

Dynamic Ferroelectric Transistor-Based Reservoir Computing for Spatiotemporal Information Processing

Ngoc Thanh Duong, Yu-Chieh Chien, Heng Xiang, Sifan Li, Haofei Zheng, Yufei Shi, and Kah-Wee Ang*

Reservoir computing (RC) architecture which mimics the human brain is a fundamentally preferred method to process dynamical systems that evolve with time. However, the difficulty in generating rich reservoir states using two-terminal devices remains challenging, which hinders its hardware implementation. Herein, the 1D array of ferroelectric field-effect transistor (Fe-FET) based on α - In_2Se_3 channel, which shows volatile memory effect for realizing various RC systems, is demonstrated. The fading effect in α - In_2Se_3 is sufficiently investigated by polarization dynamic model. The proposed Fe-FET is capable of experimentally classifying images using MNIST dataset with a high accuracy of 91%. Furthermore, time-series real-life chaotic system, for example, Earth's weather, can be accurately forecasted using our Ferro-RC based on the Jena climate dataset recorded in a 1 year period. Remarkable determination coefficient (R^2) of 0.9983 and normalized root mean square error (NRMSE) of 8.3×10^{-3} are achieved using a minimized readout network. The demonstration of integrated memory and computation opens a route for realizing a compact RC hardware system.

1. Introduction

Artificial neural networks (ANNs) are emerging in many aspects of life as a powerful tool for data-centric processing including image recognition classification,^[1] natural language learning,^[2] and security,^[3] among others. Recurrent neural network (RNN) which is commonly known as one of the major categories of ANNs is designed for chronological data processing.^[4] Unlike the convolution and feedforward neural networks which assume

that the input and output layer are independent of each other, RNNs distinguished itself utilizing the information from prior input to influence the memory states of the current input and output. Therefore, RNN is theoretically classified into different types: one-to-many, many-to-one, and many-to-many for handling different data-types. To perform such distinctive computing tasks, the neurons in the hidden layer of RNN must be cyclically connected to each other. As a sequence, these cyclic connections of neurons may cause higher training cost as well as an increasing difficulty in the training process.^[5]

Two frameworks of RNN, long-short-term memory (LSTM) and reservoir computing (RC) can solve existing problems of conventional RNN in processing temporal data effectively. In particular, by using the fading effect in physical dynamic devices to map the input layer into a high-dimensional space of features, RC has been

demonstrated as an efficient machine learning paradigm for simplifying computational works, and a necessary amendment for RNN to reduce training cost is proposed.^[6,7,2b] Figure 1a represents a dynamic reservoir concept with random and fixed neural map. The temporal inputs in conventional RC are nonlinearly transformed into a reservoir of neurons which are recurrently and randomly connected such that the original features of input can be read out by simple algorithms.^[2b,7c] Cyclic RC architecture that was introduced in 2011 utilizes time-multiplexing process with a delay loop to generate multiple reservoir states. Specifically, by dividing feedback loop into M intervals, where τ is the delay time, as shown in Figure 1b, the information in input layer (U_i) is mapped to sequential nonlinear nodes in reservoir X_i .^[8] In these compelling studies, every single paralleled reservoir is formed by different mask sequences, and thus the weight-out matrix can be obtained by simple linear regression of all virtual nodes.^[7b,9] Satisfactory performance of RC has been realized in diverse time-series information including speech recognition,^[8b,9a] chaotic systems prediction,^[7a,8b,9a] and other applications,^[2b] utilizing various devices and materials such as memristors,^[2b,6a,8b,9a] optical modules,^[9b] and spintronic oscillators.^[10] Liu et al. demonstrated multilayer reservoir using series connection of ferroelectric transistors.^[11] The involvement of many physical devices for RC execution does not only cause an increasing energy consumption and training cost, but also

N. T. Duong, Y.-C. Chien, H. Xiang, S. Li, H. Zheng, Y. Shi, K.-W. Ang
Department of Electrical and Computer Engineering
National University of Singapore
4 Engineering Drive 3, Singapore 117583, Singapore
E-mail: eleakw@nus.edu.sg

K.-W. Ang
Institute of Materials Research and Engineering
A*STAR, 2 Fusionopolis Way, Singapore 138634, Singapore

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202300009>.

© 2023 The Authors. Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202300009

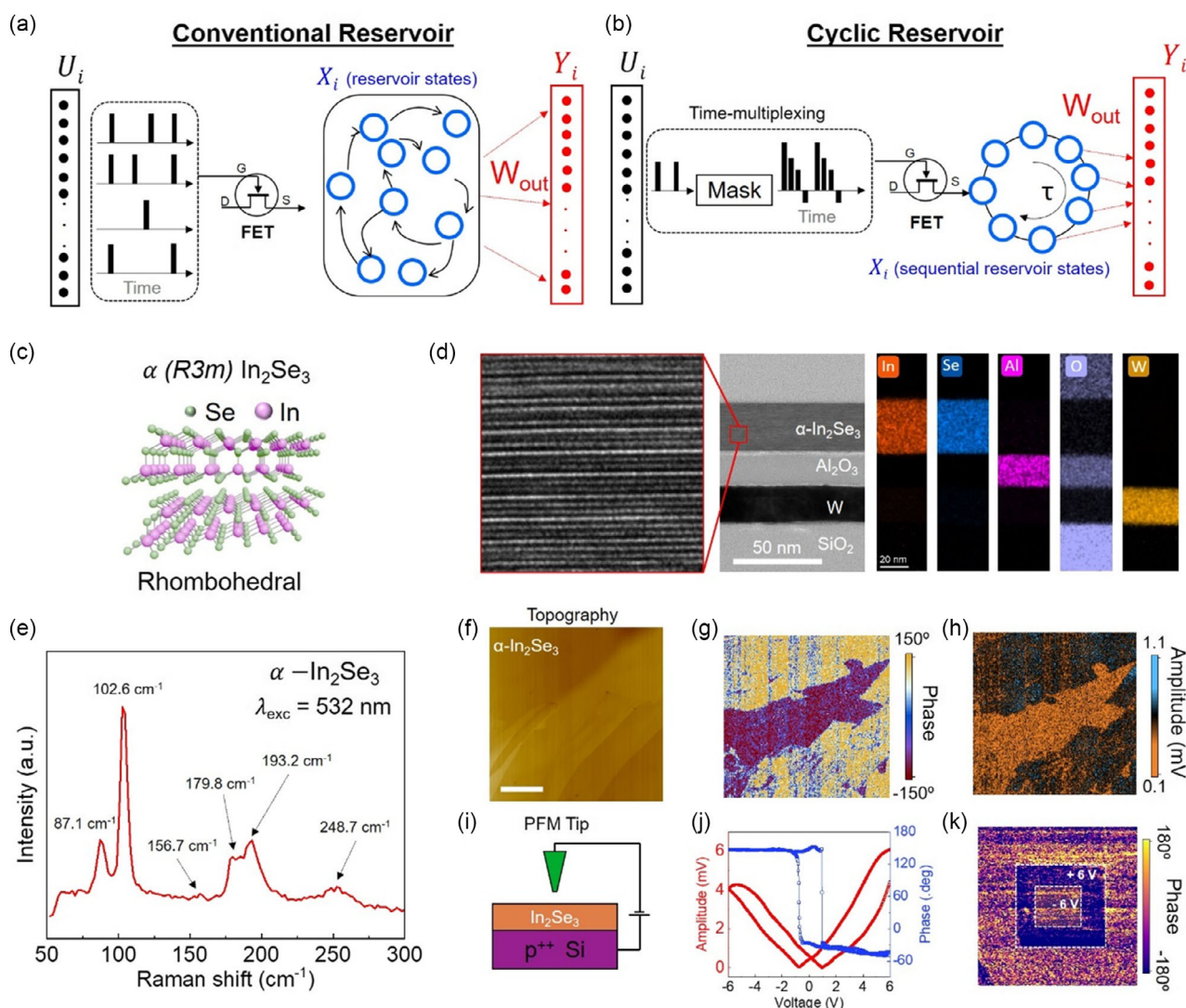


Figure 1. α -($R3m$) In_2Se_3 ferroelectric material properties. The RC architectures and correlations between input (U_i), reservoir (X_i) and output layer (Y_i) of a) Conventional reservoir, and b) Cyclic reservoir. c) Single crystalline Ferroelectric Semiconducting α ($R3m$) - In_2Se_3 . d) Layered structure indicated by HR-TEM image of α - In_2Se_3 extracted from $\text{In}_2\text{Se}_3/\text{Al}_2\text{O}_3/\text{W}$ heterostructure (second panel) and its EDS map with distribution of corresponding elements. e) Raman scattering spectrum of α ($R3m$) - In_2Se_3 . f) Topography g) Phase and h) Amplitude images of α - In_2Se_3 flake on p^{++} Si substrate, scale bar: 2 μm , showing different contrast between two oppositely oriented vertical domains. i) PFM measurement setup with conductive tip on an electrostatic cantilever. k) Equivalent hysteresis loops of phase and amplitude. l) Locally writing inner $2 \times 2 \mu\text{m}$ square (-6 V) and $4 \times 4 \mu\text{m}$ outer square (6 V) areas on α - In_2Se_3 flake.

limits accessibility to applications. In addition, the fundamental study of short retention in ferroelectric semiconductor in α - In_2Se_3 is still not enlightened. Through the dynamic and non-linear response of physical systems or devices, the experimental studies of RC are predominantly based on crossbar array of two-terminal memristors that function as a memory unit in von Neumann architecture instead of central processing modules.^[7a,12,13]

Herein, we demonstrated a 1D array of field-effect transistor (FET) based on ferroelectric α - In_2Se_3 channel material for realizing high-performance RC systems for neuromorphic computing applications. Unlike two-terminal memristors where writing/reading process is performed through one electrode, our

devices are programmed via back gate terminal while response currents are characterized by drain/source. The natural ferroelectricity of α - In_2Se_3 and its carrier transport characteristic of the FET were sufficiently examined. More appropriately, we proposed polarization dynamic in ferroelectric semiconductor FET using read-after-write model. The volatile memory reliably exhibits high endurance up to 10^4 programming/erasing cycles with negligible device-to-device variability. To verify the ability of Ferro-RC systems, pattern recognition and waveform classification tasks are carried out with excellent training and testing performance. Notably, α - In_2Se_3 Fe-FET device can classify images in MNIST database with accuracy up to 91%. Thanks to the historical capturing capability of dynamic transistors, most of the virtual node

number (by 96%) can be eliminated from training phase while still maintaining a high precision of over 80%. Meanwhile, the time-series prediction of a real-life chaotic system, daily temperature extracted from the Jena database, shows NRMSE of 8.3×10^{-3} and determination coefficient (R^2) of 0.9983. The remarkable prediction error is only one-fifth of the chaotic systems forecast based on memristor-based RC while the network size is ten times smaller.^[8b,9a]

2. Results and Discussion

Figure 1c illustrates the crystal structure of α -phase In_2Se_3 non-centrosymmetric rhombohedral, space group: $R3m$.^[14] 2D unit cell was composed of different atomic planes in the order of Se–In–Se–In–Se. To confirm the layered structure of α ($R3m$) In_2Se_3 , we performed high-resolution transmission electron microscopy (HRTEM) measurement as shown in Figure 1d, in which different atomic layers are linked by bondless electrostatic van der Waals force. Our Fe-FET device consisting of $\alpha\text{-In}_2\text{Se}_3/\text{Al}_2\text{O}_3/\text{W}$ heterostructure is clearly shown in TEM images and the corresponding elements are located accordingly in the energy-dispersive X-ray spectroscopy (EDS) map. The Raman spectroscopy is carried out using Renishaw inVia Raman equipped with 532 nm wavelength excitation laser to confirm α ($R3m$) phase of In_2Se_3 . The Raman scattering spectrum of an In_2Se_3 flake is shown in Figure 1e, exhibiting fingerprint-like peaks of typical vibration modes: E^1 (87.1 cm^{-1}), A_1^1 (102.6 cm^{-1}), E^3 (156.7 cm^{-1}), E^4 (179.8 cm^{-1}), A_1^2 (193.2 cm^{-1}), and A_1^4 (248.7 cm^{-1}), which are consistent with literatures as a signature for α ($R3m$) phase of In_2Se_3 .^[15]

We further characterized the pristine ferroelectricity in 2D $\alpha\text{-In}_2\text{Se}_3$ by the piezoresponse force microscopy (PFM) measurement.^[14,16] Figure 1f shows the topography image of an isolated $\alpha\text{-In}_2\text{Se}_3$ flake which was mechanically exfoliated from bulk crystal on bare p^{++} Si wafer. A thickness profile in the inset illustrates a significant difference in morphology ($\approx 20\text{--}50 \text{ nm}$), which can be attributed to the unintentional formation of various layers in In_2Se_3 during exfoliation. An AC voltage was applied to conductive material-coated atomic force microscope (AFM) tip with respect to sample as represented by measurement schematic in Figure 1i. The voltage established an external electric field (e-field) on the sample. Due to the ferroelectric nature that presents in the target material, the sample itself expands or shrinks according to the direction of e-field. The difference between e-fields of ferroelectric domain in $\alpha\text{-In}_2\text{Se}_3$ and the AC voltage given by the tip causes a linear mechanical deformation in the cantilever and increases the deflection, which is known as the PFM amplitude. Furthermore, the out-of-plane or in-plane polarization direction (parallel or antiparallel with external e-field) is also considered as a major factor contributing to the dephasing of AC voltage.^[16] Figure 1g,h shows the phase and amplitude contrasts of ferroelectric domains in mapping images, respectively, reflecting the sample dipole orientation. The typical hysteresis loops of phase versus applied voltage in Figure 1k imply a strong phase change of up to $\approx 180^\circ$ for a 30 nm $\alpha\text{-In}_2\text{Se}_3$ flake while the amplitude curve displays a butterfly-like shape, showing coercive voltages of $\pm 1 \text{ V}$ for identical ferroelectricity. Additional examination for the local ferroelectric

switching of $\alpha\text{-In}_2\text{Se}_3$ is carried out by areal writing and reading processes with different tip biases. Figure 1l depicts a significant contrast difference in PFM phase between inner $2 \times 2 \mu\text{m}$ square at -6 V and outer $4 \times 4 \mu\text{m}$ square at 6 V tip bias. It is noteworthy that the polarization direction of ferroelectric domain is controllably flipped by external e-field, which is also applicable for memory devices.

The $\alpha\text{-In}_2\text{Se}_3$ material used for realizing RC system in this work was vertically stacked on a series of transversal tungsten (W) bottom gate electrodes. A 20 nm Al_2O_3 was deposited as a dielectric layer by atomic layer deposition (ALD). Figure 2a shows energy band diagram of the device at equilibrium conditions for metal–insulator–semiconductor (MIS) capacitor, in which work function ϕ_M of tungsten and electron affinity χ of $\alpha\text{-In}_2\text{Se}_3$ are reported as about 4.5 and 3.7 eV, respectively.^[8a,14b,17] Detail device fabrication process and optical image are described in Experimental Section and Figure S1, Supporting Information. Figure 2b illustrates the transfer characteristics of transistors in various V_g ranges. All dual-swept $I_d\text{--}V_g$ curves show the natural n-type behavior of $\alpha\text{-In}_2\text{Se}_3$ semiconductor with an ON current density of $10 \mu\text{A } \mu\text{m}^{-1}$ and a switching ratio of over 10^6 . The clockwise hysteresis loops are attributed to the simultaneous existence of two types of charge in $\alpha\text{-In}_2\text{Se}_3$ channel, namely, mobile charge and bound charge.^[18] This unique charge distribution profile is responsible for the coexistence of semiconducting and ferroelectricity in a single-channel material. Flatband alignment under different back gate voltages is exhibited in Figure 2g,h. It is noted that the 20 nm Al_2O_3 dielectric layer used in this device has a dielectric constant (k -value) of 8.2 at room temperature.^[19] Therefore, we consider that the 2D $\alpha\text{-In}_2\text{Se}_3$ Fe-FET has a large equivalent oxide thickness (EOT) and thus the e-field induced by back gate is likely not adequate to entirely penetrate the channel, which leads to a local band bending at the semiconductor–dielectric interface. When a negative voltage below the coercive voltage is applied to back gate electrode, the polarization down switching occurs at the $\alpha\text{-In}_2\text{Se}_3/\text{Al}_2\text{O}_3$ interface, as shown in the cross-section schematic of Figure 2g. Accordingly, the semiconductor band bends downwards and causes mobile electrons in the conduction band to accumulate at the interface. This results in a low-resistance state (LRS) in the channel, which requires a greater negative voltage to switch the device off. In contrast, when a positive V_g (above coercive voltage of $\alpha\text{-In}_2\text{Se}_3$) is applied, the polarization domain switches up and $\alpha\text{-In}_2\text{Se}_3$ band bends upwards (Figure 2h). At this regime, majority carriers (electrons) are depleted from the interfaces, and the number of holes (minority carrier) in valance band is larger than that of the electrons, causing the *inversion* layer at interface. Therefore, the channel is switched to high-resistance state (HRS). The critical transport characteristics in $\alpha\text{-In}_2\text{Se}_3$ Fe-FET lead to a linear increase of memory windows as a function of V_g sweeping ranges in Figure 2c. The pulse number dependence of drain conductivity in Figure 2d,e can be explained by sequential depletion (accumulation) of majority carriers (electron) in the channel after being stimulated by positive (negative) V_g pulse, as depicted in Figure S2a,b, Supporting Information.

The transient $I\text{--}V$ measurement was performed to investigate the short-time retention in ferroelectric memory. A programming voltage pulse V_p was applied on W gate (amplitude of $-2\text{--}10 \text{ V}$,

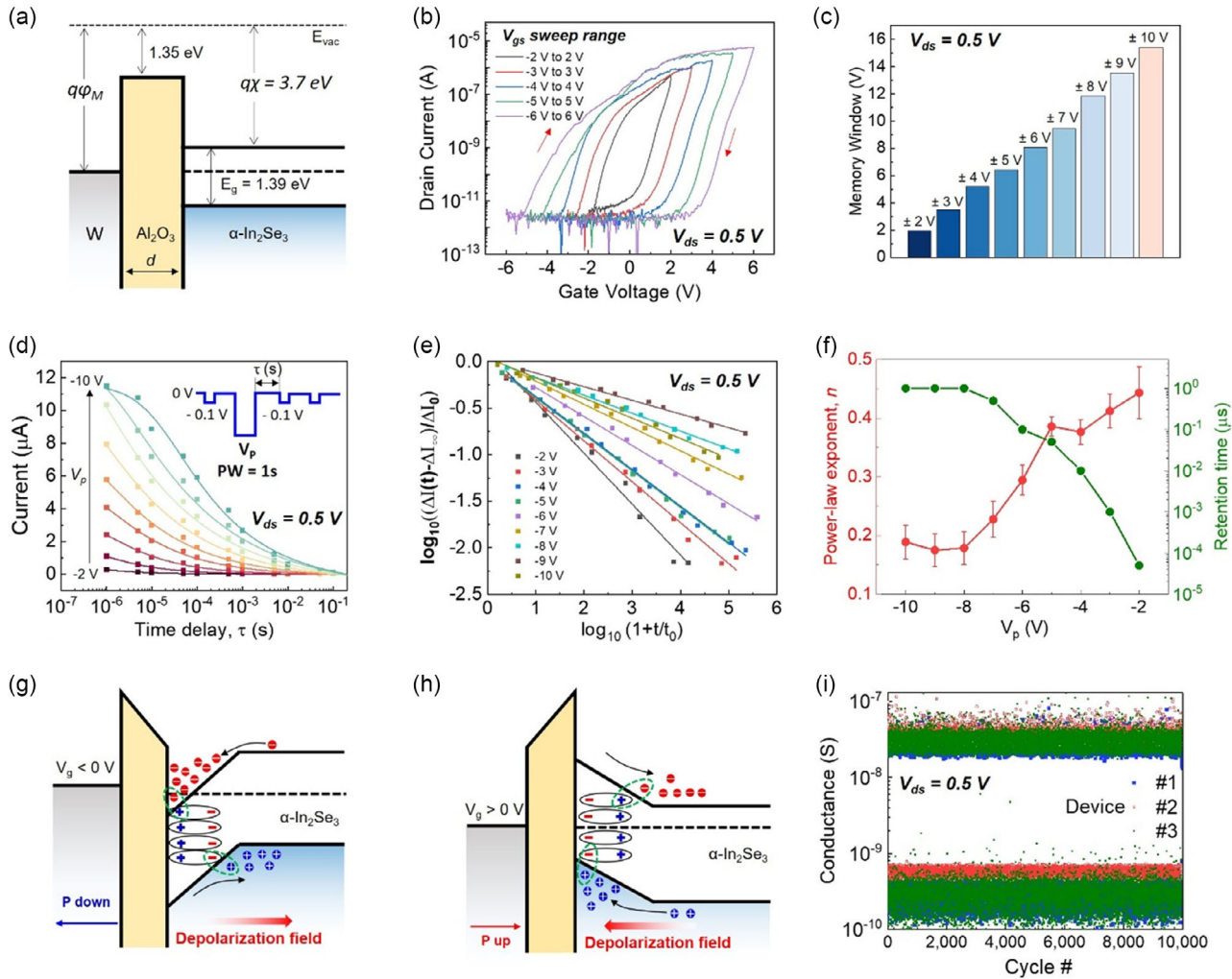


Figure 2. Electrical transport and polarization dynamic in α -In₂Se₃ field-effect transistor a) Energy band diagram of α -In₂Se₃/Al₂O₃/W ferroelectric channel field-effect transistor (Fe-FET) without bias. b) Transfer characteristics of α -In₂Se₃ Fe-FET, and c) extracted memory windows with different V_{gs} ranges. d) Retention loss of drain current indicated by read-after-write scheme with various $V_p = -2$ V to -10 V. The solid lines are fitting curves e) Retention data in log-log plot. f) Power-law factor n and retention time versus programming voltage (V_p). Band diagram of α -In₂Se₃ Fe-FET at two regimes: g) negative and h) positive V_g i) Endurance test of three different devices with 10^4 cycles.

and pulse width, PW of 1s), followed by several read voltage pulses (-0.1 V, 100 μ s) with certain time delays, τ . The measurement scheme is sketched in the inset of Figure 2d. A very fast current decay of α -In₂Se₃ channel is attributed to the polarization loss due to the presence of depolarization field.^[20] In asymmetry metal–insulator–ferroelectric–metal structure, there always exists incomplete charge compensation between two electrodes due to finite dielectric constant of insulator (Al₂O₃), which generates depolarization field. As shown in Figure 2g,h, the depolarization field tends to depolarize the dipole moment toward initial states. Thus, band bending at interface is then released, causing current drop with a very short retention time of <0.1 s. We found that the retention loss in ferroelectric channel could be well described by power-law decay function

$$\Delta P(t) = \Delta P_{\infty} + \Delta P_0 \left(1 + \frac{t}{t_0}\right)^{-n} \quad (1)$$

where polarization $\Delta P(t)$ is the difference between switched and nonswitched dipoles. ΔP_0 and ΔP_{∞} are retained values of $\Delta P(t)$ at $t=0$ and $t=\infty$, respectively, and time constant, t_0 , is a characteristic of relaxation time.^[21] Because the polarization P is proportional to surface charge density Q_p ($P = Q_p/A$, where A is surface area), the loss in polarization linearly results in current decay in channel. Figure 2d shows the experimented data of drain current, I_D , versus time delay, τ (s), and their corresponding fitting curves (solid lines) after programmed by $V_p =$ from -2 V to -10 V (step -1 V). The variation of I_D with time delay is well matched to the polarization dynamic model in Equation (1), which proves that our resistance switching mechanism is predominantly dominated by ferroelectricity in α -In₂Se₃. The Al₂O₃ oxide trap-based memory commonly showing nonvolatile retention up to 10^4 s is not considered for RC applications due to the absence of fading effect.^[22,23] We further present Equation (1) in dual-logarithmic plot of $(\Delta I(t) - \Delta I_{\infty})/\Delta I_0$ and $1 + t/t_0$ in

Figure 2e with various V_p . When V_p amplitudes increase, a gradual decrease of the power-law factor, n extracted by linear fitting slope is indicated in Figure 2f. The variation of n is clarified for the interactions of mobile charges and bound charges of α - In_2Se_3 as sketched in Figure 2g,h. The electrostatic force of attraction by Coulomb's law of mobile and bound charges that are opposed to depolarization field tends to maintain dipole moments. A larger V_p amplitude provides a higher number of electrons accumulated in conduction band, resulting in a longer retention time. Figure S3, Supporting Information shows that fading conductance effect in α - In_2Se_3 Fe-FET is observed in both negative and positive V_p , which facilitates the implementation of highly efficient RC. This feature of ferroelectric semiconductor materials is distinguished from other polarization dynamics in dielectric material such as $\text{Hf}_{0.5}\text{Zr}_{0.5}\text{O}$ (HZO) and $\text{PbZr}_{0.4}\text{Ti}_{0.6}\text{O}_3$ (PZT) where n is constant due to nonexistence of mobile charges.^[21,24] The endurance of volatile memory has been examined under 10^4 programming (-5 V , 100 ms)/erasing (5 V , 100 ms) cycles, shown in Figure 2i. The nondegradation of HRS and LRS of three devices reliably demonstrates that the devices are effectively switched with negligible cycle-to-cycle and device-to-device variability. Furthermore, we assessed the resistive switching (RS) ability of our fading ferroelectric transistor with increasing programming/erasing speed (Figure S13a, Supporting Information). It is noteworthy that a very short pulse width of $1\text{ }\mu\text{s}$ with a minimum energy of 0.78 nJ is required to give an effective switching ratio of up to 10^2 , indicating the ultrafast switching in pristine ferroelectric materials. Meanwhile, lower energy consumption is needed to read as depicted in Figure S13b, Supporting Information. Here, HRS switching energy is calculated to be about 10 fJ for a read pulse of 0.5 V , $20\text{ }\mu\text{s}$, which remains unchanged when the writing width is increased. In contrary, to probe the conductance states in LRS, the transistor expenses a higher energy of 1 pJ corresponding to a writing spike of $1\text{ }\mu\text{s}$, which may increase with larger writing widths. In general, the energy expenditure in our dynamic Ferro-RC is as low as typical excitatory/inhibitory energy consumption of synaptic devices based on α - In_2Se_3 .^[18]

2.1. Pattern Recognition

The electrical stimulation is conducted by applying discrete voltage pulses (3 V , $100\text{ }\mu\text{s}$) stream to back gate, followed by small read pulse (1 V , $500\text{ }\mu\text{s}$). Figure 3a depicts the write pulse pattern (upper panel) and response current (lower panel). If multiple positive V_g pulses are applied with short intervals, electrons are monotonically eliminated which results in a gradual decrease in channel conductance (black arrows). However, if a long discrete write pulse is introduced in between, device's conductance falls toward initial state (green arrows). This temporal dynamic is attributed to the very fast retention loss of ferroelectric domain as discussed in previous section. Conversely, when a negative V_g pulse stream was applied, electron in α - In_2Se_3 is enriched with respect to increasing pulse number, resulting in a rise of drain current. Figure 3b shows the long accumulation of response currents with various pulse amplitudes (from -2 V to -5 V). The interval-dependent increase of conductivity following a prior pulse demonstrates paired-pulse facilitation (PPF) (Figure 3c).

In this section, we performed pattern recognition task to validate the RC capability of α - In_2Se_3 Fe-FET. The pattern recognition starts with simple pixelized image of digit "9," as shown in Figure 3d, in which the gray and black pixels correspond to bit '0' and '1', respectively. The pulse amplitudes of 0.5 V and 2 V ($\text{PW} = 0.5\text{ ms}$ and $\text{interval} = 0.1\text{ ms}$) were used to linearly convert each row of pattern to input gate voltage stream (second panel). A final read pulse (0.5 V , 0.5 ms) and $V_{\text{ds}} = 0.1\text{ V}$ were applied to sense the eventual device conductance states. Due to the fading memory effect of In_2Se_3 Fe-FET, the measured read currents contain history pulsing process and can be used to map information from the input pattern. These conductance values are fed to the reservoir and serve as an individual reservoir node which is linked to other nodes in different rows to form a high-dimensional input vector X_i for software training. The algorithm for training is a simple logistic regression with sigmoid function as follows

$$Y_i = \frac{1}{1 + e^{-(W_{i,0} + W_{i,1}X_{i,1} + \dots + W_{i,n}X_{i,n})}} \quad (2)$$

where X_i is reservoir states, Y_i is output digit labels, and W_i is post-training weight. The training and testing manners are performed by Python environment to find out appropriate post-training weights (W_i) (Figure 3d). Figure 3e shows reservoir states which are reflected by the read currents of transistor after subjected to five input $[5 \times 5]$ patterns corresponding to "9" digits. As each gate electrode partially covers a certain area of In_2Se_3 flake, we are able to separate the 2D In_2Se_3 flake to different in-series transistor cells depending on the number of gate electrodes underneath. The amplitude of conductance in each row depends on the given pulse stream which makes the reservoir states not identical and distinguished from each other. We constructed a dataset of 50 patterns representing different digits, in which each digit is presented in five different patterns. The whole dataset is then converted to temporal voltage and applied to 1D Fe-FET array, where the drain current density is fed to reservoir for training and testing phases. The dataset and its corresponding reservoir states are illustrated in Figure S4, Supporting Information. To figure out the training weights W_i matrix in Equation (2), reservoir states measured from the 30 images in training dataset (Figure S4a, Supporting Information) are subjected to readout software as input vectors X_i , whereas the output layer Y_i is the consistent digit label. An accuracy of 100% is obtained after five training iterations as shown in Figure 3f, which implies that the Ferro-RC system is capable of isolating all 30 original patterns precisely. After post-training weights W_i matrix is formed, the input vectors X_i are experimentally measured from the rest of 20 patterns in the test dataset and then sent to reservoir for model testing. We noted that, in test patterns, the noise was intentionally created by adding or removing several pixels to distort the pattern as compared to training patterns (see Figure S4b, Supporting Information). Apparently, the chronological information processing of Ferro-RC is revealed with testing accuracy, reaching a maximum of 85% after five iterations (Figure 3f). The number of predicted images is shown in the confusion matrix in Figure 3g. Although interfered by noise, the test images are still classified by models with high accuracy. Three observations were incorrectly predicted over 20 test

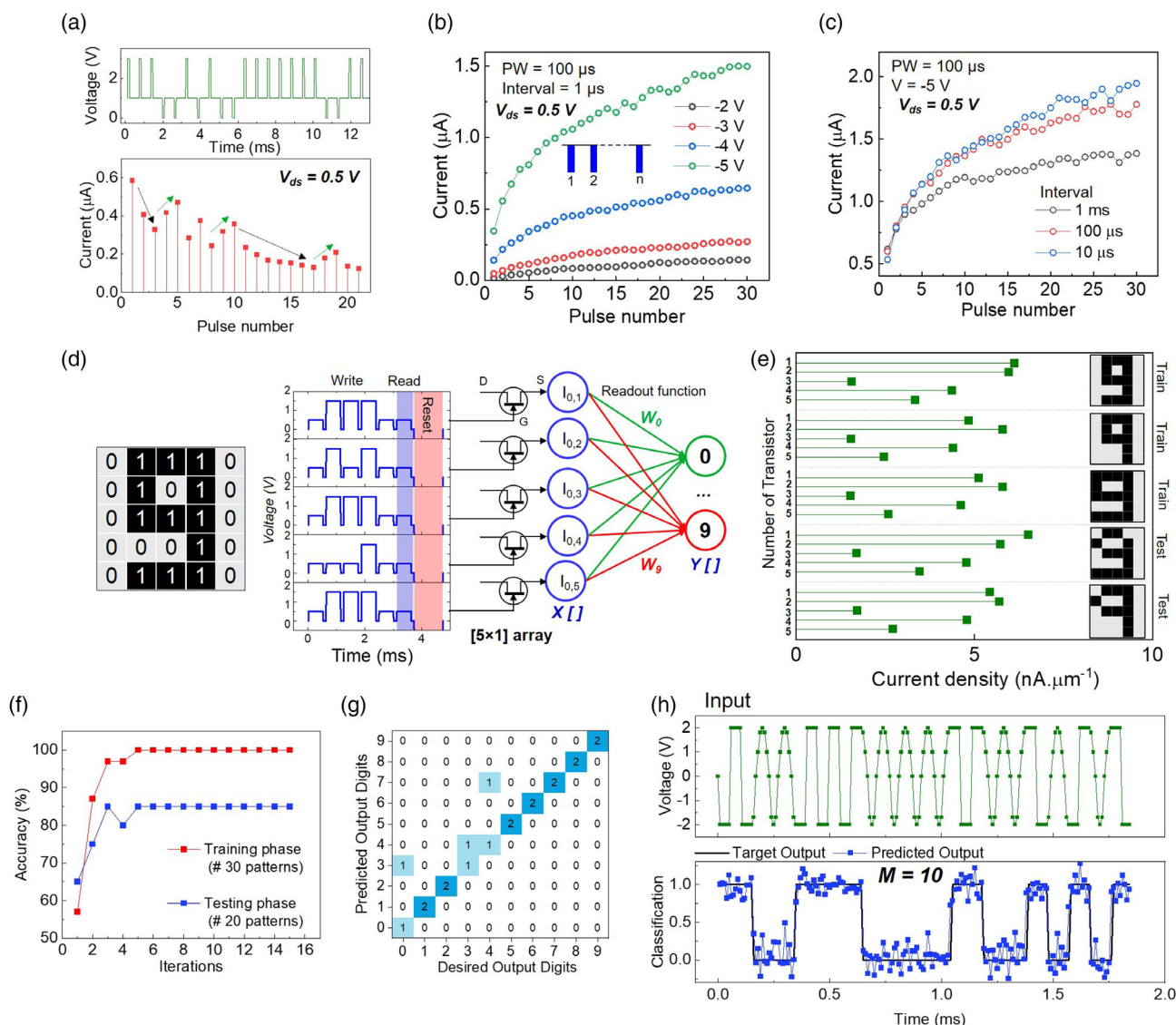


Figure 3. Temporal dynamic of fading α -In₂Se₃ memory and applications. a) Response current of dynamic α -In₂Se₃ Fe-FET on discrete V_g pulse stream b) Gradual increase of conductance with number of train pulses showing paired-pulse facilitation (PPF). c) Variation of PPF with different stimulation frequencies. d) A 5×5 input pattern representing digit “9” with binarized ‘0’ and ‘1’ bit corresponds to gray and black pixel, respectively. Each analog input voltage rows converted from pattern is experimentally applied to back gate of α -In₂Se₃ Fe-FET in the 1D array. The response currents sensed by transistor after each pulse stream are fed to reservoir for training and testing. Output layer contains 10 digits, which are labelled from 0 to 9. e) Read currents of 5 patterns corresponding to digit ‘9’. f) Variation of accuracy after iterations of train and test process. g) Interfered confusion matrices between desired output digits and predicted output digits of train and test phase. h) Waveform classification with effective cyclic reservoir systems.

images, which can be attributed to the challenge in verifying conductance differences caused by the variation of flake size and thickness between each device in 1D array. The details are described in Figure S4b, Supporting Information.

We further verified the possibility of our fading memory for cyclic reservoir concept by executing waveform classification task. The spatial or temporal input value at any time is multiplied with a binary matrix comprising only -1 and 1 elements, resulting in a time-dependent input stream in Figure S5a, Supporting Information. In this time-multiplexing process, the number of intervals in the delay loop, known as the mask length (M), is

determined by the total number of elements in the binary matrix (refer to Figure S5b, Supporting Information). The input voltage stream is then applied to our α -In₂Se₃ Fe-FET to record M reservoir nodes which are sequentially connected in a delay loop with delay time, τ . The reservoir nodes are used for training readout function by simple linear regression. Figure 3h (upper panel) shows input wave pattern (U_i) designed by random arrangement of different sine and square waveforms. The classification results show a significant improvement with increasing mask length, that is, $M = 5, 8$ (Figure S5c, Supporting Information), and 10 (Figure 3h, lower panel).

2.2. Hand-Written Digit Classification

To validate the performance of α -In₂Se₃ Ferro-RC in processing spatiotemporal information, we carried out a benchmark hand-written digits classification using 1,000 images from MNIST database as the input (see Experimental Section). Each image represents a hand-written digit pixelized to a 2D [28 × 28] matrix 8-bit grayscale that varies from 0 to 255, as shown in Figure 4a. To minimize the number of inputs, chronological data, U_i , in 2D array is preprocessed by multiplying with a [28 × 10] mask to generate temporal input voltage stream, $V(t)$, with amplitude that is applicable to gate dielectric endurance voltage of α -In₂Se₃ Fe-FET. The input voltage $V(t)$ matrix is then flattened and transformed to a time-series information with time delay τ of 120 μ s, as shown in Figure S6, Supporting Information. A single α -In₂Se₃ transistor is used to sense the response current from

250 input voltage spikes to record 250 response currents as sequential virtual nodes; see second panel in Figure 4a). These 250 virtual nodes are then served as input X_i to regress output Y_i . SoftMax regression is used as an algorithm for readout function. Four different masks simulating four-parallel reservoirs are applied to four different transistors in 1D array (Figure S6, Supporting Information). Similar to the abovementioned simple pattern recognition task, the hand-written digits classification includes two processes: training and testing. Out of the 1000 original images, 80% are unintentionally selected for training which are fed to reservoir using SoftMax regression model to determine the training weight-out (W_{out}) of readout function. Details of training and testing methods are indicated in the Experimental Section. Thereafter, the rest of dataset (200 images) is substituted to model for testing. We noted that in the entire classification procedure, the testing images are not

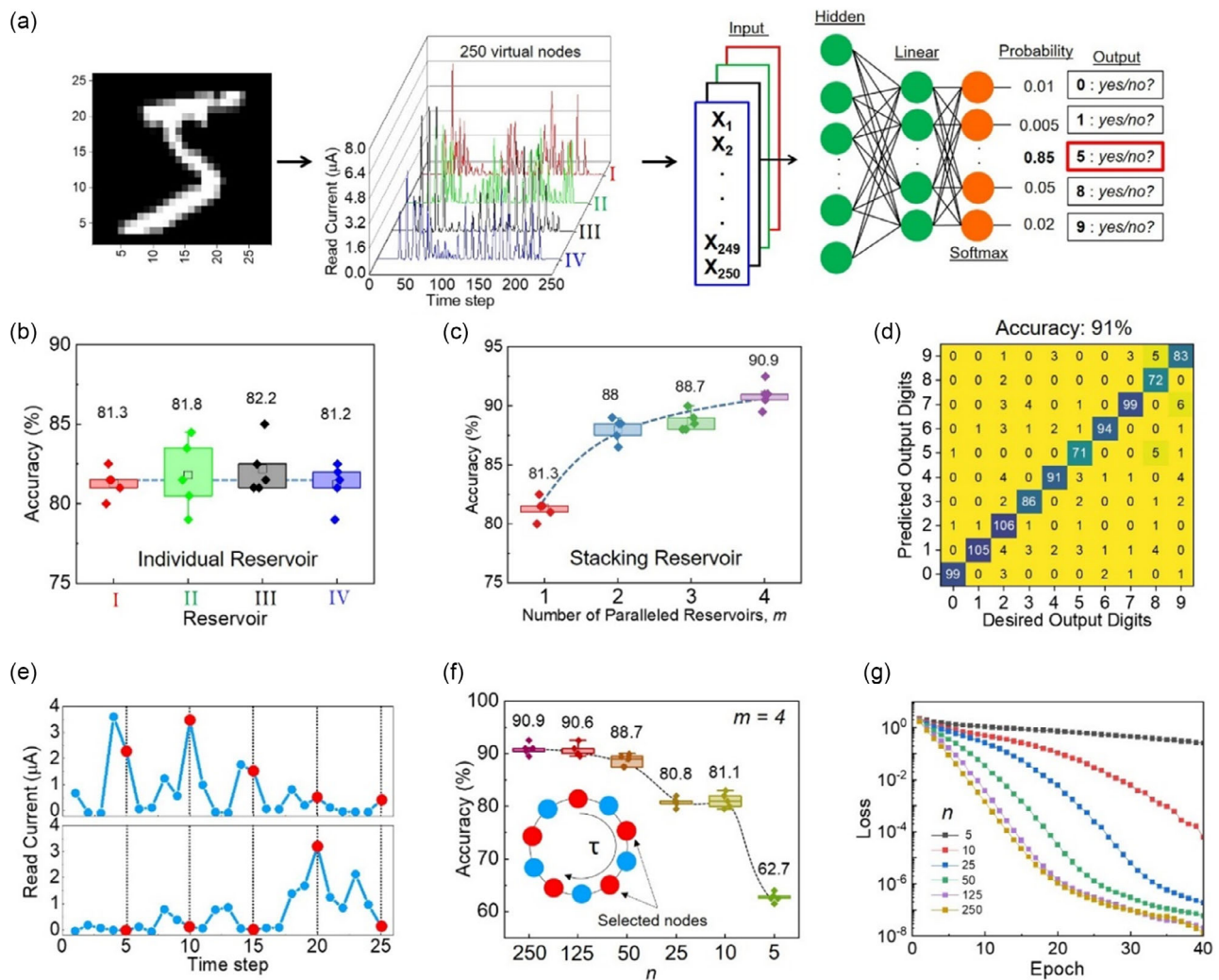


Figure 4. Hand-written digits classification. a) An example of hand-written digit labeled '5' image extracted from MNIST database. The read currents from these 250 input voltages are fed to reservoir to serve as virtual nodes with sequential dependence among nodes. b) 5-time repeated classification accuracy of test images with different masks. c) Dependence of accuracy as a function of m -parallel reservoirs. d) Confusion matrix of test images showing accuracy $\approx 91\%$ with 4 parallel reservoirs ($m=4$). e) An example of response currents of an original image with selected training virtual nodes with $n=50$. f) Accuracy and g) Training loss obtained from Ferro-RC with various n intervals.

used in the training course. Figure 4b shows the accuracy of testing processes for different masks. The mean precision values for each mask are extracted over five times of training and testing repeatedly. It is obvious that the classification accuracy does not show a significant variation among masks with an average value of about 82% in general. The interfered confusion matrices of each reservoir are presented in Figure S7, Supporting Information, which is attributed to the fundamental limitation in sensing capability of each single α -In₂Se₃ Fe-FET. Due to the minimization of input from 784 original images to 250 after mask process, the loss of information causes unsatisfactory classification results. Virtual nodes of m -paralleled reservoirs are added up to one training and testing process. The total number of virtual nodes in the training layer linearly increases to $250 \times m$ for one observation. As a result, a drastic rise of accuracy is observed in Figure 4c with increasing m . When a total of 1000 virtual nodes in four-paralleled reservoirs are supplied to readout layer, the accuracy can be optimized to 91% as shown in confusion matrix (Figure 4d). Using standard gradient-based optimization method, RMSprop, with a learning rate of 1×10^{-4} to lessen the entropy loss, the training processes show 100% accuracy after five epochs. The variation of train accuracy and cross entropy loss versus epochs are described in Figure S8, Supporting Information. Moreover, according to the accuracy and loss curves, increasing m leads to a faster training process. A matching confusion matrix is recorded over the testing process after the readout function is trained with 800 images. Overall, the training and testing are repeated five times with different observations selected randomly.

To verify the historical capturing ability of our Ferro-RC, a series of filters were applied to selectively pick up nodes in certain intervals among 250 virtual nodes. This process effectively reduces the size of reservoir layer to perform training and testing. We divide the virtual nodes corresponding to 250 time steps into n equal intervals, and the read currents after each interval are then selected and fed to the reservoir. An example of $n = 50$ is shown in Figure 4e, where 250 virtual nodes are divided into 10 columns with 25 time steps each. We plotted read current of column #5 and #8 as examples. The selected nodes are highlighted after every five steps, five which give a 5 time reduction in the total number of reservoir states fed to readout function. The diagram showing the selected nodes in feedback loop is sketched in the inset of Figure 4f. A slight decrease (2.2%) in accuracy was observed when 80% of the inputs are eliminated from the input layer at $n = 50$, implying a perfect temporal signal sensitivity of our RC system. Moreover, the dependence of training loss with different number of intervals is summarized in Figure 4f. The values of n in this test are divisible numbers of 250. We found that the high-precision recognitions can be maintained over 80% accuracy; while suffering just 4% of virtual nodes were used to readout process ($n = 10$). However, a significant reduction of accuracy to 62.7% is observed at $n = 5$, setting a lower limit on the minimum number of intervals required to classify and memorize images accurately. Confusion matrices of $n = 10$ and $n = 5$ are shown in Figure S9a,b, Supporting Information, respectively. When the intervals are reduced to $n = 5$, it implies that only 2% of the total virtual nodes in four-paralleled reservoirs are selected for training the readout function. As a result, the model becomes inefficient in figuring out the appropriate

W_{out} matrix, as reflected by the gradual decrease in accuracy curves and higher training loss in Figure S9c, Supporting Information and Figure 4g, respectively.

2.3. Time-Series Prediction of Daily Temperature

In addition to examining temporal signal processing ability of α -In₂Se₃ Ferro-RC, we carried out a benchmark task of time-series prediction using real-life daily temperature as extracted from Jena climate database. The input of this prediction is outdoor ambient temperature at time t , $T(t)$, recorded in each hour which is preprocessed with binary mask ($M = 10$) containing -1 and 1 elements. The aim of this task is to predict the actual temperature in the successive step, $t + 1$. The amplitude of the input matrix is then linearly mapped to voltage so that the voltage scale is equivalent to endurance electric field of our transistor. All ten preprocessed input voltages are equally divided into a feedback loop with a duration of $\tau = 120 \mu\text{s}$ and then applied to transistor back gate to record ten conductance states in the channel. A visualization of inherent correlation among $T(t)$, preprocessed V_g stream, and response current of transistor are presented in Figure S10, Supporting Information. The conductance values are fed to reservoir for subsequent computing with output layer set as a series of $T(t + 1)$. It is worthy to note that the single 1D mask with a mask length of 10 is used for this estimator, and the readout layer is a 10×1 network. Here, linear regression is utilized for the training algorithm (see Experimental Section). We use temperature data in degree Celsius for a period of 2000 h (from 21.07.2009, 04:00:00 to 12.10.2009, 12:10:00) in this time-series prediction task, in which 1000 data points are prepared as an input layer for training. The training process is carried out to figure out the weights and bias. After the readout function is found, the rest of 1000 temperature data are directed to the model for testing. Excellent performance was achieved in time-series prediction task for both training and testing, as shown in Figure 5a. The coefficient of determination (R^2) shows near-ideal values of 0.9901 and 0.9866 for training and testing process, respectively, which exclude the possibility of over fitting in this model. Besides, the Ferro-RC system attains small prediction errors which are reflected by a minuscule NRMSE of 0.0193 and 0.0295 for training and testing, respectively.

We further investigate the dependence of R^2 and NRMSE on increasing number of parallel reservoirs, m . In a similar fashion to the previous task, nine different random binary matrices are generated more to form ten-paralleled reservoirs. The virtual nodes measured by 9 transistors in 1D array (Figure S1, Supporting Information) are stacked to enrich the reservoir neurons in Ferro-RC systems for training and testing process over the daily temperatures of 2000 h. Diagram is sketched in Figure 5b. Herein, the training model is a $10 m \times 1$ network. The larger number of parallel reservoirs strengthens the prediction accuracy. As shown in Figure 5c, when m increases, the R^2 of training and testing process shows a drastic increase to 0.9986 and 0.9983, respectively, after all virtual nodes of ten reservoirs are simultaneously applied to the readout function. Furthermore, the NRMSE of training and testing achieves extremely low values of 7.17×10^{-3} and 8.3×10^{-3} , respectively, corroborating to excellent prediction. Figure S11, Supporting Information plots

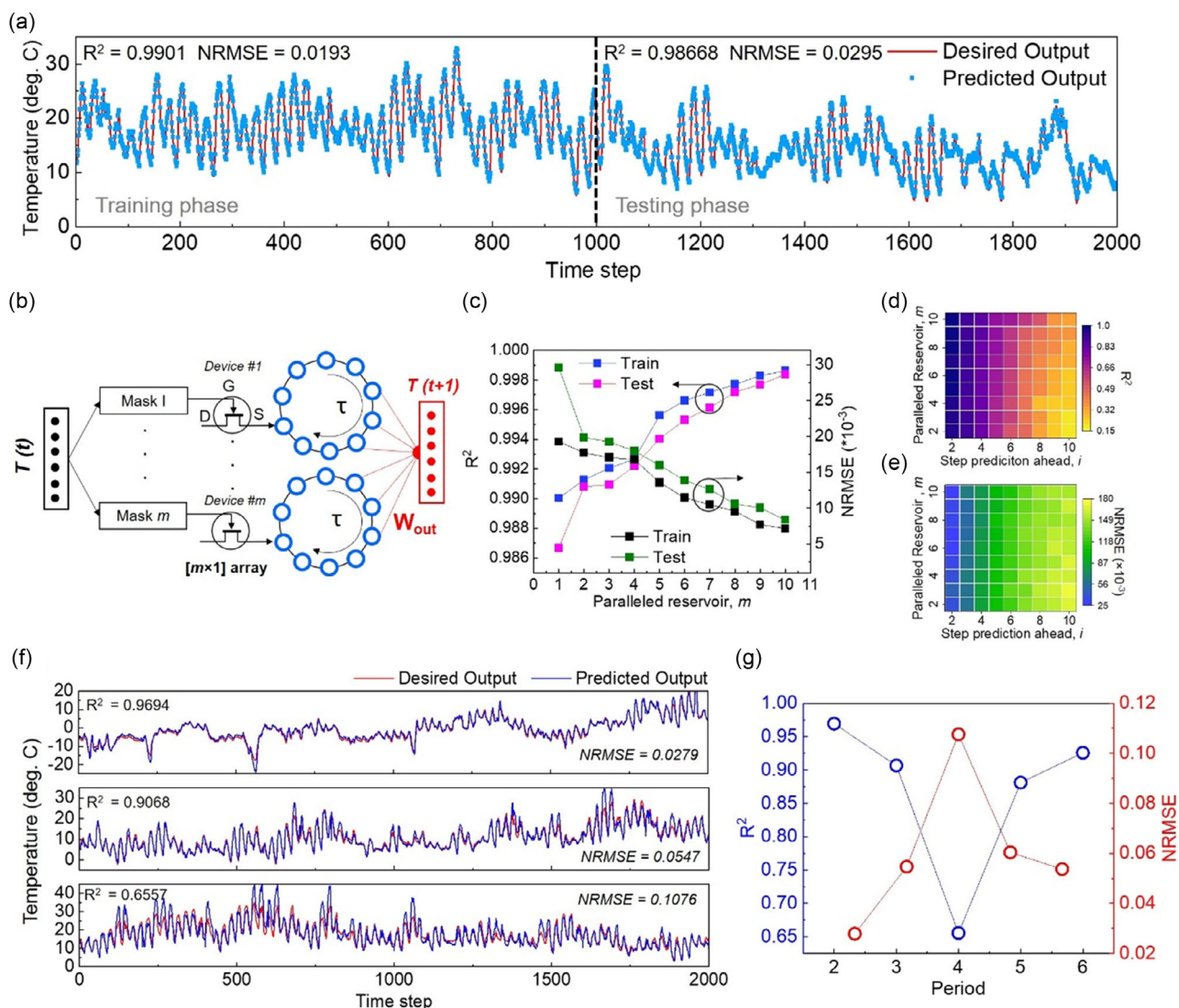


Figure 5. Time-series prediction of daily temperature in Jena climate database. a) Desired and predicted data of daily temperature recorded in a period of 2000 h experimentally obtained by α -In₂Se₃ Fe-FET based RC system. b) Diagram of multiple reservoirs architecture c) Coefficient of determination, R^2 and normalized root mean square error, NRMSE of training and testing phases varying as a function of number of paralleled reservoirs, m . d, e) Contour maps of R^2 and NRMSE, respectively, varying with m -paralleled reservoir and i -step ahead prediction. f) A long-term prediction with different periods corresponding to winter, spring and summer. g) Summarized of R^2 and NRMSE of testing process with 2000 time-steps in each period after training first period.

the predicted output from experimental test data obtained from the transistor-based RC systems with $m = 2$ and 10. For mapping numerous input data into reservoir layer, physical implementations are mostly based on two-terminal RS memristors that are reliant on metal oxide monocular in vertical orientation. In RS memristors, the external E-field controls the diffusion of ions to form conductive filaments, leading to modification of device conductance. RS memristors suffer fundamentally from various degrees of difficulty, for example, slow field-driven switching, heterogeneous performance, and low endurance may cause high prediction error (NRMSE) in RC execution. Basically, the Ferro-RC system with switchable dipole in α -In₂Se₃ activates long fading effect in volatile memory, in which the programming pulses

are applied to back gate to electrostatically adjust the polarization status in channel to sequentially modulate the conductance. Moreover, the ferroelectricity of α -In₂Se₃ can respond to ultrafast write speed up to 40 ns,^[22] and nondestructive readout, which allows the construction of high-neural-density reservoirs with low-energy consumption. We noted that our NRMSE, as quantified by the difference between target and predicted output, are the lowest values experimentally recorded in a time-series prediction of real-life chaos so far. Notably, NRMSE values obtained from the Ferro-RC are much lower than that in reported chaotic system predicted by memristor-based RC such as Mackey–Glass system^[7a,9a] and Henon map.^[8b] Furthermore, the α -In₂Se₃ FET-based Ferro-RC is employed to perform an increasingly intensive

job of predicting several steps ahead, in which the output layer is temperature T at i time steps ahead, $T(t+i)$, where $i > 1$. The contour maps presented in Figure 5d,e show the variation of R^2 and NRMSE with respect to i and number of paralleled reservoirs, m . As i increases, it becomes increasingly challenging for the system to forecast, thus resulting in a monotonic decrease of R^2 and higher NRMSE. To avoid compromising the prediction accuracy in a longer time step, multiple paralleled reservoirs with increasing number of virtual nodes are investigated. As a result, R^2 and NRMSE can be optimized to 0.3234 and 0.15, respectively, at $i = 10$ with $m = 10$. The effect of increasing m is clarified and compared in the experimentally trained data and its ground truth at $m = 2$ and 10 for $i = 10$ is shown in Figure S11b, Supporting Information.

2.3.1. Long-Term Prediction

We selected a longer period of 12 000 h, which are divided into 6 different periods with 2000 h each. The first 2000 h are used to train the readout function with $i = 1$, $m = 1$ ($M = 10$). The period is matched with autumn season in Europe, where the ambient temperature varies from -5 to 20°C . The next 5 periods that correspond to winter, spring, summer, autumn, and winter are exploited for testing. Figure 5f shows the comparison of predicted output and desired output for three distinctive seasons, that is, winter, spring, and summer. We show that the model can achieve high prediction accuracy in winter and spring with $R^2 = 0.9694$ and 0.9068 and low NRMSE of 0.0279 and 0.0547 , respectively, as the temperature ranges in these two seasons are comparable to that in trained autumn segment. However, the model suffers a compromised accuracy in summer with $R^2 \approx 0.65$ and $\text{NRMSE} \approx 0.15$ as the temperature is out of the training range. Figure 5g summarizes the variation of R^2 and NRMSE for five different continuous periods (2000 h in each) of testing process based on the training of first period. We observed a significant increase of the prediction accuracy in period 5 and 6, as the weather in these periods is in fall and winter again, leading to a reduction in daily temperature difference, which matches the amplitude of input in training database. The predicted and desired output for periods 5 and 6 are plotted in Figure S12, Supporting Information.

3. Conclusion

In conclusion, RC systems based on $\alpha\text{-In}_2\text{Se}_3$ ferroelectric memory are demonstrated. The ferroelectric properties of 2D-layered $\alpha\text{-In}_2\text{Se}_3$ are exploited to realize volatile memory with a short retention. In addition, the Ferro-RC can perform various computational tasks from simple pattern recognition and waveform classification with high sorting rate to hand-written digits classification with 91% accuracy. The historical capturing ability of our fading memory allows us to reduce the network size by up to 96% while preserving a satisfactory accuracy of over 80%. Furthermore, the Ferro-RC system can be applied to forecast real-time chaotic systems based on Jena climate time-series dataset with an ultralow NRMSE of 0.0083 using 100×1 networks, which is one-fifth of that in traditional memristor-based RC.^[7a,8b] Using 2000 h in autumn season as the training dataset, we

revealed a long-term prediction for the next 10 000 h across different seasons in a sequential order of autumn–winter–spring–summer–autumn–winter without the need for retraining. This work paves a route for the implementation of RC systems based on ferroelectric channel transistor to perform general machine learning tasks for neuromorphic computing applications.

4. Experimental Section

Device Fabrication: A 30 nm tungsten (W) was deposited by metal and dielectric sputtering, AJA system, on p^{++} Si/SiO₂ (285 nm) substrate. The patterning process was carried out by laser writer to partially protect the area of back gate electrode. The chip was then immersed in W etchant for 25 s to remove noncovered W. The back gate electrodes were rinsed by organic solvents and covered by 20 nm Al₂O₃ by 200 cycles of ALD (Picosun ALD) at 150°C for dielectric layer. After that, an exfoliated $\alpha\text{-In}_2\text{Se}_3$ flake from bulk crystal was accurately transferred on back gate electrodes. Step-by-step fabrication methods are visualized in Figure S1, Supporting Information. Drain/source electrodes were patterned by ultra-high-performance electron-beam lithography (EBL) Raith EBPG5200. Nickel (Ni) contacts were deposited by ultra-high-vacuum E-beam evaporator, AJA system, at 10^{-8} Torr.

Material Characterization: Single-crystalline $\alpha\text{-In}_2\text{Se}_3$ material was characterized by HRTEM and EDS. $\alpha\text{-(R3m)}$ rhombohedral phase was confirmed by Renishaw inVia Raman Scattering Spectroscopy dual laser: 532 and 325 nm. Piezoresponse and surface morphology measurements were performed by AFM Park System NX20 with built-in PFM function.

Datasets: The Mixed National Institute of Standards and Technology (MNIST) database was widely used for training and testing in machine learning research. The database contained training samples of 60 000 images and testing samples of 10 000 images (sized 28×28 pixels) representing digits written by high-school students and employees of the United States Census Bureau.

Jena Climate database was recorded by the Max Planck Institute for Biogeochemistry in Jena, Germany. The dataset consisted of 14 features such as temperature, pressure, humidity and etc... recorded once per 10 min for a duration of 8 years: Jan 10, 2009 – December 31, 2016. The dataset used in Figure 5a was the average temperatures (Celsius degree) recorded in a period of 2000 h from 21.07.2009 (04:00:00) to 12.10.2009 (12:10:00). For long-term prediction task, a dataset of 12 000 h from 12.10.2009 (12:00:00) to 23.02.2011 (11:50:00) was divided into six equal periods corresponding to six seasons in sequential order of fall–winter–spring–summer–fall–winter.

Measurement Setup and Readout Function Training: The electrical characterization was conducted by Keysight B5100A Semiconductor Device Analyzer in ambient condition. The transient I – V measurement setup is supported by in-series connection with B1531A RSU. For pattern recognition in Figure 3, a supervised learning algorithm, logistic regression, was used to train the readout function. Furthermore, the RC training algorithm in the time-series prediction task (daily temperatures in Figure 5, waveform classification in Figure 3h) was carried out by linear regression. These training regressions were called from Scikit-learn package in Python program. The root means square errors were calculated by built-in sklearn.metrics () tool.

MNIST Hand-Written Digits Classification: Virtual nodes from m -paralleled reservoirs were sent to readout layer with 10 outputs representing 10 different digits from 0 to 9. The readout network size was $250m \times 10$. Keras toolkit in Python which provides a high-performance programming by accessing TensorFlow was used to train the readout function. A supervised learning algorithm, softmax regression, was exploited as an activation function to calculate the probability corresponding to the different outputs. A standard gradient-based optimization method, RMSprop, with learning rate of 1×10^{-4} , was used to minimize the cost function and train the output network. The training loss and accuracy were defined by sparse categorical cross entropy and accuracy, respectively.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This research was supported by A*STAR Science and Engineering Research Council (A2083c0061), National Research Foundation, Singapore (NRF-CRP24-2020-0002), and Ministry of Education Tier-1 and Tier-2 Grant (MOE-T2EP50120-0016).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

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

Keywords

artificial neural networks, fading effect, ferroelectric field-effect transistors, in-memory computing reservoir computing, spatiotemporal information processing

Received: January 6, 2023

Revised: January 24, 2023

Published online: February 19, 2023

- [1] a) Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, 521, 436; b) V. K. Sangwan, M. C. Hersam, *Nat. Nanotechnol.* **2020**, 15, 517; c) A. Taherkhani, A. Belatreche, Y. Li, G. Cosma, L. P. Maguire, T. M. McGinnity, *Neural Networks* **2020**, 122, 253; d) S. Li, M. E. Pam, Y. Li, L. Chen, Y. C. Chien, X. Fong, D. Chi, K. W. Ang, *Adv. Mater.* **2021**, 34, 2103376; e) Z. Zhang, S. Wang, C. Liu, R. Xie, W. Hu, P. Zhou, *Nat. Nanotechnol.* **2022**, 17, 27.
- [2] a) C. Li, Z. Wang, M. Rao, D. Belkin, W. Song, H. Jiang, P. Yan, Y. Li, P. Lin, M. Hu, *Nat. Mach. Intell.* **2019**, 1, 49; b) L. Sun, Z. Wang, J. Jiang, Y. Kim, B. Joo, S. Zheng, S. Lee, W. J. Yu, B.-S. Kong, H. Yang, *Sci. Adv.* **2021**, 7, abg1455.
- [3] Y. Wen, W. Yu, *Electron. Lett.* **2019**, 55, 515.
- [4] a) Z. C. Lipton, J. Berkowitz, C. Elkan (Preprint) arXiv:1506.00019v4, v4, submitted: Oct. **2015**; b) Y. Yu, X. Si, C. Hu, J. Zhang, *Neural Comput.* **2019**, 31, 1235; c) A. Sherstinsky, *Phys. D Nonlinear Phenom.* **2020**, 404, 132306.
- [5] S. Hochreiter, J. Schmidhuber, *Neural Comput.* **1997**, 9, 1735.
- [6] a) M. Lukoševičius, H. Jaeger, *Comput. Sci. Rev.* **2009**, 3, 127; b) G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, A. Hirose, *Neural Networks* **2019**, 115, 100.
- [7] a) X. Liang, Y. Zhong, J. Tang, Z. Liu, P. Yao, K. Sun, Q. Zhang, B. Gao, H. Heidari, H. Qian, *Nat. Commun.* **2022**, 13, 1549; b) D. J. Gauthier, E. Bollt, A. Griffith, W. A. Barbosa, *Nat. Commun.* **2021**, 12, 5564; c) C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, W. D. Lu, *Nat. Commun.* **2017**, 8, 2204.
- [8] a) R. Jacobs, D. Morgan, J. Booske, *APL Mater.* **2017**, 5, 116105; b) Y. Zhong, J. Tang, X. Li, B. Gao, H. Qian, H. Wu, *Nat. Commun.* **2021**, 12, 408.
- [9] a) J. Moon, W. Ma, J. H. Shin, F. Cai, C. Du, S. H. Lee, W. D. Lu, *Nat. Electron.* **2019**, 2, 480; b) M. Nakajima, K. Tanaka, T. Hashimoto, *Commun. Phys.* **2021**, 4, 20.
- [10] M. Riou, F. A. Araujo, J. Torrejon, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, in *presented at 2017 IEEE Int. Electron Devices Meeting (IEDM)*, San Francisco, CA, December **2017**.
- [11] K. Liu, B. Dang, T. Zhang, Z. Yang, L. Bao, L. Xu, C. Cheng, R. Huang, Y. Yang, *Adv. Mater.* **2022**, 34, 2108826.
- [12] M. Mariani, H. Wang, T. Yamamoto, M. Neeley, R. C. Bialczak, Y. Chen, M. Lenander, E. Lucero, A. D. O'Connell, D. Sank, *Science* **2011**, 334, 61.
- [13] L. Appeltant, G. Van der Sande, J. Danckaert, I. Fischer, *Sci. Rep.* **2014**, 4, 3629.
- [14] a) W. Ding, J. Zhu, Z. Wang, Y. Gao, D. Xiao, Y. Gu, Z. Zhang, W. Zhu, *Nat. Commun.* **2017**, 8, 14956; b) M. Si, A. K. Saha, S. Gao, G. Qiu, J. Qin, Y. Duan, J. Jian, C. Niu, H. Wang, W. Wu, *Nat. Electron.* **2019**, 2, 580.
- [15] R. Vilaplana, S. G. Parra, A. Jorge-Montero, P. Rodríguez-Hernández, A. Munoz, D. Errandonea, A. Segura, F. J. Manjón, *Inorg. Chem.* **2018**, 57, 8241.
- [16] C. Cui, W.-J. Hu, X. Yan, C. Addiego, W. Gao, Y. Wang, Z. Wang, L. Li, Y. Cheng, P. Li, *Nano Lett.* **2018**, 18, 1253.
- [17] B. Liu, F. Lyu, B. Tang, X. Li, J. Liao, Q. Chen, *ACS Appl. Electron. Mater.* **2021**, 3, 4604.
- [18] S. Wang, L. Liu, L. Gan, H. Chen, X. Hou, Y. Ding, S. Ma, D. W. Zhang, P. Zhou, *Nat. Commun.* **2021**, 12, 53.
- [19] C. Fei, H. Liu, X. Wang, X. Fan, *Nanoscale Res. Lett.* **2015**, 10, 180.
- [20] X. Pan, T. Ma, *Appl. Phys. Lett.* **2011**, 99, 013505.
- [21] B. Kang, J.-G. Yoon, T. Noh, T. Song, S. Seo, Y. Lee, J. Lee, *Appl. Phys. Lett.* **2003**, 82, 248.
- [22] M. Specht, H. Reisinger, F. Hofmann, T. Schulz, E. Landgraf, R. Luyken, W. Rösner, M. Grieb, L. Risch, *Solid-State Electron.* **2005**, 49, 716.
- [23] E. Zhang, W. Wang, C. Zhang, Y. Jin, G. Zhu, Q. Sun, D. W. Zhang, P. Zhou, F. Xiu, *ACS Nano* **2015**, 9, 612.
- [24] Y. C. Chien, H. Xiang, Y. Shi, N. T. Duong, S. Li, K. W. Ang, *Adv. Mater.* **2022**, 35, 2204949.