite temporal memory;

*State Separation:* Different input sequences should drive the system into different trajectories. That is, if the same initial condition were driven with two different input trajectories, the resulting reservoir trajectories must

be sufficiently different. These properties could be proven analytically for the case of the memristive reservoir toy model.

These conditions ensure that the state of the reservoir becomes a function of the input trajectory and that this function carries information about the input trajectory. In the case of memristive systems, the memristive junctions act as the dynamic elements of the reservoir, and the network dynamics arise from the collective behavior of the junctions. The input signal drives the memristive junctions to switch between high and low resistance states, resulting in a temporal sequence of network states. The final state of the network can be used as a high-dimensional feature vector for further processing. In the case of silver nanowire networks, the nanowires themselves act as the memristive elements, and their collective behavior gives rise to the network dynamics. The input signal drives the nanowires to switch between metallic and insulating states, resulting in a temporal sequence of network states that can be used for further processing.

As we saw above, reservoir computing requires the fading memory property for the dynamics of a reservoir, which is equivalent to the system tending to some degree of effective ergodicity when expressed in physical terms. The scope of this manuscript is to investigate fading memory and understand how quickly a reservoir ‘forgets’ its initial conditions, using the Thirumalai-Mountain metric as our benchmark. The goal is to determine the optimal level of memory capacity for different system parameters. However, there are limitations to using ergodicity and the ergodic theorem, as ensemble averages are only defined for systems at equilibrium. Nonetheless, we can still consider the case of systems locally at equilibrium.

The original formulation of the TM metric aimed to predict the dynamics of a system numerically simulated on a computer without having to calculate the Lyapunov exponents of systems with a high number of variables. This is particularly relevant when dealing with large-scale systems, where calculating Lyapunov exponents can be computationally expensive.

Let us now briefly introduce the time series we used for our benchmarks. The Mackey-Glass time series is a widely used benchmark for testing the performance of time-series prediction models. It is a nonlinear, nonautonomous, and nonperiodic time series that exhibits chaotic dynamics. The series is defined by the following difference equation:

$$\frac{dz(t)}{dt} = r \frac{z(t - \tau)}{1 + z(t - \tau)^n} - \gamma z(t), \quad (\text{B1})$$

where  $z(t)$  is the value of the time series at time  $t$ ,  $\beta$ ,  $\gamma$ ,  $n$ , and  $\tau$  are parameters that control the dynamics of the system. The Mackey-Glass time series exhibits a range of dynamic behaviors, depending on the values of these parameters. For example, for  $r = 0.2$ ,  $\gamma = 0.1$ ,  $\tau = 17$ , and  $n = 10$ , the time series exhibits chaotic dynamics with a positive Lyapunov exponent.

## ORCID iDs

Valentina Baccetti  <https://orcid.org/0000-0002-5680-1574>

Ruomin Zhu  <https://orcid.org/0000-0002-2310-7762>

Zdenka Kuncic  <https://orcid.org/0000-0001-6765-3215>

Francesco Caravelli  <https://orcid.org/0000-0001-7964-3030>

## References

- [1] Mehonic A, Sebastian A, Rajendran B, Simeone O, Vasilaki E and Kenyon A J 2020 Memristors—from in-memory computing, deep learning acceleration, and spiking neural networks to the future of neuromorphic and bio-inspired computing *Advanced Intelligent Systems* **2** 2000085
- [2] Oliver W D 2019 Quantum computing takes flight *Nature* **574** 487
- [3] Böhm F, Verschaffelt G and der Sande G V 2019 A poor man’s coherent ising machine based on opto-electronic feedback systems for solving optimization problems *Nature Comm.* **10** 3538
- [4] Pierangeli D, Marucci G and Conti C 2019 Large-scale photonic ising machine by spatial light modulation *Phys. Rev. Lett.* **122** 213902
- [5] Vadlamani S K, Xiao T P and Yablonovitch E 2020 Physics successfully implements lagrange multiplier optimization *Proc. of the Nat. Aca. of Sci.* **117** 26639
- [6] Csaba G and Porod W 2020 Coupled oscillators for computing: A review and perspective, *App Phys. Rev.* **7** 011302
- [7] Goto H et al 2021 High-performance combinatorial optimization based on classical mechanics *Sci. Adv.* **7** eabe7953
- [8] Singh G et al 2019 Near-memory computing: past, present, and future *Micro. and Micro.* **71** 102868
- [9] Jelmini D and Wong H-S P 2018 In-memory computing with resistive switching devices *Nature Ele.* **1** 333
- [10] Traversa F L et al 2014 Dynamic computing random access memory *Nanotechnology* **25** 285201
- [11] Sebastian A et al 2020 Memory devices and applications for in-memory computing *Nature Nano.* **15** 529
- [12] Traversa F L and Di Ventra M 2015 Universal memcomputing machines *IEEE Trans. on Neur. Net. and Learn. Sys.* **26** 2702
- [13] Di Ventra M and Traversa F L 2018 Perspective: memcomputing: leveraging memory and physics to compute efficiently *J. of App. Phys.* **123** 180901
- [14] Hennessy J L and Patterson D A 2019 A new golden age for computer architecture *Comm. of the ACM* **62** 48

- [15] Sutton B *et al* 2017 Intrinsic optimization using stochastic nanomagnets *Sci. Rep.* **7** 44370
- [16] Torrejon J *et al* 2017 Neuromorphic computing with nanoscale spintronic oscillators *Nature* **547** 428
- [17] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 Optimization by simulated annealing *Science* **220** 671
- [18] Dorigo M and Stützle T 2004 *Ant Colony Optimization* (MIT)
- [19] Milano G, Porro S, Valov I and Ricciardi C 2019 Recent developments and perspectives for memristive devices based on metal oxide nanowires *Advanced Electronic Materials* **5** 1800909
- [20] Kuncic Z and Nakayama T 2021 Neuromorphic nanowire networks: principles, progress and future prospects for neuro-inspired information processing *Advances in Physics: X* **6** 1894234
- [21] Hochstetter J *et al* 2021 Avalanches and edge-of-chaos learning in neuromorphic nanowire networks *Nature Comm.* **12** 4008
- [22] Diaz-Alvarez A, Higuchi R, Sanz-Leon P, Marcus I, Shingaya Y, Stieg A Z, Gimzewski J K, Kuncic Z and Nakayama T 2019 Emergent dynamics of neuromorphic nanowire networks *Sci. Rep.* **9** 14920
- [23] Caravelli F, Sheldon F and Traversa F L 2021 Global minimization via classical tunneling assisted by collective force field formation *Science Advances* **7** 1542
- [24] Zhu R, Hochstetter J, Loeffler A, Diaz-Alvarez A, Nakayama T, Lizier J T and Kuncic Z 2021 Information dynamics in neuromorphic nanowire networks *Sci. Rep.* **11** 13047
- [25] Caravelli F, Traversa F, Bonnin M and Bonani F 2023 Projective embedding of dynamical systems: Uniform mean field equations *Physica D* **450** 133747
- [26] Palmer R 1982 Broken ergodicity *Adv. Phys.* **31** 669
- [27] Caravelli F 2017 The mise en scène of memristive networks: effective memory, dynamics and learning *Int. J. of Par., Em. and Dist. Sys.* **33** 350
- [28] Kuncic Z *et al* 2020 Neuromorphic Information Processing with Nanowire Networks 2020 *IEEE International Symposium on Circuits and Systems (ISCAS)* pp 1–5
- [29] Chua L 1971 Memristor-the missing circuit element *IEEE Trans. Circuit Theory* **18** 507
- [30] Jaeger H and Haas H 2004 Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication *Science* **304** 78
- [31] Grigoryeva L and Ortega J-P 2018 Echo state networks are universal *Neural Netw.* **108** 495
- [32] Carroll T L 2020 Do reservoir computers work best at the edge of chaos?, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **30** 121109
- [33] Tagliazucchi E, Balenzuela P, Fraiman D and Chialvo D R 2012 Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis *Front Physiol.* **3** 15
- [34] Morales G B and Muñoz M A 2021 Optimal input representation in neural systems at the edge of chaos *Biology* **10** 702
- [35] Packard N H 1988 Adaptation toward the edge of chaos *Dynamic Patterns in Complex Systems* **212** 293
- [36] Langton C G 1990 Computation at the edge of chaos: phase transitions and emergent computation *Physica D* **42** 12
- [37] Vincent E 2007 Ageing, rejuvenation and memory: the example of spin-glasses *Ageing and the Glass Transition* ed M Henkel *et al* (Springer) pp 7–60
- [38] Saccone M, Caravelli F, Hofhuis K, Parchenko S, Birkhölzer Y A, Dhuey S, Kleibert A, van Dijken S, Nisoli C and Farhan A 2022 Direct observation of a dynamical glass transition in a nanomagnetic artificial hopfield network *Nat. Phys.* **18** 517
- [39] Saccone M *et al* 2023 Real-space observation of ergodicity transitions in artificial spin ice *Nat. Commun.* **14** 5674
- [40] Dorfman J R 1999 *An Introduction to Chaos in Nonequilibrium Statistical Mechanics* (Cambridge university Press)
- [41] Ma S-K 1985 *Statistical Mechanics* (World Scientific)
- [42] Thirumalai D, Mountain R D and Kirkpatrick T R 1989 Ergodic behavior in supercooled liquids and in glasses *Phys. Rev. A* **39** 3563
- [43] Petersen K 1983 *Ergodic Theory* (World Scientific)
- [44] Mountain R D and Thirumalai D 1989 Measures of effective ergodic convergence in liquids *The Journal of Physical Chemistry* **93** 6975
- [45] Thirumalai D and Mountain R D 1993 Activated dynamics, loss of ergodicity, and transport in supercooled liquids *Phys. Rev. E* **47** 479
- [46] Bel G and Barkai E 2006 Ergodicity breaking in a deterministic dynamical system *Europhys. Lett.* **74** 15
- [47] Tiampo K F, Rundle J B, Klein W, Martins J S S and Ferguson C D 2003 Ergodic dynamics in a natural threshold system *Phys. Rev. Lett.* **91** 238501
- [48] Tiampo K F, Rundle J B, Klein W, Holliday J, Sá Martins J S and Ferguson C D 2007 Ergodicity in natural earthquake fault networks *Phys. Rev. E* **75** 066107
- [49] Stützen M 2014 Effective ergodicity in single-spin-flip dynamics *Phys. Rev. E* **90** 032141
- [50] Caravelli F 2017 Locality of interactions for planar memristive circuits *Phys. Rev. E* **96** 052206
- [51] Joglekar Y N and Wolf S J 2009 The elusive memristor: properties of basic electrical circuits *Eur. J. of Phys.* **30** 661
- [52] Prodromakis T *et al* 2011 A versatile memristor model with nonlinear dopant kinetics *IEEE Trans. on El. Dev.* **58** 3099
- [53] Caravelli F, Traversa F L and Di Ventra M 2017 Complex dynamics of memristive circuits: Analytical results and universal slow relaxation *Phys. Rev. E* **95** 022140
- [54] Li Q, Diaz-Alvarez A, Iguchi R, Hochstetter J, Loeffler A, Zhu R, Shingaya Y, Kuncic Z, Uchida K-I and Nakayama T 2020 Dynamic electrical pathway tuning in neuromorphic nanowire networks *Adv. Funct. Mater.* **30** 2003679
- [55] Loeffler A, Zhu R, Hochstetter J, Li M, Fu K, Diaz-Alvarez A, Nakayama T, Shine J M and Kuncic Z 2020 Topological properties of neuromorphic nanowire networks *Front Neuroscience* **14**
- [56] Nilsson J W and Riedel S 2011 *Electric Circuits* 9th edn (Pearson Education)
- [57] Fu K *et al* 2020 Reservoir computing with neuromemristive nanowire networks *Proceedings of the International Joint Conference on Neural Networks (IJCNN)* 20006228
- [58] Caravelli F and Sheldon F C 2020 Phases of memristive circuits via an interacting disorder approach 2009.00114 arXiv: 2009.00114
- [59] Sheldon F C, Kolchinsky A and Caravelli F 2022 The computational capacity of lrc, memristive and hybrid reservoirs *Phys. Rev. E* **106** 045310
- [60] Mackey M C and Glass L 1977 Oscillation and chaos in physiological control systems *Science* **197** 287
- [61] Caravelli F, Milano G, Ricciardi C and Kuncic Z 2023 Mean field theory of self-organizing memristive connectomes *Ann. Phys.* **535** 2300090
- [62] Manning H *et al* 2018 Emergence of winner-takes-all connectivity paths in random nanowire networks *Nat. Commun.* **9** 3219
- [63] Daniels R, Mallinson J, Heywood Z, Bones P, Arnold M and Brown S 2022 Reservoir computing with 3D nanowire networks *Neural Netw.* **154** 122
- [64] Appeltant L, Soriano M, der Sande G V, Danckaert J, Massar S, Dambre J, Schrauwen B, Mirasso C and Fischer I 2011 Information processing using a single dynamical node as complex system *Nat. Commun.* **2** 468

- [65] Caravelli F 2019 Asymptotic behavior of memristive circuits *Entropy* **21** 789
- [66] Zhu R, Hochstetter J, Loeffler A, Diaz-Alvarez A, Stieg A, Gimzewski J, Nakayama T and Kuncic Z 2020 Harnessing adaptive dynamics in neuro-memristive nanowire networks for transfer learning 2020 *International Conference on Rebooting Computing (ICRC)* pp 102–6
- [67] Milano G *et al* 2022 In materia reservoir computing with a fully memristive architecture based on self-organizing nanowire networks *Nat. Mater.* **21** 195–202
- [68] Loeffler A, Diaz-Alvarez A, Zhu R, Ganesh N, Shine J M, Nakayama T and Kuncic Z 2023 Neuromorphic learning, working memory, and metaplasticity in nanowire networks *Science Advances* **9** eadg3289
- [69] Loeffler A, Zhu R, Hochstetter J, Diaz-Alvarez A, Nakayama T, Shine J M and Kuncic Z 2021 Modularity and multitasking in neuro-memristive reservoir networks *Neuromorphic Computing and Engineering* **1** 014003
- [70] Chialvo D and Bak P 1999 Learning from mistakes *Neuroscience* **90** 1137
- [71] Carbajal J P, Martin D A and Chialvo D R 2022 Learning by mistakes in memristor networks *Phys. Rev. E* **105**
- [72] Bollobás B 1998 *Modern Graph Theory* (Springer)
- [73] Sheldon F C, Caravelli F and Coffrin C 2020 Fully analog memristive circuits for optimization tasks: a comparison *Handbook of Unconventional Computing* ed A Adamatzky (University of the West of England, Bristol, UK: World Scientific ) (WSPC book series in *Unconventional Computing*) 2
- [74] Pershin Y V, Slipko V A and Di Ventra M 2013 Complex dynamics and scale invariance of one-dimensional memristive networks *Phys. Rev. E* **87** 022116
- [75] Sillin H O, Aguilera R, Shieh H-H, Avizienis A V, Aono M, Stieg A Z and Gimzewski J K 2013 A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing *Nanotechnology* **24** 384004
- [76] Simmons J G 1963 Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film *J. Appl. Phys.* **34** 1793
- [77] Bellew A T, Manning H G, Gomes da Rocha C, Ferreira M S and Boland J J 2015 Resistance of single Ag nanowire junctions and their role in the conductivity of nanowire networks *ACS Nano* **9** 11422