An energy-efficient, CMOS-compatible physical reservoir node with post-fabrication tunable decay dynamics

IOP Publishing Neuromorph. Comput. Eng. 5 (2025) 044010 https://doi.org/10.1088/2634-4386/ae2155

NEUROMORPHIC
Computing and Engineering

PAPER

An energy-efficient, CMOS-compatible physical reservoir node with post-fabrication tunable decay dynamics

Gambali Seshasai Chaitanya[1,*], **Aditya D Arkalgud**[1], **Shubham Pande**[2] and **Ankit Arora**[1]

[1] Department of Electronics Engineering, Indian Institute of Technology (BHU) Varanasi, Varanasi 221005, India
[2] Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai 600036, India
[*] Author to whom any correspondence should be addressed.

E-mail: gseshasai.chaitanya.cd.ece22@iitbhu.ac.in

Keywords: physical reservoir node, neuromorphic computing, tunable-decay dynamics, temporal signal processing

OPEN ACCESS

RECEIVED
18 August 2025

REVISED
29 October 2025

ACCEPTED FOR PUBLICATION
19 November 2025

PUBLISHED
8 December 2025

Abstract

In reservoir computing, the memory decay rate of physical reservoir nodes governs how quickly past inputs fade, thereby determining their temporal dynamics. Optimising this rate is therefore crucial for effective temporal signal processing. However, in most reported physical reservoirs, it is fixed at the time of fabrication and cannot be altered on demand for different applications. Thus, tailoring a single node for its adaptability across tasks with diverse temporal characteristics remains challenging. In this work, we propose and computationally analyse a CMOS-compatible, tunable-decay, hybrid reservoir node that integrates a subthreshold-operated field-effect transistor (FET) with a programmable ReRAM device (a memristor) and a capacitor connected at its gate terminal. The reservoir output measured as the FET drain current simultaneously captures temporal memory and nonlinear transformation of the input, while the memory decay time constant ($\tau = R \times C$) can be modulated in real time by adjusting the ReRAM resistance. We demonstrate the effectiveness of the proposed node on two representative benchmark tasks with contrasting τ requirements, namely, the MNIST digit classification with 96% accuracy, and a chaotic Hénon map prediction with a normalised RMS error of 0.0037, matching state-of-the-art hardware reservoirs. Our design achieves ultra-low energy consumption (≈15.20 pJ/operation), at least an order of magnitude lower than state-of-the-art implementations, while maintaining reliable operation and on-demand τ tunability. This combination of mature silicon technology and adaptive memristive functionality paves the way for energy-efficient, scalable, and reliable temporal learning systems.

1. Introduction

The relentless pursuit of energy-efficient artificial intelligence hardware has driven significant interest in neuromorphic computing architectures that emulate the brain's remarkable ability to process temporal information with minimal energy consumption [1]. Among these computing paradigms, reservoir computing (RC) has emerged as an attractive framework [2]. An RC architecture, illustrated in figure 1(a), comprises an input layer that projects the incoming signal through an input weight matrix (W_{in}) into a fixed, high-dimensional nonlinear dynamical system called the reservoir. The reservoir transforms the input into rich spatiotemporal state trajectories, which are then fed to a simple, trainable linear readout layer for inference. Because only the readout layer requires training, RC offers low training complexity and minimal computational overhead, making it well-suited for hardware deployment in resource- and energy-constrained environments [3]. The physical realisation of RC systems is determined by the global reservoir topology and the intrinsic properties of individual nodes. Among reported architectures, dynamic-device RC (ddRC) offers minimal circuit complexity and high hardware efficiency, stemming from its fixed reservoir structure, simple parallel configuration, and direct exploitation of intrinsic device dynamics, while still providing the rich internal dynamics required for temporal information processing

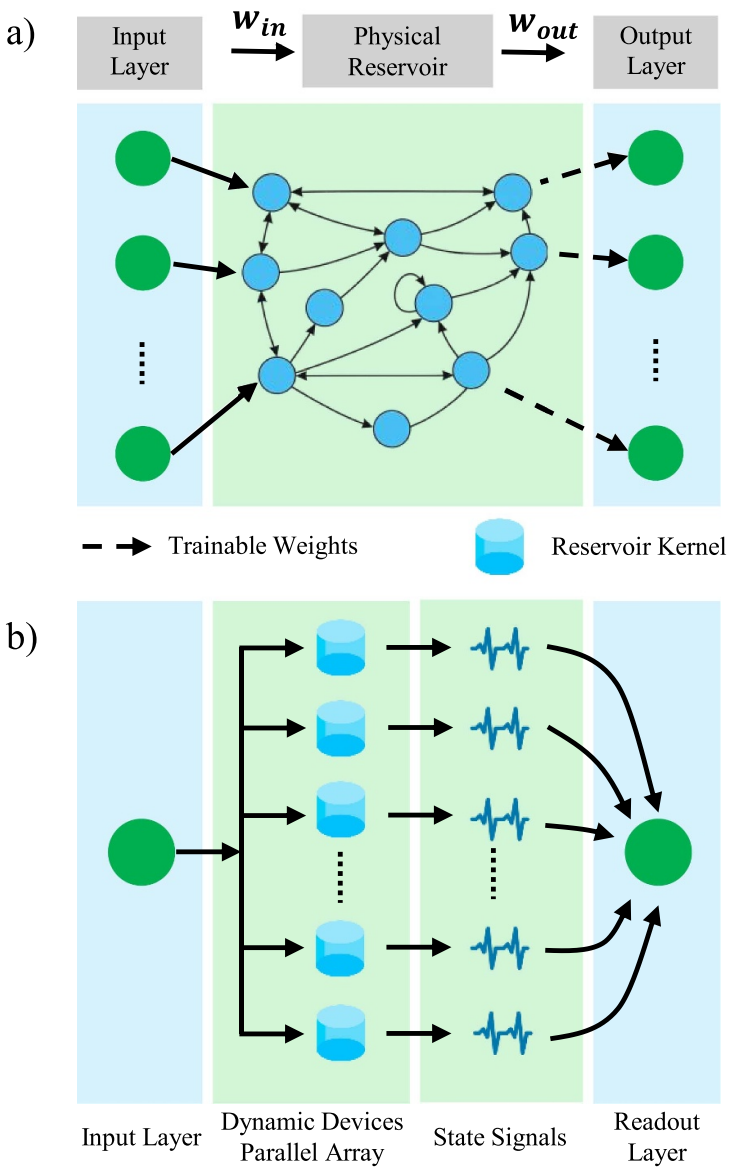

Figure 1. (a) Architecture of a typical reservoir computing system consisting of an input layer, followed by a physical reservoir, and an output layer. The input is mapped to the reservoir network via the transform W_{in}, whereas the reservoir states are mapped to the output layer by a transform W_{out}. (b) Architecture of the parallel dynamic device RC Architecture, where the reservoir layer consists of multiple dynamic devices, called reservoir nodes, that receive input signals in parallel.

[4]. For instance, in a typical ddRC configuration shown in figure 1(b), the input signal is applied in parallel to an array of dynamic devices serving as reservoir nodes, whose evolving internal states collectively form the high-dimensional feature vector for the readout layer. The performance of a ddRC system is governed by device characteristics that combine strong nonlinearity with intrinsic memory decay, allowing input sequences to be mapped into diverse transient states that embed temporal information for linear readout. By harnessing these properties, ddRC architectures achieve high computational efficiency across a broad range of temporal modelling tasks, thus making the reservoir node the key element in their physical implementation.

Field-effect transistors (FETs) and their emerging variants, such as Hafnia-based ferroelectric FETs (FeFETs) [5, 6] and $\alpha - In_2Se_3$-based FeFETs [7], are prime candidates for reservoir node devices due to their technological maturity and compatibility with industrial fabrication. Earlier reports on FET-based reservoir node approaches have leveraged threshold voltage fluctuations [8], random telegraph noise (RTN) [9], and bias temperature instability (BTI) [10, 11] to generate reservoir states. However, achieving precise wafer-scale control over these physical phenomena is challenging, as dopant fluctuations, line-edge roughness, and lithography limits introduce non-uniform device performance. Moreover, long-term reliability and aging issues such as RTN, BTI, hot-carrier injection, and gradual parameter drift can

progressively degrade operability. Collectively, these factors constrain the scalability and practicality of existing FET-based implementations for robust physical RC systems.

At the same time, a wide range of unconventional reservoir node devices have been explored for physical RC, including optical circuits [12–17], spintronic devices [18], ion-based transistors [4], solitons [19], and biological membranes [20, 21]. Although these emerging platforms exhibit promising computational capabilities, many rely on exotic or non-standard materials and specialised fabrication processes, which limit their compatibility with CMOS technology and hinder large-scale, cost-effective integration. In contrast, CMOS process compatible metal–insulator–metal devices, known as memristors, have evolved over the past decade as highly promising candidates for neuromorphic hardware [22–30]. Memristors inherently exhibit nonlinear current–voltage characteristics, enable energy-efficient operation, and possess short-term memory (STM), which together make them relatively well suited, among emerging approaches, as reservoir nodes for ddRC architectures.

Notably, most physical RC approaches have temporal dynamics fixed by intrinsic device physics. Modern edge AI tasks such as time-series classification, speech recognition, and chaotic system prediction often require different reservoir time constants, τ, which can span several orders of magnitude depending on the input. Without post-fabrication τ-tunability, adapting a single hardware platform to such varied temporal demands is challenging, limiting real-world applicability. As a result, the search for reservoir nodes with on-demand tunable τ has recently emerged as an active area of research. While a few demonstrations have been reported recently [20, 31, 32], the field remains in its early stages.

Here, we propose and computationally analyse a CMOS-compatible adaptive-dynamics reservoir node (ADRN) using circuit-level simulations that integrates a subthreshold-operated FET for rich nonlinear dynamics at ultra-low energy with a programmable ReRAM element (memristor) in a gate resistive-capacitive network for post-fabrication control of τ. The ReRAM's resistance programmability, retention, fast switching, and BEOL CMOS compatibility make it ideal for dynamic circuit tuning [33]. This hybrid node overcomes the limitations of either approach (FET-only or memristor-only) to enable energy-efficient, adaptable, and scalable physical reservoir. The remainder of this paper is organised as follows. Section 2 describes the proposed reservoir node in detail. Section 3 investigates its memory capacity and relates it to the FET subthreshold swing (SS). Section 4 presents application-level modelling using the proposed reservoir nodes, section 5 presents discussion on reservoir hyperparameters for optimal performance and section 6 concludes the work.

2. Proposed reservoir node

The circuit of the proposed ADRN, as shown in figure 2(a), was implemented using a tunable-τ network driving the gate of an NMOS FET. Input voltage pulses (V_{in}) are applied to the combination of a programmable memristor (M_1) and a capacitor (C_1), forming a circuit with a dynamically adjustable time constant $\tau = M_1 \times C_1$. The resistance of the memristor M_1 has been reported to be tunable over orders of magnitude [33–38], with typical values ranging from 50 kΩ to 25 MΩ, using standard read–write–verify programming schemes in recent experimental studies. This range of resistance corresponds to a range of τ from 50 μs to 25 ms, for $C_1 = 1$ nF, enabling the proposed ADRN to adjust its temporal response over three orders of magnitude for diverse computational tasks. The gate voltage across C_1 modulates the subthreshold drain current (I_D) of the FET, which in turn serves as the output signal representing the reservoir state.

The functionality of the ADRN circuit was validated using SPICE simulations based on the 45 nm predictive technology model [39]. The temporal evolution of I_D in response to applied pulse train V_{in}(t) for different values of M_1 is shown in figure 2(b). When a voltage pulse is applied at V_{in}, C_1 charges through M_1, increasing I_D, and when V_{in} returns to zero, C_1 discharges and I_D decays accordingly. For instance, the evolution of I_D in response to the voltage pulses V_{in} corresponding to a binary sequence '11 011' is shown in figure 2(b). The evolution of I_D changes with the change in M_1. This clearly demonstrates that the proposed reservoir node can be programmed to adjust τ, essential for on-demand tunable RC tasks with different processing rates. By leveraging a tunable τ mechanism and the inherent non-linearity of subthreshold operation, the ADRN is capable of efficient temporal processing, illustrated in figure 2(c), where the drain current evolution is shown for various binary input sequences. The resulting distinct current trajectories reflect the node's ability to transform temporally varying inputs into separable states within a high-dimensional space, making it easier for the readout layer to classify them. To further examine the influence of input parameters on this behaviour, the effect of pulse width (PW) and frequency on the reservoir dynamics was analysed (results not shown here). With increasing PW, the reservoir exhibited enhanced state separability due to prolonged activation. However, this improvement comes at the expense of possible saturation of the capacitor C_1 for longer

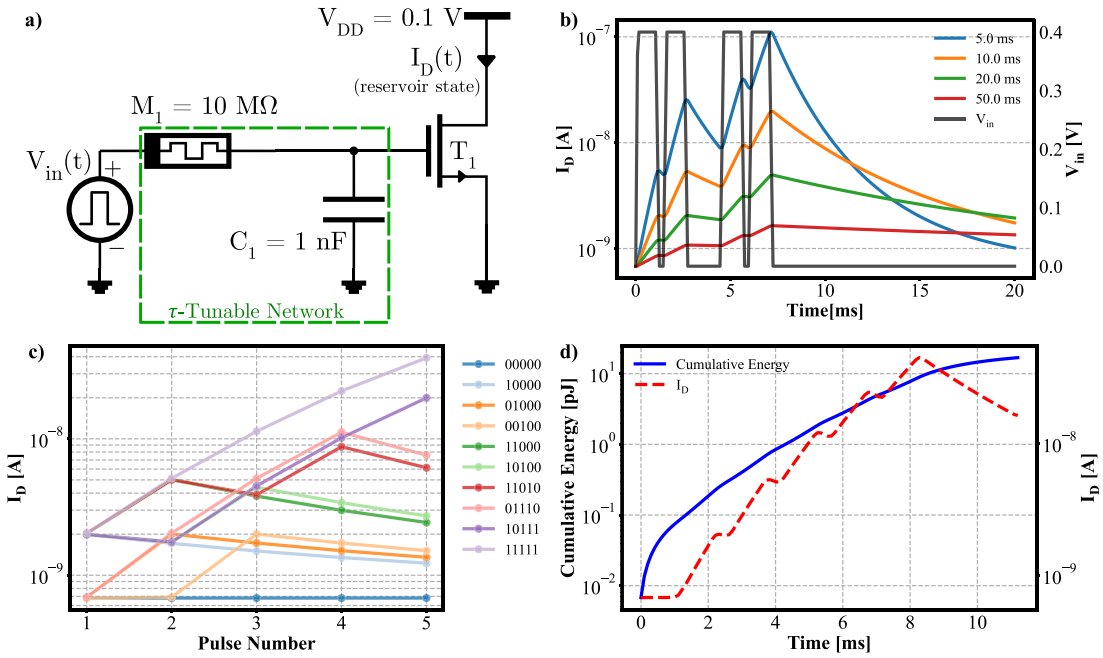

Figure 2. (a) Schematic of the ADRN circuit, where an input voltage pulse train V_{in}(t) was applied to a tunable-τ network connected to the gate of an NMOS FET. (b) Temporal evolution of the I_D in response to V_{in}(t) for different values of memristor resistance, demonstrating the tunable memory decay characteristics. (c) Distinct I_D trajectories for different binary input sequences, illustrating the reservoir's ability to map inputs into separable dynamic states. (d) Instantaneous I_D and cumulative energy consumption of the reservoir node over a sequence of input pulses.

input sequences due to the extended ON duration, indicating a trade-off between saturation and separability. A duty cycle of 2/3 was therefore chosen as the optimal compromise. Additionally, the reservoir performance was found to be dependent strongly on the input signal frequency relative to the intrinsic frequency $(1/\tau)$ of the node, as expected for an RC network. The optimal state separability was achieved when the input frequency was higher than $1/\tau$, but not too high ($<10/\tau$), allowing sufficient relaxation between successive inputs, thus leading to reservoir performance exhibiting a bell-shaped dependence on the input frequency.

During the simulations of the reservoir node operation, the ReRAM device was modelled, for simplicity, as a fixed resistor. This assumption is reasonable, as the device is pre-programmed to a target conductance state determined by the desired τ for a given application, and the voltage levels applied during operation remain well below the threshold required to induce further switching. Notably, an alternative implementation could employ a voltage-tunable capacitor, which would provide equivalent control over τ [40].

In order to evaluate its energy efficiency, the ADRN circuit was further analysed in terms of average energy consumed per time step. The energy consumed by the node was measured by applying five positive pulses, and the energy consumed at each simulation time step ($\mathrm{d}t = 10\,\mu$s) was calculated as

$$\mathrm{d}E = I_D(t) \times V_{DD} \times \mathrm{d}t, \tag{1}$$

where V_{DD} is the drain voltage. The average energy consumed was then evaluated by summing the energy consumption at each timestep and dividing by the total simulation time steps.

The cumulative energy consumption over a sequence of input pulses along with the evolution of I_D is shown in figure 2(d), which is of the order of picojoules. The average energy consumed per time step across 5 voltage pulses was calculated to be 15.20 pJ, and the average power consumed by the reservoir node was 6.97 nW. This demonstrates the energy efficiency at least one order of magnitude lower compared to other reservoir nodes [15, 22, 23, 41], thereby making it highly suitable for power-constrained applications.

3. Memory capacity and information retention

STM is a key metric for evaluating RC nodes [42], quantifying the system's ability to retain and reconstruct past inputs over a finite time window. In physical reservoirs, this fading memory emerges

naturally from the intrinsic temporal dynamics of the device. In the proposed ADRN circuit, shown in figure 2(a), the temporal dynamics or the memory behaviour is determined by the memristor and the capacitor, with the characteristic time constant given by $\tau = M_1 \times C_1$. Here, the programmable resistance of the memristor M_1 enables precise control over the memory depth of the ADRN. This tunable dynamical behaviour, combined with the exponential dependence of subthreshold I_D on a continuous-valued input V_{in}(t), introduces a nonlinear transformation that is essential for effective temporal information processing. The evolution of the gate voltage $V_G(t)$, in response to V_{in}(t), was modelled according to the dynamics of the memristor–capacitor network using the discrete-time equation in equation (2),

$$V_G(t+1) = V_G(t) + \frac{\Delta t}{\tau}\left(V_{in}(t) - V_G(t)\right). \tag{2}$$

Further, this time-varying gate voltage $V_G(t)$ modulates the transistor's subthreshold drain current which was modelled using Enz–Krummenacher–Vittoz model [43], as given in equation (3). Here V_G is the gate-to-source voltage, V_{th} is the threshold voltage, κ is the subthreshold slope factor, V_t is the thermal voltage, μ is the carrier mobility, W/L is the aspect ratio and C_{ox} is the gate oxide capacitance per unit area,

$$I_D = \left(\frac{W}{L}\right)\mu C_{ox} V_t^2 \left[\ln\left(1+\exp\left(\frac{\kappa\left(V_G - V_{th}\right)}{2V_t}\right)\right)\right]^2. \tag{3}$$

This time-varying continuos drain current $I_D(t)$, sampled discretely generate the 'state vectors' for the readout layer. Further, to emulate a high-dimensional reservoir state space without replicating multiple physical nodes, the concept of virtual nodes was employed. This was done by sampling the node's transient response at several discrete intervals within each input holding period T_p. Each sampled point, individually referred to as a virtual node, represents a distinct phase of the drain current evolution $I_D(t)$, capturing different stages of the circuit's temporal dynamics. 'V' virtual nodes are therefore defined as a set of equally spaced sampling points within the input period (T_p), such that the interval between two consecutive nodes is T_p/V. The states corresponding to these virtual nodes collectively form the reservoir state at time t and are linearly combined at the output layer. This sampling enriches the representation of reservoir nodes' temporal evolution and effectively expands the dimensionality of its state space, thereby enabling richer dynamics and improved linear separability at the readout layer.

To evaluate the STM capacity, a standard methodology for physical reservoirs was employed involving a signal reconstruction task. An input voltage–time signal was generated by creating 1000 scalar samples from a uniform distribution in the range [0, 1]. Each sample was then scaled by 0.4 to keep the FET within its subthreshold operating region, given the threshold voltage $V_{TH} = 0.46$ V for the 45 nm PTM device model [39]. The simulation time step was set to $\Delta t = 1$ ms, with the node time constant fixed at $\tau = 10$ ms. Each voltage value was held constant for 50 timesteps, forming a piecewise-constant input signal of period $T_p = 50 \times \Delta t = 50$ ms, effectively implementing a sample-and-hold operation. The resulting continuous waveform $V_{in}(t)$ was applied to the ADRN node, and the corresponding drain current $I_D(t)$ was recorded at discrete sampling points determined by the number of virtual nodes. These recorded states were subsequently used to construct reservoir state vectors. We reconstructed delayed versions of the input signal for time delays $t_\circ$ ranging from 1 to 50 input-periods (T_p). Specifically, for each $t_\circ$, an independent Ridge regression model was trained to predict $V_{in}(t - t_\circ)$ or $y_{t_\circ}$ from the instantaneous reservoir state. The reconstruction quality is quantified by the squared Pearson correlation coefficient (r^2) as given in equation (4), where $y_{t_\circ}$ is the target signal, $\hat{y}_{t_\circ}$ is the predicted signal, Cov(·) denotes covariance, and Var(·) denotes variance. The total memory capacity (MC) was then calculated as the sum of $r^2_{t_\circ}$ over a range of delays, as given in equation (5),

$$r^2_{t_\circ} = \frac{\mathrm{Cov}^2\left(y_{t_\circ}, \hat{y}_{t_\circ}\right)}{\mathrm{Var}\left(y_{t_\circ}\right)\mathrm{Var}\left(\hat{y}_{t_\circ}\right)}, \tag{4}$$

$$\mathrm{MC} = \sum_{t_\circ=1}^{t_{\circ\,\max}} r^2_{t_\circ}. \tag{5}$$

A higher memory capacity indicates improved retention and reconstruction of temporally distant inputs. Figure 3(a) illustrates how the STM capacity scales with the number of virtual nodes, created by time multiplexing the input signal on a single physical reservoir node [44, 45]. Another critical parameter influencing the memory capacity, and thereby the overall performance of the proposed ADRN, is the SS, especially given that the reservoir node is FET-based and functions in the subthreshold regime. A lower SS improves off-state current suppression without significantly changing the on-state current

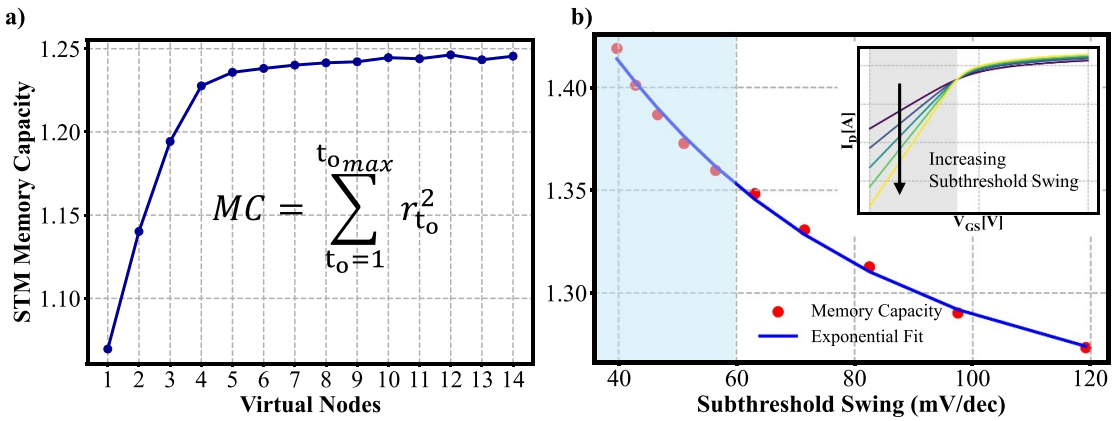

Figure 3. Memory capacity as a function of (a) the number of virtual nodes; and (b) the subthreshold swing of the FET. The shaded region in (b) indicates operation beyond the Boltzmann limit. The inset in (b) shows transfer characteristics of the FET for different SS values, with the threshold voltage kept constant.

values, as shown in the inset of figure 3(b). Since the FET is operating in the subthreshold region, the node's response is strongly affected by the nonlinearity and the range of the drain current I_D, which is exponentially dependent on the gate voltage. In order to analyse this effect, the subthreshold voltage was fixed at 0.4 V while the SS was varied by changing the value of κ in equation (3), and the memory capacity of ADRN was calculated at each step. As shown in figure 3(b), the memory capacity was observed to increase exponentially with the decrease in SS. This increase in STM capacity with decreasing SS was attributed to the enhanced readout sensitivity. In the proposed node, the linear memory of past inputs, stored in the gate voltage $V_G(t)$, evolves through the linear RC network, while the subthreshold FET acts as a nonlinear activation transforming it into the output current $I_D(t)$. A lower SS, equivalent to a steeper subthreshold slope, provides higher transconductance gain, amplifying subtle variations in $V_G(t)$ into a wider dynamic range of $I_D(t)$. This enhanced gain improves the linear separability of reservoir states, thereby increasing the measurable memory capacity. Interestingly, the memory capacity continues to improve even beyond the classical Boltzmann limit of $\sim 60\,\text{mV dec}^{-1}$ at room temperature. This result suggests that devices such as tunnel FETs, which can achieve sub-60 mV dec^{-1} SS, can further enhance memory capacity in such systems. To the best of our knowledge, the impact of SS on memory capacity in FET-based reservoir nodes has not been explicitly investigated in earlier reports.

4. Application modelling

The performance of the proposed node was evaluated on two representative benchmark tasks: (i) handwritten digit classification, and (ii) dynamic prediction of a chaotic time series generated by the Hénon map. For both tasks, the temporal evolution of the gate voltage in response to incoming input pulses was modelled using equation (2), which captures the voltage integration and decay governed by the node's intrinsic τ. The resulting drain current dynamics are then computed using equation (3). Further, the effectiveness of the proposed node was thoroughly characterised by evaluating the impact of device-level non-idealities and hyper-parameter sensitivity.

4.1. Pattern recognition—digit classification

In order to demonstrate the pattern recognition capabilities of the proposed ADRN, we performed the task of handwritten digit classification using the MNIST dataset [46]. The MNIST dataset, although static, is widely used in RC to benchmark its performance. The spatial dataset was encoded temporally and fed to the reservoir to validate the tunable memory dynamics of the proposed ADRN. The network uses a ddRC reservoir architecture with 88 parallel reservoir nodes, as shown in figure 1(b). As illustrated in figure 4(a), the processing begins at the 'Input' layer, by binarising the raw 28×28 grey-scale images using a simple thresholding operation with the threshold of 128. Further, the images were centre-cropped to 22×20 pixels by removing marginal pixel rows and columns. Next, as shown in the 'Reservoir' layer of figure 4(a), each of the 22 rows of pre-processed 22×20 binary images was divided into four non-overlapping 5 pixel segments, resulting in 88 segments. Each 5-bit segment generates a binary pulse stream, where '1' was encoded as a high-voltage pulse and '0' as no pulse. These pulse sequences are applied in parallel to the 88 reservoir nodes, enabling spatial patterns to be encoded as

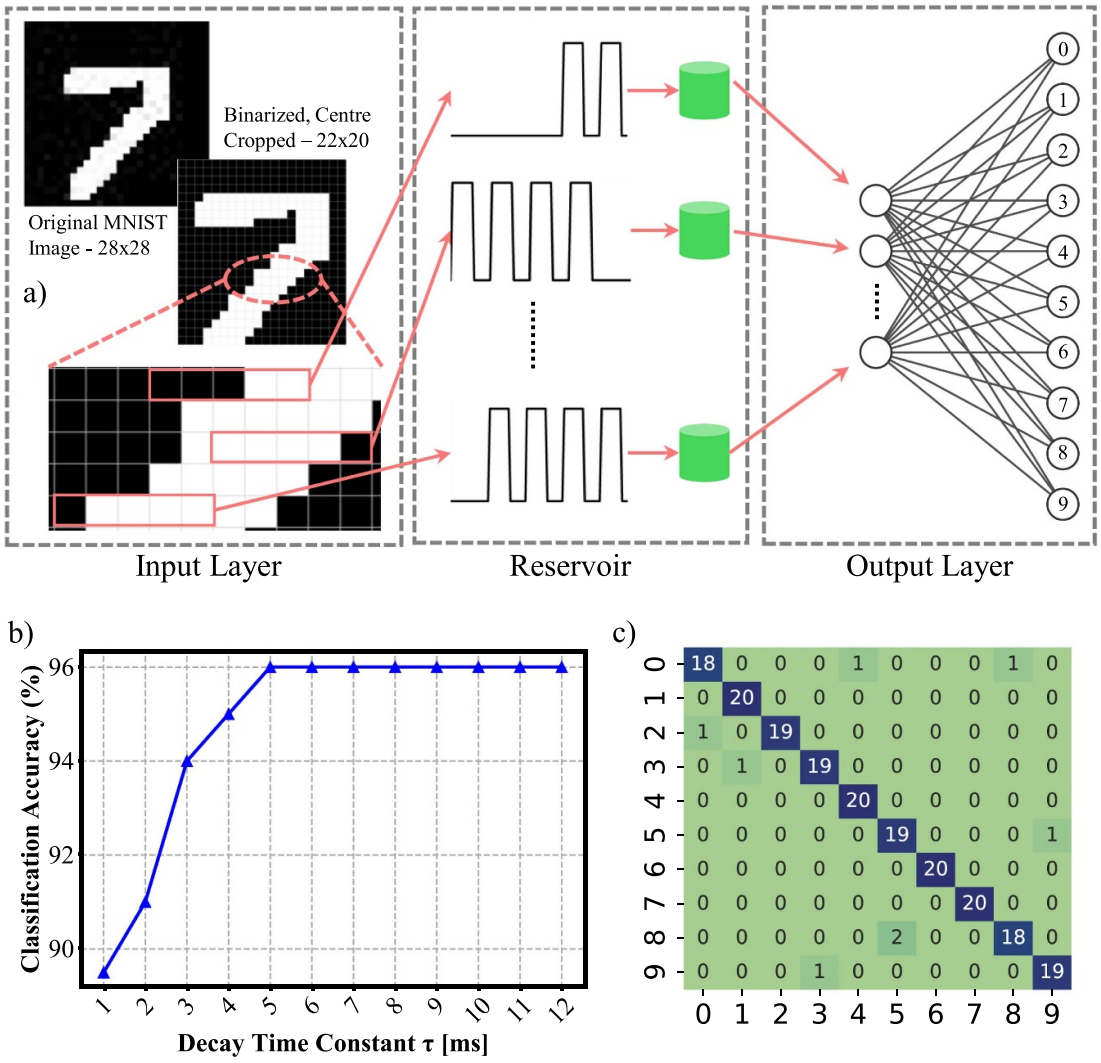

Figure 4. (a) Process flow for handwritten digit recognition using a ddRC network based on the proposed ADRN shown in three stages: first, image binarisation, cropping, and grouping of pixels; second, pixel-to-pulse encoding and reservoir processing; and third, feature extraction and classification using a regression-based readout layer. (b) Classification accuracy as a function of the reservoir decay time constant τ. (c) Confusion matrix for digit classification of digits from 0 to 9, showing true labels on x-axis and predicted labels on the y-axis.

time-varying signals. This temporal encoding allows the aforementioned spatial patterns to be processed dynamically within the reservoir framework to create a high-dimensional space for read-out layer training, shown in the 'Output' layer of figure 4(a).

Proper tuning of pulse parameters, such as PW, pulse amplitude, and the time period, is essential for optimal reservoir performance. The pulse amplitude was set to $V_{\text{pulse}} = 0.4\ V$ to ensure operation strictly within the subthreshold regime, i.e. $V_G < V_{TH}$, thereby avoiding saturation, while effectively modulating the gate voltage. Similarly, to ensure sufficient relaxation between inputs and thereby maximise state separability, the PW and the period (T) of the input pulses were set to $\tau/10$ and $1.5 \times PW$, respectively, where $\tau = M_1 \times C_1$ is the reservoir's time constant. It is important to note that for digit recognition, which is a static task, the temporal pulse parameters can be user-defined. In such cases, it is the relative magnitudes of PW, T, and τ that affect prediction accuracy and absolute timescales have little effect on the prediction accuracy. However, when these parameters are fixed by system or task constraints, the tunability of τ becomes critical for performance optimisation. For instance, with $PW = 1$ ms, $T = 1.5$ ms, the highest classification accuracy of 96% was obtained at $\tau \approx 10 \times PW = 10$ ms, as shown in figure 4(b), with performance degrading below $\tau = 5$ ms.

A dataset of 2000 images was randomly sampled from the MNIST training set, maintaining a uniform distribution across ten classes. Each image was converted into 88 voltage-time sequences, which were applied to the reservoir nodes. The final drain current from each node was sampled after the input sequence, resulting in an 88-dimensional feature vector, representing the image in the reservoir's state

space. These feature vectors collectively formed the dataset used for training the readout layer. A simple multiclass logistic regression model was used as the readout network to classify the reservoir state vectors. The model was trained using the standard cross-entropy loss function (log loss), which measures, for each sample, the divergence $\mathcal{L}^{(n)}$ between predicted probabilities $\hat{y}_i^{(n)}$ and the one-hot encoded ground truth $y_i^{(n)}$ as given by equation (6), where n is the sample index and i is the class index ranging from 1 to 10.

$$\mathcal{L}^{(n)} = -\sum_{i=1}^{10} y_i^{(n)} \log \hat{y}_i^{(n)} \quad \Rightarrow \quad \mathcal{L}^{(n)} = -\log \hat{y}_{k^{(n)}}^{(n)}. \tag{6}$$

The dataset was split into 80% for training (1600 samples) and 20% for testing (400 samples). The trained logistic regression classifier achieved an accuracy of 96%, as illustrated by the confusion matrix shown in figure 4(c), which is comparable to accuracies reported in existing hardware RC implementations table 1. It should be noted that the reported 96% accuracy was obtained from circuit-level simulations under ideal operating conditions. However, in practical hardware, additional non-idealities such as device variability and parasitic effects may slightly reduce the prediction accuracy of the reservoir.

4.2. Time series prediction—Hénon map

The advantage of τ-tunability in the proposed ADRN was evaluated on a dynamic time-series prediction task using the Hénon map [47]. It is a standard benchmark in RC due to its chaotic dynamics, sensitivity to initial conditions, and long-range temporal correlations.

The Hénon map, defined on a discrete two-dimensional state space (x_n, y_n), is governed by the equation (7), where the parameters $a = 1.4$ and $b = 0.3$ were used to set the system in its chaotic regime. A zero-mean Gaussian noise g_n with a standard deviation of 0.05 was used to model realistic perturbations [32].

$$x_{n+1} = y_n + 1 - a x_n^2, \quad y_{n+1} = b x_n + g_n. \tag{7}$$

To simplify reservoir interfacing and focus on one-dimensional temporal dependencies, the system was reformulated into a single recursive equation given by equation (8), thus generating the input signal, as shown in figure 5(a), with a sequence length of 2000 samples. Out of this, 1000 samples were used for training and the remaining 1000 for testing in a one-step-ahead prediction task.

$$x_{n+1} = b x_{n-1} - a x_n^2 + 1 + g_{n-1}. \tag{8}$$

To enhance the reservoir response, the generated input signal vector $\mathbf{V_{1\times N}}$ of length N was preprocessed using a binary masking method [44, 48]. Conceptually, the masking process serves a similar role to that of virtual nodes, in the sense that both techniques increase the effective dimensionality and dynamical richness of the reservoir. In the processing of masking, each element of the input vector $V_{1\times N}[i]$ was element-wise multiplied with a randomly generated binary mask vector $M_{1\times L}$ of length L, where each element of the mask vector $M_{1\times L}[i] \in \{-1, +1\}$, producing L temporally modulated copies of the input. This transformation can be mathematically expressed by equation (9), where $\mathbf{\tilde{V}_{1\times N\cdot L}}$ is the final expanded output sequence after the masking and concatenation process, and $[i \cdot L : (i+1) \cdot L]$ indicates the segment of the final vector $\mathbf{\tilde{V}_{1\times N\cdot L}}$ that is generated by the element-wise multiplication,

$$\mathbf{V_{1\times N}[i] \cdot M_{1\times L} = \tilde{V}_{1\times N\cdot L}[i \cdot L : (i+1) \cdot L]}. \tag{9}$$

The bipolar binary masking ($\{-1, +1\}$) is a widely used method in delay-based and dynamic-device RC systems. It was preferred over unipolar ($\{0, +1\}$) masking, as the sign-alternating modulation introduces richer variations in the input sequence, thereby increasing state diversity and improving prediction performance. Further, a multilevel mask, though capable of providing marginal performance improvement, was not adopted as it substantially increases hardware complexity and implementation overhead [49].

The masking process can be repeated n times with different masks to simulate the presence of n parallel reservoir nodes, shown in figure 5(a). The resulting masked sequences $\mathbf{\tilde{V}_{1\times N\cdot L}}$ were then normalised to the range $(0, 1)$ using equation (10) and mapped to voltages compatible with the subthreshold operation of the FET using equation (11), where $V_{\min}$ and $V_{\max}$ are the minimum and the maximum operation voltages, and $\mathbf{\tilde{v}_{1\times N\cdot L}}$ represents the final transformed voltage sequence,

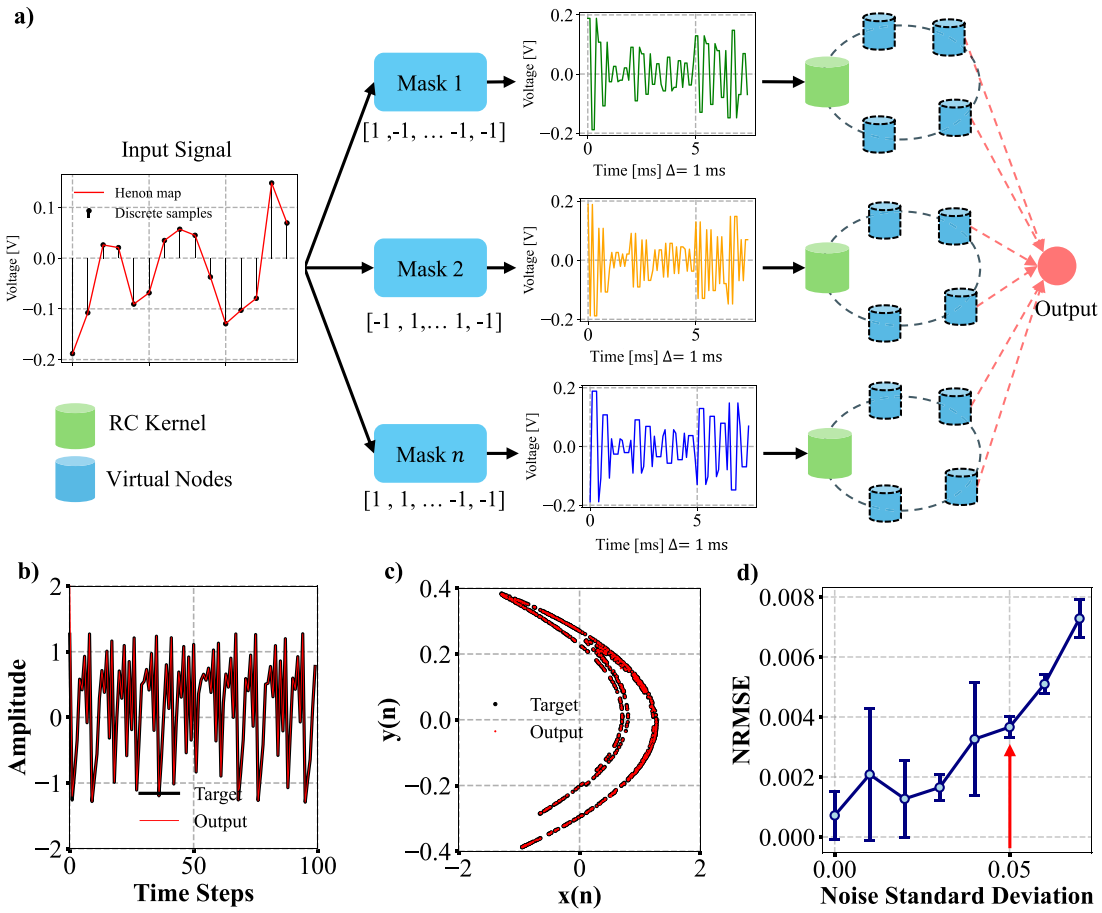

Figure 5. (a) Complete workflow of the time-series prediction task using the proposed ADRN. The input signal was first processed through a masking mechanism to generate corresponding output voltages. These voltages are then fed into the reservoir, whose states are used for target prediction through linear regression. (b) Predicted versus actual Hénon series amplitudes, with an NRMSE of 0.0037, demonstrating high prediction accuracy. (c) Reservoir output in phase space, depicting the Hénon attractor and the close conformity of prediction to the target trajectory. (d) NRMSE under the influence of different levels of Gaussian noise g_n with increasing standard deviation.

$$\tilde{\mathbf{V}}_{\mathbf{1\times N\cdot L}} = \frac{\tilde{\mathbf{V}}_{\mathbf{1\times N\cdot L}} - \min(\tilde{\mathbf{V}}_{\mathbf{1\times N\cdot L}})}{\max(\tilde{\mathbf{V}}_{\mathbf{1\times N\cdot L}}) - \min(\tilde{\mathbf{V}}_{\mathbf{1\times N\cdot L}})}, \tag{10}$$

$$\tilde{\mathbf{v}}_{\mathbf{1\times N\cdot L}} = V_{\min} + (V_{\max} - V_{\min}) \cdot \tilde{\mathbf{V}}_{\mathbf{1\times N\cdot L}}. \tag{11}$$

Each value in the sequence was held constant for a time duration $\Delta = 1$ ms, creating a piecewise-constant or 'sample-and-hold' input voltage signal thereby setting an intrinsic time scale of the signal to be processed by the ADRN network. This temporal voltage sequence was applied to the ADRN, which dynamically evolved its internal state in response to the time-varying input. The non-linear subthreshold dynamics and tunable memory properties enabled the reservoir to capture both STM and complex temporal dependencies.

Finally, the readout layer was implemented as a linear regression model to obtain the optimised weight matrix W_{out}, which was computed directly using the Moore–Penrose pseudo-inverse $W_{\text{out}} = TX^{\dagger}$ [50], where T is the target vector and X is the matrix of recorded states. This W_{out} was then used to predict the next 1000 samples for the input test sequence. The reservoir's performance was quantitatively assessed using the normalised root mean square error (NRMSE) given by equation (12), where y_t and $\hat{y}_t$ are the expected and the predicted output states, respectively.

$$\text{NRMSE} = \sqrt{\frac{\langle \| y(t) - y_{\text{target}}(t) \|^2 \rangle}{\langle \| y(t) - \langle y(t) \rangle \|^2 \rangle}} \tag{12}$$

For $n = 1$, $L = 30$, and $\tau = 3$ ms, the predicted outputs closely matched the true values, as shown in figure 5(b), achieving a minimum NRMSE of 0.0037. An exact overlap between the scatter plots of the

predicted and the original Hénon attractor shown in figure 5(c) confirms that the learned dynamics preserved the curvature of the chaotic map, indicating accurate reconstruction of the underlying behaviour. Such prediction accuracy positions the ADRN network on par with state-of-the-art physical reservoir implementations.

To evaluate the network's tolerance to noise and assess its prediction robustness, the standard deviation (σ) of the previously introduced Gaussian noise (g_n) was varied, as shown in figure 5(d). The NRMSE increased gradually with rising noise levels, while the reservoir remained stable and reproduced the chaotic sequence under moderate noise conditions. However, beyond $\sigma = 0.07$, the NRMSE exhibited a sharper rise, and occasional non-convergence of the readout layer was observed, making it increasingly difficult for the network to predict the chaotic sequence reliably.

5. Discussion

Careful tuning of several hyperparameters, i.e. the control knobs of a physical reservoir, is essential to optimise its dynamics for a given task, achieving the highest possible prediction accuracy while keeping processing time, energy consumption per operation, and hardware complexity to a minimum [2]. Here, we analyse the effect of Δ, L, τ, and memristor variability on the prediction accuracy of the ADRN.

5.1. Performance Across Tasks

Figure 6(a) shows the NRMSE of the Hénon sequence prediction as a function of Δ and τ, with the mask length fixed at $L = 10$. For $\Delta = 1$ ms, a parametric sweep of τ reduced NRMSE from 0.05 for $\tau = 10$ ms, to 0.0037 at $\tau = 3$ ms. The lower-right corner of the heatmap exhibits the highest NRMSE, as a large τ prevents the ADRN from effectively tracking fast-varying input signals. These results indicate that, to achieve minimum NRMSE across different input signal frequencies (i.e. different Δ), the reservoir's temporal dynamics (τ) must be tuned in accordance with Δ, reinforcing the importance of post-fabrication τ-tunability in the proposed ADRN.

Another important parameter in the design of a physical reservoir computer is the mask length L, which numerically corresponds to the number of virtual nodes. As illustrated in figure 6(b), for a given Δ, increasing L generally reduces the NRMSE by providing more virtual nodes for computation. For $\Delta = 1$ ms, $\tau = 10$ ms, and a parametric sweep of L reduced NRMSE from 0.05 for $L = 10$ ms, to 0.004 at $L = 30$. However, when the intrinsic τ is either too short or too long relative to Δ, the improvement from increasing L becomes marginal, offering limited scope for further NRMSE optimisation. Figure 6(c) clearly shows the variation of NRMSE with τ, for given values of Δ and L. Figure 6(d) further reveals the influence of τ on the STM capacity of ADRN (discussed earlier in section 2). When τ is too small, ADRN rapidly loses stored information, while a large value of τ leads to saturation with time, thus degrading Memory Capacity. These effects are consistent with the performance trends observed in figures 6(a) and (b).

The benefits of post-fabrication τ-tunability extend beyond accuracy to energy efficiency. By tuning the temporal dynamics to match the target input characteristics, the ADRN can achieve high prediction accuracy with fewer virtual nodes, smaller L, and shorter processing times, thereby directly reducing energy usage. For example, at $L = 10$, an NRMSE of 0.004 was achieved with an energy cost of $\sim$100 nJ, whereas increasing L to 100 yielded only a marginal NRMSE improvement to 0.0037 but increases the energy consumption to $\sim$1000 nJ. In contrast, by tuning τ at $L = 10$, a comparable NRMSE of 0.0039 was achieved, as noted earlier. For the Hénon series prediction task with optimised parameters, the cumulative energy consumed was 3.93 nJ.

Moreover, by operating the FET in its subthreshold region, the proposed ADRN circuit achieves exceptionally high intrinsic energy efficiency. The energy consumption of ADRN per operation (E/Op) was 15.20 pJ, representing at least an order-of-magnitude improvement over state-of-the-art RC implementations listed in table 1. The reported energy corresponds to the energy dissipated across the memristor M_1 and the FET T_1, obtained from circuit-level simulations. This value, however, does not account for peripheral contributions such as pulse generation and readout circuitry, and hence actual system-level energy is expected to be slightly higher, depending on the implementation of the peripheral circuitry.

5.2. Robustness to variability

Variability in memristor resistance is a key factor affecting feature generation in the proposed ADRN, as the programmed resistance directly sets the reservoir node decay constant τ. To evaluate the system's tolerance to ReRAM variability, the memristor resistance (M_1) across reservoir nodes was modelled as samples drawn from a Gaussian distribution centred at the optimal resistance for each task, with the

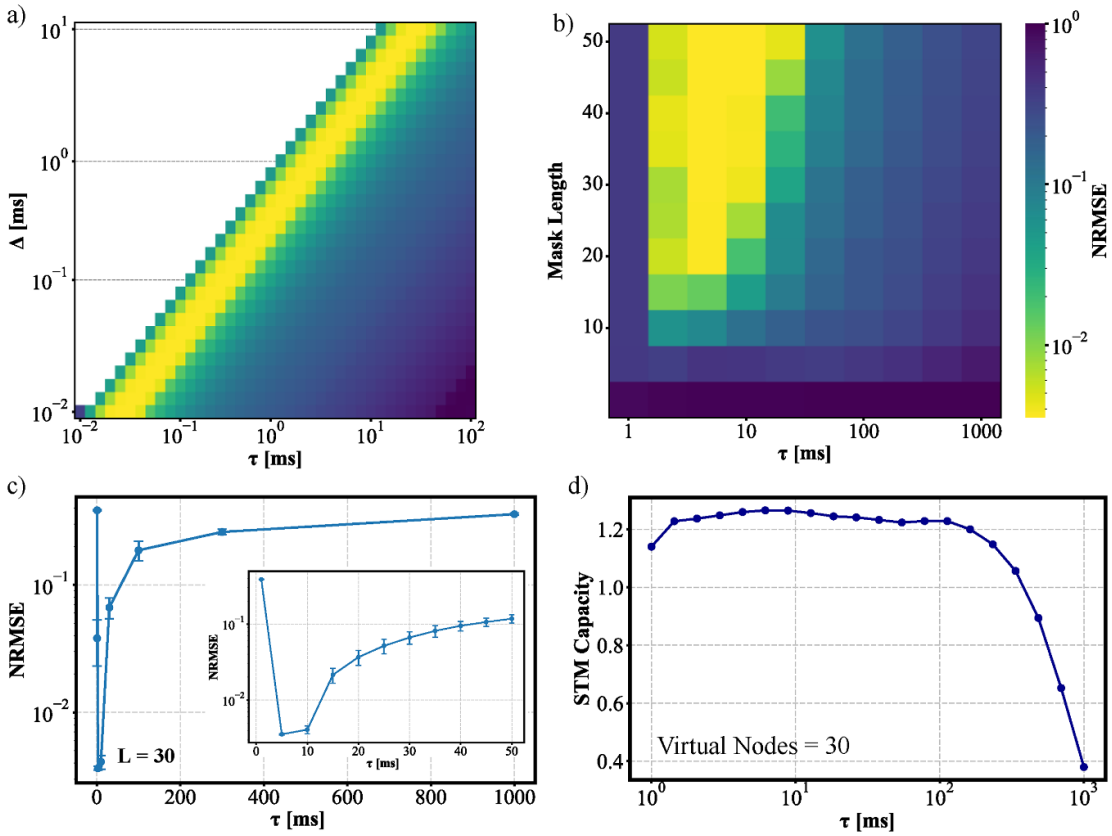

Figure 6. Heatmap of Hénon sequence NRMSE (a) as a function of Δ and τ, with mask length fixed at $L = 10$, and (b) as a function of mask length and τ for $\Delta = 1$ ms. The colour scale spans from the minimum NRMSE in yellow to the maximum in Blue, with yellow regions indicating optimal parameter combinations. (c) Variation of the NRMSE with τ for $L = 30$ and $\Delta = 1$ ms. (d) STM capacity versus τ for number of virtual nodes = 30 and $\Delta = 1$ ms.

Table 1. Comparison of energy consumption and predictive performance for various systems. The kernel presented in this work exhibits competitive energy efficiency and excellent performance. The numbers in parentheses and the references marked with an asterisk correspond to simulation based results.

		Performance				
System	Dynamics for RC	Hénon	MNIST	E/Op (nJ)	Comment	References
Oxide-ReRAM	Short-term dynamics of memristors	97.60% (98.44%)	0.046 (0.01)	10^0 (10^0)	Absence of post-fabrication τ tunability	[22, 23, 29*, 30*]
Optoelectronic	Optical delay-based	97.14% (97.80%)	—	10^{-2}	Absence of CMOS compatibilty	[15–17*]
Fe-FETs	Decay dynamics of Leaky Fe-FET	95.10%	—	10^1	Increased energy consumption due to low frequency operation	[5, 9]
Fe-Memristors	STM in Ferroelectric Devices	89.50%	0.017	10^2		[28]
Si-FETS	BTI, Random Variation in V_{th}	—	—	(10^{-1})	Non-deterministic effects employed for RC dynamics	[8*, 10*, 11]
Biomembranes	Memristive / Memcapactive behaviour of biomembrane	—	—	10^{-1} (10^{-5})	Exotic device, immature process	[20, 21*]
Si-FET + tunable ReRAM	ReRAM-Capacitor-FET (in subthreshold region)	96.00%	0.0037	10^{-3}	CMOS-compatible, Tunable τ with Deterministic Operation	This work

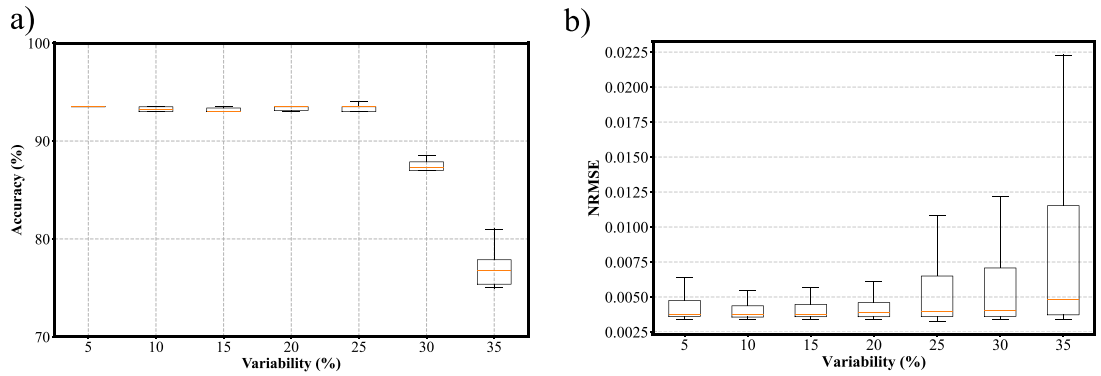

Figure 7. (a) Classification accuracy of the handwritten digit recognition task over increasing levels of variability of the memristor resistance state. (b) NRMSE values of the Hénon Map prediction task over increasing levels of variability of the memristor resistance state. Here, τ was fixed at the optimal value of 3 ms from our previous analysis.

standard deviation representing the specified variability level. The impact of this variability on performance was assessed through multiple independent random trials, where a new set of resistance values was generated in each run. Considering computational constraints, a total of 100 trials for the Hénon map prediction task and 10 for the MNIST classification task, were conducted for each value of standard deviation. The performance metrics reported in figure 7 represent the mean accuracy and the NRMSE averaged over these independent trials, with corresponding quartiles and ranges, thereby providing a statistical measure of the system's resilience to physical non-idealities. It can be seen from figure 7(a), that the mean accuracy for the handwritten digit recognition task remains stable up to 15% variability and exceeds 90% even at 25% variability. For the Hénon map prediction task (figure 7(b)), mean NRMSE increases by less than 0.01 for up to 20% variability, confirming the robustness of the proposed node to variability. In the context of ReRAM-based implementation, endurance can be a concern; however, since τ-tuning is an infrequent operation, its impact remains minimal. Also, BEOL-compatible integration enables scalable per-node control of τ.

6. Conclusion

We have demonstrated a CMOS-compatible, energy-efficient hybrid FET-ReRAM reservoir node and computationally analysed its performance using circuit-level simulations, addressing the critical need for on-the-fly optimisation of reservoir temporal dynamics. This capability, essential for processing diverse temporal tasks, is showcased using two representative benchmarks, MNIST digit classification and chaotic Hénon map prediction, achieving 96% accuracy and an NRMSE of 0.0037, respectively, comparable to state-of-the-art physical reservoirs. The ADRN achieves post-fabrication τ tuning over several orders of magnitude by programming the ReRAM conductance, enabling performance adaptation without hardware modification. Subthreshold FET operation delivers rich nonlinearity at sub-nanoampere currents with a few picojoules of energy per operation, while BEOL-compatible ReRAM provides robust, tunable memory decay. This combination establishes a scalable, adaptable, and energy-efficient hardware platform for next-generation physical RC.

Data availability statement

All data that support the findings of this study are included within the article.

Author contributions

G S C and S P developed the conceptual framework for the study. G S C and A D A performed the data analysis, modelling, numerical simulations, and benchmark study. A A and S P wrote the initial manuscript draft. All authors contributed to the interpretation of results and the review of the manuscript. The project was supervised by A A.

ORCID iDs

Gambali Seshasai Chaitanya 0009-0008-3172-4243
Aditya D Arkalgud 0009-0004-2652-9516
Shubham Pande 0000-0002-2268-4525
Ankit Arora 0000-0003-4899-1336

References

[1] Sally A M 2004 Reflections on the memory wall *Proc. 1st Conf. Computing Frontiers, (CF'04)* (Association for Computing Machinery) p 162
[2] Cucchi M, Abreu S, Ciccone G, Brunner D and Kleemann H 2022 Hands-on reservoir computing: a tutorial for practical implementation *Neuromorph. Comput. Eng.* **2** 032002
[3] Yan M, Huang C, Bienstman P, Tino P, Lin W and Sun J 2024 Emerging opportunities and challenges for the future of reservoir computing *Nat. Commun.* **15** 2056
[4] Liang X, Tang J, Zhong Y, Gao B, Qian H and Huaqiang W 2024 Physical reservoir computing with emerging electronics *Nat. Electron.* **7** 193–206
[5] Lee G, Kang C, Kim S, Park Y, Shin E J and Cho B J 2023 Physical reservoir based on a leaky-FeFET using the temporal memory effect *IEEE Electron Device Lett.* **45** 108–11
[6] Tang M, Zhan X, Shuhao W, Bai M, Feng Y, Zhao G, Jixuan W, Chai J, Hao X and Wang X *et al* 2022 A compact fully ferroelectric-FETs reservoir computing network with sub-100 ns operating speed *IEEE Electron Device Lett.* **43** 1555–8
[7] Duong N T, Chien Y-C, Xiang H, Sifan Li, Zheng H, Shi Y and Ang K-W 2023 Dynamic ferroelectric transistor-based reservoir computing for spatiotemporal information processing *Adv. Intell. Syst.* **5** 2300009
[8] Kume Y, Bian S and Sato T 2020 A tuning-free hardware reservoir based on MOSFET crossbar array for practical echo state network implementation *2020 25th Asia and South Pacific Design Automation Conf. (ASP-DAC)* (IEEE) pp 458–63
[9] Tang M, Mei J, Zhan X, Wang C, Chai J, Hao X, Wang X, Jixuan W and Chen J 2023 Fully ferroelectric-FETs reservoir computing network for temporal and random signal processing *IEEE Trans. Electron Devices* **70** 3372–7
[10] Lin X Y, Ang D S, Ju X and Li J Y 2024 Silicon FET reservoir for dynamic edge vision *2024 Smart Systems Integration Conf. and Exhibition (SSI)* (IEEE) pp 1–4
[11] Guo Y, Degraeve R, Vandemaele M, Saraza-Canflanca P, Franco J, Kaczer B, Bury E and Verbauwhede I 2024 Exploiting bias temperature instability for reservoir computing in edge artificial intelligence applications *2024 IEEE Int. Reliability Physics Symp. (IRPS)* pp 1–7
[12] Farmakidis N, Dong B and Bhaskaran H 2024 Integrated photonic neuromorphic computing: opportunities and challenges *Nat. Rev. Electr. Eng.* **1** 358–73
[13] Ferreira T D, Silva N A, Silva D, Rosa C C and Guerreiro A 2022 Reservoir computing with nonlinear optical media *J. Phys.: Conf. Ser.* **2407** 012019
[14] Ohara S, Kanno K, Uchida A and Kurokawa H 2025 Prediction of intermittent chaos in a semiconductor laser with optical feedback using reservoir computing *Jpn. J. Appl. Phys.* **64** 012001
[15] Tsurugaya T, Hiraki T, Aihara T, Nakajima M, Diamantopoulos N-P, Segawa T and Matsuo S 2024 Reservoir computing using on-chip XGM-based nonlinear processing by membrane SOAs on SI-MZI *J. Lightwave Technol.* **42** 2859–67
[16] Sun Y, Li Q, Zhu X, Liao C, Wang Y, Li Z, Liu S, Hui X and Wang W 2023 In-sensor reservoir computing based on optoelectronic synapse *Adv. Intell. Syst.* **5** 2200196
[17] Yue D, Hou Y, Chunxia H, Zang C and Kou Y 2023 Handwritten digits recognition based on a parallel optoelectronic time-delay reservoir computing system *Photonics* **10** 236
[18] Marrows C H, Barker J, Moore T A and Moorsom T 2024 Neuromorphic computing with spintronics *npj Spintronics* **2** 12
[19] Silva N A, Ferreira T D and Guerreiro A 2021 Reservoir computing with solitons *New J. Phys.* **23** 023013
[20] Armendarez N X, Mohamed A S, Dhungel A, Hossain Md R, Hasan Md S and Najem J S 2024 Brain-inspired reservoir computing using memristors with tunable dynamics and short-term plasticity *ACS Appl. Mater. Interfaces* **16** 6176–88
[21] Hossain Md R, Mohamed A S, Armendarez N X, Najem J S and Hasan Md S 2023 Biomembrane-based memcapacitive reservoir computing system for energy-efficient temporal data processing *Adv. Intell. Syst.* **5** 2300346
[22] Zhong Y, Tang J, Xinyi Li, Gao B, Qian H and Huaqiang W 2021 Dynamic memristor-based reservoir computing for high-efficiency temporal signal processing *Nat. Commun.* **12** 408
[23] Moon J, Wen M, Shin J H, Cai F, Chao D, Lee S H and Wei D L 2019 Temporal data classification and forecasting using a memristor-based reservoir computing system *Nat. Electron.* **2** 480–7
[24] Jaafar A H, Shao Li, Dai P, Zhang T, Han Y, Beanland R, Kemp N T, Bartlett P N, Hector A L and Huang R 2022 3d-structured mesoporous silica memristors for neuromorphic switching and reservoir computing *Nanoscale* **14** 17170–81
[25] Wang T, Huang H-M, Wang X-X and Guo X 2021 An artificial olfactory inference system based on memristive devices *InfoMat* **3** 804–13
[26] Yang J, Cho H, Ryu H, Ismail M, Mahata C and Kim S 2021 Tunable synaptic characteristics of a ti/tio2/si memory device for reservoir computing *ACS Appl. Mater. Interfaces* **13** 33244–52
[27] Zhu X, Wang Q and Wei D L 2020 Memristor networks for real-time neural activity analysis *Nat. Commun.* **11** 2439
[28] Chen Z, Wenjie Li, Fan Z, Dong S, Chen Y, Qin M, Zeng M, Xubing L, Zhou G and Gao X *et al* 2023 All-ferroelectric implementation of reservoir computing *Nat. Commun.* **14** 3585
[29] Wu X, Lin Z, Deng J, Li J and Feng Y 2024 Nonmasking-based reservoir computing with a single dynamic memristor for image recognition *Nonlinear Dyn.* **112** 6663–78
[30] Zyarah A and Kudithipudi D 2024 Time-series forecasting and sequence learning using memristor-based reservoir system *ACM Trans. Embed. Comput. Syst.* **24** 1–17
[31] Liu K, Zhang T, Dang B, Bao L, Liying X, Cheng C, Yang Z, Huang R and Yang Y 2022 An optoelectronic synapse based on α-in2se3 with controllable temporal dynamics for multimode and multiscale reservoir computing *Nat. Electron.* **5** 761–73
[32] Yongfei Li, Tang W, Zhiyuan Li, Niu W, Han S, Miao X and Yang R 2025 Homogeneous memristors with tunable decay dynamics for self-adaptive reservoir computing *ACS Nano* **19** 28186–96

[33] Zahoor F, Zulkifli T Z A and Khanday F A 2020 Resistive random access memory (RRAM): an overview of materials, switching mechanism, performance, multilevel cell (MLC) storage, modeling and applications *Nanoscale Res. Lett.* **15** 90
[34] Yang J J, Strukov D B and Stewart D R 2013 Memristive devices for computing *Nat. Nanotechnol.* **8** 13–24
[35] Kim H, Mahmoodi M R, Nili H and Strukov D B 2021 4K-memristor analog-grade passive crossbar circuit *Nat. Commun.* **12** 5198
[36] Sharma D, Rath S P, Kundu B, Korkmaz A, Harivignesh S, Thompson D, Bhat N, Sreebrata Goswami R S W and Goswami S 2024 Linear symmetric self-selecting 14-bit kinetic molecular memristors *Nature* **633** 560–6
[37] Ielmini D and Pedretti G 2025 Resistive switching random-access memory (RRAM): applications and requirements for memory and computing *Chem. Rev.* **125** 5584–625
[38] Shen Z, Zhao C, Yanfei Q, Wangying X, Liu Y, Mitrovic I Z, Yang Li and Zhao C 2020 Advances of RRAM devices: resistive switching mechanisms, materials and bionic synaptic application *Nanomaterials* **10** 1437
[39] Cao Y 2013 *Predictive Technology Model for Robust Nanoelectronic Design* (Springer)
[40] Lee M *et al* 2024 Hafnium oxide-based ferroelectric devices for in-memory computing: resistive and capacitive approaches *ACS Appl. Electron. Mater.* **6** 5391–401
[41] Chen Z-L, Xiao Y, Huang W-Y, Jiang Y-P, Liu Q-X and Tang X-G 2023 In-sensor reservoir computing based on optoelectronic synaptic devices *Appl. Phys. Lett.* **123** 09
[42] Jaeger H 2001 *Short Term Memory in Echo State Networks* (GMD Forschungszentrum Informationstechnik)
[43] Bucher M, Lallement C, Enz C, Theodoloz F and Krummenacher F 1997 The Epfl-Ekv Mosfet model equations for simulation, model version 2.6 *Electronics Laboratories* (Swiss Federal Institute of Technology (EPFL))
[44] Appeltant L, Soriano M C, der Sande G V, Danckaert J, Massar S, Dambre J, Schrauwen B, Mirasso C R and Fischer I 2011 Information processing using a single dynamical node as complex system *Nat. Commun.* **2** 468
[45] Takagi S-ichi, Kasidit Toprasertpong E N, Rikuo Suzuki S-Y M, Takenaka M and Nakane R 2023 Physical reservoir computing using HZO-based FeFETs for edge-AI applications *2023 Int. Electron Devices Meeting (IEDM)* (IEEE) pp 1–4
[46] Li D 2012 The mnist database of handwritten digit images for machine learning research *IEEE Signal Process. Mag.* **29** 141–2
[47] Hénon M 1976 A two-dimensional mapping with a strange attractor *Commun. Math. Phys.* **50** 69–77
[48] Tanaka G, Yamane T, Héroux J B, Nakane R, Kanazawa N, Takeda S, Numata H, Nakano D and Hirose A 2019 Recent advances in physical reservoir computing: a review *Neural Netw.* **115** 100–23
[49] Soriano M C, Ortín S, Brunner D, Larger L and Mirasso C R 2013 Ingo Fischer and Luıs Pesquera. Optoelectronic reservoir computing: tackling noise-induced performance degradation *Opt. Express* **21** 12–20
[50] Lukoševičius M and Jaeger H 2009 Reservoir computing approaches to recurrent neural network training *Comput. Sci. Rev.* **3** 127–49