IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 7, 2020, accepted December 18, 2020, date of publication December 25, 2020, date of current version January 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047491

A Process-Aware Memory Compact-Device Model Using Long-Short Term Memory

ALBERT S. LIN[1], (Member, IEEE), SPARSH PRATIK[1], JUN OTA[1,2], TEJENDER SINGH RAWAT[1], TZU-HSIANG HUANG[1], CHUN-LING HSU[1], WEI-MING SU[1], AND TSEUNG-YUEN TSENG[1], (Fellow, IEEE)

[1]Department of Electronics Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan, R.O.C.
[2]Technology Division, Crystal Engineering Department, Imari Crystal Engineering Section, SUMCO Corporation, Saga 849-0506, Japan

Corresponding authors: Albert S. Lin (hdtd5746@gmail.com) and Tseung-Yuen Tseng (tseng@nctu.edu.tw)

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 109-2221-E-009-060.

ABSTRACT With the immense increase in the processing data during the scaling down of semiconductor devices by Moore's Law, it is in urgent need to use data analytics to meet the state of the art performance in both manufacturing and device compact modeling. In particular, managing the fabrication cost and promptly providing compact device models, especially for new or emerging devices, is challenging. To ease out these issues, we propose a unified, general-purpose, process-aware machine learning (ML) based compact model (CM) for resistive random-access memory (RRAM), and the same methodology can be used for any memory devices with hysteresis. A long short-term memory (LSTM) ML model is used to fit the RRAM current-voltage (I-V) characteristics. The memorizing capability of LSTM ensures one model can fit both RRAM low resistance state (LRS) and high resistance state (HRS). The fitted dataset is based on the fabricated RRAM samples using TaN/HfO_2/Pt/Ti/SiO_2/Si structure. The resultant fitting error is 0.0096 in sinusoidal wave input voltage and 0.0148 in random walk voltage sequences. In the process-aware demonstration, we use post-oxide annealing dataset from 300°C to 500°C. The root mean squared error (RMSE) in the process-aware RRAM compact model is 0.0028. Thus, the LSTM-based CM has the potential to compete with the conventional compact device models in terms of shorter developing time, better fitting capability in emerging devices and a large number of devices, easily incorporated process-aware models, and one unified model accounting for LRS and HRS ensuring differentiability. We propose that the LSTM based memory CM can be useful in intelligent manufacturing, process tuning, and simulation program with integrated circuit emphasis (SPICE) modeling in circuit simulation.

INDEX TERMS Intelligent manufacturing systems, machine learning, process aware modeling, RRAM, semiconductor process modeling.

I. INTRODUCTION

A memory unit is an integral part of any computing system and in 1967, Kahng and Sze revolutionized the semiconductor memory industry by introducing the concept of a floating gate in a non-volatile type memory device [1]. This concept brought a revolution in the field of the semiconductor memory industry. Chua [2] came up with a new two-terminal device called a memristor, and its first fabrication was done by the HP lab [3]. With excellent read/write operation, better endurance, and retention [4], [5], this device has the potential to meet the expectations of memory industries. Due to its fast operations, low power consumption, and many other properties [6], resistive random-access memory (RRAM) is potentially very efficient to be an integral part of the electronics industry. It is a two-terminal device in which an insulating layer is sandwiched between two metal layers, a typical metal-insulator-metal (MIM) structure. There is a wide variety of insulating materials, such as HfO_x [7]–[10], ZnO [11], [12], and ZrO_x [13], [14], available in the fabrication of RRAM. The operation depends upon the formation and destruction of conducting filament in these insulating layers, and this corresponds to two important states of this device, i.e., low resistance state (LRS) and high resistance state (HRS). The transition from HRS to LRS corresponds to the set process, and vice-versa corresponds to the reset

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

process. In bipolar RRAM, there is a continuous change in the states by applying the voltage of different polarities. The performance of the switching behavior is controlled by the oxygen vacancies and the flow of oxygen ions present in the RRAM devices [15]. Despite the rising popularity of RRAM, the compact device model for RRAM is still at a less developed stage compared to metal oxide semiconductor field-effect transistors (MOSFETs) [16]. In the field of compact device modeling, machine learning-based compact models (CM) have emerged as an alternative to physics-based conventional CMs [16]. While the physics-based CMs have the advantage of capturing the essential physics, the downside is that they require a lot of development time. In addition, one physical model can only fit one kind of device. Taking RRAM as an example, RRAM with different dielectric materials or electrodes may not use the same physical CM model due to the different physics involved. Previously, plenty of work related to the physics-based compact modeling for RRAM [17]–[24] has been done. The shortcoming of physics-based CM is that different kinds of RRAMs have different physics, which, in turn, require different CMs. Even for two states, LRS and HRS, there exist different physics, and two equations or models are required in many cases. Also, in most conventional physics-based CMs, an additional model is needed to account for the state variable or state switching. The machine learning (ML) based CM [25], [26] is a promising alternative, and it avoids the physics dependent parameters and can be used to analyze different RRAMs made of various materials in circuits [27]. In addition, since one single voltage value can have two different current values corresponding to either HRS or LRS, it will be shown that one long short term memory (LSTM) model can fit both RRAM HRS and LRS branches at the same voltage value. This eliminates the need for separate models for two switching states.

In addition to the CM for circuit simulation, intelligent manufacturing in process tuning is a related field that is also showing promising significance. The semiconductor manufacturing industry is going through the phase of automation. Industry 4.0, a fourth industrial revolution, proposes the idea of intelligent manufacturing by implementing the Internet of things (IoT) and Artificial intelligence (AI) technologies for smart fabrication [28]–[30]. The conjunction between the simulation program with integrated circuit emphasis (SPICE), compact modeling, and process tuning or optimization is termed process-aware compact device modeling [31], [32]. In this regime, a single model encompasses the prediction from the process, material, and device parameters to device current-voltage (I-V) characteristics. Specific to RRAM, the main challenge toward commercialization is the interpretation of device operation, which arises from the variation in process parameters during fabrication and measurement. These variations in the process parameter affect the device characteristics and their switching behavior [33], [34]. Thus, it becomes crucial to examine the impact of process parameters on device I-V that, in turn, affects circuit simulation. To optimize the process parameters and to arrive at the optimum performance, an extensive and accurate process-aware CM is required. While in conventional CM models, incorporating process parameters is difficult, in ML-based CM, process parameters and device parameters are not different in terms of modeling.

To date, to our best knowledge, we do not have any LSTM-related SPICE models on RRAM. In this paper, we present a unified, reliable, technology-independent, accurate, differentiable ML-based CM, which can easily extend into a process-aware CM model. Here, we are defining a single LSTM-based ML model that can demonstrate all types of device physics, and it is also capable of predicting accurately for all LRS and HRS sets of data. Finally, the process-aware capability is demonstrated using the thermal annealing effect on RRAM I-V. Initial work related to designing of RRAM CM using ML techniques can be found in a student thesis [35]. Our proposed ML-based process-aware SPICE model can be considered as an integral subset of the intelligent manufacturing in fabrication [30] of RRAM. Our model is quite capable of handling huge amounts of RRAM I-V training data. Thus, different RRAM I-V with different processes or device parameters can be fitted using a single LSTM network. Finally, the same methodology in this work can be used to fit any memory device [36] and provide a process-aware spice model for intelligent foundry manufacturing, process tuning, and circuit simulation.

The flow of this work is shown in Fig. 1: (1) the fabrication, (2) the measurement, (3) data generation techniques for the multilayer perceptron (MLP) and the long-short term memory (LSTM) model using sinusoidal wave and random walk signals, and (4) the training and testing of MLP and LSTM. The as-deposited samples are fabricated, and annealing is carried out for part of the samples. The I-V for the different samples is measured following fabrication. For the data set generation in the MLP model, the measured IV data are separated into HRS and LRS. For LSTM modeling, sequences of voltage and current values are generated, representing the IV sweeping history on the devices. Due to the memory device nature, the measured current at this time instant depends upon the previous states of the devices, which in turn depends on the I-V sweeping history. This memory behavior can be modeled using the LSTM networks in the ML field since the LSTM output depends on the previous outputs and states through the recurrent gate. The transformation from regular voltage and current sequences to the format suitable for feeding into LSTM, as demonstrated in Fig. 1(c)-2, can be referenced at common natural language processing tasks [37], [38]. The data generation techniques for different scenarios are explained thoroughly in Section II. Our process-aware compact model facilitates the prediction of RRAM performance in fabrication in terms of process tuning. In addition, the LSTM based CMs can be useful for circuit simulation, especially for new RRAM materials and devices where conventional physics-based CMs are unavailable. Finally, we want to emphasize that process-aware nature

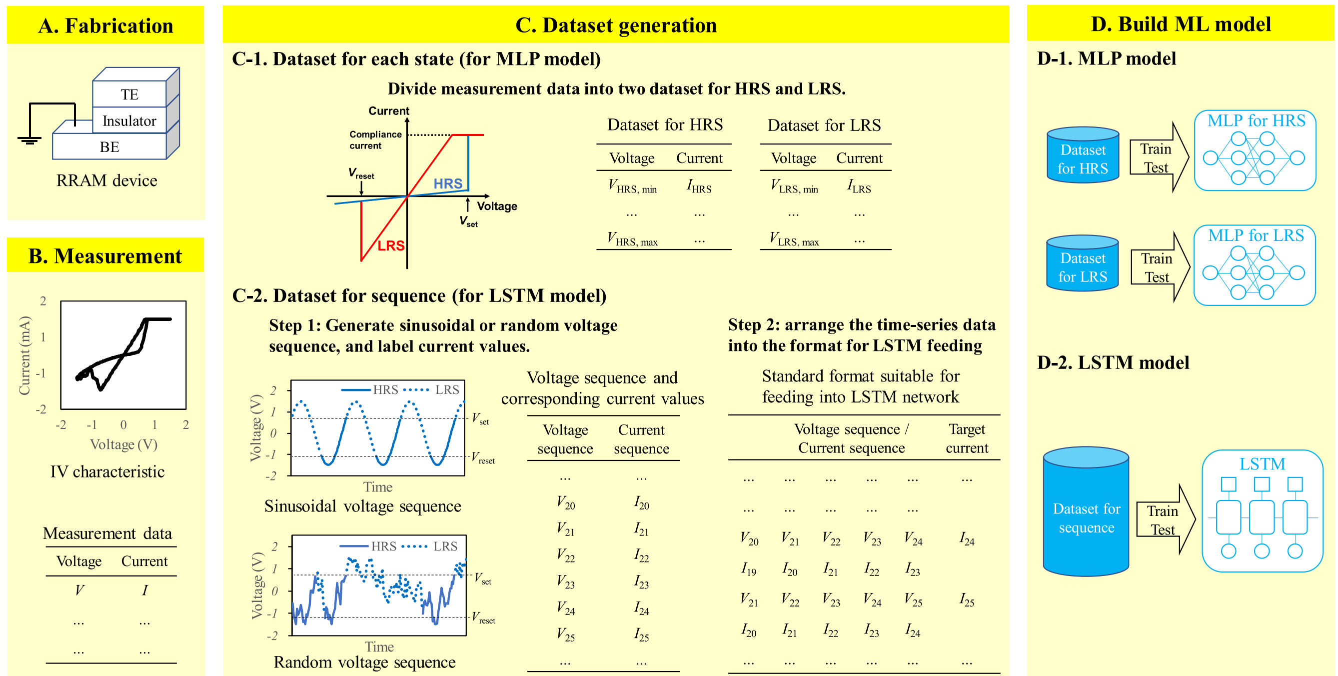

FIGURE 1. The flowchart for RRAM compact device modeling using machine learning (a) fabrication of RRAM sample (b) measurement of RRAM sample (c) generation of dataset (d) training and test of MLP and LSTM.

facilitates the co-optimization of circuits and fabrications. It is not easy to build a process-aware CM using a conventional physics-based approach.

II. METHOD: DESCRIPTION OF MACHINE LEARNING BASED COMPACT MODEL FOR RRAM

A. SAMPLE FABRICATION AND MEASUREMENT

The stack of TaN/ HfO_2/Pt/Ti/ SiO_2/Si has been selected for fabrication, where tantalum nitride (TaN) acts as the top electrode, HfO_2 as an insulating layer, and Pt as the bottom electrode. The transmission electron microscopy (TEM) image of this structure is shown in Fig. 2. A 4-inch silicon wafer is put in a wet bench for RCA cleaning. It removes the different types of contaminations present on the wafer. To protect the sample from leakage current, a 200nm of SiO_2 is deposited through a furnace SJ-CA1200-D4, and it acts as an isolation layer. The entire isolation layer deposition is done at a temperature of 950°C through the wet oxidation method. The stack of Pt(65nm)/Ti(20nm) [39] is deposited by Japan ULVAC EBX-10C Electron beam deposition. Ti layer provides better adhesion to the Pt layer. Afterward, an active layer of 10nm HfO_2 is deposited through Atomic layer deposition (ALD) Fjji F202DC at 250°C. The precursors used for this deposition are Tetrakis(Dimethylamido)hafnium($Hf(NMe_2)_4$) TDMAH and H_2O. The heating process for TDMAH is carried out at 75°C, and the gaseous flow for argon carrier and argon plasma is 60 sccm and 200 sccm, respectively. A pulse of 0.055sec is applied to maintain the change in the pressure value between 25-35 mtorr for H_2O precursor, and for TDMAH precursor, 1-sec pulse regulated the change in the pressure value in 16-20 mtorr. The quality of ALD HfO_2 is better than other [40] HfO_2 deposition machines. A stable electrical property of HfO_2 is achieved by the reaction of both the precursors with the substrate, and the deposition rate is 1.01 Å/cycle.

Post oxide annealing technique has been incorporated to demonstrate the effect of temperature on the RRAM. The heating of the sample at 300°C, 400°C, and 500°C is carried out in the vacuum environment, and the pressure is kept at 5 mtorr, as shown in Fig. 3. The heating rate of the heating chamber is 10°C/min, and the duration of the annealing process is set for 30 minutes. A slow and continuous cooling process is used where the sample reaches to room temperature inside the chamber by itself. TaN top electrode (TE) is patterned using a shadow mask, and its diameter is 400 microns. KD-SPUTTER A-09L08 DC sputter machine is used for the deposition of TE. The base pressure, the DC sputtering power, and temperature are set at 1.4×10^{-6} torr, 800W, and room temperature, respectively. This process is carried out in the vacuum environment with argon plasmons. During the deposition, a strong electric field is applied to generate argon plasma, and it hits the target material that results in the deposition of 100nm TaN. The rate of Argon gas flow is 100 sccm, and N_2 is 10 sccm. The deposition thickness is estimated as 1.00 Å/sec. Once the as-deposited and annealed samples are ready, the characterization is done through Agilent Keysight B1500A.

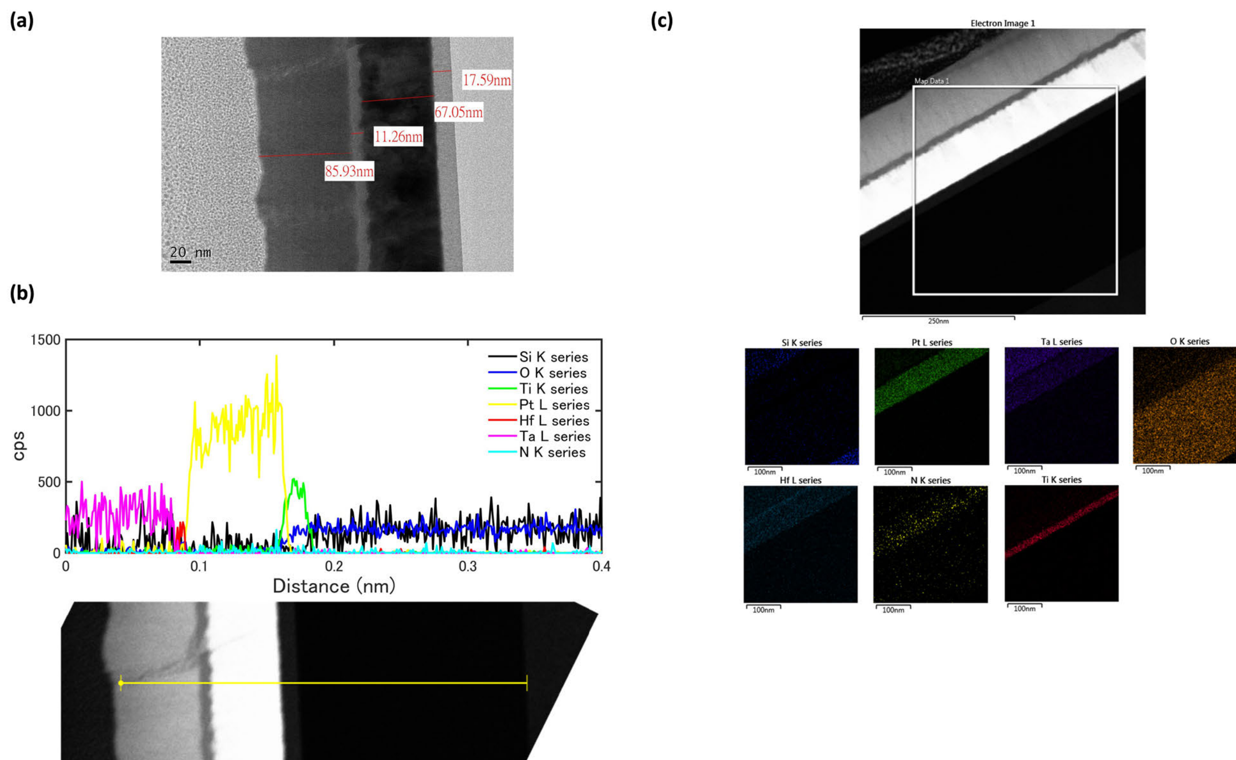

FIGURE 2. (a) TEM image for the RRAM sample. (b) TEM-EDS line scan, and (c) TEM-EDS 2D mapping.

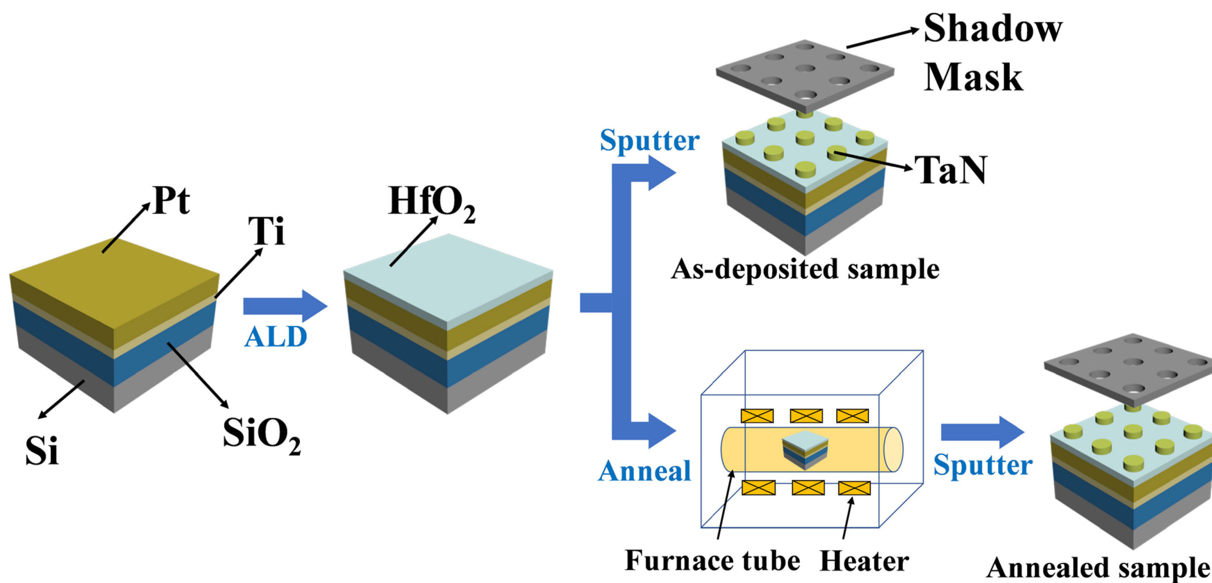

FIGURE 3. The device structure of $TaN/HfO_2/Pt/Ti/SiO_2/Si$ RRAM and its fabrication flow.

B. DATA GENERATION

Data set generation is straightforward for MLP model since HRS and LRS are fitted separately using two MLP networks. The data set generation for LSTM is more complex, but the basic idea is illustrated in Fig.1 and Fig. 4. The RRAM output current depends on the state, i.e., LRS or HRS, which in turn depends on the sweeping history. For example, at $V_{reset} < V < V_{set}$, the value of the output current is determined from the previous state of the RRAM. If previously RRAM is at LRS, then the current values will be known from the LRS branch, and vice versa. Therefore, to use one model to take into account both LRS and HRS, the sweeping voltage history,

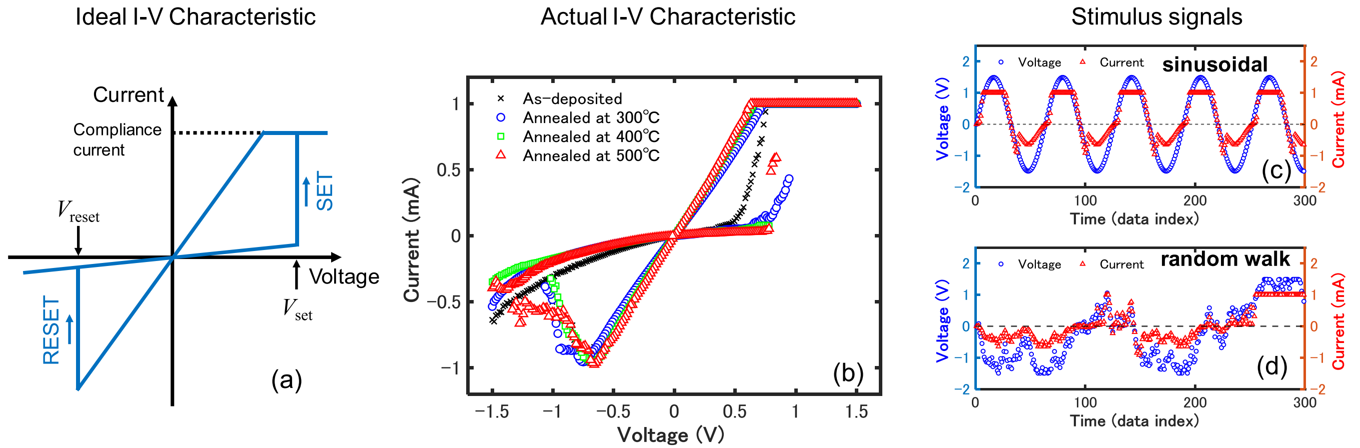

FIGURE 4. (a) ideal I-V characteristics for RRAM (b) actual I-V characteristics for RRAM sample at As-deposit, 300 °C, 400 °C, and 500°C (c) Voltage and current for a sinusoidal input voltage signal (as-deposit) (d)Voltage and current cycle for a random walk input voltage signal (as-deposit).

i.e., the voltage input values at previous time steps, need to be fed into LSTM in addition to the voltage input value at this time step. The input voltage, either sinusoidal or random walk signals, are generated. Based on the measured IV as shown in Fig. 1(b) or Fig. 4(b), the corresponding current values are generated. The measured IV consists of 400 data points ranging from -1.5V to 1.5V, and interpolation is conducted if the input signal values do not exactly reside on the probed voltage values. This way, the entire input voltage sequence and output current sequence is generated, and the compact modeling becomes a standard time-series machine learning problems where the goal is to predict the output values in the future steps. To have simple terminology, and since our measurement data points are dense enough, we do not distinguish between measured values and interpolated values from the measurement in the following write-up. Similar to the earlier benchmark RRAM CMs [21]–[23], the slight hysteresis after reset is not taken into account. The slight hysteresis after set does not exist in our case or in most of the literature due to current compliance (c.c.). The time series, i.e., V_{seq} vs. I_{seq}, needs to be transformed into the format suitable for LSTM feeding, and this aspect is explained in many standard time-series machine learning examples, such as machine translation, in Tensorflow manual [37], [38]. As illustrated in Fig. 1(c)-2, we transform the input data into the common form suitable for LSTM feeding taking into account the voltage (V) and current (I) values at previous time steps.

1) DATASET GENERATION FOR MLP

For MLP implementation, a separate MLP model for each state is implemented, and it is shown in Fig. 6(a)

$$f_{MLP}(V_{LRS}) = I_{LRS} \tag{1}$$

$$f_{MLP}(V_{HRS}) = I_{HRS} \tag{2}$$

where V_{LRS} and V_{HRS} are different voltages corresponding to LRS and HRS states, and I_{LRS} and I_{HRS} are the LRS and

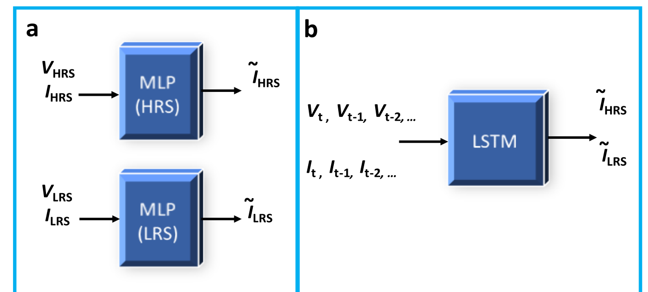

FIGURE 5. (a) MLP: Two model architecture. (b) LSTM: One model architecture.

HRS current values, which are the functions of V_{LRS} and V_{HRS}, respectively. The selection of hyper-parameters is done in accordance with Table 1.

2) DATASET GENERATION FOR SINUSOIDAL INPUT VOLTAGE

Firstly, we test the LSTM-based RRAM CM using sinusoidal input voltage sequences. The equation for output current is represented by (3)

$$f(V_{seq}, I_{seq}) = I_{target} \tag{3}$$

where I_{seq} and V_{seq} are the voltage and current sequence in the past, and I_{target} is the current at present. V_{seq} is generated by

$$\begin{aligned} V_{seq} &= V_{offset} + A\sin(t+\varphi) \\ &= \frac{V_{max}+V_{min}}{2} + A\sin(t+\varphi) \end{aligned} \tag{4}$$

$$A = \frac{V_{max}-V_{min}}{2} \tag{5}$$

where t is time, V_{max} and V_{min} are the maximum and minimum values in voltage sweeping, A is amplitude, and φ is the phase. The total length of the voltage sequence is 10000 in the training and test sets. The dimension for training the model

TABLE 1. List of models and hyperparameters.

	MLP	LSTM (Sine)	LSTM (RandomWalk)	LSTM (Anneal)
Number of Model block	2	1	1	1
Number of neurons	100	100	100	200
Activation Function, LSTM	relu	tanh	tanh	tanh
Batch size	500	100	1000	2000
Number of samples, training	200	10000	10000	50000
Number of samples, testing	-	10000	10000	10000
Number of Epoch Cycle	2000	2000	2000	2000
Input Shape	(Batch size, 1)	(Batch size, 5,2)	(Batch size,5,2)	(Batch size,5,3)
Output Shape	(Batch size,1)	(Batch size, 1)	(Batch size, 1)	(Batch size, 1)
Length of V_{seq}, I_{seq}	-	5	5	5

will be the voltage sequences along with the labeled current value, i.e. (Batch size, 5,2) at the input and (Batch size, 1) at the output. Different values of φ are also considered to generate a test set such that proper evaluation of LSTM-based CM can be implemented. In this case, a single LSTM model manages both LRS and HRS states. The hyper-parameters selection for the model training for this scenario is based on Table 1.

3) DATASET GENERATION FOR RANDOM WALK INPUT VOLTAGE

The shape of the waveform should not be strictly a sinusoidal signal in circuit simulation. To ensure the model comprehensiveness, i.e., our LSTM model can withstand any undue input conditions, we also consider random walk signals in the test of LSTM-based CM. Similar to the LSTM handling sine signals, we use the same set-up to handle random walk input voltage

$$f(V_{seq}, I_{seq}) = I_{target} \tag{6}$$

$$V_{seq} = V_{start} + dv \times random(\text{-}1,1) \tag{7}$$

Here V_{start} and dv can be considered as any random value of the measured data. Our values are 0 and 0.3, respectively. *random*(-1,1) will give a random number lying between -1 and 1.We are also clipping the amplitude of the random signal using logical loops. This clipping is required to make the sequence voltage lie in the maximum and minimum values of the measured data. The total size of the input training data will become (Batch size, 5, 2). These conditions work very efficiently and effectively for any random walk signals.

4) DATASET GENERATION FOR PROCESS-AWARE COMPACT MODEL

Data generation for this condition is similar to the approach in previous sub-section 3). As we are using one LSTM model for all types of RRAM devices annealed at various temperatures, we need to concatenate the as-deposited, 300°C, 400°C, and 500°C measured RRAM I-V data. Additionally, we have to extend the input parameters of the LSTM model to include the annealing temperature (T), as represented by (8)

$$f(V_{seq}, I_{seq}, T) = I_{target} \tag{8}$$

V_{seq} and other parameters are generated in a similar way using (7). The size of the input training dataset will become (Batch size, 5, 3). This model is also evaluated on the generated test dataset.

C. MACHINE LEARNING ALGORITHMS

In an MLP [41], the input layer is well connected with the output layer through a layered network of hidden layers, and every layer consists of many neurons, as shown in Fig. 6(a). From (9), the incoming signals are multiplied with the corresponding weight and further added with the bias. Further, different types of activation functions [42] are used to add some nonlinear effects in the model, and it also gives a better fitting to the output.

$$y = a(\sum xw + b) \tag{9}$$

where x is the incoming input to a neuron, y is the output from a neuron, a is the activation function, w is the weight, and b is the bias. In contrast to MLP, LSTM [43], [44], as shown in Fig.6(b) consists of input, output, and forget gate, which are just a combination of neurons with the activation function. Sigmoid and hyperbolic tangent (tanh) are used in this model. Cell state values are the most important aspect of the LSTM unit, and it acts as a repository. Forget gate decides which values need to be deleted from the cell state. The mathematical operation is shown in (10) [43]. The range of sigmoid activation is between 0 to 1. Values closer to

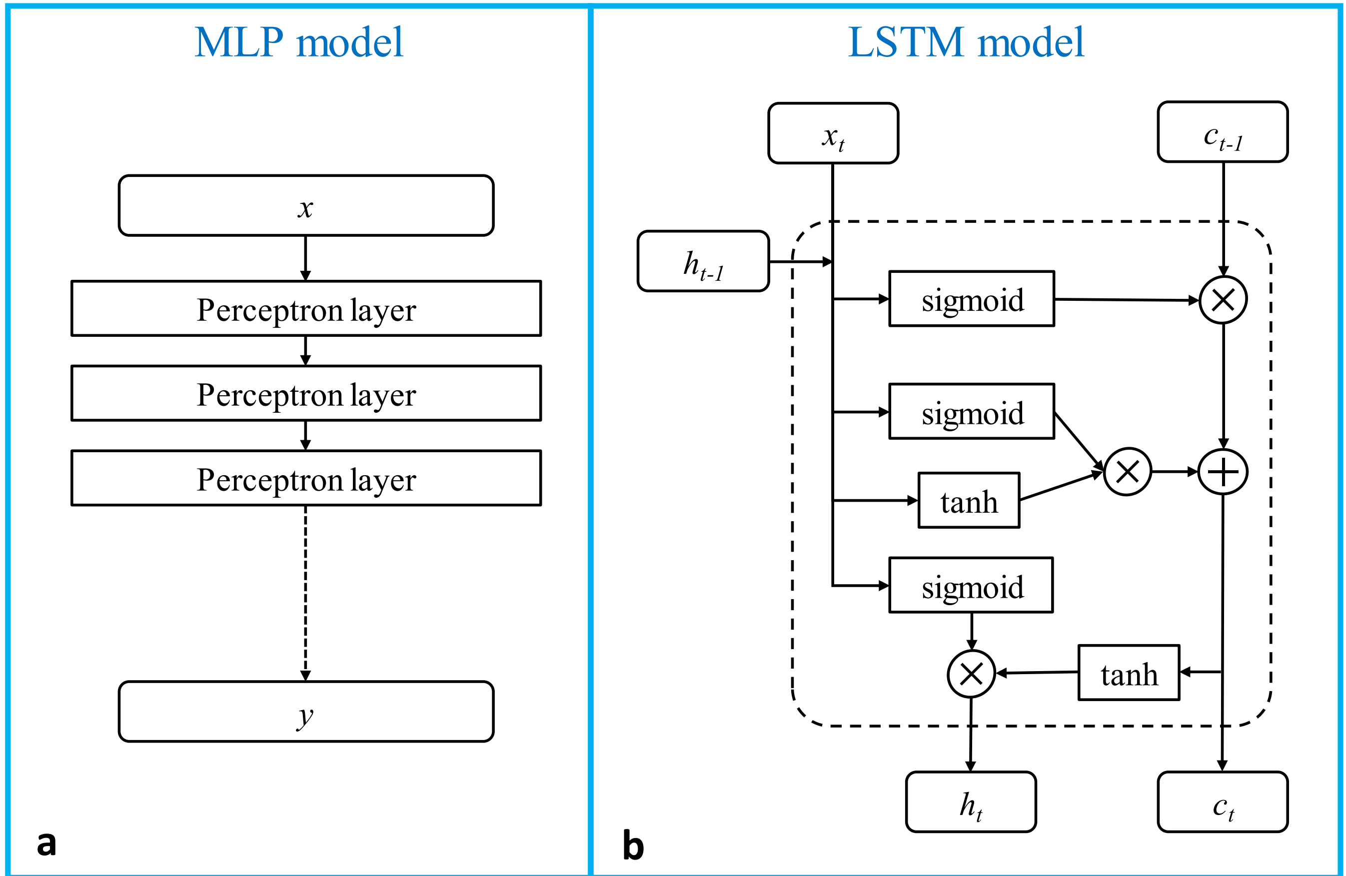

FIGURE 6. ML algorithms (a) MLP model [41], (b) LSTM [44].

0 corresponds to delete, and closer to 1 corresponds to retain.

$$f_t = \sigma(x_t w_f^x + h_{t-1} w_f^x + b_f) \tag{10}$$

Adding new values in the cell state is done by using the tanh function. Its range varies from -1 to 1. A mathematical operation for adding new values is calculated through (11) [44].

$$i_t = \sigma(x_t w_i^x + h_{t-1} w_i^x + b_i) \tag{11}$$

Equation (12) [44] denotes the updated state of the cell state.

$$c_t = f_t c_{t-1} + i_t \tanh(x_t w_c^x + h_{t-1} w_c^x + b_c) \tag{12}$$

The hidden state stores information for the previous input, and the updated state is calculated using [44]

$$o_t = \sigma(x_t w_o^x + h_{t-1} w_o^x + b_o) \tag{13}$$
$$h_t = o_t \times \tanh(c_t) \tag{14}$$

In (10)-(14), the symbols, f corresponds to the forget gate, i corresponds to the input gate, o corresponds to the output gate, σ is the sigmoid activation, c is cell state, h is the hidden state, and t is the present time step. w is the weight, and b is the bias.

We are proposing a single LSTM based CM device that can be used in process tuning and SPICE simulation for RRAM. Hyper-parameters are tuned in such a way that the overall training loss present in the model is minimized. The memorizing capability of LSTM is achieved by the recurrence, in which the output at the current time step will be fed back as the input at the next time step. Because the historical or past outputs before the current time step are continuously fed back to the neural network as inputs, the output at the current time step is not only a function of the input at the current time step but also of the inputs at earlier time steps.

III. RESULTS AND DISCUSSION

Two different ML algorithms, MLP and LSTM, have been implemented in the design of the semiconductor device CM for RRAM. The code is written using MATLAB [45], Python TensorFlow 2.0 with Keras [37], [38], and Scikit-Learn library [46]. The evaluation metrics for all the models are the root mean squared error (RMSE), R-2 score, mean absolute error (MAE), and relative absolute error (RAE). In a regression problem, RMSE is defined as the square root of the mean squared error (MSE), where MSE is expressed as the summation of the square of the difference between the predicted values and the true values, divided by the total number of data points. It measures the deviation of the predicted values from

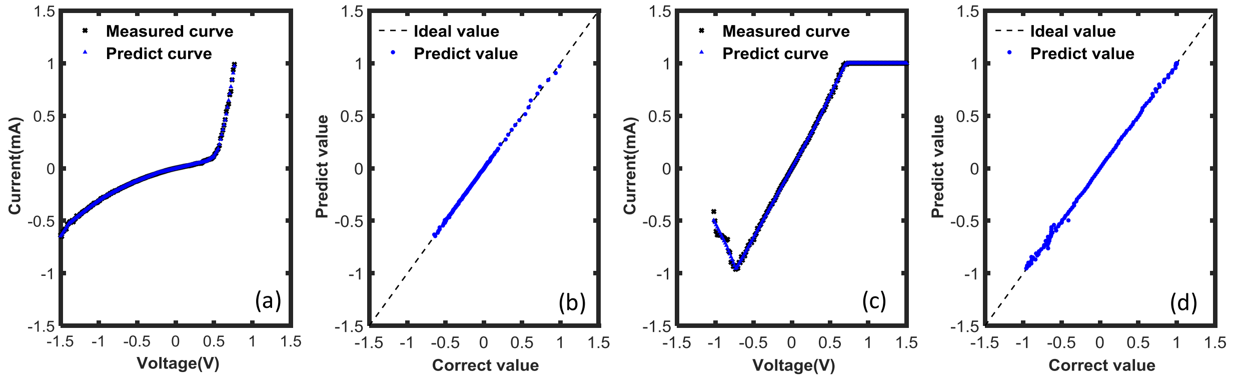

FIGURE 7. MLP based RRAM compact model fitting. (a) I-V characteristics on HRS, (b) scatter plot on HRS, (c) I-V characteristics on LRS, and (d) scatter plot on LRS.

the true values:

$$RMSE = \sqrt{\sum_{i=0}^{n} \frac{(Y_{original,i} - Y_{predict,i})^2}{n}} \tag{15}$$

where $Y_{original}$ is the current value obtained from measurement, $Y_{predict}$ is the current value obtained from different ML models, and n is the total number of data points.

A. MLP

Similar to some conventional physics or equivalent circuit-based CM RRAM models where separate equations are used for LRS and HRS, here we also use two MLP models to fit LRS data and HRS branches in RRAM I-V, respectively. Certainly, a unified model is desired where one model is capable of predicting both LRS and HRS current values. One model also ensures model differentiability where the separate models can lead to discontinuity at transition points.

In MLP-based CM, each MLP has two hidden layers with 100 neurons in each layer. Preprocessing of measured data is required to obtain better fitting with fast convergence, which includes the normalization of the input voltage and output current variables and separation of the RRAM I-V into HRS and LRS branches. Sklearn library is used for normalization purposes. The model is trained for 2000 epoch cycles with a batch size of 500, and its fitting performance for HRS and LRS are shown in Fig. 7. The model predicted values are shown with the sample I-V characteristic in Fig. 7(a), (c) and with scatter plot in Fig. 7(b), (d). In the MLP separate LRS/HRS modeling, the output current is only a function of input voltage and does not depend on the history and the trace of the sweeping. The HRS I-V characteristic for our RRAM sample is smooth in behavior, and thus, the ML model fits well on this branch of data with RMSE of 0.0072. On the other hand, the MLP model for LRS is less accurate because of the small fluctuation present in the current value during the reset process of −0.7 V to −1.5 V. Therefore, the MLP model for LRS has an RMSE of 0.0172.

The neurons present in these algorithms are triggered using rectified linear unit (ReLU) activation function, and the overall current value is calculated using (16)

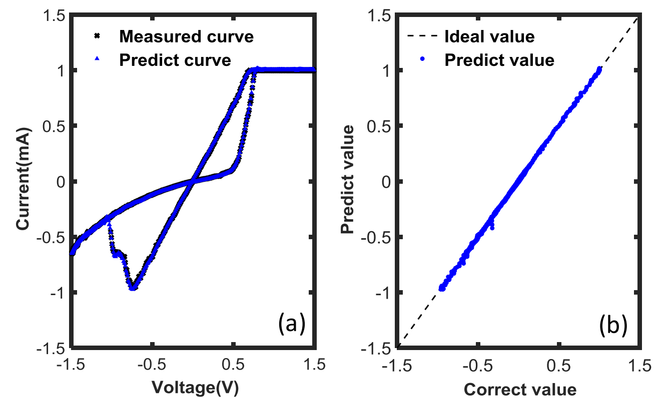

FIGURE 8. The train set for LSTM based RRAM compact model in response to the sinusoidal input voltage. (a) I-V on the train set, and (b) the scatter plot on the train set.

$$I = \sum_k \left\{ f_{ReLU2} \left\{ \sum_j f_{ReLU1} \left(V \times w_i + b_i \right) \times w_j + b_j \right\} \times w_k + b_k \right\} \tag{16}$$

where w_i, w_j, w_k, b_i, b_j, and b_k are weights and biases for the first hidden layer, the second hidden layer, and the output layer. f_{ReLU1} and f_{ReLU2} is ReLU function for the first and second hidden layers, respectively.

B. LSTM MODEL WITH SINUSOIDAL DISTRIBUTION

In the previous section, two MLP models are used to fit the LRS and HRS branches in RRAM I-V. To proceed a step further, we use the LSTM model with memory to fit both of the branches at once. MLP is a feedforward network. Feedforward networks are static and acyclic in nature. As the RRAM output depends on sweeping history, MLP does not match intrinsically with the RRAM behavior. On the other hand, while LSTM uses recurrence, LSTM can handle sequence, and its output depends on previous input in addition to the

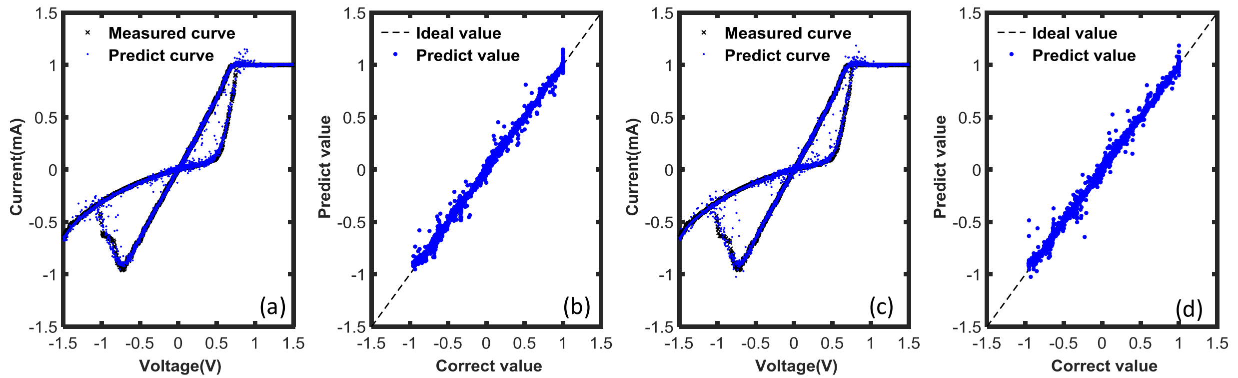

FIGURE 9. The test set for LSTM based RRAM compact model in response to sinusoidal input voltage (a) I-V on test dataset for $\varphi = -1$, (b) scatter plot on test dataset for $\varphi = -1$ (c) I-V on test dataset for $\varphi = 1$, (d) scatter plot on test dataset for $\varphi = 1$.

current input. Therefore, MLP has poor performance on temporal or time-series data while LSTM, in general, performs better. One way to make the MLP output depends on the previous input signals is to expand the input neuron number. Nonetheless, this leads to very large models, especially for complex problems such as process-aware CMs.

The LSTM is applied to the dataset that has sinusoidal input voltage. It should be emphasized that the sinusoidal signals are the basic waveforms that can be used to synthesize other waveforms using the Fourier series. In this setup, multiple cycles of sinusoidal voltages with a constant time period (T) and amplitude (A) are generated. Specific to RRAM or other memory devices, during one cycle of endurance, a single voltage has two different current values. One corresponds to LRS current and another to HRS current. This is the most difficult aspect to be modeled, especially if only one model can be used. The historical sequence in voltage or voltage trace, i.e., voltage sweeping in the previous steps, differentiates these two states. In this work, we use the memory capability of LSTM to model this aspect. The memorizing capability can be achieved by using recurrence, i.e., feedback, in ML models. Since the output at the previous time instant (t-1) is re-used, through recurrence, as the input at the current time instant (t), the output at the current time instant is a function of the current input at (t) and previous inputs. Similar concepts exist in circuit design and communication systems where feedback exists. In conclusion, when we have recurrence or equivalently feedback loop, the system more or less memorizes the previous inputs, and this is the reason this model is called "long-short term memory (LSTM)". In our cases, the previous inputs are the previous voltage sweeping values across the RRAM. Fig. 4(c) shows the current characteristics for a given sinusoidal voltage signal. The training data with zero phases (φ) is generated using (4), and its fitting result using the LSTM model is shown in Fig. 8. Two sinusoidal signals with different phases named Test-1 and Test-2 are also incorporated for testing the LSTM-based compact model capability. Fig. 9 shows the fitting results of the LSTM model on the Test-1 and Test-2 dataset.

The RMSE of 0.0094 and 0.0092 is coming for Test-1 and Test-2, respectively. Even though we have different phases for both test signals, Fig. 9 shows a perfect fitting with respect to training data for both cases. The other performance metrics R-2 score, mean absolute error (MAE), and relative absolute error (RAE) for Test-1 are 0.9998, 0.0066, and 0.0105, respectively. These values are also very close to their ideal values, indicating satisfactory prediction. From Table 2, we can observe that all the evaluation parameters for Train, Test-1, and Test-2 are showing satisfactory fitting and prediction and close to their respective ideal values. Thus, we can generalize that this LSTM based CM for RRAM can handle the sinusoidal distribution model, giving an excellent response for any sinusoidal input signals.

C. LSTM FOR RANDOM WALK DISTRIBUTION

It is not desired that our proposed LSTM based approach is limited to only sinusoidal signals, and the transient input voltage signals can actually be in any shape. Thus, to increase the applicability of our approach up to the state-of-the-art, we have considered the fitting and prediction accuracy when the input voltage is a random walk signal. One example of the random walk signal is depicted in Fig. 4(d). If the LSTM-based RRAM CM can handle arbitrary random signals in voltage, the usability of it can be widely expanded.

The LSTM model is trained for the 2000 epoch cycles, and its corresponding RMSE value is shown in Table 2. To decrease the computational time or to increase the fitting accuracy, training batch sizes are altered. In Tensorflow and specific to our RRAM ML CM problem, decreased batch size leads to better fitting in the training set while increased batch size leads to shorter training time for a fixed number of epochs. The effect of the training set batch size on the test set prediction accuracy is less pronounced, but if the fitting on the training deteriorates significantly, test set accuracy will be affected. In the calculation in this sub-section, we use batch size = 1000 to arrive at the trade-off between training set fitting accuracy and training time. LSTM [47], [48] has a strong dependency on the previous data, and this property makes LSTM a perfect model for predicting RRAM ML-based CM

TABLE 2. Performance table.

		RMSE Value (Ideal Value=0)	R2-Score (Ideal Value=1)	MAE (Ideal Value=0)	RAE (Ideal Value=0)
MLP	LRS	0.0172	0.9994	0.0090	0.0139
	HRS	0.0072	0.9994	0.0040	0.0191
Sinusoidal Distribution (LSTM)	Train φ=0	0.0096	0.9998	0.0066	0.0105
	Test-1 φ=-1	0.0094	0.9998	0.0066	0.0105
	Test-2 φ=1	0.0092	0.9998	0.0065	0.0105
Random Walk (LSTM)	Train, Seed=1	0.0148	0.9993	0.0093	0.0173
	Test-1, Seed=2	0.0182	0.9990	0.0100	0.0198
	Test-2, Seed=3	0.0195	0.9989	0.0102	0.0192
	Test-3, Seed=4	0.0187	0.9990	0.0102	0.0188
	Test-4, Seed=5	0.0173	0.9991	0.0096	0.0177
Process-aware Annealing	Train, Seed=1	0.0028	0.9999	0.0017	0.0031
	Test As-deposit, Seed=2	0.0098	0.9997	0.0035	0.0069
	Test 300°C, Seed=2	0.0092	0.9997	0.0018	0.0042
	Test 400°C, Seed=2	0.0118	0.9995	0.0019	0.0041
	Test 500°C, Seed=2	0.0155	0.9992	0.0031	0.0064
	Test As-deposit, Seed=3	0.0121	0.9995	0.0036	0.0069
	Test 300°C, Seed=3	0.0099	0.9996	0.0019	0.0040
	Test 400°C, Seed=3	0.0127	0.9994	0.0020	0.0041
	Test 500°C, Seed=3	0.0104	0.9996	0.0028	0.0056

through its regulated feedback mechanism. The random walk signal presents more challenges than sinusoidal signals since the randomness nature requires better handling of the voltage trace information and the agiler response of the ML model to different input voltage sequences. It is demonstrated here that even for random walk signals, LSTM can fit RRAM IV well in both HRS and LRS branches, using only one model. RMSE value on the training data is 0.0148. For Test-1, Test-2, Test-3, Test-4, RMSE is 0.0182, 0.0195, 0.0187, 0.0173 respectively. Fig. 10(b) is showing the scatter plot on the training data.

Generally, the problem of overfitting is more pronounced in the regression type problem [49] relative to classification problems. Thus, to verify the model's performance on unseen data, we have used four different random walk signals for testing purposes. From Table 2, we can observe that the training set RMSE is less than the test set RMSE by a small margin, verifying the overfitting is not present. On the other hand, there is no underfitting problem, evident from the satisfactory

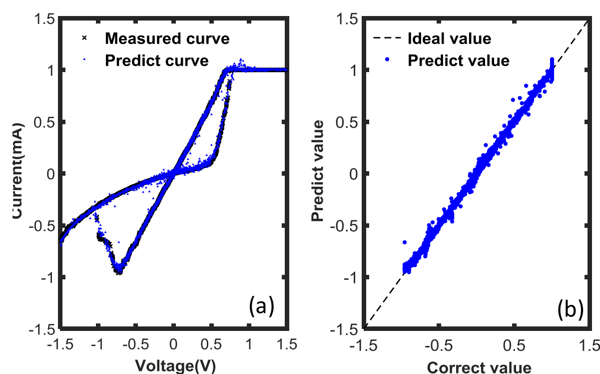

FIGURE 10. The train set for LSTM based RRAM compact model in response to the random walk signal (a) I-V on train dataset, (b) scatter plot on train dataset.

matching between predicted and true values in the training set. With these evaluation metrics values, we believe we are proposing a suitable LSTM-based CM for any type of

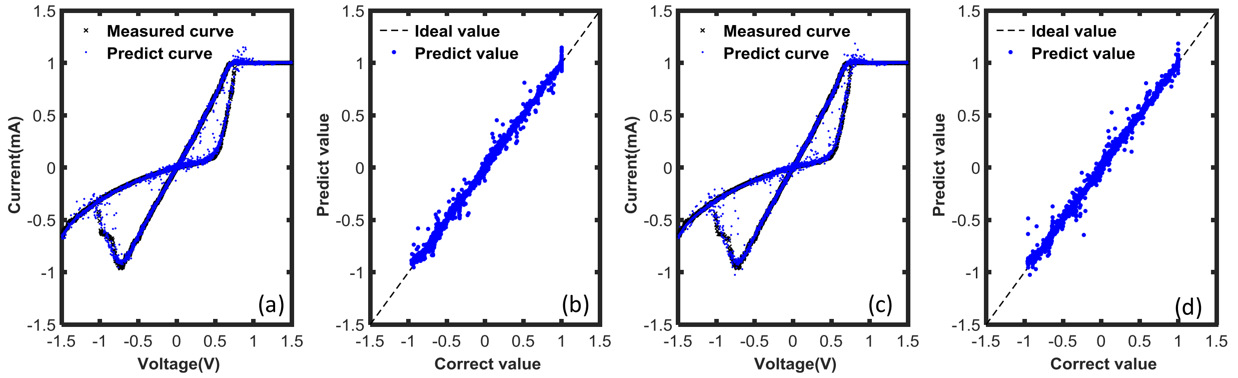

FIGURE 11. The test set for LSTM based RRAM compact model in response to the random walk signals. (a) I-V on test random signal-1 dataset, (b) scatter plot on test random signal-1 dataset (c) I-V on test random signal-2 dataset, (d) scatter plot on test random signal-2 dataset.

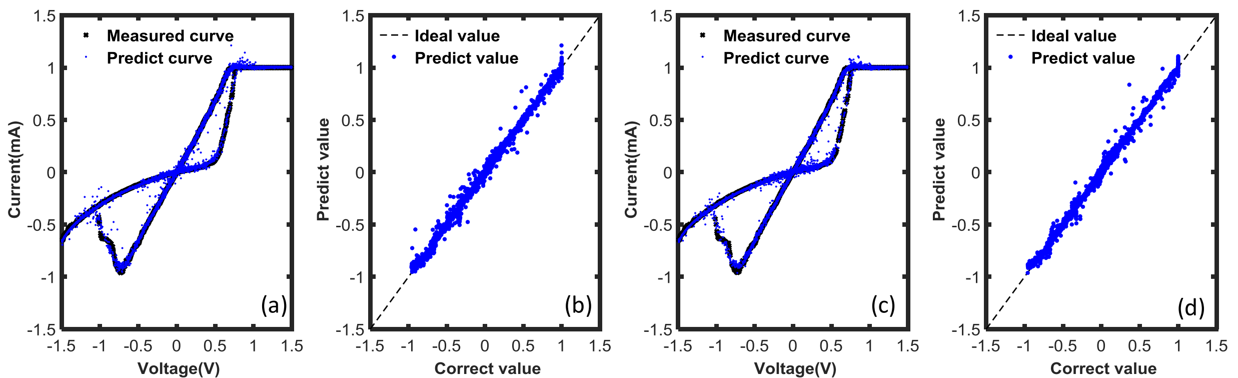

FIGURE 12. The test set for LSTM based RRAM compact model in response to the random walk signals (a) I-V on test random signal-3 dataset, (b) scatter plot on test random signal-3 dataset (c) I-V on test random signal-4 dataset, (d) scatter plot on test random signal-4 dataset.

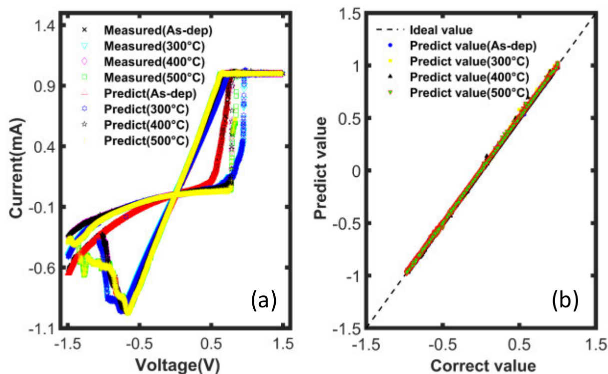

FIGURE 13. Train set for process-aware LSTM model applied to anneal experiments. (a) I-V on train dataset, (b) the scatter plot on train dataset.

input voltage stimulus, which can be useful in process tuning, intelligent manufacturing, and circuit simulation.

D. LSTM FOR TEMPERATURE-DEPENDENT PROCESS AWARE MODEL

The variation on the I-V of RRAM on changing the process parameters is the key to judge its performance, robustness, and yield. Incorporating process parameters in the conventional physics-based SPICE model is not very easy since the relationship between the semiconductor fabrication process parameters and the device I-V characteristics are complex, and in some cases, the underlying physics has not been fully understood. Given the fact that the compact device model relating V to I is already complex by itself, adding process physics such as plasma, fluid dynamics, and thermal physics into the device CM is in practice impossible. On the other hand, using ML as the building block for device CMs does not suffer from the difficulty of incorporating process parameters into the model. This is because there is little difference between pure CM relating V to I, or process-aware CM relating V, annealing temperature, gas flow rate and partial pressure, and plasma power to I. To demonstrate the effectiveness of using LSTM for process-aware CMs, we investigate the effect of annealing temperature on the device performance and construct a process-aware CM whose inputs include annealing temperature, in addition to voltage and current sequences. The annealing process-aware CM has not been incorporated into the RRAM SPICE model to date to our best knowledge. Here we are inculcating the effect of different post-oxide [50] annealing temperatures on I-V characteristics of the RRAM sample directly to the

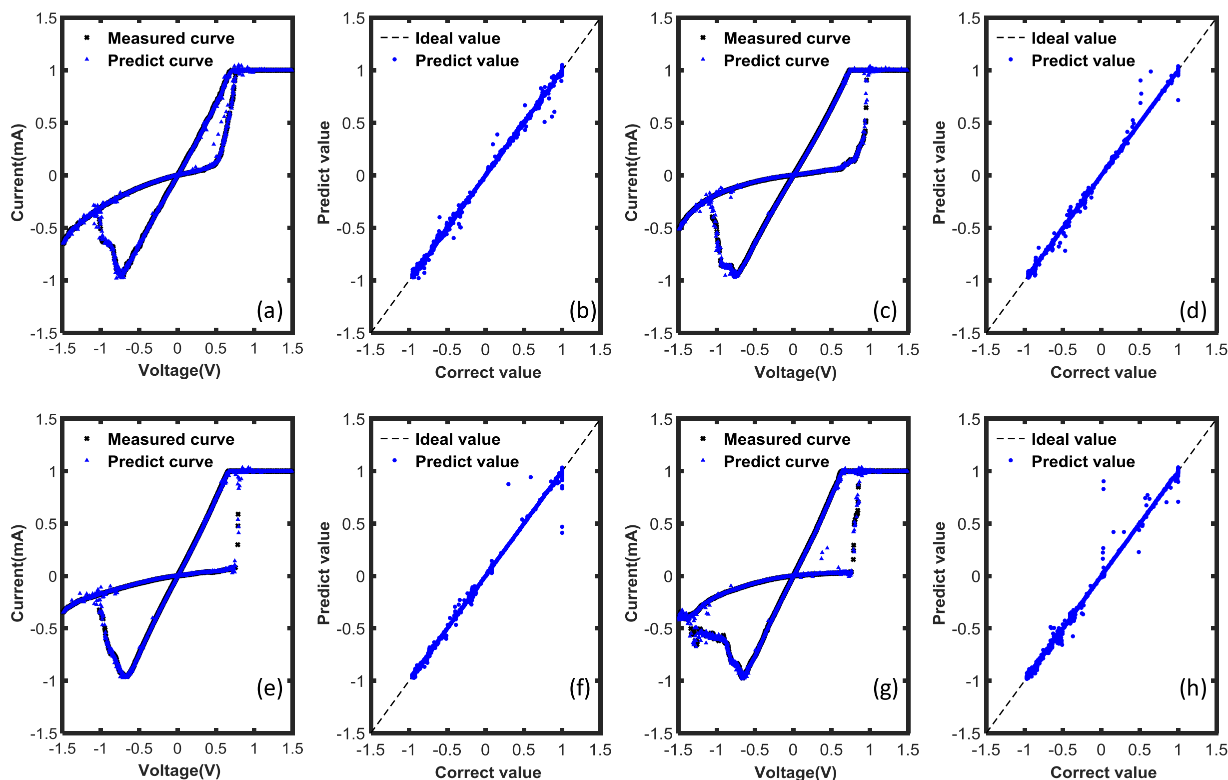

FIGURE 14. Test set process-aware LSTM model applied to anneal experiments for random seed 2 test dataset. (a) I-V for As- deposit, (b) scatter plot for As- deposit (c) IV characteristics on 300 °C, (d) scatter plot on 300 °C, (e) I-V characteristics on 400 °C, (f) scatter plot on 400 °C (g) I-V characteristics on 500 °C, (h) scatter plot on 500 °C.

device compact model (CM). Similar to the previous subsection, LSTM with a random walk signal is used to show the device I-V dependency on the annealing temperature. Model training is carried out on the I-V dataset generated from as-deposited, 300°C, 400°C, and 500°C sample data, and its shape is. (Batch size, 5,3) with totally 50000 data points for each temperature. Depending upon the different random seed values, test signals are generated to study the effect of annealing on RRAM I-V. Test signals for as-deposited, 300°C, 400°C, and 500°C are generated separately, and their shape is (Batch size, 5,3) with 10000 data points for each temperature. In the annealing section of Table 2, we observe that RMSE in the training set is 0.0028. The RMSE testing results for as-deposited, 300°C, 400°C, and 500°C are 0.0098, 0.0092, 0.0118, 0.0155, respectively, and it is calculated on random seed -2 dataset. The same approach is applied for on random seed -3 dataset, and values are 0.0121, 0.0099, 0.0127, 0.0104 for as-deposited, 300°C, 400°C, and 500°C, respectively. There is also no overfitting and underfitting problem observed in the temperature-dependent process-aware model, evident from the small train RMSEs and the train and the test RMSEs with the same order of magnitude. Fitted I-V characteristic on training data is shown in Fig. 13, and the test results are shown in Fig. 14. The simple explanation on the physics is that annealing alters the crystallinity of HfO_2 and thus changes the resistive switching characteristics of the sample, which in turn leads to the shifting in the set and reset voltage. This shift is accurately modeled by LSTM-based CM, evident in Fig. 14. The process-aware compact device model in this section demonstrate multiple input capability in LSTM-based RRAM CMs. More inputs can be easily incorporated into LSTM since inherently LSTM handles multiple input features.

IV. CONCLUSION

Inspired by the demand of electronic design automation (EDA) industries and the semiconductor foundries for accurate process-aware compact models, we are proposing a powerful LSTM-based ML RRAM model. The recurrent nature and the state-dependent operation of LSTM properly take into account the RRAM I-V with hysteresis. Using the real data from the fabricated RRAM samples with post-oxide annealing, we demonstrate the feasibility of a single, differentiable, purely ML-based, and process-aware RRAM compact model that can fit experimental I-V well, and the same approach can be easily generalized to any memory devices without changing the architectures. RMSE, R-2 score, MAE, and RAE results show that the single LSTM model can handle both sinusoidal and random walk input voltage sequences properly and possess process-aware capability. Specifically, the fitting RMSE values are 0.0096 for sinusoidal signals, 0.0148 for random walk signals, and 0.0028 for

process-aware CM, and the RMSE is at the same order of magnitude in testing. The advantages and potentials of LSTM RRAM CM, compared to conventional physics-based CM, lies in shorter development time, capability in fitting different memory devices and a large number of devices, incorporation of the process parameters into CM, and a single model for LRS/HRS branches.

ACKNOWLEDGMENT

The code of this work can be downloaded from https://github.com/albertlin11/rram_lstm. *(Albert S. Lin and Sparsh Pratik contributed equally to this work.)*

REFERENCES

[1] D. Kahng and S. M. Sze, "A floating gate and its application to memory devices," *Bell Syst. Tech. J.*, vol. 46, no. 6, pp. 1288–1295, Jul. 1967.

[2] L. Chua, "Memristor-The missing circuit element," *IEEE Trans. Circuit Theory*, vol. CT-18, no. 5, pp. 507–519, Sep. 1971.

[3] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80–83, May 2008.

[4] S. Yu, *Resistive Random Access Memory (RRAM): From Devices to Array Architectures*. Morgan & Claypool: San Rafael, CA, USA, 2016.

[5] S. Munjal and N. Khare, "Advances in resistive switching based memory devices," *J. Phys. D, Appl. Phys.*, vol. 52, no. 43, Oct. 2019, Art. no. 433002.

[6] K. M. Kim, D. S. Jeong, and C. S. Hwang, "Nanofilamentary resistive switching in binary oxide system; a review on the present status and outlook," *Nanotechnology*, vol. 22, no. 25, Jun. 2011, Art. no. 254002.

[7] K.-L. Lin, T.-H. Hou, J. Shieh, J.-H. Lin, C.-T. Chou, and Y.-J. Lee, "Electrode dependence of filament formation in HfO2 resistive-switching memory," *J. Appl. Phys.*, vol. 109, no. 8, Apr. 2011, Art. no. 084104.

[8] P. Lorenzi, R. Rao, and F. Irrera, "Conductive filament evolution in HfO2 resistive RAM device during constant voltage stress," *Microelectron. Rel.*, vol. 55, nos. 9–10, pp. 1446–1449, Aug. 2015.

[9] Y. Qi, C. Zhao, Y. Fang, Q. Lu, C. Liu, L. Yang, and C. Z. Zhao, "Compliance current effect on switching behavior of hafnium oxide based RRAM," in *Proc. IEEE 24th Int. Symp. Phys. Failure Anal. Integr. Circuits (IPFA)*, Jul. 2017, pp. 1–4.

[10] A. S. Sokolov, Y.-R. Jeon, S. Kim, B. Ku, D. Lim, H. Han, M. G. Chae, J. Lee, B. G. Ha, and C. Choi, "Influence of oxygen vacancies in ALD HfO_{2-x} thin films on non-volatile resistive switching phenomena with a $Ti/HfO_{2-x}/Pt$ structure," *Appl. Surf. Sci.*, vol. 434, pp. 822–830, Mar. 2018.

[11] F. M. Simanjuntak, P. Singh, S. Chandrasekaran, F. J. Lumbantoruan, C.-C. Yang, C.-J. Huang, C.-C. Lin, and T.-Y. Tseng, "Role of nanorods insertion layer in ZnO-based electrochemical metallization memory cell," *Semicond. Sci. Technol.*, vol. 32, no. 12, Dec. 2017, Art. no. 124003.

[12] X. Huang, K.-C. Chang, T.-C. Chang, T.-M. Tsai, C.-C. Shih, R. Zhang, S.-Y. Huang, K.-H. Chen, J.-H. Chen, H.-J. Wang, W.-J. Chen, F. Zhang, C. Chen, and S. M. Sze, "Controllable set voltage in bilayer $ZnO:SiO_2/ZnO_x$ resistance random access memory by oxygen concentration gradient manipulation," *IEEE Electron Device Lett.*, vol. 35, no. 12, pp. 1227–1229, Dec. 2014.

[13] S. Chandrasekaran, F. M. Simanjuntak, and T.-Y. Tseng, "Controlled resistive switching characteristics of ZrO2-based electrochemical metallization memory devices by modifying the thickness of the metal barrier layer," *Jpn. J. Appl. Phys.*, vol. 57, no. 4S, Apr. 2018, Art. no. 04FE10.

[14] C.-L. Lin and T.-Y. Lin, "Superior unipolar resistive switching in stacked $ZrO_x/ZrO_2/ZrO_x$ structure," *AIP Adv.*, vol. 6, no. 3, Mar. 2016, Art. no. 035103.

[15] S. Munjal and N. Khare, "Multilevel resistive and magnetization switching in $Cu/CoFe_2O_4/Pt$ device: Coexistence of ionic and metallic conducting filaments," *Appl. Phys. Lett.*, vol. 113, no. 24, Dec. 2018, Art. no. 243501.

[16] T. H. M. N. Paydavosi, D. D. Lu, W. M. Yang, M. V. Dunga, X. J. Xi, J. He, W. Liu, K. M. Cao, X. Jin, J. J. Ou, M. Chan, A. M. Niknejad, and C. Hu, "BSIM4v4.8.0 MOSFET model user's manual," Dept. Elect. Eng. Comput. Sci., Berkeley, CA, USA, Tech. Rep. BSIM4v4.8.0 Manual, 2013.

[17] J. Reuben, M. Biglari, and D. Fey, "Incorporating variability of resistive RAM in circuit simulations using the Stanford–PKU model," *IEEE Trans. Nanotechnol.*, vol. 19, pp. 508–518, 2020.

[18] D. Panda, P. P. Sahu, and T. Y. Tseng, "A collective study on modeling and simulation of resistive random access memory," *Nanosc. Res. Lett.*, vol. 13, no. 1, p. 8, Dec. 2018.

[19] J. Blasco, P. Jančovič, K. Fröhlich, J. Suñé, and E. Miranda, "Modeling of the switching I-V characteristics in ultrathin (5 nm) atomic layer deposited HfO_2 films using the logistic hysteron," *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron., Mater., Process., Meas., Phenomena*, vol. 33, no. 1, Jan. 2015, Art. no. 01A102.

[20] S. Long, C. Cagli, D. Ielmini, M. Liu, and J. Suñé, "Analysis and modeling of resistive switching statistics," *J. Appl. Phys.*, vol. 111, no. 7, Apr. 2012, Art. no. 074508.

[21] P.-Y. Chen and S. Yu, "Compact modeling of RRAM devices and its applications in 1T1R and 1S1R array design," *IEEE Trans. Electron Devices*, vol. 62, no. 12, pp. 4022–4028, Dec. 2015.

[22] A. J. Lohn, P. R. Mickel, and M. J. Marinella, "Modeling of filamentary resistive memory by concentric cylinders with variable conductivity," *Appl. Phys. Lett.*, vol. 105, no. 18, Nov. 2014, Art. no. 183511.

[23] P. Sheridan, K.-H. Kim, S. Gaba, T. Chang, L. Chen, and W. Lu, "Device and SPICE modeling of RRAM devices," *Nanoscale*, vol. 3, no. 9, pp. 3833–3840, 2011.

[24] G. Gonzalez-Cordero, J. B. Roldan, and F. Jimenez-Molinos, "SPICE simulation of RRAM circuits. A compact modeling perspective," in *Proc. Spanish Conf. Electron Devices (CDE)*, Feb. 2017, pp. 1–4.

[25] M. Li, O. Irsoy, C. Cardie, and H. G. Xing, "Physics-inspired neural networks for efficient device compact modeling," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 2, pp. 44–49, 2016.

[26] D. Chiozzi, M. Bernardoni, N. Delmonte, and P. Cova, "A neural network based approach to simulate electrothermal device interaction in SPICE environment," *IEEE Trans. Power Electron.*, vol. 34, no. 5, pp. 4703–4710, May 2019.

[27] X. Guan, S. Yu, and H.-S.-P. Wong, "A SPICE compact model of metal oxide resistive switching memory with variations," *IEEE Electron Device Lett.*, vol. 33, no. 10, pp. 1405–1407, Oct. 2012.

[28] R. Y. Zhong, X. Xu, E. Klotz, and S. T. Newman, "Intelligent manufacturing in the context of industry 4.0: A review," *Engineering*, vol. 3, no. 5, pp. 616–630, Oct. 2017.

[29] S. Kang, "Joint modeling of classification and regression for improving faulty wafer detection in semiconductor manufacturing," *J. Intell. Manuf.*, vol. 31, no. 2, pp. 319–326, Feb. 2020.

[30] C.-F. Chien, H.-K. Wang, and W. Fu, "Industry 3.5 framework of an advanced intelligent manufacturing system: Case studies from semiconductor intelligent manufacturing," *Manage. Review*, vol. 37, pp. 105–121, Jan. 2018.

[31] D. V. K. A. P. A. R. Karmarkar Saha, G. K. S. Braun, and L. X-W, "Process aware compact model parameter extraction for 45 nm process," presented at the Tech. Proc. NSTI Nanotechnol. Conf. Trade Show, Boston, MA, USA, 2008.

[32] V. M. Y. Mahotin, S. Tirumala, and X.-W. Lin, "Process aware hybrid SPICE models based on TCAD and silicon data," presented at the NSTI-Nanotech, 2007.

[33] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Understanding switching variability and random telegraph noise in resistive RAM," in *IEDM Tech. Dig.*, Dec. 2013, p. 31.

[34] F. M. Puglisi, L. Larcher, A. Padovani, and P. Pavan, "A complete statistical investigation of RTN in HfO2-based RRAM in high resistive state," *IEEE Trans. Electron Devices*, vol. 62, no. 8, pp. 2606–2613, Aug. 2015.

[35] T.-H. Huang, "RRAM SPICE compact model based on machine learning," M.S. thesis, Sci., Electron. Eng., Nat. Chiao Tung Univ., Hsinchu, Taiwan, 2020.

[36] P. Mazumder, S. M. Kang, and R. Waser, "Memristors: Devices, models, and applications," *Proc. IEEE*, vol. 100, no. 6, pp. 1911–1919, Jun. 2012.

[37] P. Bailey, S. Bouaziz, S. Carter, J. Gordon, C. Häne, A. Mordvintsev, J. Valentin, and M. Wicke, "Differentiable graphics with TensorFlow 2.0," in *Proc. ACM SIGGRAPH Courses*, Jul. 2019, pp. 1–211.

[38] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv.org/abs/1603.04467

[39] X.-D. Huang, Y. Li, H.-Y. Li, K.-H. Xue, X. Wang, and X.-S. Miao, "Forming-free, fast, uniform, and high endurance resistive switching from cryogenic to high temperatures in W/AlOx/Al2O3/Pt bilayer memristor," *IEEE Electron Device Lett.*, vol. 41, no. 4, pp. 549–552, Apr. 2020.

[40] X. Liu, S. Ramanathan, A. Longdergan, A. Srivastava, E. Lee, T. E. Seidel, J. T. Barton, D. Pang, and R. G. Gordon, "ALD of hafnium oxide thin films from Tetrakis(ethylmethylamino)hafnium and ozone," *J. Electrochem. Soc.*, vol. 152, no. 3, p. G213, 2005.

[41] E. B. Baum, "On the capabilities of multilayer perceptrons," *J. Complex.*, vol. 4, no. 3, pp. 193–215, Sep. 1988.

[42] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, Mar. 1996.

[43] F. A. Gers, "Learning to forget: Continual prediction with LSTM," in *Proc. 9th Int. Conf. Artif. Neural Network, ICANN*, 1999, pp. 1–19.

[44] S. Hochreiter and J. J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[45] *MATLAB, 11.16.2.1 (R2019b) Ed*, The MathWorks Inc., MathsWorks, Natick, MA, USA, 2019.

[46] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: Experiences from the scikit-learn project," 2013, *arXiv:1309.0238*. [Online]. Available: http://arxiv.org/abs/1309.0238

[47] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," 2018, *arXiv:1809.04356*. [Online]. Available: http://arxiv.org/abs/1809.04356

[48] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, *arXiv:1508.01991*. [Online]. Available: http://arxiv.org/abs/1508.01991

[49] F. E. Harrell, *Regression Modeling Strategies*. New York, NY, USA: Springer-Verlag, 2006.

[50] T.-L. Tsai, H.-Y. Chang, F.-S. Jiang, and T.-Y. Tseng, "Impact of post-oxide deposition annealing on resistive switching in HfO2-based oxide RRAM and conductive-bridge RAM devices," *IEEE Electron Device Lett.*, vol. 36, no. 11, pp. 1146–1148, Nov. 2015.