scientific reports

OPEN

Data-driven RRAM device models using Kriging interpolation

Imtiaz Hossen[1], Mark A. Anders[2], Lin Wang[3] & Gina C. Adam[1]✉

A two-tier Kriging interpolation approach is proposed to model jump tables for resistive switches. Originally developed for mining and geostatistics, its locality of the calculation makes this approach particularly powerful for modeling electronic devices with complex behavior landscape and switching noise, like RRAM. In this paper, a first Kriging model is used to model and predict the mean in the signal, followed up by a second Kriging step used to model the standard deviation of the switching noise. We use 36 synthetic datasets covering a broad range of different mean and standard deviation Gaussian distributions to test the validity of our approach. We also show the applicability to experimental data obtained from TiO_x devices and compare the predicted vs. the experimental test distributions using Kolmogorov–Smirnov and maximum mean discrepancy tests. Our results show that the proposed Kriging approach can predict both the mean and standard deviation in the switching more accurately than typical binning model. Kriging-based jump tables can be used to realistically model the behavior of RRAM and other non-volatile analog device populations and the impact of the weight dispersion in neural network simulations.

Many research advances in neuromorphic systems with emerging devices rely on the availability of accurate device models to quantify their algorithmic performance due to the limited availability and high cost of RRAM/CMOS tape outs. However, RRAM modeling is challenging given its complex multi-physics behavior[1]. Significant progress has been made to determine models relevant for different applications. For example, SPICE models use an underlying physical model and fitting parameters to experimental data to simulate the current vs. voltage characteristics useful for circuit design[2–8]. Some models focus entirely on physical principles[9–13], but might be incomplete if they model only some specific behaviors or do not include the entire set of principles required to capture the multi-physics operation of the device. In general, physical model development for new devices requires assumptions regarding the underlying physical phenomena and the shape of the filaments[14–17], which takes time to uncover and thus can delay the investigation in algorithmic simulations. Atomistic models based on first-order principles can provide significant insight into the device behavior and variability[18–26], but are very computationally intensive and thus typically restricted to device investigations, not large scale neural network simulations.

By comparison, jump table models (or their variant increment plots) are derived only from experimental data and are agnostic to the underlying switching mechanism. They are phenomenological, stochastic lookup tables that define the probability of moving from one weight state to another. Since they are purely derived from data, they can represent a broad range of device non-idealities in network simulations[27,28]. While this jump table methodology can only predict the next conductance state and is not applicable for circuit modeling, it is particularly pertinent for modeling the weight update in neural network simulations. Switching noise leads to weight dispersion, which can impact negatively the accuracy and performance of the overall system[29,30]. Modeling this switching noise is therefore significant for providing a realistic estimate of the training of neuromorphic systems implemented with real devices.

Jump tables have been traditionally derived from experimental data using binning. However, this approach might not be statistically optimal and can introduce unwanted artifacts or exclude key device behaviors. We propose the use of a two-tier Kriging interpolation approach to model RRAM in the jump table framework and perform a statistical investigation into the validity of the proposed approach. Kriging interpolation, also known as Gaussian Process Regression process (GPR) has been previously used to separate signal and noise for analog memory elements for neuromorphic computing[31]. GPR was used to predict the noise-free signal (mean) in order to separate total variability into two parts: device-to-device variability and inherent switching variability for a device. The uncertainties in the readout and in the programming can also be modeled using linear fitting for readout and a sliding window and statistical correction for the programming data respectively[32].

[1]Department of Electrical and Computer Engineering, The George Washington University, Washington, DC 20052, USA. [2]National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. [3]Department of Statistics, The George Washington University, Washington, DC 20052, USA. ✉email: ginaadam@gwu.edu

By comparison, our proposed two-tier approach models both the mean and the standard deviation for the device switching using Kriging interpolation to fully model the jump table. Our contributions are summarized below:

- The Kriging interpolation is functionally more suitable for continuous data, e.g., RRAM measurements, i.e., it does not introduce artifacts and artificial constraints on the data by comparison with binning and nearest neighbor approaches.
- A two-tier methodology was introduced, which models both the mean in the signal (first Kriging model) and the standard deviation of the switching noise (second Kriging model).
- The validity of the approach is tested by using 36 synthetic Gaussian data models. Our proposed Kriging interpolation approach can predict the synthetic distributions of the mean and standard deviation of device behavior with lower root mean square error (RMSE) than the binning and interpolation approach.
- The Kriging vs. binning approaches were also compared on experimental RRAM TiO_x device data using Kolmogorov–Smirnov and Maximum Mean Discrepancy tests. The experimental data was split into a modeling set and a test set to validate the approach. These tests compare whether two data sets are drawn from the same distribution and are critical for ultimately determining the utility of these approaches.

The remainder of the paper is organized as follows. "Device modeling" section introduces background information. "Methods" section III describes our method and "Results" section its evaluation on synthetic and experimental data sets. "Discussion" section concludes the discussion.

Device modeling

Jump table basics. Jump tables are cumulative distribution functions (CDF) of the change in device conductance per voltage (programming) pulse (ΔG/pulse) vs. initial device conductance (G). They can model the stochastic nature of the device programming and the nonideal conductance response ΔG as a function of initial conductance G[27–30]. The variability is represented by the standard deviation (SD) around the mean. An alternative representation in the resistance space (ΔR vs. R) called increment plots, is also possible[32]. However, for the remainder of this paper, the jump table in ΔG vs. G is used since the weight update is naturally mapped to the conductance space in a neural network implemented with real RRAM devices. When training a neural network, voltage pulses are applied to a device to adjust the conductance according to an equivalent desired weight during backpropagation[29]. Two jump tables are needed in simulation—one for potentiation (increase in G) and one for depression (decrease). For a device of initial conductance G, a probability value from 0 to 100% is generated from a uniform distribution using a random number generator. This probability value is used to determine ΔG/pulse from the CDF at conductance G. The device conductance is updated to G + ΔG and the cycle is repeated for follow-up pulses.

Binning and interpolation. In typical jump tables, data binning is used to group individual data into bins of equal width or equal frequency[27]. From this binned data, the mean and standard deviation information can be extracted through a simple Gaussian fit. The mean and standard deviation values are then linearly interpolated across the G bins to obtain the jump table model.

However, the appropriate number of bins for a given dataset may be difficult to optimize[33–35] and prone to artifacts as Fig. 1 shows equal width binning for 3 representative synthetic datasets. The constant model requires a low G bin count for lowest RMSE, while the random one requires a high count. The underlying distribution of experimental data is unknown. Edge effects, empty bins and excessive smoothing due to the linear interpolation are other visible issues. These challenges motivate the search for a statistically sound alternative for jump table modeling, as well as methods to compare the predicted distribution against the measured dataset.

Kriging/GPR modeling. The theoretical background for data modeling and prediction using Gaussian processes has been under development since the 1940's[36]. In geostatistics, it is known as *Kriging* and applied particularly for two- and three- dimensional input variables[37–39]. However, it is also known more generally as Gaussian Process Regression (GPR) and applied for prediction purposes to a broad range of applications, from computer experiments[40] to optimized machine learning algorithms[41].

In Kriging/GPR, any collection of the responses of the data are assumed to be jointly Gaussian distributed with a mean function μ(x) defining the mean of the response y at any point x in the input space and a covariance function defining the covariance between the responses at any two points. Kriging predicts a function via a weighted average of the observed points using a set of considered functions as defined by the kernel of choice[42]. This technique has been used in RF device and analog circuit modeling[43–45] as well as to extract mean behavior in phase change memory devices[31].

While several types of Kriging modeling techniques exist, we utilize the ordinary Kriging in this paper, where the trend is constant. For interpolation situations, prior work[40,46,47] shows that the constant trend plus a stationary Gaussian process in ordinary kriging can model complex systems as well as universal kriging where the trend is dependent on the variable. This paper assumes modeling only for interpolation purposes since typically the minimum and maximum conductance range for RRAM is fixed in neural network simulation. Since ordinary Kriging is computationally simpler, it is recommended to use in interpolation situations, such as our case.

Consider a RRAM measurement of n sampled points $(G_1, \Delta G_1), \ldots, (G_n, \Delta G_n)$, where for $i = 1, \ldots, n$. G_i is the initial device conductance and ΔG_i is the corresponding conductance change. The ordinary kriging model with a noise term is described as

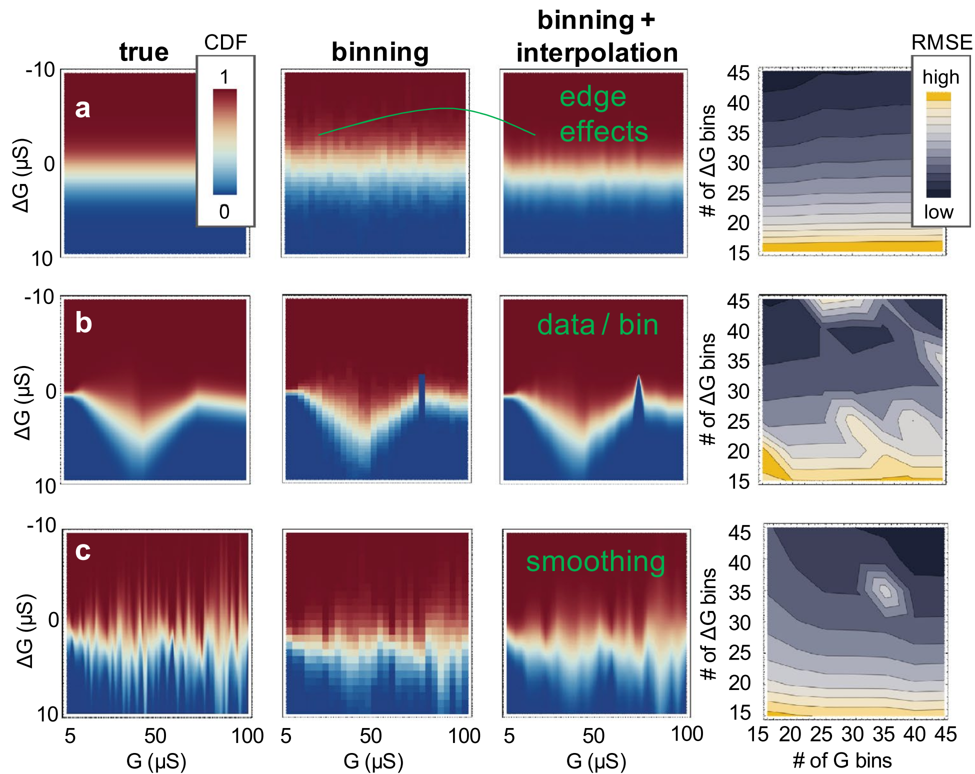

Figure 1. Binning challenges shown on synthetic distributions (**a**) with constant mean and SD; (**b**) an approximation of the model from ref[27]; (**c**) randomly generated mean and SD.

$$\Delta G_i = \mu + Z(G_i) + \epsilon_i, \tag{1}$$

where μ is the constant trend, $\epsilon_i \sim N(0, \tau^2)$[48], and $Z(\cdot)$ is a stationary Gaussian process with zero mean and covariance function governed by a kernel k:

$$k(G_i, G_j) = Cov(Z(G_i), Z(G_j)) = \sigma^2 r(G_i, G_j). \tag{2}$$

Here σ^2 is the process variance and $r(G_i, G_j)$ is the correlation between the conductance change corresponding to two sampling points G_i and G_j for $i, j = 1, \ldots, n$. The correlation function is typically chosen from a family of kernels functions, based on the application at hand. We chose an exponential kernel[42] since it provided us with the lowest RMSE of the predicted mean. By the model in (1), we can obtain the parameter estimation $\hat{\mu} = \left(1^T C^{-1} 1\right)^{-1} 1^T C^{-1} \mathbf{\Delta G}$, where 1 is a column vector of unity, $C = K + \tau^2 I_n$, K is the covariance matrix with entries $k(G_i, G_j)$, I_n is an identity matrix of order n, and $\mathbf{\Delta G}$ is the vector of n sampled values for conductance change. Then the prediction of conductance change at a new value of conductance G_0 is given by

$$\Delta \widehat{G}_0 = \hat{\mu} + R^T C^{-1} (\mathbf{\Delta G} - \hat{\mu}) \tag{3}$$

where $R = (k(G_0, G_1), \ldots, k(G_0, G_n))$ is the row vector measuring the covariance between conductance change at G_0 and at the sampled points.

Methods

Datasets. This work demonstrates the proposed two-tier Kriging modeling vs. Binning approaches on two types of data (Fig. 2). Experimental data from a 10 nm edge device with 2.5 nm Al_2O_3/15 nm TiO_x/5 nm Ti/30 nm Pt, was obtained utilizing two fast SMUs, a probe station, and text-based programming. The forming was done with monotonically increasing voltage pulses until the device was put into the high conductance state. Current–voltage measurements were made with a semiconductor parameter analyzer and are shown in Fig. 2a. The jump table data showed in Fig. 2b, c used for the modeling were collected after forming and then cycling the device 30 times for set and reset. The data collection algorithm is as follows. Before the algorithm starts, the device conductance is measured once. (1) The device is programmed to a random conductance value within a given range. (2) A write pulse is applied. The write pulse voltage is chosen randomly from a list. In this case, the list consists of ± 1.35 V, ± 1.5 V, ± 1.65 V and ± 1.8 V. All write pulses had a base voltage of 0 V, 500 ns high time, and 100 ns rise and fall time. (3) A subsequent read pulse of 100 mV is applied about 100 µs after the write pulse (during this time, the device is held at 0 V). The read pulse is held for about 20 ms and the current is measured and averaged for that time. G and ΔG values are recorded on each write pulse. 4) Steps 2 and 3 are repeated until

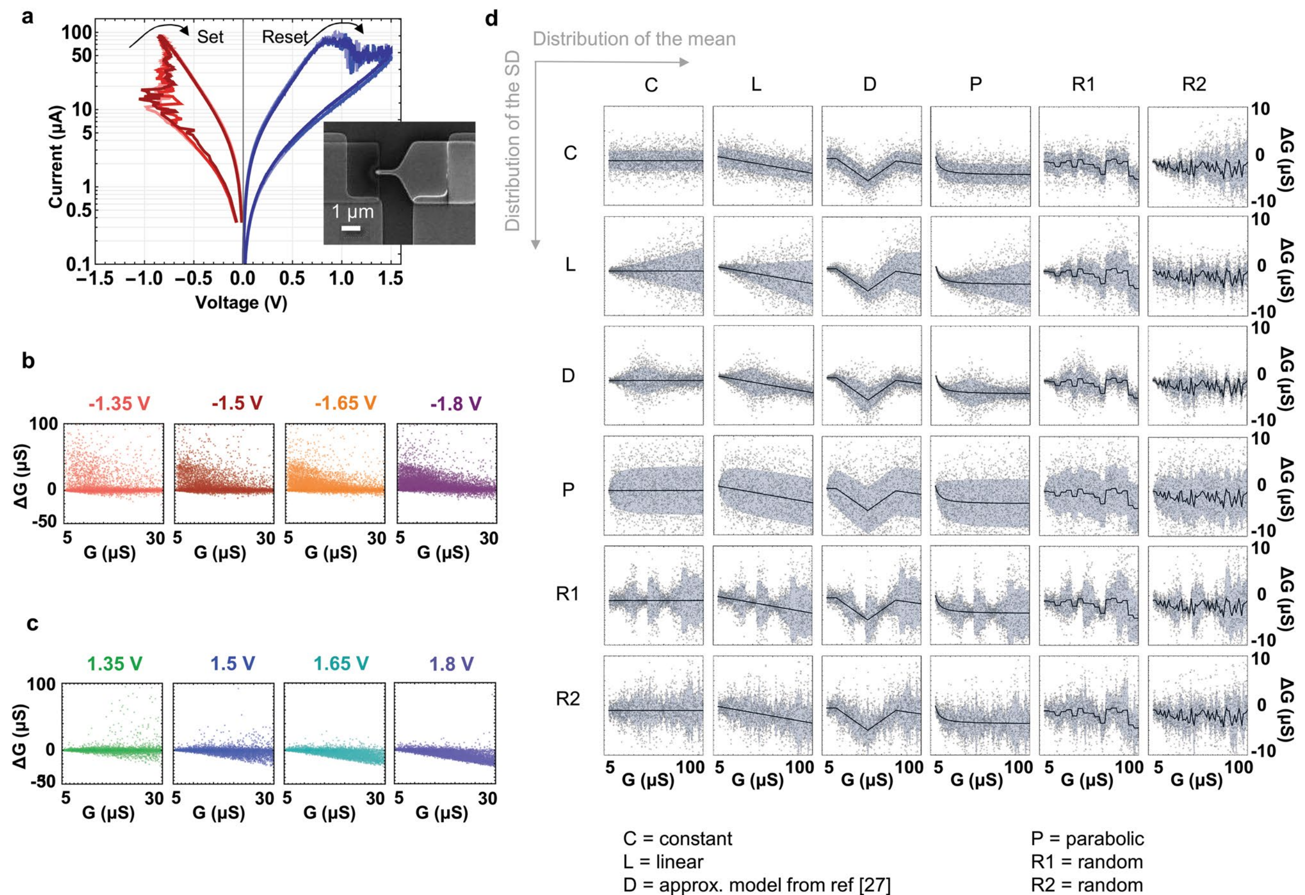

Figure 2. Device data. (**a**) RRAM used for experimental data gathering with current–voltage characteristics and SEM image; and ΔG vs. G data at different pulse amplitudes for (**b**) set and (**c**) reset. (**d**) Synthetic device data encompassing a wide range of distributions of mean and SD from constant to random.

either all pulses in the list have been applied or the device conductance exceeds some defined limit. 5) Return to step 1 and repeat the cycle. This algorithm was developed as it has some benefits over some other algorithms. For example, cyclically setting and resetting a device with monotonically increasing or decreasing write pulse voltage steps typically results in a sparse data set for high write voltages as the device typically will set or set before the higher voltages are applied. The algorithm described in this paper results in data sets of roughly equal number of points (~10,000) for each write pulse voltage. This work is based on experimental data obtained at a fixed reading voltage of 100 mV. The reason for that is because, in neuromorphic circuits, the read voltage is typically fixed for synaptic device programming[32]. However, RRAM devices present non-linear characteristics with the conductance change profile dependent on other factors beyond the initial conductance G. In the future, additional experimental data can be gathered, e.g. ΔG vs. G vs. write pulse width vs. write pulse amplitude vs. readout voltage, etc. in order to devise a multi-dimensional model that can support other programming schemes.

However, it is important to point out that the proposed modeling approach and testing methodology can be applied to a broad range of RRAM devices based on other materials and with different (G, ΔG) switching properties. Jump table/increment plots modeling has already been applied to some phase change memory devices[27,29] and to some TaO_x RRAM devices[28,30] and to TiO_x.

RRAM devices[32]. Kriging GPR was applied for signal vs. noise extraction in HfO_x RRAM devices and Ge_2Sb_2Te – based phase change devices[31] as well as RF devices[43] and nano CMOS thermal sensors[44]. Our proposed methodology can not only be applied for a more realistic modeling of RRAM jump tables, but it can also be applied beyond the ΔG vs. G switching of RRAM devices for other emerging devices where variability significantly affects behavior, e.g. spin torque transfer RAM[49], Ferroelectric RAM[50], conductive-bridge RAM[51], two-dimensional materials based devices[52], electrolyte transistor based synapses[53], etc. and to other novel devices yet to be developed.

In order to be inclusive of the potential switching characteristics of such devices, we propose the use of synthetic data distributions with known means and standard deviations covering the full spectrum of statistics from constant to random distributions. In this work, 36 synthetic distributions with known means and standard deviations shown in Fig. 2d are used, exemplifying combinations of Gaussian distributions with 6 variable means and 6 variable standard deviation profiles for reset (depression) switching. For set (potentiation), positive mean values would be used instead. These profiles are as follows: constant (labeled as C); linear (L); piecewise linear (labeled D and inspired by IBM's model from[27]); parabolic (P) and two random models (R1 and R2). For

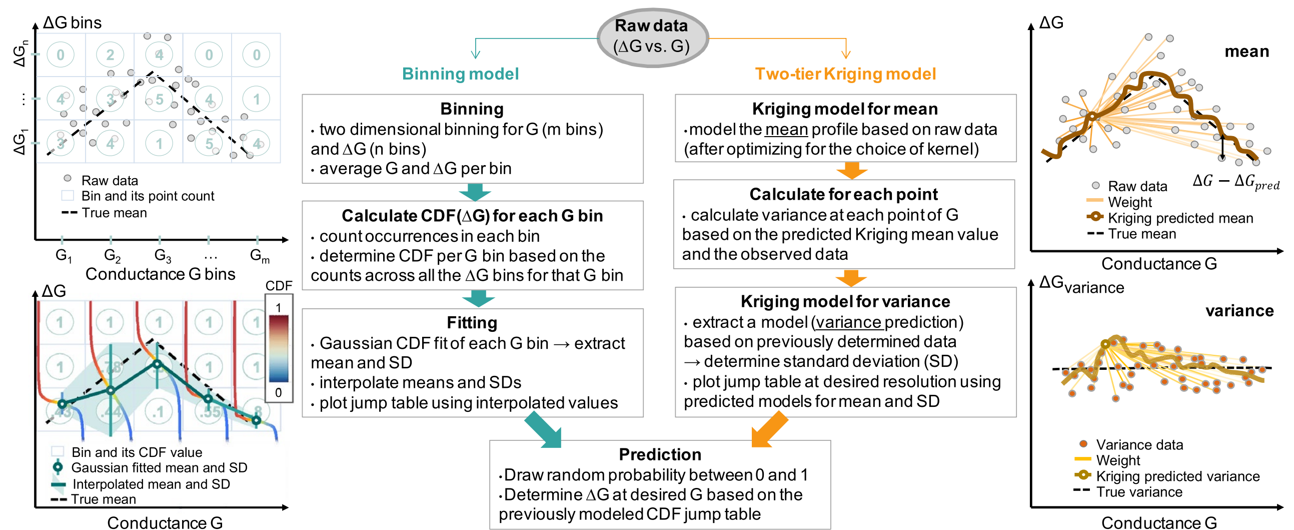

Figure 3. Algorithmic comparison. A binning modeling approach vs. the proposed two-tier Kriging modeling approach.

example, the constant model assumes a value of the mean ΔG of −1 µS and of the standard deviation of 2 µS. At the other extreme, the random R2 model has randomly generated points between −4.5 and −0.5 µS for the mean profile and between 0.5 and 7 µS for the standard deviation profile. The synthetic (G, ΔG) data points are drawn from these Gaussian distributions, using a random number generator for a desired number of points. The detailed algorithm is listed below.

Our work explores a broad spectrum of synthetic device data, from constant to random profiles which goes well beyond actual devices that have been explored in existing literature. However, it is important to note that some of these profile shapes we discussed here seem to have been already observed experimentally. For example, the Ag:a-Si conductive bridge RAM modeled in the NeuroSim neural network simulator[54,55] has a linear mean profile for both the set and the reset switching operations. The phase change memory device from IBM seems to have a piecewise linear behavior for the mean and standard deviation (see Fig. 17 in[27]). The TaOx RRAM devices modeled in[30] seems to have a parabolic mean profile and a small constant standard deviation profile for the set, and somewhat random profiles for reset (see Fig. 12 in[30]).

While the focus of our work is on RRAM device modeling, the broader idea behind our synthesized data is to study how robust Kriging interpolation would be as a general modeling technique for various types of unconventional devices. Since the true profile of the mean and standard deviation is unknown for experimental data, these synthetic models were chosen to test the performance of the modeling approach across the entire range from constant to random. A constant profile should be easiest to predict, while a random profile the hardest. The chosen modeling approach should perform the best across the entire spectrum of profiles.

Algorithm 1. Synthetic data generation

```
1:   procedure generate_synthetic_data (in int n, int N, list<float> Gn, list<float> μn, list<float> σn; out
     list<tuple<float, float>> synthetic_dataset)
2:     Gmin, Gmax = Gn.first(), Gn.last()
3:     f = linear_interpolate(x = Gn, y = μn)
4:     g = linear_interpolate(x = Gn, y = σn)
5:     for i in range [1, N] do
6:       Gs = random_float(min = Gmin, max = Gmax)
7:       μs= f(Gs), σs= g(Gs)
8:       PDF(Gs) = 1/(σ√(2π)) e^(-1/2((Gs-μs)/σs)^2)
9:       gaussian_distribution = PDF(Gs)
10:      ΔGs = random variate from gaussian_distribution(Gs)
11:      # additional limitations might be imposed on ΔGs to ensure physical realism
12:      synthetic_dataset[i] = (Gs, ΔGs)
13:    end for
14:  end procedure
```

Modeling and testing approach. Figure 3 shows the algorithmic comparison between the typical binning modeling and the proposed two-tier Kriging modeling approach. As a first step, a one-dimensional ordinary Kriging is used to predict the mean at each G based on the raw data. The variance of the sampled data to the Kriging mean is determined for each point and used to extract the variance $(\Delta G - \Delta G_{Krig})^2$ at each point of G in the dataset. The second step consists of another one-dimensional ordinary Kriging, used to predict the standard deviation based on the previously calculated variance data. This modeling was performed with the DiceKriging package[56] from R. By comparison, data binning is used to group individual data into regions of the

designated size and extract mean and standard deviation information through a simple Gaussian fit. Based on Fig. 1, 30 bins for G and 40 bins for ΔG are used for binning in the remainder of the paper.

For the synthetic datasets with known mean and standard deviation, the root mean square error (RMSE) is calculated for the predicted mean and standard deviation, respectively for both approaches for various sample sizes. For the experimental data, the true underlying mean and standard deviation profiles are unknown. Therefore, it is not possible to calculate the mean of RMSE and SD from the experimental data. We used the Kolmogorov–Smirnov (K–S) test and the maximum mean discrepancy (MMD) test to determine the goodness of fit between the observed distribution and the predicted distribution. The K-S test is a non-parametric test that can compare a sample drawn from an unknown distribution to a known reference continuous distribution[57,58] or it can estimate the maximum distance $D_n = \sup_x |F_P(x) - F_Q(x)|$ between two empirical cumulative distributions F in our case of the predicted dataset based on the modeled experimental data P vs. the experimental test dataset Q. Since the K-S test tends to be less sensitive at the tails of the distributions among other limitations[59], we utilized MMD as a complementary measure. MMD[60] represents the distance between empirical distributions as the distance between their mean embeddings $\mathbb{E}$ determined via feature maps φ (or reproducing kernel Hilbert space $\mathcal{H}$) $MMD(P(x), Q(y)) = \sup_{\|\varphi\|_{\mathcal{H}} \leq 1} |\mathbb{E}_{\boldsymbol{P}}[\varphi(x)] - \mathbb{E}_{\boldsymbol{Q}}[\varphi(y)]|$. The kernel determines the type of distance computed. Some, e.g. Gaussian kernel, lead to the MMD distance being zero only if the two datasets are drawn from the same underlying empirical distribution. MMD provides the maximum distance across the test kernels. Two R packages[61,62] are used.

Results

Figure 4 shows the impact of the sample size on the RMSE of the mean and standard deviation prediction of the two approaches on two synthetic models—constant mean and constant standard deviation (C) and random mean/random standard deviation (R2). The insets are showing worst case scenario for the Kriging prediction of the mean and standard deviation profiles. Overall, Kriging predicts the mean and the standard deviation profiles better than the Binning approach, particularly at higher sample count. Kriging has difficulty robustly predicting the desired profile with a low sample count as shown by the large error bars for the estimates with less than 2000 points. This is most evident in the mean/standard deviation predictions of the random model for 500 points in the sampled dataset (Fig. 4c, d). For that case, the Kriging model is predicting them wrongly as (close to) constant although the true mean and standard deviation profile have random behavior (dashed black line). However, at larger sample sizes, the Kriging interpolation consistently estimates the mean and the standard deviation profile better than the binning approach. For example, for the random model at 5000 data points, the Kriging RMSE for the mean prediction is 0.345 ± 0.033 µS vs. for the binning 0.76 ± 0.021 µS and the RMSE for the standard deviation Kriging prediction is 0.326 ± 0.024 µS vs. the RMSE of 0.072 ± 0.06 µS for the Binning. This indicates the existence of a minimum number of points required to generate an accurate model for high randomness in the mean and SD. Large datasets are desirable, but difficult to obtain in practice and computationally intensive for the Kriging approach. Our results indicate that samples with at least 2000 points should be used.

Figure 5 shows the results across all 36 synthetic data models. As the complexity of the mean/standard deviation profile increases, the RMSE increases as expected for both binning and Kriging. The lowest errors are seen for the constant, linear and the device model from[27]. The models with parabolic mean and/or standard deviation distributions also seem to perform poorly. This may be because both approaches are based on linear interpolation, weighted or not. Kriging interpolation consistently performs the best for the mean prediction, but there are a few instances when it underperforms for the parabolic and random standard deviation cases.

Both methods were applied to the experimental data of TiO_x RRAM devices for various pulse amplitudes. A random sub-sample of 4000 data points (modeling set) was used to generate the Binning and Kriging models for set and reset respectively. Another non-overlapping 4000 points sub-sample (test set) was compared with the predicted models. The K–S and MMD tests were used to provide a quantitative estimate of the difference between the predicted and experimental test sets, as reference. The experimental data (as shown in Fig. 2) was randomly sub-sampled for the modeling and testing sets and modeled using this methodology 20 times to provide statistically significant and computationally accessible results. Both the set and reset results are shown in Fig. 6. The lower value of K–S and MMD indicate the better model fit to the experimental test data. On average, the K–S and MMD values of Kriging reconstructed data are less than those of binning + interpolation based reconstructed data. Both binning and Kriging models give a better prediction for reset data (Fig. 6b) than the set data (Fig. 6a). Kriging K–S values for reset.

are almost half in comparison with the K–S values for set. However, the standard deviation of K–S and MMD values for Kriging is lower than the binning + interpolation which can indicate that the reconstructed data of the Kriging model is more consistent than the reconstructed data of binning + interpolation method.

It is also important to notice that for the set data, there are 18 cases out of 80 for which the binning + interpolation method leads to slightly better results than the Kriging model, particularly at lower pulse voltage amplitudes. This might be because the standard deviation prediction is likely affected since our models assume a Gaussian distribution whereas the set data seems to be intrinsically skewed. To test this hypothesis, Fig. 6c, d includes reconstructed data vs. experimental data for set and reset respectively at pulse amplitudes ± 1.8 V. As shown in Fig. 6c, both binning + interpolation and Kriging models have some difficulty reproducing the overall shape of the set data due to skewness. However, the binning underestimates the standard deviation by a larger margin than the Kriging, as seen in the two jump tables (insets). Based on these quantitative and qualitative results, the reconstructed data from the Kriging model seems to cover better the test experimental data.

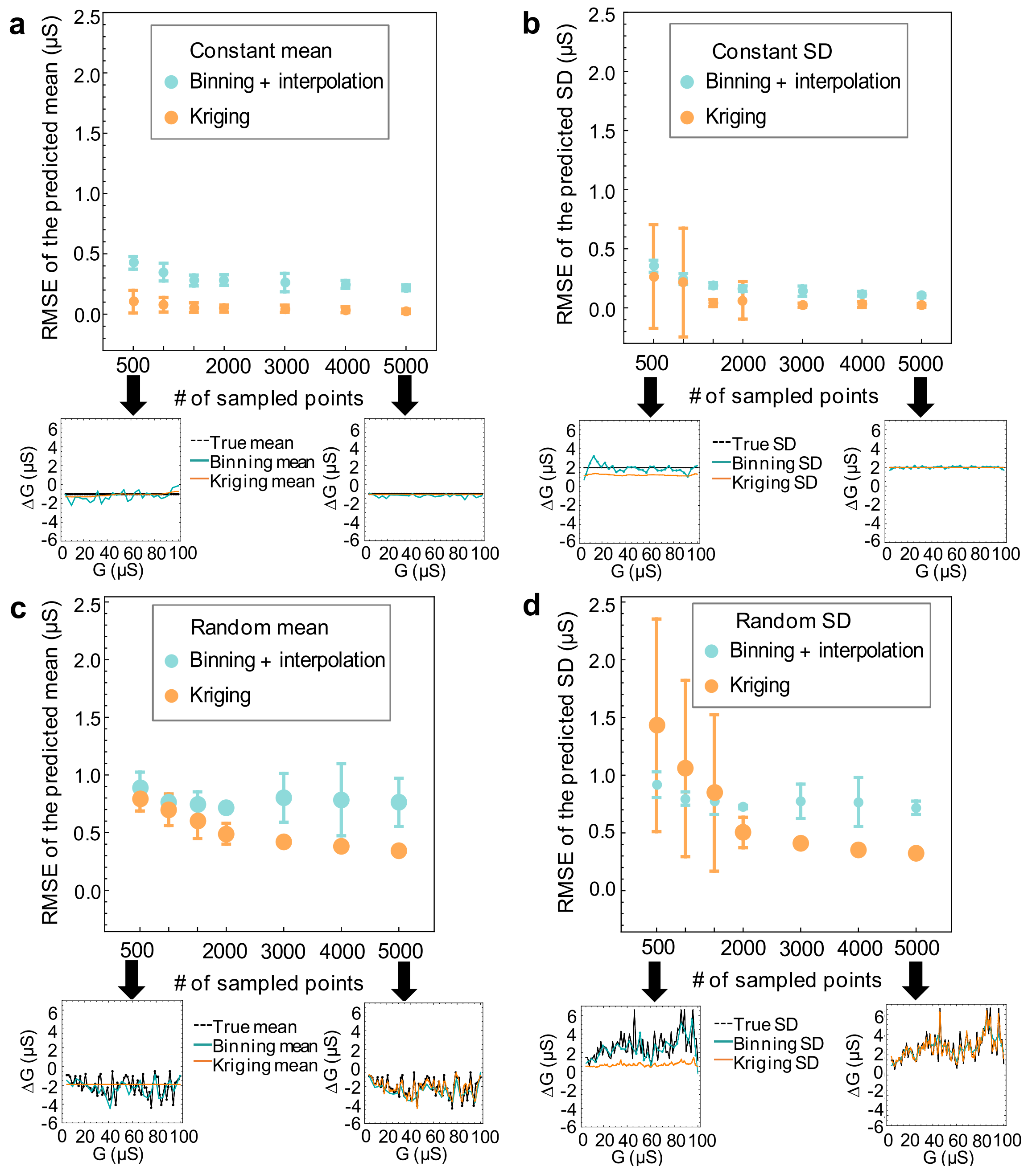

Figure 4. Prediction RMSE of Binning and Kriging model vs. number of sampled points in the dataset. (**a**, **b**) prediction of the mean and SD for the constant distribution model and (**c**, **d**) for the random distribution model R2. # of iterations = 20 and the error bar indicates one standard deviation.

Discussion

The applicability of the proposed modeling approach has to be discussed in the broader context of RRAM device modeling. Three broad device model categories can be considered: physics-based models, semi-empirical device models and empirical (data-based) models, in accordance with existing categorizations in the field at large[63,64]. The proposed methodology belongs in the last category of data-driven models and its advantages and limitations are highlighted in Fig. 7. Physically-based models are supported by first-order principles and fundamental calculations. They can provide accurate representations via atomistic simulations or closed form solutions of coupled nonlinear partial differential equations. This can significantly improve the understanding of the internal physical phenomena and support robust device design[65]. However, these approaches are typically computationally intensive and might be incomplete and unable to fully explain the experimental reality. By comparison, semi-empirical (or physically-based compact) models are particularly useful for circuit design. These compact models are typically based on a simplified physical model with fitting parameters to make the model conform to experimental current vs. voltage data. Since only a few equations have to be solved at each step, the computational efficiency is

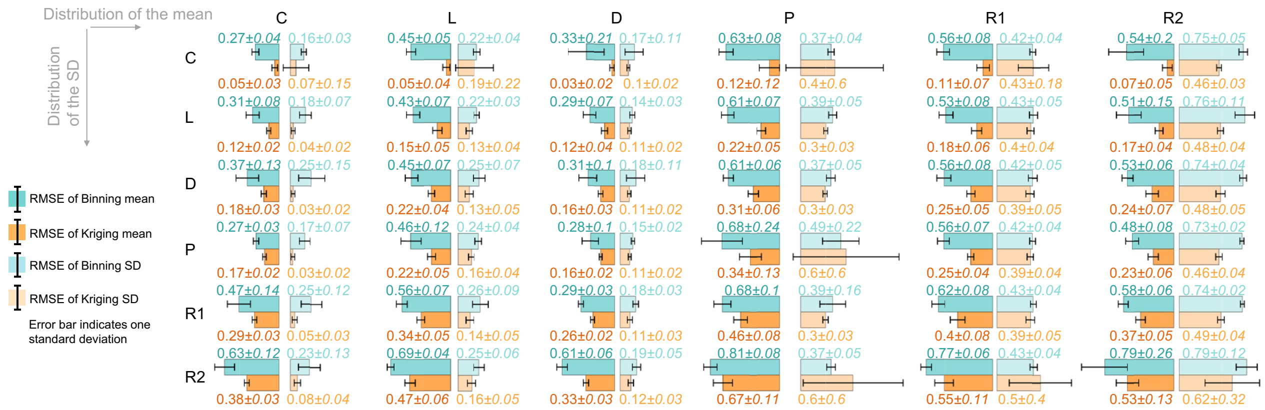

Figure 5. Comparison of mean and SD prediction results for the 36 synthetic models. The Kriging interpolation for the mean is consistently better than binning approach. # Points = 2000, # G bins = 30, # ΔG bins = 40, # of iterations = 20.

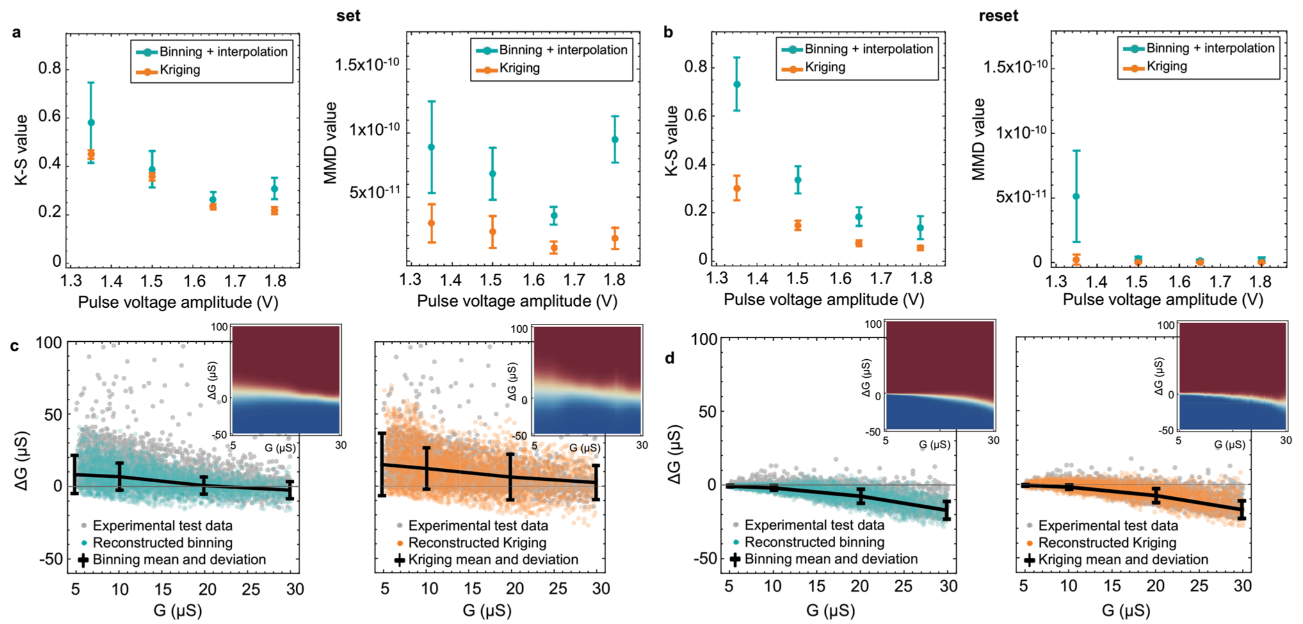

Figure 6. Results for experimental test data. K–S test and MMD metrics of test data vs. reconstructed data for four pulse amplitudes (**a**) set and (**b**) reset. # Points = 4000 experimental/reconstructed data points, #iterations = 20. Reconstructed data based on Binning + interpolation and Kriging modeling for (**c**) set (**d**) reset at pulse amplitudes ± 1.8 V.

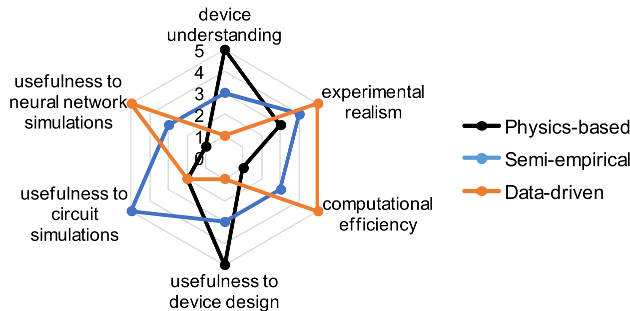

Figure 7. Comparison of key considerations for the major categories of device modeling approaches. The proposed two-tier Kriging approach can be classified as a data-driven device modeling approach.

improved and with careful fitting, these models are particularly useful in circuit simulators, e.g. SPICE[66]. However, for large neural network simulations that require constant device update modeling, these compact models can be insufficient. Large-scale models can require thousands to millions of individual conductance updates per training epoch depending on the batch size and network size, which requires accessing the device model the same number of times. In these situations, semi-empirical models can be too computationally intensive[45]. Data-based models, such as the one proposed, aim to provide a fast approach to derive realistic models of novel devices that might otherwise be difficult to represent by physical functions.

Since data-based models are generated directly from measured data without knowledge of the device physics, the fitting methodology used is entirely responsible for capturing the features of the data. In this work, we have showed that suitable statistical methods, e.g. Kriging/GPR should be used for continuous RRAM data since they tend to perform better than binning approaches. Moreover, the validity of the approach has to be tested using appropriate methods, e.g. on synthetic datasets with known distributions, using nonparametric tests for sample comparison, etc. It is important to point out that our two-tier modeling approach based on ordinary Kriging is suitable for interpolation, in the range between minimum and maximum conductance of the experimental data, as needed for neural network simulations using these device models. For extrapolation situations, the functional type assumed for the trend model is very important, so if extrapolation needs to be considered, more sophisticated Kriging (GPR) methods such as Universal Kriging have to be considered.

This work also highlights a potential limitation of the proposed approach for skewed data and the need to expand to more general non-Gaussian interpolation methods. Existing literature models variability as a Gaussian distribution, for example the jump tables by Burr et al. seem to also have this assumption[27,29]. The uncertainty model of switching noise by Stathopoulos et al. also assumes Gaussian data distributions for its RRAM devices[32]. Gong et al. established a practical method to separate the signal and noise components of analog NVM elements based on the Gaussian regression process[31]. Perez et al. also assumes a Gaussian distribution for the readout current vs. read voltage of RRAM devices[67]. However, our results show that the skewness can be a challenge, particularly for set operation without compliance. The skew can have many potential origins, an obvious example being the physical bounds on the maximum amount that the conductance can be changed, e.g. the conductance cannot go above the ceiling set by the series parasitic resistance. Or if the initial G is 35 μS, the reset ΔG cannot be < -35 μS, which we considered in how we reconstructed the Binning/Kriging sets. Non-Gaussian methods that support the incorporation of physical constraints would be desirable for the modeling of non-volatile memory devices.

Another limitation is the fact that current–voltage characteristics are not modeled and cannot be reconstructed as part of this work. The lack of physical insight into the actual device can also be considered a drawback. However, it is important to note that GPR has been previously used to predict the mean signal and separate device-to-device variability and cycle-to-cycle variability[31]. The series resistance could also be potentially estimated based on the lower data bound in reset data. These can be critical insights drawn entirely from measured data and useful for device optimization. However, if voltage-behavior is incorporated in the model, it can be useful for fitting important parameters for compact models, e.g. the evolution curve of read resistance vs. pulse amplitude typically fitted with ad-hoc functions and fitting parameters[1]. The resulting model of read resistance change could model the dynamic behavior of the device and together with a physically inspired static model could reproduce current–voltage RRAM characteristics. In addition, the proposed technique has potential to also be applied in contexts relevant to SPICE modeling where the focus is on current vs. voltage characteristics for circuit simulation of inference operations. For example, a behavioral model was recently proposed for multi-state HfO_2-based memristors by modeling CDFs of the readout currents at five conductance states via simulation in LTspice[67]. In that work, experimental CDFs of the measured readout currents were fitted with a Gaussian distribution, while the means and standard deviations as a function of read voltage were linearly fitted. According to the authors, the specific implementation of this model is a SPICE sub-circuit with two terminal connections and a parameter for the selection of the device conductance state. The model was successfully used in circuit simulations of neural networks[68]. Kriging modeling could be potentially applied to model the means and standard deviations of the readout current as a function of the read voltage instead of the simple linear fit. Moreover, the Kriging/GPR can be expanded to multi-dimensional data that include pulse width, pulse amplitude, read out voltage, temperature, etc. thus providing a comprehensive model capturing the device population behavior. This has direct applicability to identifying the most optimal programming scheme for the devices which can reduce the experimental variability observed[68], and thus more realistically modeling RRAM-implemented weight updates in neural network simulations.

Nevertheless despite the targeted applicability to memristive weight update modeling, it is important to point out that the proposed modeling approach and testing methodology can be extended to modeling a broad range of novel non-volatile memory devices in a prompt fashion conducive to large-scale neuromorphic simulations and experimental demonstrators. As soon as the (G, ΔG) switching properties are measured, the two-tier Kriging model can be obtained and incorporated in a simulator, e.g. NeuroSim[55]. The proposed methodology for testing can be utilized to determine the goodness of fit between the observed experimental distribution and the predicted distribution based on the model. A small distance between these two distributions is desired to ensure that the simulation results are indicative of potential performance results on a hardware prototype.

Conclusions

This paper proposes a two-tier Kriging/GPR approach for modeling jump tables of RRAM devices and tests it comparatively to the traditional binning approach using a broad range of synthetic Gaussian datasets with known mean and standard deviation profiles, as well as experimental data. Binning introduces artifacts and artificial constraints to continuous data. The Kriging modeling can determine the data trends more reliably providing a better prediction than the binning for all the investigated mean models and almost all standard deviation models.

The work also demonstrates the use of statistical tests e.g. K-S and MMD, to determine how far the reconstructed points based on the proposed models are from the underlying experimental data. This work also highlights that for skewed experimental data, the Kriging model can fail when the assumption of Gaussian distribution is no longer valid.

Future work will expand to non-Gaussian methods that consider physical constraints to better predict the experimental data and to generate reliable statistical models of analog device dynamics for use in neuromorphic simulations. The performance advantage is expected to be larger for higher dimensionalities of the parameter space which will also be explored in the future. Multivariate distribution-free two sample tests will be essential for determining the suitability of such methods.

Received: 21 December 2021; Accepted: 16 March 2022
Published online: 08 April 2022

References

1. Merrikh Bayat, F., Hoskins, B. & Strukov, D. B. Phenomenological modeling of memristive devices. *Appl. Phys. A* **118**, 779–786 (2015).
2. Biolek, Z., Biolek, D. & Biolková, V. SPICE model of memristor with nonlinear dopant drift. *Radioengineering* **18**, 6 (2009).
3. Yakopcic, C., Taha, T. M., Subramanyam, G. & Pino, R. E. Memristor SPICE model and crossbar simulation based on devices with nanosecond switching time. In *The 2013 International Joint Conference on Neural Networks (IJCNN)* 1–7 (2013). doi:https://doi.org/10.1109/IJCNN.2013.6706773.
4. Jiang, Z. *et al.* A compact model for metal–oxide resistive random access memory with experiment verification. *IEEE Trans. Electron Dev.* **63**, 1884–1892 (2016).
5. Puglisi, F. M., Pacchioni, L., Zagni, N. & Pavan, P. Energy-efficient logic-in-memory 1-bit full adder enabled by a physics-based RRAM compact model. in *2018 48th European Solid-State Device Research Conference (ESSDERC)* 50–53 (2018). doi:https://doi.org/10.1109/ESSDERC.2018.8486886.
6. Bengel, C. *et al.* Variability-aware modeling of filamentary oxide-based bipolar resistive switching cells using SPICE level compact models. *IEEE Trans. Circuits Syst. I: Regul. Pap.* **67**, 4618–4630 (2020).
7. Ambrogio, S., Balatti, S., Gilmer, D. C. & Ielmini, D. Analytical modeling of oxide-based bipolar resistive memories and complementary resistive switches. *IEEE Trans. Electron Dev.* **61**, 2378–2386 (2014).
8. Messaris, I. *et al.* A data-driven verilog—A ReRAM model. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **37**, 3151–3162 (2018).
9. Gao, B., Kang, J., Liu, L., Liu, X. & Yu, B. A physical model for bipolar oxide-based resistive switching memory based on ion-transport-recombination effect. *Appl. Phys. Lett.* **98**, 232108 (2011).
10. Strukov, D. B., Alibart, F. & Stanley Williams, R. Thermophoresis/diffusion as a plausible mechanism for unipolar resistive switching in metal–oxide–metal memristors. *Appl. Phys. A* **107**, 509–518 (2012).
11. Kim, S., Choi, S., Lee, J. & Lu, W. D. Tuning resistive switching characteristics of tantalum oxide memristors through Si doping. *ACS Nano* **8**, 10262–10269 (2014).
12. La Torre, C., Zurhelle, A. F., Breuer, T., Waser, R. & Menzel, S. Compact modeling of complementary switching in oxide-based ReRAM devices. *IEEE Trans. Electron Dev.* **66**, 1268–1275 (2019).
13. Kim, S. *et al.* Physical electro-thermal model of resistive switching in bi-layered resistance-change memory. *Sci. Rep.* **3**, 1680 (2013).
14. Bersuker, G. *et al.* Metal oxide resistive memory switching mechanism based on conductive filament properties. *J. Appl. Phys.* **110**, 124518 (2011).
15. Ielmini, D., Nardi, F. & Cagli, C. Physical models of size-dependent nanofilament formation and rupture in NiO resistive switching memories. *Nanotechnology* **22**, 254022 (2011).
16. Larentis, S., Nardi, F., Balatti, S., Gilmer, D. C. & Ielmini, D. Resistive switching by voltage-driven ion migration in bipolar RRAM—Part II: Modeling. *IEEE Trans. Electron Dev.* **59**, 2468–2475 (2012).
17. González-Cordero, G. *et al.* A new compact model for bipolar RRAMs based on truncated-cone conductive filaments—A Verilog-A approach. *Semicond. Sci. Technol.* **31**, 115013 (2016).
18. Butcher, B. *et al.* Connecting the physical and electrical properties of Hafnia-based RRAM. In *2013 IEEE International Electron Devices Meeting* 22.2.1–22.2.4 (2013). doi:https://doi.org/10.1109/IEDM.2013.6724682.
19. Guan, X., Yu, S. & Wong, H.-S.P. On the switching parameter variation of metal-oxide RRAM—Part I: Physical modeling and simulation methodology. *IEEE Trans. Electron Dev.* **59**, 1172–1182 (2012).
20. Pan, F., Yin, S. & Subramanian, V. A detailed study of the forming stage of an electrochemical resistive switching memory by KMC simulation. *IEEE Electron Dev. Lett.* **32**, 949–951 (2011).
21. Menzel, S. Comprehensive modeling of electrochemical metallization memory cells. *J. Comput. Electron.* **16**, 1017–1037 (2017).
22. Aldana, S. *et al.* Resistive switching in HfO_2 based valence change memories, a comprehensive 3D kinetic Monte Carlo approach. *J. Phys. D: Appl. Phys.* **53**, 225106 (2020).
23. Abbaspour, E., Menzel, S. & Jungemann, C. Studying the switching variability in redox-based resistive switching devices. *J. Comput. Electron.* **19**, 1426–1432 (2020).
24. Abbaspour, E., Menzel, S., Hardtdegen, A., Hoffmann-Eifert, S. & Jungemann, C. KMC simulation of the electroforming, set and reset processes in redox-based resistive switching devices. *IEEE Trans. Nanotechnol.* **17**, 1181–1188 (2018).
25. Padovani, A., Larcher, L., Pirrotta, O., Vandelli, L. & Bersuker, G. Microscopic modeling of HfOx RRAM operations: From forming to switching. *IEEE Trans. Electron Dev.* **62**, 1998–2006 (2015).
26. Padovani, A., Gao, D. Z., Shluger, A. L. & Larcher, L. A microscopic mechanism of dielectric breakdown in SiO_2 films: An insight from multi-scale modeling. *J. Appl. Phys.* **121**, 155101 (2017).
27. Burr, G. W. *et al.* Experimental demonstration and tolerancing of a large-scale neural network (165 000 synapses) using phase-change memory as the synaptic weight element. *IEEE Trans. Electron Dev.* **62**, 3498–3507 (2015).
28. Niroula, J. *et al.* Piecewise empirical model (PEM) of resistive memory for pulsed analog and neuromorphic applications. *J. Comput. Electron.* **16**, 1144–1153 (2017).
29. Sidler, S. *et al.* Large-scale neural networks implemented with non-volatile memory as the synaptic weight element: Impact of conductance response. In *2016 46th European Solid-State Device Research Conference (ESSDERC)* 440–443 (IEEE, 2016). doi:https://doi.org/10.1109/ESSDERC.2016.7599680.
30. Marinella, M. J. *et al.* Multiscale co-design analysis of energy, latency, area, and accuracy of a ReRAM analog neural training accelerator. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **8**, 86–101 (2018).
31. Gong, N. *et al.* Signal and noise extraction from analog memory elements for neuromorphic computing. *Nat. Commun.* **9**, 2102 (2018).
32. Stathopoulos, S., Serb, A., Khiat, A., Ogorzałek, M. & Prodromakis, T. A memristive switching uncertainty model. *IEEE Trans. Electron Dev.* **66**, 2946–2953 (2019).

33. Bollen, K. A. & Barb, K. H. Pearson's R and coarsely categorized measures. *Am. Sociol. Rev.* **46**, 232 (1981).
34. Virkar, Y. & Clauset, A. Power-law distributions in binned empirical data. *Ann. Appl. Stat.* **8**, 89–119 (2014).
35. van Leeuwen, J., Smeets, J. B. J. & Belopolsky, A. V. Forget binning and get SMART: Getting more out of the time-course of response data. *Atten. Percept. Psychophys.* **81**, 2956–2967 (2019).
36. Kolmogoroff, A. Interpolation und extrapolation von stationaeren zufaelligen Folgen. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya* **5**, 3–14 (1941).
37. Krige, D. G. A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. South. Afr. Inst. Min. Metall.* **52**, 119–139 (1951).
38. Matheron, G. Principles of geostatistics. *Econ. Geol.* **58**, 1246–1266 (1963).
39. Montero, J.-M., Fernández-Avilés, G. & Mateu, J. *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging* (Wiley, 2015).
40. Sacks, J., Welch, W. J., Mitchell, T. J. & Wynn, H. P. Design and analysis of computer experiments. *Stat. Sci.* **4**, 409–423 (1989).
41. Rasmussen, C. E. *Evaluation of Gaussian Processes and Other Methods for Non-linear Regression* (University of Toronto, 1997).
42. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning* (MIT Press, 2005).
43. Passos, F. *et al.* Physical vs. surrogate models of passive RF devices. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)* 117–120 (IEEE, 2015). doi:https://doi.org/10.1109/ISCAS.2015.7168584.
44. Okobiah, O., Mohanty, S. & Kougianos, E. Geostatistical-inspired fast layout optimisation of a nano-CMOS thermal sensor. *IET Circuits Dev. Syst.* **7**, 253–262 (2013).
45. You, H., Yang, M., Wang, D. & Jia, X. Kriging Model combined with latin hypercube sampling for surrogate modeling of analog integrated circuit performance. In *2009 10th International Symposium on Quality Electronic Design* 554–558 (2009). doi:https://doi.org/10.1109/ISQED.2009.4810354.
46. Journel, A. G. & Rossi, M. E. When do we need a trend model in kriging?. *Math. Geol.* **21**, 715–739 (1989).
47. Chen, H., Loeppky, J. L., Sacks, J. & Welch, W. J. Analysis methods for computer experiments: How to assess and what counts?. *Stat. Sci.* **31**, 40–60 (2016).
48. Xiao, Q., Wang, L. & Xu, H. Application of kriging models for a drug combination experiment on lung cancer. *Stat. Med.* **38**, 236–246 (2019).
49. Song, J., Dixit, H., Behin-Aein, B., Kim, C. H. & Taylor, W. Impact of process variability on write error rate and read disturbance in STT-MRAM devices. *IEEE Trans. Magn.* **56**, 1–11 (2020).
50. Wu, D., Kunishima, I., Roberts, S. & Gruverman, A. Spatial variations in local switching parameters of ferroelectric random access memory capacitors. *Appl. Phys. Lett.* **95**, 092901 (2009).
51. Be Belmonte, A. *et al.* Origin of the current discretization in deep reset states of an Al_2O_3/Cu-based conductive-bridging memory, and impact on state level and variability. *Appl. Phys. Lett.* **104**, 233508 (2014).
52. Lanza, M., Smets, Q., Huyghebaert, C. & Li, L. J. Yield, variability, reliability, and stability of two-dimensional materials based solid-state electronic devices. *Nat. Commun.* **11**, 1–5 (2020).
53. Rasheed, F., Hefenbrock, M., Beigl, M., Tahoori, M. B. & Aghassi-Hagmann, J. Variability modeling for printed inorganic electrolyte-gated transistors and circuits. *IEEE Trans. Electron Dev.* **66**, 146–152 (2018).
54. Jo, S. H. *et al.* Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* **10**, 1297–1301 (2010).
55. Chen, P.-Y., Peng, X. & Yu, S. NeuroSim: A circuit-level macro model for benchmarking neuro-inspired architectures in online learning. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **37**, 3067–3080 (2018).
56. Roustant, O., Ginsbourger D., Deville Y. DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization. *J. Stat. Softw.* **51**(1), 1–55 (2012).
57. Massey, F. J. The Kolmogorov–Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **46**, 68–78 (1951).
58. Lilliefors, H. W. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* **62**, 399–402 (1967).
59. Steinskog, D. J., Tjøstheim, D. B. & Kvamstø, N. G. A cautionary note on the use of the Kolmogorov-Smirnov test for normality. *Mon. Weather Rev.* **135**, 1151–1157 (2007).
60. Gretton, A., Borgwardt, K. M., Rasch, M., Scholkopf, B. & Smola, A. J. A kernel method for the two-sample-problem. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **19**, 513–520 (2006).
61. Xiao, Y. A fastalgorithm for two-dimensional Kolmogorov–Smirnov two sample tests. *Computational Statistics & Data Analysis* **105**, 53–58 (2017). R package available at https://cran.r-project.org/web/packages/Peacock.test/Peacock.test.pdf.
62. You, K. *maotai: Tools for Matrix Algebra, Optimization and Inference* (2021).
63. Muhammad, E.-S. *Transport of Information-Carriers in Semiconductors and Nanodevices* (IGI Global, 2017).
64. Khandelwal, S. Physics-based compact models: An emerging trend in simulation-based GaN HEMT power amplifier design. In *2019 IEEE 20th Wireless and Microwave Technology Conference (WAMICON)* 1–4 (IEEE, 2019).
65. Sun, W. *et al.* Understanding memristive switching via in situ characterization and device modeling. *Nat. Commun.* **10**, 3453 (2019).
66. Hajri, B., Aziza, H., Mansour, M. M. & Chehab, A. RRAM device models: A comparative analysis with experimental validation. *IEEE Access* **7**, 168963–168980 (2019).
67. Pérez-Ávila, A. J. *et al.* Behavioral modeling of multilevel HfO2-based memristors for neuromorphic circuit simulation. In *2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS)* 1–6 (2020). https://doi.org/10.1109/DCIS51330.2020.9268652.
68. Pérez, E. *et al.* Optimization of multi-level operation in RRAM arrays for in-memory computing. *Electronics* **10**, 1084 (2021).

Acknowledgements

This work has been supported by the DARPA/ONR Grant No. N00014-20-1-2031, the GW University Facilitating Fund and the GW Cross-Disciplinary Research Fund. We thank Osama Yousuf for useful discussions.

Author contributions

I.H. processed the data and performed the simulations. M.A. helped with the experimental results. L.W. provided support. G.A. supervised the work. All authors participated in data analysis, discussed the results, and co-edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to G.C.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022