Machine Learning for Statistical Modeling: The Case of Perpendicular Spin-Transfer-Torque Random Access Memory

URMIMALA ROY and TANMOY PRAMANIK, Microelectronics Research Center, The University of Texas at Austin, USA

SUBHENDU ROY, Cadence Design Systems, USA

AVHISHEK CHATTERJEE, Department of Electrical Engineering, Indian Institute of Technology Madras, India

LEONARD F. REGISTER and SANJAY K. BANERJEE, Microelectronics Research Center, The University of Texas at Austin, USA

We propose a methodology to perform process variation-aware device and circuit design using fully physics-based simulations within limited computational resources, without developing a compact model. Machine learning (ML), specifically a support vector regression (SVR) model, has been used. The SVR model has been trained using a dataset of devices simulated a priori, and the accuracy of prediction by the trained SVR model has been demonstrated. To produce a switching time distribution from the trained ML model, we only had to generate the dataset to train and validate the model, which needed ~ 500 hours of computation. On the other hand, if 10^6 samples were to be simulated using the same computation resources to generate a switching time distribution from micromagnetic simulations, it would have taken ~ 250 days. Spin-transfer-torque random access memory (STTRAM) has been used to demonstrate the method. However, different physical systems may be considered, different ML models can be used for different physical systems and/or different device parameter sets, and similar ends could be achieved by training the ML model using measured device data.

CCS Concepts: • **Computing methodologies** $\rightarrow$ *Support vector machines*; • **Hardware** $\rightarrow$ *Modeling and parameter extraction*; *Spintronics and magnetic technologies*;

Additional Key Words and Phrases: Spin-transfer-torque random access memory, process variation, machine learning, support vector regression

ACM Reference format:

Urmimala Roy, Tanmoy Pramanik, Subhendu Roy, Avhishek Chatterjee, Leonard F. Register, and Sanjay K. Banerjee. 2021. Machine Learning for Statistical Modeling: The Case of Perpendicular Spin-Transfer-Torque Random Access Memory. *ACM Trans. Des. Autom. Electron. Syst.* 26, 3, Article 24 (January 2021), 17 pages. <https://doi.org/10.1145/3440014>

U. Roy was with the Microelectronics Research Center, The University of Texas at Austin, Austin, TX, 78758. She is now with Intel Corporation, Santa Clara, CA, 95052.

Authors' addresses: U. Roy and T. Pramanik, Microelectronics Research Center, The University of Texas at Austin, Austin, Texas, 78758; email: urmimala@utexas.edu; S. Roy, Cadence Design Systems, San Jose, CA, 95134; email: subhendu@utexas.edu; A. Chatterjee, Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai, India-600036; email: avhishek@utexas.edu; L. F. Register and S. K. Banerjee, Microelectronics Research Center, The University of Texas at Austin, Austin, Texas, 78758; emails: register@austin.utexas.edu, banerjee@ece.utexas.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Association for Computing Machinery.

1084-4309/2021/01-ART24 \$15.00

<https://doi.org/10.1145/3440014>

1 INTRODUCTION

All electronic designs, digital, analog, and mixed signal, use compact models to describe the constituent devices, semiconductor or otherwise. Compact models are descriptions of the device-level behavior in a mathematical form that can be solved in “reasonable” computation time [26]. Ideally, such compact models should be analytical equations [42]. Moreover, all device fabrication technologies suffer from process-induced variations, more so for extremely scaled technologies. Hence, circuit designers need to take into account process variations during device fabrication in their designs using compact models that can capture the process variation. Extreme process variation at the advanced technology nodes affects fabrication yield. To be able to achieve high yield, the designers should be able to estimate yield and then design accordingly. Design for manufacturability is an active area of research in recent years [24]. For CMOS circuits, process-induced variation is most commonly taken care of by introducing process corners to MOSFET compact models (typical, slow, and fast MOSFETs). Even if a MOSFET is represented in terms of a compact model, such as BSIM [26], searching the high-dimensional design space to look for an optimized design functional under the process corners within a limited computational load is a challenge. Machine learning (ML) has been used previously for high-dimensional statistical modeling, and even commercial software packages exist that employ ML techniques to reduce the computational load needed in a brute-force Monte Carlo simulation covering all process corners, all design parameters, and all operating conditions [20]. Also, in many cases coming up with a computationally light compact model might not be straightforward, such as for emerging devices where a full physics-based simulation might be the only way to predict device-level performance. Even with a physics-based compact model available, simulation of a large number of devices can be too computationally expensive, prompting researchers to explore use of ML to extract “features” from device-level characteristics and use it to predict performance of memory arrays, such as that of a resistive random access memory (RRAM) [18].

In this work, we look for a scheme to allow running statistical simulation without having to use conventional compact models. Thus, we focus only on the statistical modeling sub-space of compact modeling. We use spin-transfer-torque random access memory (STT RAM) as an example device to demonstrate that statistical simulations can be performed within reasonable computation load using ML techniques, a support vector regression (SVR) method for the particular system considered here, without resorting to (often over-simplified) compact models. Reference datasets of device behavior as a function of various parameters and the thermal fluctuations within individual devices have been obtained by micromagnetic simulation. However, different devices may be considered using different ML models and/or different device parameter sets, and similar ends could be achieved by training the ML model using experimental device data.

STT RAM is one of the promising non-volatile memories being considered as a replacement for traditional semiconductor device-based volatile memories such as SRAM and DRAM, and non-volatile ones such as flash, because of the scaling challenges that the conventional memories face. Capability of fast read and write operations together with non-volatile storage and fairly high endurance makes STT RAM an attractive choice for the so-called universal memory [23]. Several prototype STT RAM chips already have been demonstrated from leading semiconductor companies [17, 35].

Being a relatively new device compared to its CMOS competitors, there are limited or few attempts at variation-aware STT RAM circuit design. Most of the variation-aware STT RAM circuit design methods proposed so far involve Monte-Carlo simulation of the basic memory element of the STT RAM cell, the magnetic tunnel junction (MTJ) [5, 19, 29]. The design space exploration for the STT RAM memory array needs to be variation aware for both the MTJ and the access transistor

[7]. For this exploration, first, a SPICE macromodel [12, 25] has to be developed for MTJ, allowing one to simulate hybrid CMOS/MTJ circuits. The development of a computationally cheap SPICE macromodel for the STT RAM bit involves the so-called macrospin approximation that assumes that the storage layer magnet, the “free layer” (FL), switches coherently as a single spin. Even within the macrospin approximation, attempts have been made to reduce the computational load by modeling variation of parameters in a multi-dimensional space in terms of linear analytical equations that describe the stochastic magnetization dynamics [8]. However, it has been shown that a macrospin approximation cannot always capture the complicated physics arising from incoherent switching [33], such as sub-volume excitation [37, 38], anomalous branching of write error variation with increasing write voltage or current [21], and effects coming from higher-order spin wave modes [9]. Thus, full micromagnetic simulations may be needed for a more accurate study.

However, variation-aware stochastic micromagnetic simulations are prohibitively expensive in terms of computational load, thus thwarting most attempts to perform statistical modeling of STT RAM while simultaneously taking into account the physics of incoherent switching and thermal fluctuations. Moreover, the previous attempts that have been made to model a variation-aware circuit using full micromagnetic simulations assumed only a variation of the oxide thickness, which allowed the simulation to be run on a single FL magnet for varying strengths of spin polarized current, and the variation in the stochastic magnetization dynamics (as described in terms of mean, standard deviation, and skewness of a skewed Gaussian probability distribution function) to be stored as a function of injected current a priori [30–32, 34]. If process-induced variation of more parameters, such as the FL magnet dimensions and/or anisotropy energy, were taken into account, the a priori creation and use of such a lookup table would become much more difficult since the parameter space would become multi-dimensional.

In this work, we propose a scheme to “learn” the variation of the switching performance of an STT RAM bit as a function of its material and geometric parameters, using a set of devices simulated a priori, using techniques from ML. Our key contributions include performance prediction of an STT RAM bit taking into account micromagnetic effects and process-induced variation to material and geometric parameters, *without resorting to macrospin SPICE macromodels and analytical expressions*. We use STT RAM as an example device, though we believe the technique should be usable for other novel electronic and spintronic devices.

Understandably, an ML model can be termed as a “black box,” since no analytical physics-based equation is being derived. The main motivation for ML-driven black-box models is the following. In practice, we often know all the parameters that impact a particular performance metric, but we do not know the precise quantitative dependence. A black-box model can be trained to learn this dependence using the available experimental/simulation data. Definitely, when the underlying mechanism (or physics) of the system is analytically or numerically tractable, a physics-based theoretical model is likely to be much more accurate and efficient. However, black-box models are efficient alternatives in many scenarios where unfortunately the theoretical models are intractable.

2 METHOD

The simplest possible bit-cell structure in an STT RAM memory array consists of a 1-transistor, 1-MTJ cell as shown in Figure 1(a). Here, we ignore any process-induced variations in the accompanying transistor for a simple proof-of-concept approach and assume that the transistor drives a constant current through the MTJ, irrespective of the MTJ resistance. In an actual scenario, the current to the MTJ has to be calculated self-consistently where both the MTJ and the transistor are subject to process-induced variation, and changing the MTJ resistance changes the current flow. In addition, only variations in the material and geometric parameters of the FL magnet in the MTJ structure will be considered. Any variations in the reference layer (RL) magnet in the MTJ is

Fig. 1. (a) One-bit cell of basic 1-transistor, 1-MTJ memory cell, and schematic illustration of an MTJ with perpendicular magnets in free layer (FL) and reference layer (RL). The geometric parameters subject to process variation are marked. (b) Magnetization (space-averaged out-of-plane magnetization of the FL magnet $\langle m_z \rangle$) vs. time trajectories for two different STT-induced switching events of the same FL magnet, as calculated from micromagnetic simulations including thermal effects at $T = 300$ K. Snapshots show magnetization texture inside the FL magnet at time instances corresponding to the dots marked on the trajectories, with the local magnetization indicated by the color bar. One trajectory shows a quasi-uniform macrospin-like switching mode, while the other exhibits a domain wall-like state during switching. (c) and (d) compare the probability distribution of switching times obtained from micromagnetic and macrospin simulations, for two different FL magnet parameters. In each case, 10^3 independent simulations are performed with different random thermal fields. Depending on the current overdrive J_{ov} , the probability distributions obtained from micromagnetic simulations (red histogram) can be very different from the macrospin prediction (blue histogram). The solid lines in (c) and (d) are log-normal fits to the histogram.

ignored, assuming it to be an ideal source of spin-polarized current. A more detailed model could include, for example, variations in the RL magnet and its effect on the magnetostatic coupling with the FL. We also assume no variations in the oxide thickness in the MTJ and corresponding variations in the resistance and tunnel-magneto-resistance (TMR). The magnetization dynamics of the FL magnet then are studied under STT and random thermal fluctuations at a temperature (T) of 300 K for a set of five different material and geometry parameters of the FL magnet: thickness (t_{FL}) and planar dimensions D_x and D_y , as marked on Figure 1(a), and uniaxial anisotropy energy (K_u) and Gilbert damping constant (α_G).

Two example magnetization trajectories from micromagnetic simulations are shown in Figure 1(b) for a set of FL magnet parameters. Magnetization snapshots at selected time instants reveal a quasi-uniform macrospin-like magnetization distribution for one trajectory and a very

non-uniform domain-wall-like magnetization state for the other trajectory. Such non-uniform switching modes require micromagnetic simulations for accurate description of the switching process and the switching time of the FL magnet. Here, we define the switching time (t_{sw}) as the time taken by the space-averaged out-of-plane magnetization of the FL magnet ($\langle m_z \rangle$) to fall below zero during switching from $\langle m_z \rangle \approx 1$ to $\langle m_z \rangle \approx -1$.

Figures 1(c) and 1(d) compare the probability distribution of t_{sw} obtained from micromagnetic and macrospin simulations for two representative devices (detailed on the figure). In each case, 10^3 instances of the switching are simulated to obtain the distribution resulting from thermal fluctuation. As is evident, probability distributions obtained from micromagnetic simulations can be very different from ones obtained from macrospin simulations. Average switching times obtained from micromagnetic simulations are always found to be lower than that obtained from macrospin simulations, in agreement with previous studies [33, 34]. The probability distribution function (PDF) of t_{sw} resulting from the thermal fluctuation has been described previously in terms of a skew-normal PDF, where the PDF is characterized by the mean, standard deviation, and skewness [30–32] or Pearson IV form [34]. Here, however, we describe the PDF of t_{sw} for a given STTRAM bit in terms of a log-normal PDF following [8]. Note that, within the macrospin assumption (coherent switching), the switching probability is described by the well-known double exponential function derived from the Fokker-Planck equation [2]. This description captures the effect of device geometry and material parameters in terms of current overdrive and energy barrier for switching. However, such a treatment is not available for the micromagnetic case (incoherent switching). With a log-normal assumption, $\log(t_{sw})$ is roughly approximated as a Gaussian distribution, allowing the PDF to be described in terms of the mean, $\mu_{\log(t_{sw})}$, and standard deviation, $\sigma_{\log(t_{sw})}$, of the log-normal distribution, without having to take the skew into account. The log-normal fits are shown in Figures 1(c) and 1(d). For later reference, we also define the mean and standard deviation of t_{sw} for a given STTRAM bit as $\mu_{t_{sw}}$ and $\sigma_{t_{sw}}$, respectively. We also will use a normalization factor of 1 ns when calculating $\log(t_{sw})$ from t_{sw} and vice versa.

2.1 Features and Datasets for ML

All of the input data to the method consists of three datasets (training set, validation set, and test set, to be described shortly), each with the above-described five features (t_{FL} , D_x , D_y , K_u , α_G) for each MTJ, and the corresponding $\mu_{\log(t_{sw})}$ and $\sigma_{\log(t_{sw})}$ obtained from the micromagnetic simulation of STT-induced switching. These micromagnetic simulations are performed using either of the two simulators, MuMax3 [41] or OOMMF [6], both of which have options to include a thermal fluctuation field in the Landau-Lifshitz-Gilbert-Slonczewski (LLGS) equation. The two micromagnetic simulators have been thoroughly benchmarked for consistency against each other. The total number of sample devices in the dataset (to capture process variations of the five features) and the number of independent micromagnetic simulations per device (to capture the switching time distributions due to thermal fluctuations) are selected keeping in mind a reasonable simulation time on our computing cluster. However, results show that this selection does ensure enough sampling for training and testing the ML model with reasonable accuracy. Figures 2(a) to 2(e) show histograms for the five features for a total of 1,880 devices. The targeted (mean) μ of design parameters for the FL magnet are chosen as $t_{FL} = 1$ nm, $D_x = D_y = 60$ nm, $K_u = 8 \times 10^5$ Jm⁻³, and $\alpha_G = 0.01$. Choice of material parameters is consistent with experimentally measured values for thin CoFeB alloys [15, 16]. The FL magnet dimension assumed here is also similar to reported STTRAM devices targeted for cache replacement [14, 44]. Each of these parameters is assumed to have a Gaussian distribution due to process variation. We consider standard deviations due to process variation of $\sigma = 1\%$ (t_{FL}), 2.5% (for K_u), and 10% (for D_x , D_y , and α_G) of their respective means μ for each of the five features. However, the datasets for simulation here were chosen initially from roughly

Fig. 2. (a)-(e) Histograms of five features in the dataset of 1,880 devices. Blue histograms represent the first group with features sampled uniformly from within a $\pm 3\sigma$ window around their respective means μ . Red histograms represent the second group of samples where the selection has been skewed for features D_x , D_y , K_u , and α_G to generate more samples with higher average switching times. (f) The probability distribution of average switching times $\mu_{t_{sw}}$ for the devices in the uniform (blue samples) and the skewed (red samples) group showing skewed characteristics. $\mu_{t_{sw}}$ for each sample device has been calculated from 10 micromagnetic simulations of the switching event at a given applied current. Solid lines in (f) are non-parametric fits to the histograms.

uniform distributions (Figure 2, blue histograms) spanning $\pm 3\sigma$ windows around the μ values. The uniform distributions have been chosen to ensure that the trained model predicts the switching times equally well for devices near the targeted design, as well as for devices far from the targeted design. However, as shown in Figure 1, the PDF of t_{sw} (resulting from thermal fluctuation) is intrinsically skewed such that there is a tail towards higher t_{sw} . Figure 2(f) (blue histogram) shows the probability distribution of $\mu_{t_{sw}}$ (averaged over 10 micromagnetic simulations for each sample device) for individual devices selected with features from the blue histograms (Figures 2(a) to 2(e)). Interestingly, this PDF of $\mu_{t_{sw}}$ is also found to be skewed towards the lower switching times, making it difficult to train the model for devices with high $\mu_{t_{sw}}$. To account for this skew, a second set of training samples with more high-switching-time samples was created. This second set of samples was again from within a $\pm 3\sigma$ window, but with a skewed sampling scheme for D_x , D_y , K_u , and α_G (Figures 2(a) to 2(f), red histograms). The total of 1,880 samples from these two types of sampling together is sufficient to train and validate the model for $\mu_{t_{sw}}$ up to ~ 15 ns (Figure 2(f)), representative of the practical range of interest for cache applications [40].

For reference, within a macrospin description, the critical switching current (I_c) and the current overdrive (J_{ov}) for each sample device is calculated as [11, 28, 36, 43]

$$I_c = \frac{4\alpha_G q}{\hbar P} V \left[K_u - \frac{1}{4} \mu_0 M_s^2 (2N_{zz} - N_{xx} - N_{yy}) \right] \quad (1)$$

Fig. 3. Visual representation of training, validation, and test sets. (a) Comparison of $\mu_{t_{sw}}$ varying with J_{ov} from macrospin and micromagnetic simulations. Each device has been subject to 10^3 micromagnetic (macrospin) simulation times to capture thermal fluctuation effects. (b) Selection of the three datasets, with $\mu_{t_{sw}}$ and $\sigma_{t_{sw}}$ for each device calculated from 10 micromagnetic simulations.

and

$$J_{ov} = \frac{4}{\pi D_x D_y} (I_{\text{applied}} - I_c), \quad (2)$$

respectively, where, $V = (\pi/4)D_x D_y t_{FL}$ is the volume of the FL; N_{xx} , N_{yy} , N_{zz} are the demagnetization factors of the FL calculated using [1]; $M_s = 1,100$ kA/m is the saturation magnetization of the FL material; $P = 0.4$ is the spin polarization of the spin-polarized current injected to the FL; q is electronic charge; and μ_0 is the vacuum permeability. For micromagnetic simulations, an exchange constant of 20 pJ/m has been used [45].

Figure 3(a) shows the $\mu_{t_{sw}}$ obtained from micromagnetic (10 trials) and macrospin simulations (10^3 trials) for each of the 1,880 devices with inverse of their J_{ov} . A constant applied current of $I_{\text{applied}} = 84.2 \mu\text{A}$ is used, twice the critical current of the nominal device (that having the target values of parameters) within a macrospin model. As is evident, $\mu_{t_{sw}}$ cannot be described by a simple function of J_{ov} even for the macrospin case. We then sampled 50 devices (highlighted in Figure 3(a) as blue squares) out of these 1,880 cases such that they cover the J_{ov} vs. $\mu_{t_{sw}}$ space roughly uniformly. Each of these latter 50 devices then is subject to 10^3 micromagnetic simulations to obtain a better estimate of the t_{sw} distribution resulting from thermal fluctuations. It is because of the 10^3 micromagnetic simulations per device and associated computational demands that this latter set is limited to 50 devices. This latter dataset then is used to verify the final trained model, and henceforth is referred to as the “test set.”

Two more datasets of 915 devices each are then created by randomly selecting samples from the remaining population of 1,830 devices, a “training set” and a “validation set.” For each device in these two datasets, the mean, $\mu_{\log(t_{sw})}$, and standard deviation, $\sigma_{\log(t_{sw})}$, of the log-normal PDF of t_{sw} are estimated from the 10 micromagnetic simulations corresponding to different thermal fields. The training set is then used to train the ML models to predict $\mu_{\log(t_{sw})}$ and $\sigma_{\log(t_{sw})}$ given the set of five device parameters. The validation set is used to determine parameter values needed in the ML models. Together, the training and validation sets are used to optimize the model parameters, as will be discussed shortly. Finally, the “test set” provides an independent assessment of the final optimized models. The 10^3 thermal simulations per test set device provide a higher confidence in the mean, $\mu_{\log(t_{sw})}$, and standard deviation, $\sigma_{\log(t_{sw})}$, of the log-normal distribution for each device in the set. Figure 3(b) shows the mean, $\mu_{t_{sw}}$, and standard deviation, $\sigma_{t_{sw}}$, of the switching times

Table 1. Comparison of Different Machine Learning Models for Prediction of the Mean Switching Time

Method	Parameter, if any	Training set MAPE (%)	Validation set MAPE (%)	Test set MAPE (%)
Linear regression		4.26	4.45	2.53
Lasso regression	$\alpha = 0.05$	20.81	20.37	19.12
Ridge regression	$\alpha = 0.05$	4.25	4.47	2.65
Polynomial fit on J_{ov}		5.91	6.13	6.41
SVR (RBF kernel)	$C = 100, \gamma = 0.1$	3.41	3.65	0.85

for all 1,880 devices plotted as a function of J_{ov} and categorized into the training, validation, and test sets.

Using 16 nodes in parallel on a supercomputing cluster [4], it required ~ 350 hours to run all $50 \times 10^3 = 5 \times 10^4$ of the test-set simulations, and ~ 110 hours to run all $1830 \times 10 = 18300$ training and validation set simulations combined.

2.2 Selection of ML Model

We have explored several supervised learning models for predicting $\mu_{\log(t_{sw})}$ and $\sigma_{\log(t_{sw})}$ by using the Python-based Scikit-learn package [27], as discussed in detail in following paragraphs. Table 1 summarizes the resulting mean absolute percentage error (MAPE) of the predicted $\mu_{\log(t_{sw})}$ for devices in the training, validation, and test set across various ML models. Based on the results, as elaborated on below, for the considered system and device parameter set, we ultimately employ the SVR model. However, we note that for other physical systems and/or with other device parameter sets, other ML models may be appropriate.

First, we attempted to use the various variants of linear regression models. The objective function for minimization in a regularized regression can be generalized as

$$F(\theta) = ||(Y - X(\theta))||^2 + \alpha ||\theta||^p, \quad (3)$$

where $Y - X(\theta)$ is the difference in the actual and predicted value of the model, θ are the coefficients of the feature set X , and α is the “regularization” parameter. For simple linear regression, $\alpha = 0$. The values of p are respectively 1 and 2 for Lasso and Ridge regression. Simple linear regression is the simplest form of regression and does not need parameter tuning. With simple linear regression, we achieved a MAPE of 4.26% and 4.45% for the training set and validation set, respectively. Lasso and Ridge regression do not improve the model accuracy, as seen from the MAPE. Notably, we varied α for Lasso and Ridge regression and observed no significant variation in MAPE, and so we have reported results for only one value of α . The error for Lasso regression is much higher than that for Ridge or linear regression. However, we note that Lasso regression typically works well in feature selection when the number of features is very high, whereas the number of features is only five in this work.

For each device, the five features may be mapped onto a single variable J_{ov} using the macrospin Equations (1) and (2). However, as seen in Table 1, attempting to model $\mu_{\log(t_{sw})}$ in terms of only J_{ov} by using a polynomial fit results in even higher MAPE than either simple linear or Ridge regression. Much higher MAPE for this J_{ov} fit agrees with our earlier discussion of Figure 3(a), indicating that switching times cannot be described by a simple single-valued function of J_{ov} .

Next, we explored the SVR model with the radial basis function (RBF) kernel [27]. Two parameters, C and γ , are used in this model. By the proper selection of these parameters, we could achieve the best accuracy among all models. In the next subsection, the tuning of these two parameters is discussed in detail.

Fig. 4. SVR model tuning for $\mu_{\log(t_{sw})}$: (a) Heat map for error of validation set, and tuning of (b) C and (c) γ for SVR model using the validation dataset.

2.3 Tuning the SVR Model

The parameters of the SVR model are tuned using a grid search technique [13, 27], where we look for the lowest MAPE of the validation set as the parameters C and γ are varied over a wide range of values. Figure 4(a) shows the heat map of the MAPE of the validation set. The error is found to be minimum for $C = 100$ and $\gamma = 0.1$. Figures 4(b) and 4(c) show the MAPE of both training and validation sets around the optimum C and γ . With $\gamma = 0.1$, increasing C reduces the MAPE of both training and validation sets (Figure 4(b)). On the other hand, with $C = 100$, varying γ leads to an optimum point where the MAPE for both training and validation sets is minimized (Figure 4(c)). Increasing γ beyond this point causes the MAPE of the validation set, in particular, to increase rapidly. The inverse of the parameter γ can be thought of as the radius of influence of the training samples. With increased γ , the model presumably produces an overfit of the training set, causing the model to fit the training set samples not just closely but too closely, so that it fails to generalize well to the (previously) “unseen” inputs from the validation set. This is similar to the well-known “bias-variance trade-off” where the prediction error of the validation set increases while the prediction error of the training set does not [10]. As MAPE for both the training and validation set are at the minimum, we set C and γ to be 100 and 0.1, respectively, for the optimized SVR model for $\mu_{\log(t_{sw})}$. With these parameters, the MAPEs of the validation set and test set are found to be 3.65% and 0.85%, respectively, which is the lowest across all ML models as listed in Table 1. The process of optimizing C and γ can be repeated with 1,830 devices randomly reshuffled into new training and validation sets. However, we observed that the MAPE obtained from the optimized model after reshuffling does not change significantly.

The above exercise is then repeated to predict standard deviation $\sigma_{\log(t_{sw})}$ with the SVR model. For $\sigma_{\log(t_{sw})}$, the optimum values for C and γ are found to be 10 and 0.1, respectively.

3 RESULTS

Figure 5 shows the actual $\mu_{\log(t_{sw})}$ (calculated from micromagnetic simulations) vs. the $\mu_{\log(t_{sw})}$ predicted by the optimized SVR model with parameters $C = 100$ and $\gamma = 0.1$ for all three datasets. Actual vs. predicted $\mu_{\log(t_{sw})}$ follows a 1:1 trend, thereby demonstrating the correctness of prediction by SVR, qualitatively. The model performs well for validation set as well as the test set. Similarly, Figure 6 shows actual $\sigma_{\log(t_{sw})}$ vs. predicted $\sigma_{\log(t_{sw})}$, again with optimized parameters $C = 10$ and $\gamma = 0.1$. The accuracy of prediction of $\sigma_{\log(t_{sw})}$ is much worse than that of $\mu_{\log(t_{sw})}$, as

Fig. 5. Actual $\mu_{\log(t_{sw})}$ calculated from micromagnetic simulations vs. $\mu_{\log(t_{sw})}$ predicted by SVR model for training, validation, and test datasets. Actual $\mu_{\log(t_{sw})}$ for each device is calculated from 10 micromagnetic simulations for the training and validation sets, and 10^3 micromagnetic simulations for the test set.

evidenced by Figure 6, with the test set MAPE for $\sigma_{\log(t_{sw})}$ being 5.35%, presumably because of worse confidence in the training data from only 10 micromagnetic simulations per sample device. However, the correlation of predicted $\sigma_{\log(t_{sw})}$ to the actual $\sigma_{\log(t_{sw})}$ for the test set (10^3 runs per device) is indeed much better (see Figure 6 inset) than for the training or validation dataset (both 10 runs per device).

The probability distribution of t_{sw} , $p(t_{sw})$ resulting from thermal fluctuations can now be obtained from the predicted $\mu_{\log(t_{sw})}$ and $\sigma_{\log(t_{sw})}$ using the expression for log-normal PDF:

$$p(t_{sw}) = \frac{1}{\sigma_{\log(t_{sw})} t_{sw} \sqrt{2\pi}} e^{-\frac{[\log(t_{sw}) - \mu_{\log(t_{sw})}]^2}{2\sigma_{\log(t_{sw})}^2}}. \quad (4)$$

In Figure 7, we plot the log-normal PDF given by Equation (4) for a few selected devices from the test set, using the predicted $\mu_{\log(t_{sw})}$ and $\sigma_{\log(t_{sw})}$. When compared against the actual histograms of t_{sw} (plotted with t_{sw} obtained from 10^3 micromagnetic simulations), the predicted probability distributions show good agreement. Also, to quantify the accuracy of prediction by the ML model in terms of t_{sw} instead of only $\log(t_{sw})$, we calculate mean, $\mu_{t_{sw}}$, and standard deviation, $\sigma_{t_{sw}}$, of t_{sw} from $\mu_{\log(t_{sw})}$ and $\sigma_{\log(t_{sw})}$, using the following expressions:

$$\mu_{t_{sw}, \text{predicted}} = e^{\mu_{\log(t_{sw})} + 0.5\sigma_{\log(t_{sw})}^2} \quad (5)$$

$$\sigma_{t_{sw}, \text{predicted}}^2 = [e^{\sigma_{\log(t_{sw})}^2} - 1] e^{2\mu_{\log(t_{sw})} + \sigma_{\log(t_{sw})}^2}. \quad (6)$$

For the test set, the MAPE in the thus-predicted $\mu_{t_{sw}}$ with respect to actual $\mu_{t_{sw}}$ is 1.64%, and the MAPE in the predicted $\sigma_{t_{sw}}$ with respect to the actual $\sigma_{t_{sw}}$ is 8.45%. We do not attempt to estimate the skew in t_{sw} because of lack of enough statistics in a dataset of just 10^3 samples for each device in the test set. The non-monotonicity of the dependence of $\mu_{t_{sw}}$ on J_{ov} , as evident in Figure 7, again illustrates that a simple fitting based on J_{ov} will not accurately predict the PDF of t_{sw} .

In the context of stochastic switching of STTRAM bits, the more commonly discussed quantity is write error rate (WER), which is the probability of a bit not switching at a given time after the write

Fig. 6. Actual $\sigma_{\log(t_{sw})}$ calculated from micromagnetic simulations vs. $\sigma_{\log(t_{sw})}$ predicted by SVR model. Actual $\sigma_{\log(t_{sw})}$ for each device is calculated from 10 micromagnetic simulations for the training and validation sets, and 10^3 micromagnetic simulations for the test set. Inset shows the predicted vs. actual comparison for the test set only. Prediction on the test set appears to be much better as compared to the prediction on the training and validation set.

pulse is applied. With that in mind, Figure 8 compares the WER calculated from the trained ML model and from micromagnetic simulations for three devices. The micromagnetic simulations for each device contain 10^3 trials. Device 2 is the nominal device (diameter 60 nm) used in this study, device 1 is smaller (diameter ~ 45 nm), and device 3 is larger (diameter ~ 75 nm) than the nominal device. The other parameters are similar for these three devices. The WER predicted by the model (solid lines) agrees well with the micromagnetic simulations (squares), demonstrating the model accuracy far from the nominal device. Figure 8 also confirms that log-normal approximation does capture the lower WER tails quite well.

Depending on the error-correcting capabilities, a WER in the range of 10^{-6} to 10^{-9} is required for practical applications. In the Figure 8 inset we show that the model-predicted log-normal distribution is capable of predicting the WER accurately down to 10^{-9} . The true WER (orange line) plotted in the inset has been calculated from a “rare-event-enhanced” micromagnetic simulation following a previous study [33]. Note that, if the rare events with non-switching probability in the range of 10^{-9} can still be captured by a log-normal probability distribution of t_{sw} , the proposed method should work in predicting those rare events. If not, there has to be special modeling schemes first to study the switching probability in the range of 10^{-9} . The same “rare-event-enhanced” micromagnetic simulation scheme [33] can be used for this purpose. We believe if the training data could be generated using such a method, an ML model can in principle be trained to accurately predict the low-probability tails of switching time distributions. Alternatively, experimental WER data could be used to train the model. Further validation using the REE technique or using experimentally collected data remains an interesting open question for future study.

Experimentally measured WER is usually plotted as a function of applied write current (I_{applied}) or write voltage to the bit for a given pulse width, as opposed to plotting as a function of t_{sw} for a given applied current ($I_{\text{applied}} = 84.2 \mu\text{A}$ in Figure 8). To plot WER as a function of I_{applied} , I_{applied} will have to be varied in the training set and could be included as a sixth feature besides t_{FL} , D_x ,

Fig. 7. Calculated histogram of t_{sw} from 10^3 micromagnetic simulations vs. predicted log-normal PDF $p(t_{sw})$ using SVR (solid lines) for selected devices from the test set. The device parameters are as follows: (a) $D_x = 75.3$ nm, $D_y = 67.1$ nm, $t_{FL} = 1.00$ nm, $K_u = 8.1 \times 10^5$ Jm $^{-3}$, $\alpha_G = 0.013$; (b) $D_x = 73.8$ nm, $D_y = 62.9$ nm, $t_{FL} = 1.00$ nm, $K_u = 8.4 \times 10^5$ Jm $^{-3}$, $\alpha_G = 0.010$; (c) $D_x = 66.9$ nm, $D_y = 62.9$ nm, $t_{FL} = 1.01$ nm, $K_u = 8.5 \times 10^5$ Jm $^{-3}$, $\alpha_G = 0.010$; (d) $D_x = 62.3$ nm, $D_y = 49.1$ nm, $t_{FL} = 0.98$ nm, $K_u = 8.6 \times 10^5$ Jm $^{-3}$, $\alpha_G = 0.013$; (e) $D_x = 53.5$ nm, $D_y = 69.8$ nm, $t_{FL} = 0.99$ nm, $K_u = 8.3 \times 10^5$ Jm $^{-3}$, $\alpha_G = 0.009$; (f) $D_x = 67.5$ nm, $D_y = 57.3$ nm, $t_{FL} = 1.00$ nm, $K_u = 7.7 \times 10^5$ Jm $^{-3}$, $\alpha_G = 0.008$. The wide range of geometric and material parameters in the devices shown for which the trained ML model correctly predicts the distribution of t_{sw} goes to demonstrate the wide range (in terms of t_{FL} , D_x , D_y , K_u , α_G) of applicability of the trained ML model without needing a re-tuning.

D_y , K_u , and α_G . Any variation in the transistor drive current, tunnel barrier oxide thickness, and MTJ spin polarization could also be translated as a variation in the applied current.

We further demonstrate the use of the trained ML model to predict the t_{sw} distribution of an STTMRAM cell subject to process variation and thermal fluctuation. Figure 9 shows the probability distributions of t_{sw} calculated from the trained SVR model for three different cases: (I) nominal device (mean feature values μ : $t_{FL} = 1$ nm, $D_x = D_y = 60$ nm, $K_u = 8 \times 10^5$ Jm $^{-3}$, $\alpha_G = 0.01$) with no process variation ($\sigma = 0$), (II) nominal device with process variation as in Figure 2 ($\sigma = 1\%$, 10% , 10% , 2.5% and 10% of their respective means μ), and (III) a shifted process with mean feature values of $\mu + 0.8 \times \sigma$ and standard deviation of $\sigma \times 0.8$. For case I (the nominal device with no process variations), Figure 9 shows the distribution of t_{sw} resulting from thermal fluctuations alone. The mean and standard deviation of the log-normal distribution is predicted by the ML model. For cases II and III, a set of sample device parameters is generated first for the given process variations. Then for each of the sample devices, the log-normal distribution due to thermal fluctuation is obtained from the trained ML model. Finally, the t_{sw} distribution for all sample devices is plotted together in Figure 9. Comparing the red histogram (case I) to the blue (case II), it's clear that both sources of variation are important, although thermal fluctuation contributes to more variation under these particular assumptions.

Fig. 8. The write error rates calculated from the trained ML model compared to micromagnetic simulations (10^3 trials) for three devices. The solid lines are cumulative probability of non-switched fraction given by the log-normal distribution with parameters predicted by the ML model. The device parameters are as follows: (1) $D_x = 44.97$ nm, $D_y = 45.30$ nm, $t_{FL} = 1.01$ nm, $K_u = 7.86 \times 10^5$ Jm $^{-3}$, $\alpha_G = 0.011$; (2) $D_x = 60.0$ nm, $D_y = 60.0$ nm, $t_{FL} = 1.00$ nm, $K_u = 8 \times 10^5$ Jm $^{-3}$, $\alpha_G = 0.01$; (3) $D_x = 75.74$ nm, $D_y = 74.10$ nm, $t_{FL} = 0.99$ nm, $K_u = 8.06 \times 10^5$ Jm $^{-3}$, $\alpha_G = 0.012$. The inset compares WER predicted by the ML model with respect to that calculated using REE [33] for Device 2, which is the nominal device.

Fig. 9. Predicted histogram of t_{sw} that includes variations due to thermal fluctuation and process variation of the five selected features. The predicted t_{sw} is within the training range of t_{sw} shown in Figure 3. Solid lines are non-parametric fit to the histograms.

For each distribution in Figure 9, 10^6 samples are plotted. Using the same computational resources as noted earlier, calculation of switching times of 10^6 samples from micromagnetic simulations would have taken ~ 250 days, compared to only ~ 500 hours of simulations to build the ML model.

As has been done with macrospin models [12], ML models built from micromagnetic simulations can also be integrated into Electronic Design Automation (EDA) tools. Recently, the major EDA vendors are using ML models to improve (1) the tool’s underlying algorithms, which they call machine learning inside, and (2) productivity of the tools by the efficient exploration of the design space, which is known as ML outside [3]. A recent paper from Google Brain [22] has demonstrated the usage of ML models to improve the chip placement. The integration of ML models inside the tool is often done as a two-step process. In the first step, the models are trained using Python or other programming languages, just like what we demonstrated in this work, which saves the models into readable formats like .h5 or .csv or .pb (for tensor flow). EDA tools are typically written in C/C++, and the tools read these models (there is C/C++ support for reading these models [39]) and can call the native predict functions to get the predictions based on ML models. In our work, we focused on the first step and have shown that a reliable model can be built using ML. The actual integration into EDA tools would need access to the source code of the tools, which is beyond the scope of this work. However, the models that have been developed can be easily incorporated if the EDA vendors intend to do so with the saved models.

4 CONCLUSION

We demonstrate that statistical simulation can be performed using ML when there is no obvious method to describe the device-level behavior in a “compact” mathematical form. We use perpendicular STT RAM as an example device and show that the stochastic device performance, as predicted from full micromagnetic simulations, can be reproduced well by a carefully trained ML model. In doing so, no assumption had to be made for the region of operation for the device, such as precessional or thermal regime. However, different physical systems may be considered, different ML models can be used for different physical systems, and similar ends could be achieved by training the ML model using measured device data.

ACKNOWLEDGMENTS

The authors thank the Texas Advanced Computing Center at the University of Texas at Austin for providing high-performance computing resources that have contributed substantially to the research results reported within this article. TP thanks X. Fong for help with the OOMMF extension for non-zero temperature simulations. Experimental work motivating this was performed in part at the University of Texas Microelectronics Research Center, a member of the National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the National Science Foundation (grant NNCI-2025227).

REFERENCES

- [1] Amikam Aharoni. 1998. Demagnetizing factors for rectangular ferromagnetic prisms. *Journal of Applied Physics* 83, 6 (1998), 3432–3434. DOI: <https://doi.org/10.1063/1.367113> arXiv:<https://doi.org/10.1063/1.367113>
- [2] W. H. Butler, T. Mewes, C. K. A. Mewes, P. B. Visscher, W. H. Rippard, S. E. Russek, and R. Heindl. 2012. Switching distributions for perpendicular spin-torque devices within the macrospin approximation. *IEEE Transactions on Magnetics* 48, 12 (2012), 4684–4700.
- [3] Cadence Design Systems. [n.d.]. AI/Machine Learning Solutions. Retrieved October 14, 2020, from https://www.cadence.com/en_US/home/solutions/machine-learning.html.
- [4] Texas Advanced Computing Center. [n.d.]. Retrieved from <https://portal.tacc.utexas.edu/user-guides/stampede2>.
- [5] Y. Chen, X. Wang, H. Li, H. Xi, Y. Yan, and W. Zhu. 2010. Design margin exploration of spin-transfer torque RAM (STT-RAM) in scaled technologies. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 18, 12 (Dec. 2010), 1724–1734. DOI: <https://doi.org/10.1109/TVLSI.2009.2032192>
- [6] M. J. Donahue and D. G. Porter. 1999. *OOMMF User’s Guide, Version 1.0, Interagency Report NISTIR 6376*. Technical Report. National Institute of Standards and Technology, Gaithersburg, MD.

- [7] R. Dorrance, F. Ren, Y. Toriyama, A. A. Hafez, C. K. K. Yang, and D. Markovic. 2012. Scalability and design-space analysis of a 1T-1MTJ memory cell for STT-RAMs. *IEEE Transactions on Electron Devices* 59, 4 (April 2012), 878–887. DOI: <https://doi.org/10.1109/TED.2011.2182053>
- [8] E. Eken, L. Song, I. Bayram, C. Xu, W. Wen, Y. Xie, and Y. Chen. 2016. NVSim-VXs: An improved NVSim for variation aware STT-RAM simulation. In *2016 53rd ACM/EDAC/IEEE Design Automation Conference (DAC'16)*. 1–6. DOI: <https://doi.org/10.1145/2897937.2898053>
- [9] Eric R. Evarts, Ranko Heindl, William H. Rippard, and Matthew R. Pufall. 2014. Correlation of anomalous write error rates and ferromagnetic resonance spectrum in spin-transfer-torque-magnetic-random-access-memory devices containing in-plane free layers. *Applied Physics Letters* 104, 21 (2014), 212402. DOI: <https://doi.org/10.1063/1.4879847> arXiv:<https://doi.org/10.1063/1.4879847>
- [10] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Springer.
- [11] J. Grollier, V. Cros, H. Jaffrès, A. Hamzic, J. M. George, G. Faini, J. Ben Youssef, H. Le Gall, and A. Fert. 2003. Field dependence of magnetization reversal by spin transfer. *Physical Review B* 67, 17 (May 2003), 174402. DOI: <https://doi.org/10.1103/PhysRevB.67.174402>
- [12] J. D. Harms, F. Ebrahimi, X. Yao, and J. P. Wang. 2010. SPICE macromodel of spin-torque-transfer-operated magnetic tunnel junctions. *IEEE Transactions on Electron Devices* 57, 6 (June 2010), 1425–1430. DOI: <https://doi.org/10.1109/TED.2010.2047073>
- [13] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. 2003. A practical guide to support vector classification. Retrieved from <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- [14] G. Hu, J. J. Nowak, M. G. Gottwald, S. L. Brown, B. Doris, C. P. DâEmic, P. Hashemi, D. Houssameddine, Q. He, D. Kim, J. Kim, C. Kothandaraman, G. Lauer, H. K. Lee, N. Marchack, M. Reuter, R. P. Robertazzi, J. Z. Sun, T. Suwannasiri, P. L. Trouilloud, S. Woo, and D. C. Worledge. 2019. Spin-transfer torque MRAM with reliable 2 ns writing for last level cache applications. In *2019 IEEE International Electron Devices Meeting (IEDM'19)*. 2.6.1–2.6.4.
- [15] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno. 2010. A perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction. *Nature Materials* 9, 9 (2010), 721–724. DOI: <https://doi.org/10.1038/nmat2804>
- [16] Jodi M. Iwata-Harms, Guenole Jan, Huanlong Liu, Santiago Serrano-Guisan, Jian Zhu, Luc Thomas, Ru-Ying Tong, Vignesh Sundar, and Po-Kang Wang. 2018. High-temperature thermal stability driven by magnetization dilution in CoFeB free layers for spin-transfer-torque magnetic random access memory. *Scientific Reports* 8, 1 (Sept. 2018), 1–7. DOI: <https://doi.org/10.1038/s41598-018-32641-6>
- [17] G. Jan, L. Thomas, S. Le, Y. Lee, H. Liu, J. Zhu, R. Tong, K. Pi, Y. Wang, D. Shen, R. He, J. Haq, J. Teng, V. Lam, K. Huang, T. Zhong, T. Torng, and P. Wang. 2014. Demonstration of fully functional 8Mb perpendicular STT-MRAM chips with sub-5ns writing for non-volatile embedded memories. In *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*. 1–2. DOI: <https://doi.org/10.1109/VLSIT.2014.6894357>
- [18] Z. Jiang, P. Huang, L. Zhao, S. Kvatinisky, S. Yu, X. Liu, J. Kang, Y. Nishi, and H. S. P. Wong. 2015. Performance prediction of large-scale 1S1R resistive memory array using machine learning. In *2015 IEEE International Memory Workshop (IMW'15)*. 1–4. DOI: <https://doi.org/10.1109/IMW.2015.7150302>
- [19] J. Li, C. Augustine, S. Salahuddin, and K. Roy. 2008. Modeling of failure probability and statistical design of spin-torque transfer magnetic random access memory (STT MRAM) array for yield enhancement. In *2008 45th ACM/IEEE Design Automation Conference*. 278–283.
- [20] Mentor Graphics. [n.d.]. Solido Design Automation. Retrieved from https://www.mentor.com/products/ic_nanometer_design/solido-solutions/.
- [21] T. Min, Q. Chen, R. Beach, G. Jan, C. Horng, W. Kula, T. Torng, R. Tong, T. Zhong, D. Tang, P. Wang, M. M. Chen, J. Z. Sun, J. K. Debrose, D. C. Worledge, T. M. Maffitt, and W. J. Gallagher. 2010. A study of write margin of spin torque transfer magnetic random access memory technology. *IEEE Transactions on Magnetics* 46, 6 (June 2010), 2322–2327. DOI: <https://doi.org/10.1109/TMAG.2010.2043069>
- [22] Azalia Mirhoseini, Anna Goldie, Mustafa Yazgan, Joe Jiang, Ebrahim Songhori, Shen Wang, Young-Joon Lee, Eric Johnson, Omkar Pathak, Sungmin Bae, Azade Nazi, Jiwoo Pak, Andy Tong, Kavya Srinivasa, William Hang, Emre Tuncer, Anand Babu, Quoc V. Le, James Laudon, Richard Ho, Roger Carpenter, and Jeff Dean. 2020. Chip Placement with Deep Reinforcement Learning. arxiv:cs.LG/2004.10746
- [23] A. Nigam, C. W. Smullen, V. Mohan, E. Chen, S. Gurumurthi, and M. R. Stan. 2011. Delivering on the promise of universal memory for spin-transfer torque RAM (STT-RAM). In *IEEE/ACM International Symposium on Low Power Electronics and Design*. 121–126. DOI: <https://doi.org/10.1109/ISLPED.2011.5993623>
- [24] Michael Orshansky, Sani Nassif, and Duane Boning. 2010. *Design for Manufacturability and Statistical Design: A Constructive Approach* (1st ed.). Springer Publishing Company, Incorporated.
- [25] G. D. Panagopoulos, C. Augustine, and K. Roy. 2013. Physics-based SPICE-compatible compact model for simulating hybrid MTJ/CMOS circuits. *IEEE Transactions on Electron Devices* 60, 9 (Sept. 2013), 2808–2814. DOI: <https://doi.org/10.1109/TED.2013.2275082>

- [26] N. Paydavosi, S. Venugopalan, Y. S. Chauhan, J. P. Duarte, S. Jandhyala, A. M. Niknejad, and C. C. Hu. 2013. BSIM-SPICE models enable FinFET and UTB IC designs. *IEEE Access* 1 (2013), 201–215. DOI : <https://doi.org/10.1109/ACCESS.2013.2260816>
- [27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [28] N. Perrissin, S. Lequeux, N. Strelkov, A. Chavent, L. Vila, L. D. Buda-Prejbeanu, S. Auffret, R. C. Sousa, I. L. Prejbeanu, and B. Dieny. 2018. A highly thermally stable sub-20 nm magnetic random-access memory based on perpendicular shape anisotropy. *Nanoscale* 10, 25 (2018), 12187–12195. DOI : <https://doi.org/10.1039/C8NR01365A>
- [29] A. Raychowdhury, D. Somasekhar, T. Karnik, and V. De. 2009. Design space and scalability exploration of 1T-1STT MTJ memory arrays in the presence of variability and disturbances. In *2009 IEEE International Electron Devices Meeting (IEDM'09)*. 1–4. DOI : <https://doi.org/10.1109/IEDM.2009.5424242>
- [30] R. De Rose, M. Lanuzza, F. Crupi, G. Siracusano, R. Tomasello, G. Finocchio, and M. Carpentieri. 2017. Variability-aware analysis of hybrid MTJ/CMOS circuits by a micromagnetic-based simulation framework. *IEEE Transactions on Nanotechnology* 16, 2 (March 2017), 160–168. DOI : <https://doi.org/10.1109/TNANO.2016.2641681>
- [31] R. De Rose, M. Lanuzza, F. Crupi, G. Siracusano, R. Tomasello, G. Finocchio, M. Carpentieri, and M. Alioto. 2017. A variation-aware simulation framework for hybrid CMOS/spintronic circuits. In *2017 IEEE International Symposium on Circuits and Systems (ISCAS'17)*. 1–4. DOI : <https://doi.org/10.1109/ISCAS.2017.8050920>
- [32] R. De Rose, M. Lanuzza, F. Crupi, G. Siracusano, R. Tomasello, G. Finocchio, M. Carpentieri, and M. Alioto. 2018. A variation-aware timing modeling approach for write operation in hybrid CMOS/STT-MTJ circuits. *IEEE Transactions on Circuits and Systems I: Regular Papers* 65, 3 (March 2018), 1086–1095. DOI : <https://doi.org/10.1109/TCSI.2017.2762431>
- [33] U. Roy, T. Pramanik, L. F. Register, and S. K. Banerjee. 2016. Write error rate of spin-transfer-torque random access memory including micromagnetic effects using rare event enhancement. *IEEE Transactions on Magnetics* 52, 10 (Oct. 2016), 1–6. DOI : <https://doi.org/10.1109/TMAG.2016.2580532>
- [34] G. Siracusano, R. Tomasello, M. d'Aquino, V. Puliafito, A. Giordano, B. Azzerboni, P. Branganca, G. Finocchio, and M. Carpentieri. 2017. Description of statistical switching in perpendicular STT-MRAM within a numerical micromagnetic framework. In *2017 IEEE International Magnetism Conference (INTERMAG'17)*. 1–1. DOI : <https://doi.org/10.1109/INTMAG.2017.8007797>
- [35] Y. J. Song, J. H. Lee, S. H. Han, H. C. Shin, K. H. Lee, K. Suh, D. E. Jeong, G. H. Koh, S. C. Oh, J. H. Park, S. O. Park, B. J. Bae, O. I. Kwon, K. H. Hwang, B. Y. Seo, Y. K. Lee, S. H. Hwang, D. S. Lee, Y. Ji, K. C. Park, G. T. Jeong, H. S. Hong, K. P. Lee, H. K. Kang, and E. S. Jung. 2018. Demonstration of highly manufacturable STT-MRAM embedded in 28nm logic. In *2018 IEEE International Electron Devices Meeting (IEDM'18)*. 18.2.1–18.2.4. DOI : <https://doi.org/10.1109/IEDM.2018.8614635>
- [36] J. Z. Sun. 2000. Spin-current interaction with a monodomain magnetic body: A model study. *Physical Review B* 62, 1 (Jul 2000), 570–578. DOI : <https://doi.org/10.1103/PhysRevB.62.570>
- [37] J. Z. Sun, S. L. Brown, W. Chen, E. A. Delenia, M. C. Gaidis, J. Harms, G. Hu, Xin Jiang, R. Kilaru, W. Kula, G. Lauer, L. Q. Liu, S. Murthy, J. Nowak, E. J. O'Sullivan, S. S. P. Parkin, R. P. Robertazzi, P. M. Rice, G. Sandhu, T. Topuria, and D. C. Worledge. 2013. Spin-torque switching efficiency in CoFeB-MgO based tunnel junctions. *Physical Review B* 88, 10 (Sep 2013), 104426. DOI : <https://doi.org/10.1103/PhysRevB.88.104426>
- [38] J. Z. Sun, R. P. Robertazzi, J. Nowak, P. L. Trouilloud, G. Hu, D. W. Abraham, M. C. Gaidis, S. L. Brown, E. J. O'Sullivan, W. J. Gallagher, and D. C. Worledge. 2011. Effect of subvolume excitation and spin-torque efficiency on magnetic switching. *Physical Review B* 84, 6 (Aug 2011), 064413. DOI : <https://doi.org/10.1103/PhysRevB.84.064413>
- [39] tensorflow.org. [n.d.]. TensorFlow C++ Reference. Retrieved October 14, 2020, from https://www.tensorflow.org/api_docs/cc.
- [40] L. Thomas, G. Jan, S. Serrano-Guisan, H. Liu, J. Zhu, Y. Lee, S. Le, J. Iwata-Harms, R. Tong, S. Patel, V. Sundar, D. Shen, Y. Yang, R. He, J. Haq, Z. Teng, V. Lam, P. Liu, Y. Wang, T. Zhong, H. Fukuzawa, and P. Wang. 2018. STT-MRAM devices with low damping and moment optimized for LLC applications at Ox nodes. In *2018 IEEE International Electron Devices Meeting (IEDM'18)*. 27.3.1–27.3.4.
- [41] Arne Vansteenkiste, Jonathan Leliaert, Mykola Dvornik, Mathias Helsen, Felipe Garcia-Sanchez, and Bartel Van Waeyenberge. 2014. The design and verification of MuMax3. *AIP Advances* 4, 10 (2014), 107133. DOI : <https://doi.org/10.1063/1.4899186> arXiv:<https://doi.org/10.1063/1.4899186>
- [42] A. F. Vincent, N. Locatelli, J. Klein, W. S. Zhao, S. Galdin-Retailleau, and D. Querlioz. 2015. Analytical macrospin modeling of the stochastic switching time of spin-transfer torque devices. *IEEE Transactions on Electron Devices* 62, 1 (Jan. 2015), 164–170. DOI : <https://doi.org/10.1109/TED.2014.2372475>

- [43] G. Wang, Y. Zhang, J. Wang, Z. Zhang, K. Zhang, Z. Zheng, J. Klein, D. Ravelosona, Y. Zhang, and W. Zhao. 2019. Compact modeling of perpendicular-magnetic-anisotropy double-barrier magnetic tunnel junction with enhanced thermal stability recording structure. *IEEE Transactions on Electron Devices* 66, 5 (2019), 2431–2436.
- [44] L. Xue, C. Ching, A. Kontos, J. Ahn, X. Wang, R. Whig, H. Tseng, J. Howarth, S. Hassan, H. Chen, M. Bangar, S. Liang, R. Wang, and M. Pakala. 2018. Process optimization of perpendicular magnetic tunnel junction arrays for last-level cache beyond 7 nm node. In *2018 IEEE Symposium on VLSI Technology*. 117–118.
- [45] M. Yamanouchi, A. Jander, P. Dhagat, S. Ikeda, F. Matsukura, and H. Ohno. 2011. Domain structure in CoFeB thin films with perpendicular magnetic anisotropy. *IEEE Magnetism Letters* 2 (2011), 3000304–3000304.

Received August 2020; revised October 2020; accepted November 2020