



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/200221/>

Preprint:

Dale, Matthew, Evans, Richard F. L., Jenkins, Sarah et al. (2021) Reservoir Computing with Thin-film Ferromagnetic Devices. [Preprint]

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

RESERVOIR COMPUTING WITH THIN-FILM FERROMAGNETIC DEVICES

Matthew Dale^{1,*}, Richard F. L. Evans², Sarah Jenkins², Simon O’Keefe¹, Angelika Sebald³, Susan Stepney¹,
Fernando Torre¹, and Martin Trefzer⁴

¹Department of Computer Science, University of York, UK

²Department of Physics, University of York, UK

³Department of Chemistry, University of York, UK

⁴Department of Electronic Engineering, University of York, UK

*matt.dale@york.ac.uk

ABSTRACT

Advances in artificial intelligence are driven by technologies inspired by the brain, but these technologies are orders of magnitude less powerful and energy efficient than biological systems. Inspired by the nonlinear dynamics of neural networks, new unconventional computing hardware has emerged with the potential for extreme parallelism and ultra-low power consumption. Physical reservoir computing demonstrates this with a variety of unconventional systems from optical-based to spintronic [1]. Reservoir computers provide a nonlinear projection of the task input into a high-dimensional feature space by exploiting the system’s internal dynamics. A trained readout layer then combines features to perform tasks, such as pattern recognition and time-series analysis. Despite progress, achieving state-of-the-art performance without external signal processing to the reservoir remains challenging. Here we show, through simulation, that magnetic materials in thin-film geometries can realise reservoir computers with greater than or similar accuracy to digital recurrent neural networks. Our results reveal that basic spin properties of magnetic films generate the required nonlinear dynamics and memory to solve machine learning tasks. Furthermore, we show that neuromorphic hardware can be reduced in size by removing the need for discrete neural components and external processing. The natural dynamics and nanoscale size of magnetic thin-films present a new path towards fast energy-efficient computing with the potential to innovate portable smart devices, self driving vehicles, and robotics.

Main

Performing machine learning at ‘the edge’ is a growing area of interest, where inference is performed locally in real time [2, 3, 4]. Embedded devices that can perform complex information processing without the need to outsource to remote servers are ideal for real-time applications. However, current systems are limited by processing speeds, memory, size, and power consumption. Unconventional hardware is a potential alternative to classical computing hardware, with low-energy consumption, inherent parallelism, and no separation between processor and memory (the von Neumann bottleneck) [5]. Neuro-inspired hardware [6] is one route to embed machine learning at the edge, another is to exploit embodied computation in novel dynamical systems.

By design, neural-based hardware implements the abstract behaviour of neurons and their connectivity at the lowest circuit level, e.g. weighted summation, threshold functions, synapses. This typically requires a combination of simpler components to implement the model. For example, a single neuron with conventional complementary metal–oxide–semiconductor technology takes 10s to 100s of transistors to replicate a neuron-synapse circuit [7, 8]. Another option is to force the neuron model directly onto the material to improve energy-efficiency and reduce the physical footprint, yet model constraints may require removal of useful natural properties (e.g. variability in components) or require additional engineering [9]. Here we demonstrate an alternative approach exploiting the dynamical behaviours of neural systems without the direct implementation of neural units, allowing further reductions in size and efficiency.

Dynamical properties that occur naturally within complex materials, such as memory, nonlinear oscillation, and chaos can be directly exploited for computation, with less top-down engineering of the material. However, the discovery and control of intractable or unknown material properties raises new challenges.

Two novel approaches have been proposed to exploit the embodied computation of materials: evolution *in materio* and reservoir computing. Miller and Downing [10] proposed using artificial evolution as a mechanism to exploit and configure materials, arguing natural evolution is the method *par excellence* for exploiting the physical properties of materials.

Evolution *in materio* uses computer-controlled manipulation of external stimuli to configure the material and its input-output mapping, using digital computers to directly evolve physical material configurations. A range of materials have been evolved to perform classification, real-time robot control and pattern recognition [11, 12, 13, 14].

Reservoir computing is a neuro-inspired framework that harnesses the high-dimensionality and temporal properties of recurrent networks and novel systems [15, 16]. Physical implementations of the reservoir model are diverse [17, 18, 19] with recent spintronic reservoirs showing some key advantages compared to other systems, combining GHz+ operating frequencies, ultra-compact size and ultra-low-energy consumption [20, 21, 22, 23, 24, 25, 26, 27].

Here we demonstrate material computation with ferromagnetic materials in thin nano-film geometries, combining both evolution *in materio* and reservoir computing methods. The reservoir model is used to harness the propagation of information through magnetic films, and artificial evolution is used to optimise reservoir parameters. Using open-source simulation software, we evolve three ferromagnetic materials to solve three time-dependent tasks of increasing complexity. All materials are evaluated at various film sizes with direct comparisons to equivalent-sized recurrent neural networks. The magnetic system is then characterised by metrics to understand the dynamical properties of each material. Lastly, the effects of temperature and film size are explored to inform future physical implementations.

1 Magnetic Reservoir System

Reservoir computers are composed of three layers: input, reservoir and output layer (Fig. 1a). A reservoir, typically a fixed random network of discrete processing nodes with recurrent connections, features non-linear characteristics and a short-term memory. The reservoir network is driven by a time-varying input u that propagates through a random input mapping via connection weights \mathbf{W}_{in} (see Methods). The non-linear reservoir provides a high-dimensional projection of the input from which a subsequent linear readout layer can extract features relevant to the problem task. Training occurs only at the readout through trained weighted connections \mathbf{W}_{out} connecting observable states to the final output. Typically, one-shot learning is used through linear regression, making learning extremely fast.

Fig. 1b details the layout of the proposed magnetic system and its reservoir representation. The film does not possess any discrete processing nodes; our representation of the system defines discrete “cells” for the purpose of input and output locations. The film is conceptually divided into a grid of magnetic cells; each cell is connected to a time-varying input signal source and a bias source via weighted connections \mathbf{W}_{in} . The output of each cell is represented by a three-dimensional magnetisation vector X_{xyz} . This approach models a grid of nano-contacts across the film, measuring a low-resolution snapshot of the film’s magnetic state.

The reservoir thin films are simulated micromagnetically where the atomistic detail is coarse grained into 5 nm cells (see Methods). Here we consider three simple ferromagnetic metals: Cobalt (Co), Nickel (Ni) and Iron (Fe). The atomic magnetic properties of these materials are well understood from first principle calculations [28], providing a detailed insight into microscopic and macroscopic magnetic behaviour. These metals are abundant in nature, inexpensive and highly stable.

As a thin film, the reservoir is highly-structured. The influence each cell has on its nearest neighbours is determined by the physical properties of exchange, anisotropy, and dipole Hamiltonian (see Methods). The exchange interactions dominate over short lengthscales, meaning that cells have finite time- and spatial correlations over the total sample size. Fig. 1c shows a typical simulated micromagnetic response to three input pulses at the films centre. When perturbed, spin waves propagate through the film inducing reflections, oscillations and interference patterns. At the edges, a similar characteristic response is seen per impulse but with some contributions from previous stimuli.

To exploit the fast spin dynamics of the ferromagnetic materials, data inputs are applied at 10ps intervals (100 GHz). Selecting a suitable input timescale depends on the material’s dynamics. An input faster or slower than the system’s intrinsic timescale alters the temporal dynamics and thus can affect settling times, refractory periods and memory in the system. The inherent volatility and nonlinear dynamics of the spin precession provides a temporal mapping of the input into different reservoir states.

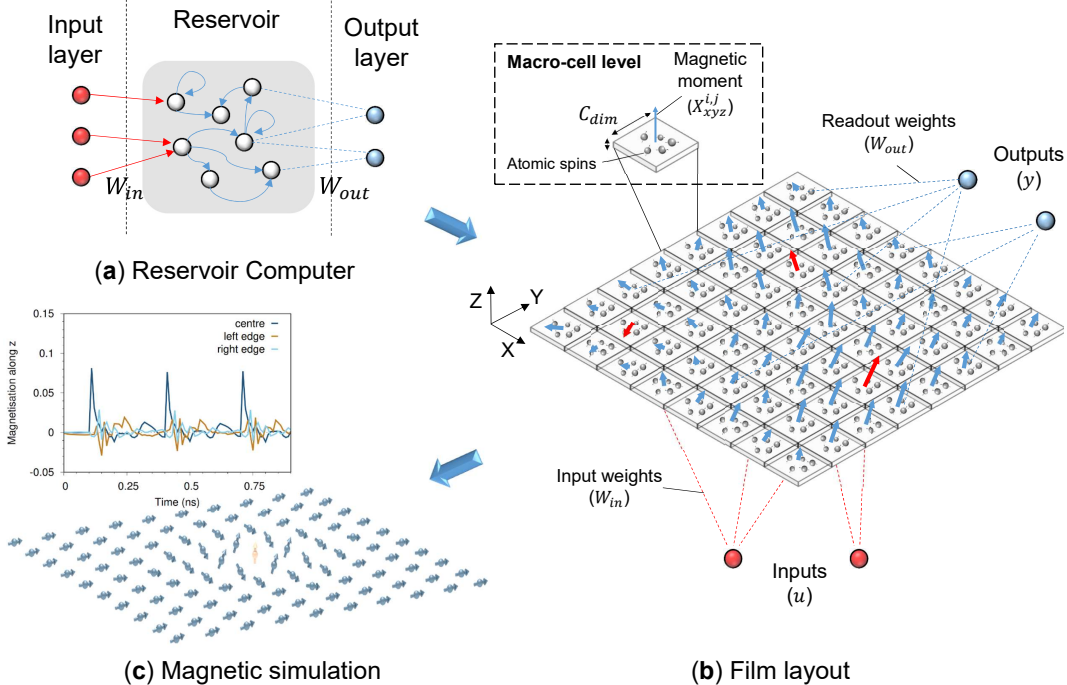


Figure 1: a) Reservoir computing model split into input, reservoir and output layers connected by adjustable weights. The reservoir is self-contained, typically featuring a sparse, recurrent network of processing nodes. b) Schematic of our simulated thin-film magnetic reservoir system, consisting of micromagnetic cells derived from atomistic values. Global input sources u connect via weights W_{in} to drive local magnetisation fields inducing spin oscillations. Each cell’s average magnetic moment produces a 3-d orientation vector X_{xyz} forming a reservoir state. States are then combined via a linear readout function W_{out} to produce the final system output y . c) Impulse response of micromagnetic spin system. Signal injected in the centre of the film via the z -axis at 25 time-step intervals with 10 ps scanning frequency.

To evaluate the materials, three temporal tasks are evaluated. The time series prediction Santa Fe chaotic laser data set [29] is chosen for its nonlinear properties and periodic structure, and the nonlinear autoregressive moving average model (NARMA) with lags of 10 (NARMA-10) and 30 (NARMA-30) time-steps are chosen to evaluate the film’s ability to manage the nonlinearity-memory trade off [30]. Each benchmark increases in difficulty, demonstrating the film’s dynamic range and ability to perform increasingly complex tasks.

2 Results

Our experimental results show the investigated materials are competitive to state-of-the-art reservoir networks, and typically outperform small networks with equivalent reservoir size. Fig. 2 shows the performance of each material at three film sizes. Four types of recurrent neural networks are provided for the comparison, including random and evolved networks, and networks with limited connectivity. As reservoir-internal connections are typically random, baseline comparisons to random networks are included. Highly-structured networks, such as a lattice, more accurately model the material crystal structure. Lattice networks with recurrent connections have been shown to be dynamically similar to less restrictive recurrent neural networks, but often have to compensate with larger network size [31, 32, 33].

For the laser task (Fig. 2, top row), all materials significantly outperform random networks at small sizes. At the largest size (225 nodes, right column), only Co outperforms random networks, however, Ni and Fe remain statistically similar. At the smallest film size, all materials outperform evolved networks. At 100 nodes, only Co outperforms evolved networks with a normalised mean square error (NMSE) of roughly 3.5×10^{-3} , the smallest error found. At 100 nodes, Ni and Fe remain statistically similar to evolved networks. For the laser task, even the smallest magnetic reservoirs here outperform larger material reservoirs reported in the literature [34, 35].

For the NARMA-10 task (Fig. 2, middle row), all materials outperform random networks at small sizes. At 225 nodes, all materials are statistically similar to random lattices but worse than other networks. In some cases, materials are

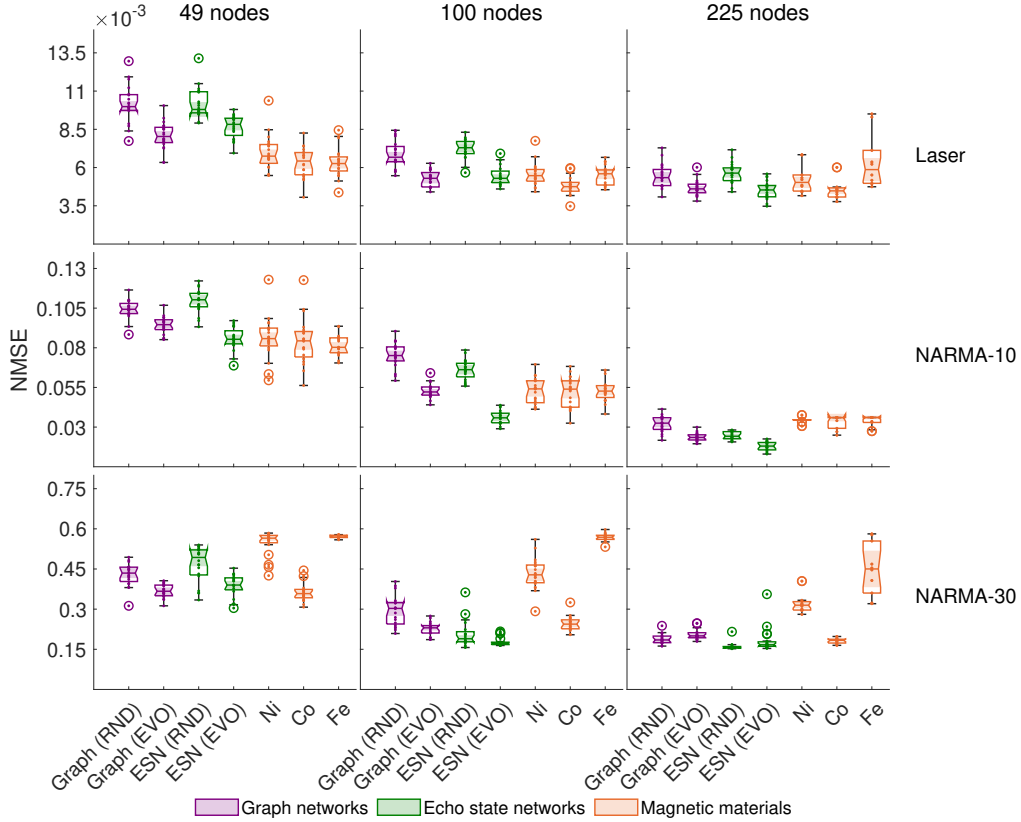


Figure 2: Performance of materials and simulated reservoir networks on benchmark tasks. Normalised mean square error (NMSE) is used to compare equivalent-sized reservoirs. Multiple reservoir sizes are displayed in columns, and each task is divided into rows. Each type of system is represented by colour (lattice reservoir=purple; echo state reservoir=green; material=orange). The method used to create the reservoirs is given on the x -axis (random or evolved). For random search, the best reservoir from a batch of 2000 instances is shown, repeated over 20 batches. For evolved, the final evolved reservoirs are given from 20 evolutionary runs of 2000 evaluations each.

better than, or similar to, evolved networks, which have unrestricted access to long-distance connections. The lowest material errors found on this task are $NMSE = 0.056$ (Co, 49 nodes), 0.032 (Co, 100) and 0.025 (Co, 225). These are highly competitive to, or outperform other, material reservoirs reported in the literature, such as optoelectronic ($NMSE \approx 0.168$, 50 nodes [36]) and digital reservoirs ($NMSE \approx 0.023$, 400 node delay-line [17]).

For the NARMA-30 task (Fig. 2, bottom row), the difference between materials becomes clear. Each material performs differently, with Co being able to better match the dynamics of the task. Across all sizes, Co is competitive to random and evolved networks. The lowest error found is $NMSE \approx 0.165$ at 225 nodes. Ni and Fe struggle to compete with other networks at small sizes; nevertheless, as size increases, $NMSE$ decreases. This suggests that these materials require larger films to exhibit the necessary dynamics to perform the tasks.

The NARMA-30 task results show a strong distinction between the materials, despite their similar performances on other tasks. To understand this further, task-independent measures are used to assess non-linearity and memory. These measures better determine the general underlying dynamics of the system than tasks can achieve alone. They have been used to qualitatively assess the dynamical range of materials for reservoir computing [37, 31] and to determine a system's total information processing capacity [38]. Here, the non-linear projection and short-term memory are measured, using the kernel rank (KR) [39] and linear memory capacity (MC) [40] of the reservoir (see Supplementary Material). Fig. 3 shows values of these measures for each of the material reservoirs used in the NARMA-30 task (see Supplementary Material for all tasks). The Co material (orange) tends to cluster around a normalised $KR \approx 0.25$ and an $MC \approx 30$. This suggests it is exploiting a weak non-linearity and a large memory to perform the task, which corresponds to the known dynamics of the task (see eq. 13). Ni (green) typically has smaller memory than Co but larger than Fe (black), explaining its intermediate performance. Fe features small values in both KR and MC across all

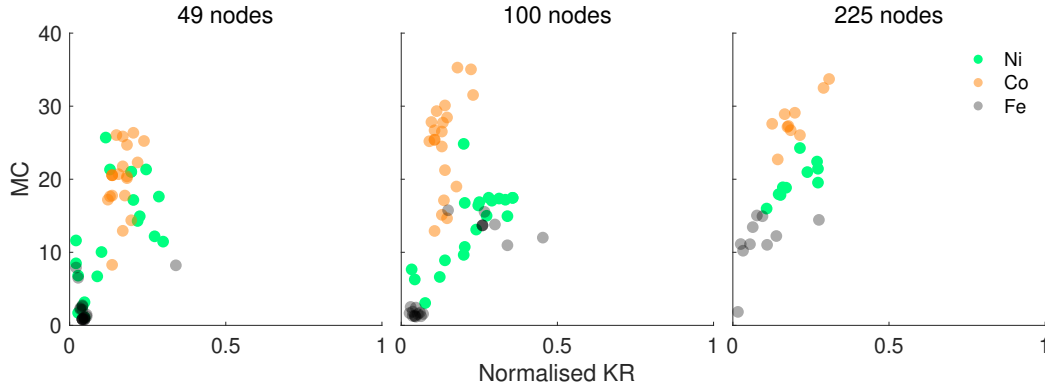


Figure 3: Normalised kernel rank (KR) and linear memory capacity (MC) of evolved films across three sizes, for the NARMA-30 task. Materials are separated by colour – Nickel (red), Cobalt (blue), and Iron (black). For the NARMA-30 task, to perform well, MC should be close to the driving equation’s time-lag of 30, which in turn requires more linear behaviour (i.e., a low KR). The Ni and Co materials do this well, however Fe does not. Only at larger film sizes does Fe grow in memory capacity.

sizes; however as size increases both measures slowly move towards values representative of more desirable dynamics. This change, relative to increase in size, mirrors the gradual decrease in error shown in Fig. 2.

Task performances and KR/MC measure assessment indicate that several trade-offs exist. First, smaller films generally show better performance than similarly sized digital reservoirs. This suggests properties of small films, such as shorter distances between edges, may improve performance. Interference and reflection from edges of travelling spin waves are likely to increase as size decreases. The geometry of the film is also likely to have an effect. In our experiments, only square films are used; other shapes can provide greater asymmetry at the boundaries. Second, depending on the material, larger films can boost desirable dynamical properties such as memory. A large surface area enables signals to persist unperturbed away from rapidly changing input sources. Exploiting geometry, size, and inputs to control these trade-offs are of great interest for future work.

3 Paths to Physical Realisation

The simulated platform is realisable in physical hardware. Fig. 4a shows a proposed 5×5 input-output interface. The device consists of a nanoscale thin-film, encapsulated by point contacts (yellow) that measure the local tunnelling magnetoresistance in different regions of the film. Underlying magnetic field sources (grey) provide locally controllable magnetic field inputs $B(t)$ to each region of the device.

With any new reservoir system, an ability to scale hardware components and reduce error is desired. In our experiments, each material exhibits a significant improvement as film size increases, despite its restrictive lattice topology and no predefined discrete processing nodes. The greatest improvements relate to the difficulty of the task, where distinct trade-offs in non-linearity and memory are required. The most significant differences between material and size are shown for the NARMA tasks, where memory is a strong indicator of performance.

To assess scaling potential, additional evolutionary searches are conducted with the Co material for larger systems. In order to compare material scaling with digital reservoirs, equivalent-sized networks are evolved as well. Fig. 4b shows NARMA-10 task performance as film and reservoir size increases. Scaling begins at 25 material cells/network nodes up to 900-cells/nodes, representing film dimensions (D) of 25nm^2 up to 150nm^2 : $D = (\sqrt{\text{Num cells}}) \times \text{cell size}$. The results show that up to 400 cells/nodes there is a significant reduction in the average error as size increases. After this, the median error is no longer significantly different, however lower errors continue to be found in the best runs. This could indicate that larger films with lower errors are more challenging to discover, or that potentially beneficial properties of small films are lost, such as interaction of reflections from edges.

At the nanoscale, thermal noise is a limiting factor. Maintaining performance close to room temperature is desirable for practical implementations. Stability and reproducibility can be adversely affected by thermal noise. In our experiments, temperature is set to absolute zero kelvin to observe pure magnetic behaviour without thermal effects. Methods to control and reduce thermal fluctuations have been proposed using spin transfer torque to modify thermal activation

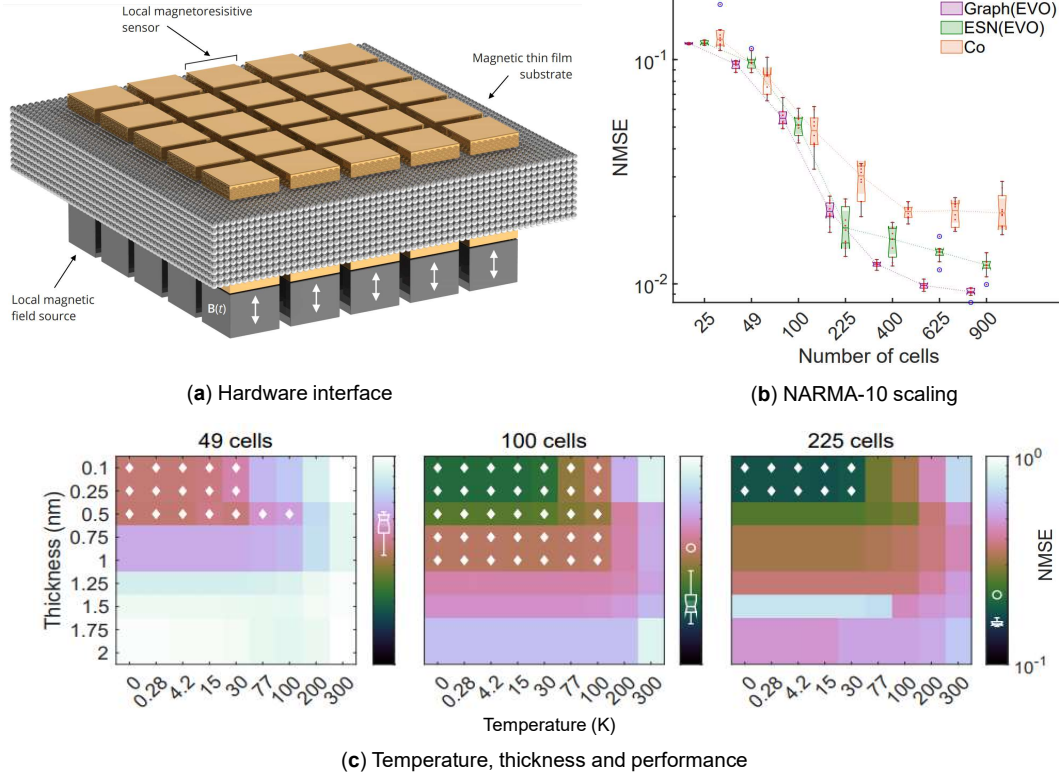


Figure 4: **a)** Proposed hardware interface to realise a thin-film reservoir computing device. **b)** Performance of Co material on the NARMA-10 task as number of cells increases. Performance of the material remains competitive to scaled simulated reservoirs. **c)** Grid sweep of film temperatures (K) and film thickness (nm). The NMSE of the evolved Co configuration is shown using colour. Errors are for the NARMA-30 task. Temperature ranges from 0 K (original experiments) to more practical temperatures including room temperature (300 K). White box-plots in colour bar display performances of the 20 best random ESNs at the respective size. A white diamond in a cell signals task error is within the ESN range.

rates [41]. This suggests different paths towards room temperature computing with thin-films without cooling are plausible.

To demonstrate the effect of temperature on our films, additional experiments are conducted. Fig. 4c shows reservoir performance at various temperatures on the NARMA-30 task. The temperature range includes: millikelvin (0.28 K), liquid helium (4.2 K), liquid nitrogen (77 K), and room temperature (300 K). The top-left shows the original experimental setup (temperature = 0 K and thickness = 0.1nm) for an evolved Co reservoir. As temperature increases along the x -axis, thermal noise dominates and degrades performance. A similar pattern is present across all film sizes, tasks and materials (see Supplementary Material).

Film thickness is also investigated to see whether thickness can compensate for a rise in temperature. On the y -axis of Fig. 4c, film thickness varies from 0.1–2nm. In general, performance is maintained with thicknesses up to 0.5nm and temperatures up to 30–77 K. Between 0.5–1nm, the change in error slows as temperature rises (30 to 200 K), however errors are higher than for thinner films. Beyond 1nm, thicker films tend to degrade performance, but this varies depending on material and film size (see Supplementary Material). The results show that films with sub-nanometer thickness at temperatures up to 100 K work best, outperforming or matching equivalent-sized random reservoir networks.

4 Conclusion

Our spintronic-based system provides an exceptional platform for machine learning with analogue hardware. By combining two frameworks, evolution *in materio* and reservoir computing, novel magnetic computing devices are demonstrated.

Without the need for discrete neural components, physical reservoirs are possible with smaller footprints than other neuromorphic devices, e.g., memristors, spin torque oscillators, photonics [42, 22, 24]. The evolved devices operate at frequencies of 100 GHz and require no special preprocessing to emulate network structures [17, 22]. The basic materials used are inexpensive and feature a large dynamical range that can be reconfigured externally to solve different machine learning tasks.

With this generic platform, other complex magnetic materials such as alloys, oxides, skyrmion fabrics, and antiferromagnetic reservoirs [43] can be optimised and exploited. Furthermore, simulations of complex atomic structures are possible. With atomistic simulations, desirable hetero-structures or defects can be introduced to add more reservoir complexity and greater physical realism.

The natural dynamics and nanoscale size of the proposed magnetic substrates presents a new path towards fast energy-efficient computing platforms enabling new innovations in smart technologies.

5 Methods

5.1 Spin Model

For a generic atomistic model with n nearest neighbour interactions, the Curie temperature T_C can be calculated from the atomistic exchange J_{ij} by the mean-field expression. This sums over every exchange that occurs in each cell to calculate the total exchange [44].

$$T_C = \frac{\varepsilon}{3k_B N_c} \sum_{i=0}^{N_c} \sum_{j=0}^n J_{ij} \quad (1)$$

where k_B is the Boltzmann constant, N_c is the number of atoms per cell, and ε is a correction factor from the usual mean-field expression which arises due to spin waves in the 3D Heisenberg model.

The anisotropy k_u and the spontaneous magnetisation M_s are calculated as a sum of the atomic anisotropies and spin moments within each cell. The gyromagnetic ratio γ and the damping constant α are calculated as an average of the atomic parameters for each cell.

The energetics of the micromagnetic system are described using a spin Hamiltonian neglecting non-magnetic contributions and given by:

$$\mathcal{H}_{eff} = \mathcal{H}_{app} + \mathcal{H}_{ani} + \mathcal{H}_{exc} + \mathcal{H}_{dip} \quad (2)$$

where \mathcal{H}_{app} is the applied field, \mathcal{H}_{ani} is the anisotropy field, \mathcal{H}_{exc} is the intergranular exchange, and \mathcal{H}_{dip} is the dipole field.

The anisotropy Hamiltonian describes the directional dependence of the materials magnetisation, in this case the anisotropy is uniaxial along z and is described by:

$$\mathcal{H}_{ani} = KV(\mathbf{m}_x^2 + \mathbf{m}_y^2) \quad (3)$$

The exchange field is calculated as a sum of the exchange interactions between neighbouring cells, the micromagnetic exchange constant A is a sum over all atoms which have a neighbours in another cell. The summation over all the interactions gives a total interaction from cell i to cell j . From this the micromagnetic exchange constant is calculated by multiplying by the distance between the atomistic atoms.

$$\mathbf{H}_{ex}^i = A_{ij} \frac{2}{M_S \Delta^2 m_e^2} \sum_{n_{cells}} (\mathbf{m}_j - \mathbf{m}_i) \quad (4)$$

$$\mathbf{H}_{dip} = \frac{\mu_0}{4\pi} \frac{3(\mathbf{m} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}}{|\hat{\mathbf{r}}|^3} - \frac{\mu_0 \mathbf{m}}{3V} \quad (5)$$

The atomistic Landau–Lifshitz–Gilbert (LLG) equation is used to model the time-dependent behaviour of the magnetic films given by:

$$\frac{\partial \mathbf{m}_i}{\partial t} = -\frac{\gamma}{(1 + \lambda^2)} [\mathbf{m}_i \times \mathbf{H}_{eff}^i + \lambda \mathbf{m}_i \times (\mathbf{m}_i \times \mathbf{H}_{eff}^i)] \quad (6)$$

where \mathbf{m}_i is a unit vector representing the direction of the magnetic spin moment of cell i , γ is the gyromagnetic ratio and $\mathbf{H}_{\text{eff}}^i$ is the net magnetic field on each cell and is equal to the derivative of the spin Hamiltonian:

$$\mathbf{H}_{\text{eff}}^i = -\frac{1}{\mu_s} \frac{\partial \mathcal{H}_{\text{eff}}}{\partial \mathbf{S}_i} \quad (7)$$

5.2 Reservoir Model

The reservoir dynamics of simulated networks are given by the state update equation:

$$x(t) = (1 - a)x(t - 1) + af(b\mathbf{W}_{in}[u(t); u_{bias}] + c\mathbf{W}x(t - 1)) \quad (8)$$

where x is the internal state at time-step t , f is the non-linear neuron activation function (a \tanh function), u is the input signal, and u_{bias} is a bias source. \mathbf{W}_{in} and \mathbf{W} are weight matrices giving the connection weights to inputs and internal neurons respectively. The parameters b and c control the global scaling of the input weights and internal weights. Input scaling b affects the non-linear response of the reservoir and relative effect of the current input. Internal scaling c controls the reservoir's stability as well as the influence and persistence of the input: low values dampen internal activity and increase response to input, and high values lead to chaotic behaviour. A leakage filter a is used to match the internal timescales of the film to the characteristic timescale of the task. This is similar to adding a low-pass filter before the output. The leak rate controls the time-scale mismatch between the input and reservoir dynamics; when $a = 1$, the previous states do not leak into the current states.

For both random and evolved reservoir networks, \mathbf{W}_{in} and \mathbf{W} are initialised as sparse normally distributed random matrices (input sparsity = 0.1, internal sparsity = 0.1, mean = 0, variance = 1). For the lattice network, we define a square grid of neurons each connected to its nearest neighbours in its Moore neighbourhood [45]. Each non-perimeter node has eight connections to neighbours and one self-connection, resulting in each node having a maximum of nine adaptable weights in \mathbf{W} .

The final trained output $y(t)$ is given when the reservoir states $x(t)$ are combined with the trained readout weight matrix \mathbf{W}_{out} :

$$y(t) = \mathbf{W}_{out}x(t) \quad (9)$$

Readout training is performed using ridge regression [46] and occurs within the evolutionary loop during the training phase. A validation and testing phase is carried out to evaluate the generalisation of the readout to new data. This approach is similar to previous work [47, 48].

5.3 Experimental Setup

During the simulation, material parameters such as exchange interaction, anisotropies, and atomic moments are defined by the material and remain unaltered. Parameters controlling the input mapping, field intensity b , intrinsic magnetic damping α , and a post-state collection filter a are tuned.

The material is interpreted as a reservoir in the following way:

$$\mathbf{X}(t) = \sigma(b\mathbf{W}_{in}[u; u_{bias}], \alpha) \quad (10)$$

$$\mathbf{X}_f(t) = (1 - a)\mathbf{X}(t - 1) + a\mathbf{X}(t) \quad (11)$$

$$y(t) = \mathbf{W}_{out}\mathbf{X}_f(t) \quad (12)$$

where \mathbf{X} is the global material state comprising each cell's local X_{xyz} 3d magnetisation vector, σ represents the material function, a is the leakage parameter, and \mathbf{X}_f is an external filter layer with a one-step memory implemented after the observation of material state \mathbf{X} and before the readout weights are applied.

The input mapping \mathbf{W}_{in} consists of weighted connections from the input u and a bias u_{bias} source to each cell. The input search space is typically large and grows with film size. Field intensity ($0 < b \leq 2$) is a global scaling factor applied to the input mapping. This suppresses or raises the overall magnitude of the locally applied fields promoting varying dynamical behaviours.

The magnetic damping parameter ($0 < \alpha \leq 1$) controls the speed of information propagation and oscillation. Damping describes the non-linear spin relaxation across the film, controlling the rate at which magnetisation spins reach equilibrium.

To optimise magnetic reservoirs, artificial evolution is applied. To reduce convergence time, linear regression is also used to train the readout rather than evolving it (see Methods). The evolutionary goal is to find parameters that optimise the efficiency and ability of the readout layer to perform its function.

	Ni	Co	Fe	unit
Crystal structure	fcc	fcc	bcc	–
Unit cell size a	3.524	2.507	2.866	Å
Atomic spin moment μ_s	0.606	1.72	2.22	μ_B
Exchange energy J_{ij}	2.757×10^{-21}	6.064×10^{-21}	7.050×10^{-21}	J/link
Anisotropy k	5.47×10^{-26}	6.69×10^{-24}	5.65×10^{-25}	J/atom
Temp. rescaling exponent	2.322	2.369	2.876	–
Rescaling Curie temperature	635	1395	1049	–

Table 1: Parameters used to simulate each ferromagnetic material in VAMPIRE. These parameters are static in our work and are not affected by the evolutionary algorithm.

Many heuristics can be used to optimise reservoirs [49], but here the microbial genetic algorithm (MGA) [50] is chosen for its simplicity. The MGA allows individuals to survive across many generations, provides elitism for free, and offers a simple mechanism for selection, recombination and mutation.

Parameters for the MGA include: population size = 100, number of generations = 2000, mutation rate = 0.05(5%), recombination rate = 0.5(50%), deme size (species separation) = 0.1(10% of population), and number of runs = 20. These parameters were used for all experiments involving an evolutionary algorithm.

To conduct the experiments, the VAMPIRE source code was adapted to construct a dynamic input-output mechanism. Important parameters for the VAMPIRE simulation include input frequency, integration time-step, initial spin direction, and macro-cell size (micromagnetic simulation). The input frequency chosen – 10ps / 100 GHz – was based on qualitative experiments in search of characteristic behaviours, such as fast response and a short settling time. The input frequency has to closely match the internal timescales and dynamics of the system.

To optimise the evaluation process and reduce computational cost an integration timestep of 100fs was used. This provides a less accurate model compared to an integration timestep of 1fs but provides manageable computational run times. Details about how this parameter choice minimally affects performance are provided in the supplementary material.

The initial spin direction was aligned with the x -axis, and input signals were injected in the z -direction. The macro-cell size for each simulation was fixed at 5nm.

Simulation parameters for each material are given in Table 1. These include exchange constants and second-order uniaxial anisotropy constants. To conduct accurate temperature calculations, rescaling exponents and curie temperature information are also included.

5.4 Benchmark Tasks

The chosen tasks are widely used benchmarks for different reservoir systems and methods [51, 33, 52, 36, 34, 30]. The laser task predicts the next value of the Santa Fe time-series Competition Data (dataset A) [29]. The dataset holds original source data recorded from a Far-Infrared-Laser in a chaotic state. The training and testing uses the first 2,000 values of the dataset, divided into three sets: 1200 (training set), 400 (validation set), and 400 (test set). The first 50 output values of each sub-set are discarded as an initial washout period.

The NARMA task originates from work on training recurrent networks [53]. It evaluates a reservoir’s ability to model an n -th order highly non-linear dynamical system where the system state depends on the driving input as well as its own history. The challenging aspect of the NARMA task is that it contains both non-linearity and long-term dependencies created by the n -th order time-lag.

An n -th ordered NARMA experiment is carried out by predicting the output $y(n+1)$ given by eq.(13) when supplied with $u(n)$ from a uniform distribution of interval [0, 0.5]. For the 10-th and 30-th order systems $\alpha = 0.3$, $\beta = 0.05$, $\delta = 10$ and $\gamma = 0.1$.

$$y(n+1) = \alpha y(n) + \beta y(n) \left(\sum_{i=0}^{\delta} y(n-i) \right) + 1.5u(n-\delta)u(n) + \gamma \quad (13)$$

The NARMA equation is simulated for 5,000 values and split into: 3,000 training, 1,000 validation and 1,000 test for both versions. The first 50 values of each sub-set are discarded as an initial washout period.

Acknowledgements

This work is part of the SpInspired project, funded by EPSRC Grant EP/R032823/1. All experiments were carried out using the University of York's Super Advanced Research Computing Cluster (Viking).

References

- [1] Tanaka, G. *et al.* Recent advances in physical reservoir computing: A review. *Neural Networks* (2019).
- [2] Shi, W., Cao, J., Zhang, Q., Li, Y. & Xu, L. Edge computing: Vision and challenges. *IEEE internet of things journal* **3**, 637–646 (2016).
- [3] Chen, J. & Ran, X. Deep learning with edge computing: A review. *Proceedings of the IEEE* **107**, 1655–1674 (2019).
- [4] Wang, X. *et al.* Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials* **22**, 869–904 (2020).
- [5] Adamatzky, A. (ed.) *Advances in Unconventional Computing: Volume 2 Prototypes, Models and Algorithms* (Springer, 2016).
- [6] Young, A. R., Dean, M. E., Plank, J. S. & Rose, G. S. A review of spiking neuromorphic hardware communication systems. *IEEE Access* **7**, 135606–135620 (2019).
- [7] Indiveri, G. *et al.* Neuromorphic silicon neuron circuits. *Frontiers in neuroscience* **5**, 73 (2011).
- [8] Jo, S. H. *et al.* Nanoscale memristor device as synapse in neuromorphic systems. *Nano letters* **10**, 1297–1301 (2010).
- [9] Xia, Q. & Yang, J. J. Memristive crossbar arrays for brain-inspired computing. *Nature materials* **18**, 309–323 (2019).
- [10] Miller, J. F. & Downing, K. Evolution in materio: Looking beyond the silicon box. In *NASA/DoD Conference on Evolvable Hardware 2002*, 167–176 (IEEE, 2002).
- [11] Mohid, M. *et al.* Evolving solutions to computational problems using carbon nanotubes. *International Journal of Unconventional Computing* **11**, 245–281 (2015).
- [12] Massey, M. *et al.* Evolution of electronic circuits using carbon nanotube composites. *Scientific Reports* **6** (2016).
- [13] Bose, S. *et al.* Evolution of a designless nanoparticle network into reconfigurable boolean logic. *Nature nanotechnology* doi:10.1038/nnano.2015.207 (2015).
- [14] Chen, T. *et al.* Classification with a disordered dopant-atom network in silicon. *Nature* **577**, 341–345 (2020).
- [15] Schrauwen, B., Verstraeten, D. & Van Campenhout, J. An overview of reservoir computing: theory, applications and implementations. In *Proceedings of the 15th European symposium on artificial neural networks* (Citeseer, 2007).
- [16] Verstraeten, D. & Schrauwen, B. On the quantification of dynamics in reservoir computing. In *Artificial Neural Networks–ICANN 2009*, 985–994 (Springer, 2009).
- [17] Appeltant, L. *et al.* Information processing using a single dynamical node as complex system. *Nature Communications* **2**, 468 (2011).
- [18] Caravelli, F. & Carbajal, J. Memristors for the curious outsiders. *Technologies* **6**, 118 (2018).
- [19] Dion, G., Mejaouri, S. & Sylvestre, J. Reservoir computing with a single delay-coupled non-linear mechanical oscillator. *Journal of Applied Physics* **124**, 152132 (2018).
- [20] Prychynenko, D. *et al.* Magnetic skyrmion as a nonlinear resistive element: A potential building block for reservoir computing. *Physical Review Applied* **9**, 014034 (2018).
- [21] Pinna, D., Bourianoff, G. & Everschor-Sitte, K. Reservoir computing with random skyrmion textures. *Phys. Rev. Applied* **14**, 054020 (2020). URL <https://link.aps.org/doi/10.1103/PhysRevApplied.14.054020>.
- [22] Torrejon, J. *et al.* Neuromorphic computing with nanoscale spintronic oscillators. *Nature* **547**, 428–431 (2017).
- [23] Nakane, R., Tanaka, G. & Hirose, A. Reservoir computing with spin waves excited in a garnet film. *IEEE Access* **6**, 4462–4469 (2018).
- [24] Romera, M. *et al.* Vowel recognition with four coupled spin-torque nano-oscillators. *Nature* **563**, 230–234 (2018).

- [25] Zheng, Q., Zhu, X., Mi, Y., Yuan, Z. & Xia, K. Recurrent neural networks made of magnetic tunnel junctions. *AIP Advances* **10**, 025116 (2020).
- [26] Watt, S. & Kostylev, M. Reservoir computing using a spin-wave delay-line active-ring resonator based on yttrium-iron-garnet film. *Physical Review Applied* **13**, 034057 (2020).
- [27] Zahedinejad, M. *et al.* Two-dimensional mutually synchronized spin hall nano-oscillator arrays for neuromorphic computing. *Nature Nanotechnology* **15**, 47–52 (2020).
- [28] Pajda, M., Kudrnovský, J., Turek, I., Drchal, V. & Bruno, P. Ab initio calculations of exchange interactions, spin-wave stiffness constants, and curie temperatures of fe, co, and ni. *Physical Review B* **64**, 174402 (2001).
- [29] Weigend, A. *The Santa Fe Time Series Competition Data: Data set A, Laser generated data* (1991 (accessed March, 2016)). URL <http://www-psych.stanford.edu/~andreas/Time-Series/SantaFe.html>.
- [30] Inubushi, M. & Yoshimura, K. Reservoir computing beyond memory-nonlinearity trade-off. *Scientific Reports* **7**, 10199 (2017).
- [31] Dale, M. *et al.* The role of structure and complexity on reservoir computing quality. In *International Conference on Unconventional Computation and Natural Computation*, 52–64 (Springer, 2019).
- [32] Dale, M., O’Keefe, S., Sebald, A., Stepney, S. & Trefzer, M. A. Reservoir computing quality: connectivity and topology. *Natural Computing* (2020). doi:10.1007/s11047-020-09823-1.
- [33] Rodan, A. & Tiño, P. Simple deterministically constructed recurrent neural networks. In *International Conference on Intelligent Data Engineering and Automated Learning*, 267–274 (Springer, 2010).
- [34] Larger, L. *et al.* Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing. *Optics Express* **20**, 3241–3249 (2012).
- [35] Hou, Y. *et al.* Prediction performance of reservoir computing system based on a semiconductor laser subject to double optical feedback and optical injection. *Optics Express* **26**, 10211–10219 (2018).
- [36] Paquot, Y. *et al.* Optoelectronic reservoir computing. *Scientific Reports* **2** (2012).
- [37] Dale, M., Miller, J. F., Stepney, S. & Trefzer, M. A. A substrate-independent framework to characterize reservoir computers. *Proceedings of the Royal Society A* **475**, 20180723 (2019).
- [38] Dambre, J., Verstraeten, D., Schrauwen, B. & Massar, S. Information processing capacity of dynamical systems. *Scientific Reports* **2** (2012).
- [39] Legenstein, R. & Maass, W. Edge of chaos and prediction of computational performance for neural circuit models. *Neural Networks* **20**, 323–334 (2007).
- [40] Jaeger, H. *Short term memory in echo state networks* (GMD-Forschungszentrum Informationstechnik, 2001).
- [41] Demidov, V. E. *et al.* Magnetic nano-oscillator driven by pure spin current. *Nature materials* **11**, 1028–1031 (2012).
- [42] Du, C. *et al.* Reservoir computing using dynamic memristors for temporal information processing. *Nature communications* **8**, 1–10 (2017).
- [43] Kurenkov, A., Fukami, S. & Ohno, H. Neuromorphic computing with antiferromagnetic spintronics. *Journal of Applied Physics* **128**, 010902 (2020).
- [44] Jiles, D. *Introduction to magnetism and magnetic materials* (CRC press, 2015).
- [45] Adamatzky, A. *Game of life cellular automata*, vol. 1 (Springer, 2010).
- [46] Lukoševičius, M. A practical guide to applying echo state networks. In *Neural Networks: Tricks of the Trade*, 659–686 (Springer, 2012).
- [47] Dale, M., Miller, J. F., Stepney, S. & Trefzer, M. A. Evolving carbon nanotube reservoir computers. In *International Conference on Unconventional Computation and Natural Computation*, 49–61 (Springer, 2016).
- [48] Dale, M. Neuroevolution of hierarchical reservoir computers. In *Proceedings of the Genetic and Evolutionary Computation Conference*, 410–417 (ACM, 2018).
- [49] Bala, A., Ismail, I., Ibrahim, R. & Sait, S. M. Applications of metaheuristics in reservoir computing techniques: a review. *IEEE Access* **6**, 58012–58029 (2018).
- [50] Harvey, I. The microbial genetic algorithm. In *European Conference on Artificial Life*, 126–133 (Springer, 2009).
- [51] Jaeger, H. The “echo state” approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report* **148**, 34 (2001).

- [52] Tran, S. D. & Teuscher, C. Memcapacitive reservoir computing. In *2017 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, 115–116 (IEEE, 2017).
- [53] Atiya, A. F. & Parlos, A. G. New results on recurrent network training: unifying the algorithms and accelerating convergence. *IEEE Transactions on Neural Networks* **11**, 697–709 (2000).
- [54] Ganguli, S., Huh, D. & Sompolinsky, H. Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences* **105**, 18970–18975 (2008).

Supplementary Material

Optimised Integration Time-step

To reduce computational time simulating thin-films a large integrator time-step was used. Ideally, small time steps are preferable to more accurately capture spin precession and general dynamics between input pulses, however, this comes with a large computational cost.

A characterisation of the how the integrator time step affects task performance is given in Fig. 5. Here, we show the chosen 100fs integrator time-step compared to the more accurate 1fs time-step. These results show that, in general, our chosen time-step is statistically similar, representing a reasonably accurate model of the driven dynamics.

To test whether the medians of both time-steps were significantly different, the non-parametric two-sided Wilcoxon rank sum test was used. This tests the null hypothesis that both samples are from the same distribution with equal medians. A rejection of the null hypothesis at the 95% significance level is indicated by a p -value > 0.05 .

The p -values for each task are: $p = 0.23$ (laser), $p = 0.42$ (NARMA-10), and $p = 0.57$ (NARMA-30). This indicates that performance is not significantly affected by the change in time-step, however, computational time is reduced dramatically from hours to minutes.

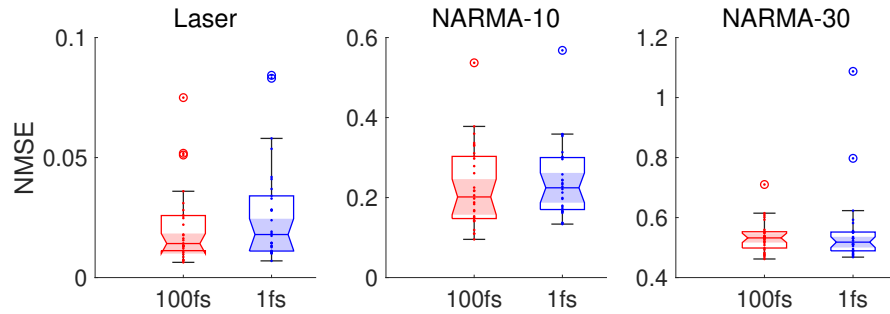


Figure 5: Comparing integrator time-steps across three tasks. 100fs provides a less accurate model compared to 1fs but dramatically reduces run-time. Only the Co material used in this experiment. Each boxplot shows a total of 30 random configurations for each task compared at both time-steps.

Kernel Rank and Memory Capacity

In this work, two property measures are used to assess the underlying dynamics of the reservoir system. These measures are the kernel rank (KR) and linear memory capacity (MC).

Kernel rank is a measure of the reservoir’s ability to separate distinct input patterns [39]. It measures a reservoir’s ability to produce a rich non-linear representation of the input u and its history $u(t-1), u(t-2), \dots$. This is closely linked to the *linear separation property*, measuring how different input signals map onto different reservoir states. As many practical tasks are linearly inseparable, reservoirs typically require some non-linear transformation of the input. KR is a measure of the complexity and diversity of these non-linear operations performed by the reservoir.

Reservoirs in ordered dynamical regimes typically have a low ranking value of KR, and in chaotic regimes, it is high. The maximum value of KR is relative to the number of observable states. In our experiments, KR is normalised to observe the underlying non-linearity of the task without distortion from reservoir size.

Another important property for reservoir computing is memory, as reservoirs are typically configured to solve temporal problems. A simple measure for reservoir memory is the *linear short-term memory capacity* (MC). This was first outlined in [40] to quantify the echo state property. For the echo state property to hold, the dynamics of the input driven reservoir must asymptotically wash out any information resulting from initial conditions. This property therefore implies a fading memory exists, characterised by the short-term memory capacity.

A full understanding of a reservoir’s memory capacity, however, cannot be encapsulated through a linear memory measure alone, as a reservoir will possess some non-linear memory. Other memory measures proposed in the literature quantify other aspects of memory, such as the quadratic and cross-memory capacities, and total memory of reservoirs

using the Fisher Memory Curve [54, 38]. The linear measure is used here as a simple benchmark. More sophisticated measures are unnecessary to identify the differences in the following tasks.

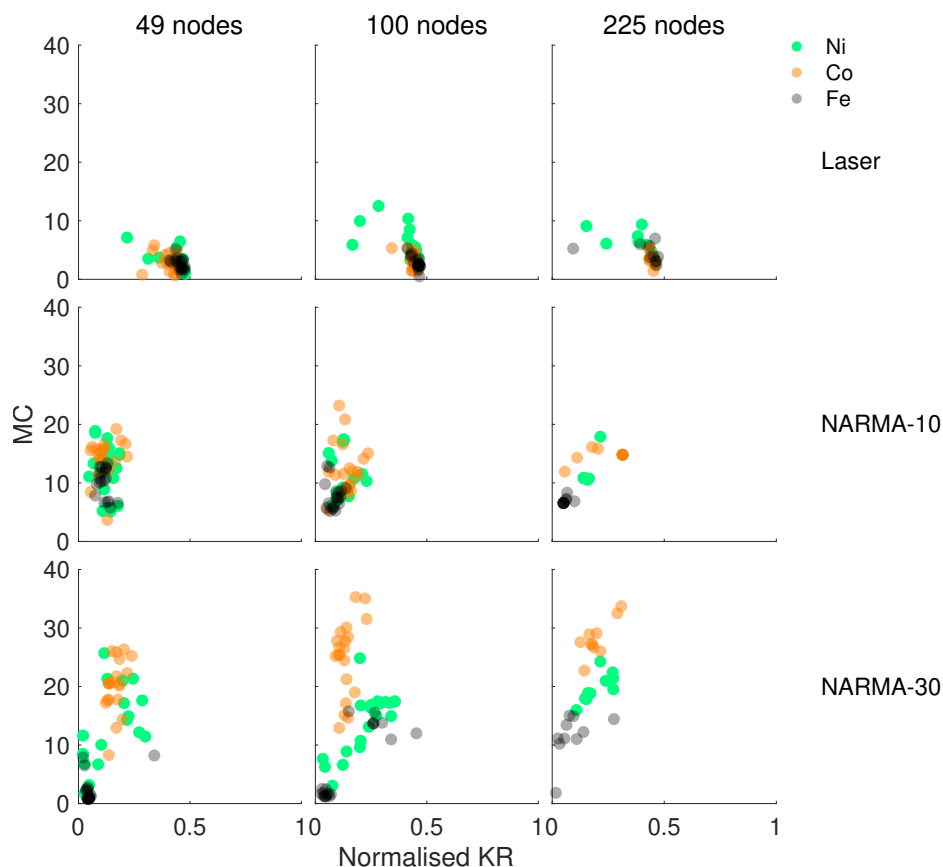


Figure 6: Kernel rank (normalised) versus linear memory capacity for all materials, sizes and tasks. Each column refers to a film size (left, middle, right: 49, 100, 225) and rows to a task (top, middle, bottom: laser, NARMA-10, NARMA-30). Material reservoirs shown are those displayed in Fig. 2.

In Fig. 3, just the results for the NARMA-30 task are given. The results for all tasks and sizes are provided in Fig. 6. From these results, it is possible to determine the difference in dynamics required for each task.

The laser task requires very little memory, and is mainly driven by non-linear dynamics. The normalised KR of 0.5 is relatively high when taking into account that many of the magnetic materials observable states are highly correlated, e.g., from the x and y dimension of the spins.

The NARMA-10 task features more linear tendencies. We see memory capacity clusters around the value of 10, relating of course to the 10 time-step time-lag present in the system being modelled. Irrespective of size, the same characteristic dynamics have converged during evolutionary selection, and all materials are able to exhibit the same dynamics.

The NARMA-30 task requires a notable increase in memory capacity. At the smallest size, no material meets the necessary criteria ($MC = 30$) to perform well at the task, however Co and Ni attempt to maximise their MC. Fe struggles to exhibit any memory. As size increases, Co and Ni gradually reach $MC = 30$ and this is reflected in their performance. The MC of Fe also increases but at a slower rate proportional to size.

Temperature Effect and Film Thickness

To build practical computing systems it is desirable for the materials to function close to room temperature. In addition, thicker films put less strain on the fabrication process. In our main experiments, each material film was evolved at 0K to evaluate performance without thermal fluctuations. Here, we show how temperature affects performance at all film sizes (number of cells) and across each task.

For the laser task (Fig. 7), performance is stable and competitive – to random ESNs of equivalent node size – at higher temperatures typically up to 100K, depending on the material and number of cells. The most stable material and film size is Fe at 100 cells. In this configuration, only a small change in performance is present as thickness is increased up to 1nm.

For the NARMA-10 task performance is again stable, in some cases up to 100K, e.g., Co with 100 cells. As temperature increases, performance tends to drop off slightly faster than the laser task. This could be due to degradation in memory quality as thermal noise increases. In general, the results suggest the Co material responds better to increased temperatures. However, thicker films tend to be more detrimental to performance. The same trends are seen for the NARMA-30 task.

Interference and Reflective Boundaries

The proposed system exploits the nonlinear interactions of spins when perturbed by local magnetic fields. As information propagates, local coupled spins form wave crests and troughs that interact, creating interference patterns. At the boundaries, waves are reflected back into the film. Figs. 10 and 11 provide a visualisation of this dynamical behaviour for two Co films (49 cell and 900 cell). At $t = 10$, a single input pulse is supplied to two separate input locations. At $t = 11-13$, waves appear and propagate. The smaller film (Fig. 10) interacts almost instantly with the boundaries, and waves reverberate around the film for some time. In the larger film (Fig. 11), signals propagate for longer, undisturbed, until the wave crests reach each other and the boundaries. At $t > 20$, interference and reflected waves begin to dominate; however, memory of past inputs are still recoverable according to the memory capacity measure.

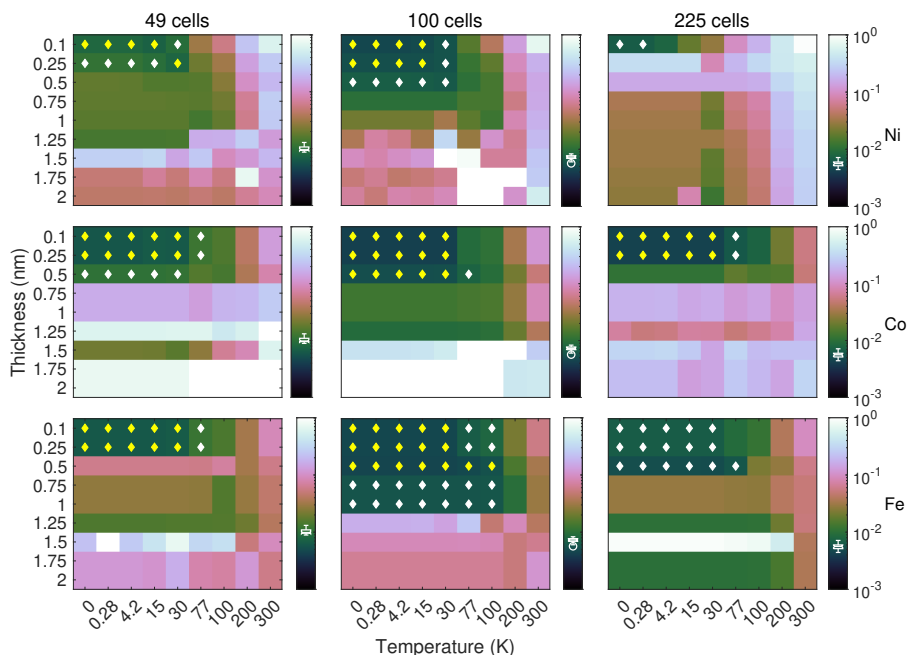


Figure 7: Temperature and film thickness sweeps for laser task. Normalised mean square error (colour) of an evolved configuration for Ni, Co, Fe. Box plots (white) display performances of 20 best random ESNs for this task. A diamond (white) signals that performance of film is within the ESN range. A diamond (yellow) signals that performance of film is better than ESN

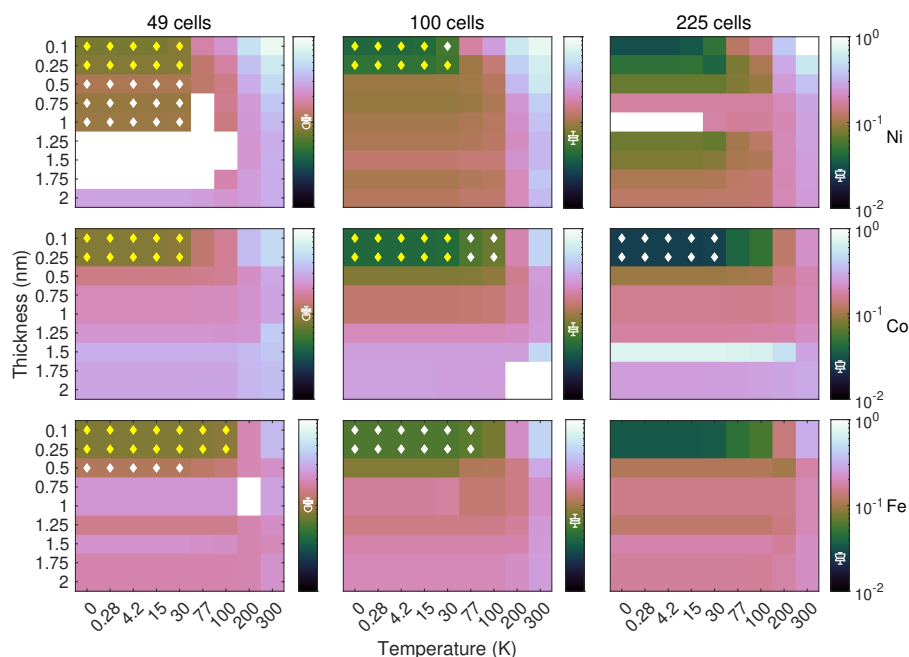


Figure 8: Temperature and film thickness sweeps for NARMA-10 task. Normalised mean square error (colour) of an evolved configuration for Ni, Co, Fe. Box plots (white) display performances of 20 best random ESNs for this task. A diamond (white) signals that performance of film is within the ESN range. A diamond (yellow) signals that performance of film is better than ESN.

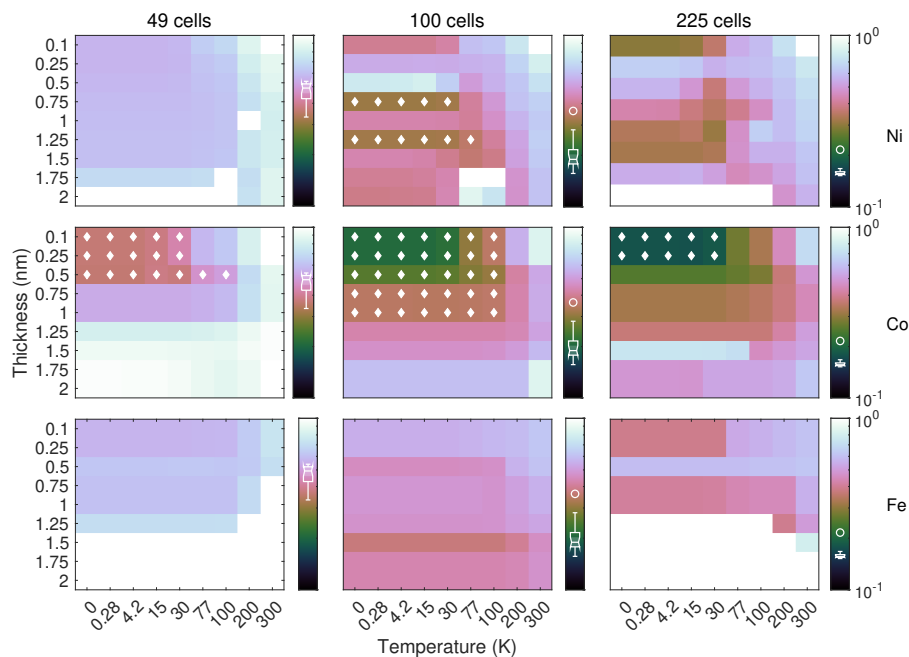


Figure 9: Temperature and film thickness sweeps for NARMA-30 task. Normalised mean square error (colour) of an evolved configuration for Ni, Co, Fe. Box plots (white) display performances of 20 best random ESNs for this task. A diamond (white) signals that performance of film is within the ESN range. A diamond (yellow) signals that performance of film is better than ESN.

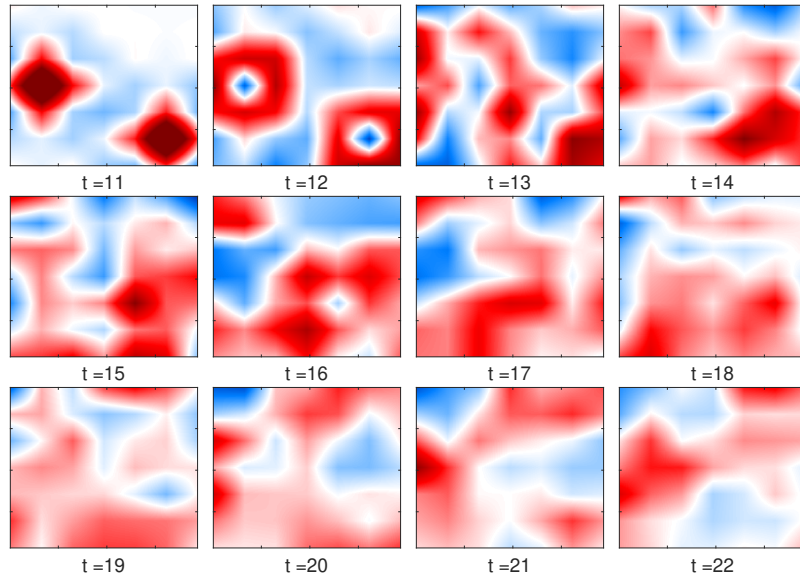


Figure 10: Dynamics of Co magnetic film with 49 cells. An input pulse is supplied at two locations on the film at $t = 10$. Red indicates a positive magnetisation, and blue, negative.

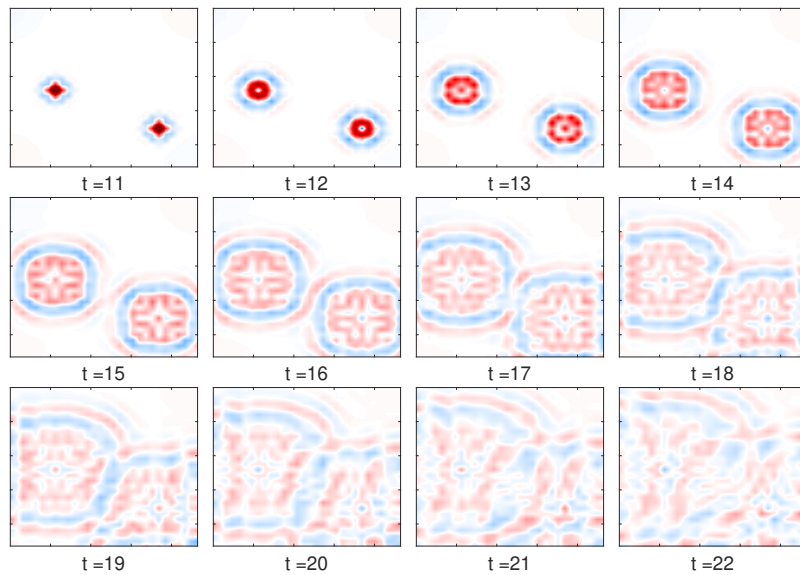


Figure 11: Dynamics of Co magnetic film with 900 cells. An input pulse is supplied at two locations on the film at $t = 10$. Red indicates a positive magnetisation, and blue, negative.